Statistica Sinica **27** (2017), 373-388
doi:http://dx.doi.org/10.5705/ss.2014.162

ROBUST ESTIMATION OF DISPERSION PARAMETER IN DISCRETELY OBSERVED DIFFUSION PROCESSES

Junmo Song

Jeju National University

Abstract: In this paper, we consider robust estimation of the dispersion parameter in discretely observed diffusion processes. To construct a robust estimator, we first approximate the transition density of the diffusion process to the Gaussian density by using Kessler (1997) approach and then employ Basu et al. (1998) minimum density power divergence (MDPD) estimation method. It is shown that, under regularity conditions, the MDPD estimator is strongly consistent and asymptotically normal. Through a simulation study, we compared the performances of the MDPD estimator and the quasi-maximum likelihood (QML) estimator based on the approximated transition density. Numerical results demonstrate that the proposed estimator has strong robust properties with little loss in asymptotic efficiency relative to the QML estimator.

Key words and phrases: Diffusion processes, dispersion parameter, minimum density power divergence estimator, robust estimation.

1. Introduction

Diffusion processes are popular in such fields as the physical and biological sciences, finance, and engineering. In particular, they are widely used in finance for pricing options and other derivatives. During the past decades, many articles have been devoted to statistical inference for diffusion processes. Especially, estimation in discretely observed diffusion processes has received a great deal of attention. Authors such as Dacunha-Castelle and Florens-Zmirou (1986), Yoshida (1992), Kessler (1997), and Aït-Sahalia (2002, 2008) approximated diffusion models or transition densities to estimate the models. Shephard and Pitt (1997), Liesenfeld and Richard (2006), and Richard and Zhang (2007) proposed various importance sampling methods. Simulation-based estimation methods were also developed by Pedersen (1995), Durham and Gallant (2002), and Beskos, Papaspiliopoulos and Roberts (2009). Recently, Kleppe, Yu and Skaug (2014) combined the closed form approach of Aït-Sahalia (2008) and the importance sampling technique of Richard and Zhang (2007). The statistical inference and some basic results for diffusion processes are well summarized in Prakasa Rao (1999), Kutoyants (2004), and Phillips and Yu (2009).

As is well known, estimators based on likelihood, especially Gaussian likelihood, are strongly influenced by outliers or extreme values. To overcome this problem, various robust estimation methods have been developed. However, to the best of our knowledge, few works have addressed the problem of estimating diffusion models in the presence of outliers. Recently, Lee and Song (2013) studied this problem on the Ornstein-Uhlenbeck type processes. Recall that many estimation methods for diffusion processes rely on the maximum likelihood (ML) approach based on an approximated transition density. Since the transition distributions approach the normal distribution when the sampling interval is short, as in high-frequency sampling cases, it can be easily surmised that a similar problem is likely to occur in the estimation procedures of the diffusion processes. Actually, our simulation study shows that the quasi-ML estimator of Kessler (1997) is severely damaged by outliers.

The purpose of this paper is to propose a robust estimator for the dispersion parameter in discretely observed diffusion processes. For this task, we consider the estimation method based on divergence that evaluates the discrepancy between any two probability distributions. The divergence-based estimation methods have been used successfully in constructing robust estimators. For a review, we refer to Pardo (2006), Cichocki and Amari (2010), and the references therein. In this paper, we employ Basu et al. (1998) (henceforth, BHHJ) density power divergence

$$d_\alpha(g,f) = \begin{cases} \displaystyle\int \left\{ f^{1+\alpha}(z) - (1+\frac{1}{\alpha})\, g(z)\, f^\alpha(z) + \frac{1}{\alpha}\, g^{1+\alpha}(z) \right\} dz & ,\alpha > 0, \\ \displaystyle\int g(z)\left\{\log g(z) - \log f(z)\right\} dz & ,\alpha = 0, \end{cases} \tag{1.1}$$

where f and g are probability densities. This divergence includes Kullback-Leibler divergence and L_2-distance as special cases. BHHJ proposed the minimum density power divergence (MDPD) estimator by minimizing the empirical version of the density power divergence:

$$\hat{\theta}_{\alpha,n} = \underset{\theta\in\Theta}{\operatorname{argmin}} \ \frac{1}{n}\sum_{i=1}^{n} H_\alpha(\theta; X_i)\,,$$

where

$$H_\alpha(\theta; X_i) = \begin{cases} \displaystyle\int f_\theta^{1+\alpha}(z)dz - \left(1+\frac{1}{\alpha}\right) f_\theta^\alpha(X_i) & ,\alpha > 0, \\ -\log f_\theta(X_i) & ,\alpha = 0, \end{cases}$$

and $X_1, \ldots, X_n$ are i.i.d. random variables from an unknown density g. BHHJ showed that $\hat{\theta}_{\alpha,n}$ is weakly consistent for $T_\alpha(g) := \operatorname{argmin}_{\theta\in\Theta} d_\alpha(g, f_\theta)$, which is the essential target parameter, and asymptotically normal. Further, they demonstrated that the estimator possesses strong robust properties with little loss in asymptotic efficiency relative to the ML estimator. Indeed, α controls the trade-off between robustness and asymptotic efficiency in the estimation procedure.

Compared to other robust methods, such as minimum Hellinger distance estimation, BHHJ's method does not require any smoothing methods. Hence, it can avoid such difficulties as bandwidth selection for the nonparametric density estimation. For this reason, the method can be conveniently applied to any parametric models (see, e.g., Juárez and Schucany (2004), Fujisawa and Eguchi (2006), and Kim and Lee (2013)).

The remainder of the paper is organized as follows. In Section 2, we construct the MDPD estimator for diffusion process and present the asymptotic properties of the proposed estimator. In Section 3, we report on a simulation study that compared the performance of the MDPD estimator and the quasi-ML estimator. Section 4 concludes the paper. Proofs of the main results in Section 2 are provided in the online supplementary.

2. Main Result

Consider the univariate time-homogeneous diffusion process $\{X_t : t \geq 0\}$ defined by

$$dX_t = a(X_t)dt + b(X_t, \sigma)dW_t, \quad X_0 = x_0, \tag{2.1}$$

where $\sigma \in \Theta$, a compact subset of $\mathbb{R}$, and $\{W_t : t \geq 0\}$ is a standard Wiener process. The real valued functions a and b are assumed known apart from σ and smooth enough to admit a unique solution. We assume that a sample $\{X_{t_i^n} : 0 \leq i \leq n\}$ is discretely observed, where $t_i^n = ih_n$ and $\{h_n\}$ is a sequence of positive numbers with $h_n \to 0$ and $nh_n \to \infty$.

To construct the MDPD estimator for (2.1), we consider the density power divergence for a time-homogeneous Markov chain similar to the analogue for regression models (see page 555 in Basu et al. (1998)). Let $\{X_t\}$ be a Markov chain with a transition density of $p(y|x)$ and assume that one intend to approximate the transition density with certain parametric conditional density of $p_\theta(y|x)$. Then, substituting $g(z)$ and $f_\theta(z)$ in (1.1) with $p(y|x)$ and $p_\theta(y|x)$, respectively, one can obtain the divergence for the Markov chain $\{X_t\}$. Based on this, given observations $X_0, X_1, \ldots, X_n$, the MDPD estimator for the Markov chain is

$$\hat{\theta}_{\alpha,n} = \operatorname*{argmin}_{\theta\in\Theta} \begin{cases} \displaystyle\sum_{t=1}^{n} \int p_\theta^{1+\alpha}(y|X_{t-1})dy - \Big(1+\frac{1}{\alpha}\Big) \sum_{t=1}^{n} p_\theta^{\alpha}(X_t|X_{t-1}) & , \alpha > 0, \\ \displaystyle -\sum_{t=1}^{n} \log p_\theta(X_t|X_{t-1}) & , \alpha = 0. \end{cases} \tag{2.2}$$

We use this estimator for the parameter estimation in (2.1). To define a contrast function, we consider the approximation technique of Kessler (1997) in which he proposed an asymptotically efficient estimator based on the Gaussian approximation of the transition density. More specifically, we approximate the conditional distribution of $X_{t_i^n}|X_{t_{i-1}^n}$ to the conditional normal with mean $m(X_{t_{i-1}^n},\sigma)=E\big(X_{t_i^n}|X_{t_{i-1}^n}\big)$ and variance $v(X_{t_{i-1}^n},\sigma)=E\big((X_{t_i^n}-m(X_{t_{i-1}^n},\sigma))^2|X_{t_{i-1}^n}\big)$, which results in the contrast function for (2.1) given by

$$\begin{cases}\displaystyle\sum_{i=1}^{n}\frac{1}{v(X_{t_{i-1}^n},\sigma)^{\alpha/2}}\left\{\frac{1}{\sqrt{1+\alpha}}-\Big(1+\frac{1}{\alpha}\Big)\exp\Big(-\frac{\alpha}{2}\frac{(X_{t_i^n}-m(X_{t_{i-1}^n},\sigma)^2}{v(X_{t_{i-1}^n},\sigma)}\Big)\right\}, & \alpha>0,\\ \displaystyle\sum_{i=1}^{n}\left\{\frac{(X_{t_i^n}-m(X_{t_{i-1}^n},\sigma))^2}{v(X_{t_{i-1}^n},\sigma)}+\log v(X_{t_{i-1}^n},\sigma)\right\} & ,\alpha=0.\end{cases}\tag{2.3}$$

We replace m and v with the closed-form approximations as in Kessler (1997). Let L_σ denote the generator of the diffusion process

$$L_\sigma g(x)=a(x)\frac{\partial g}{\partial x}(x)+\frac{1}{2}b(x,\sigma)^2\frac{\partial^2 g}{\partial x^2}(x)\,,\quad \text{for } g\in C^2(\mathbb{R}),\tag{2.4}$$

where $a(x)$ and $b(x,\sigma)$ are assumed to be differentiable with respect to x up to order $2k$. Hereafter, we denote $\partial^r/\partial\sigma^r$ (resp. $\partial^r/\partial x^r$) by ∂_σ^r (resp. ∂_x^r). For any $k_0\in\{1,\ldots,k\}$, we substitute m and v in (2.3) with

$$r_{k_0}(h_n,X_{t_{i-1}^n},\sigma):=X_{t_{i-1}^n}+\sum_{i=1}^{k_0}\frac{h_n^i}{i!}L_\sigma^i\,X_{t_{i-1}^n},$$
$$b(X_{t_{i-1}^n},\sigma)^2\big\{1+\bar\Gamma_{k_0+1}(h_n,X_{t_{i-1}^n},\sigma)\big\},$$

respectively. For the explicit form of $\bar\Gamma_l$, refer to Kessler (1997), pages 214-215. To avoid technical difficulties, the Taylor expansions of $\big(1+\bar\Gamma_{k_0+1}\big)^{-1}$, $\big(1+\bar\Gamma_{k_0+1}\big)^{-\alpha/2}$, and $\log\big(1+\bar\Gamma_{k_0+1}\big)$ are considered, and we denote the coefficient of h_n^j in each expansion by d_j, d_j^α, and e_j, respectively. Then, the MDPD estimator for the dispersion parameter in (2.1) is

$$\hat\sigma_{\alpha,n}(k_0,k_1)=\underset{\sigma\in\Theta}{\operatorname{argmin}}\ \frac{1}{n}\sum_{i=1}^{n}V_{n,i}^\alpha(\sigma;k_0,k_1),\tag{2.5}$$

where $k_0\in\{1,\ldots,k\}$, $k_1\in\{0\}\cup\mathbb{N}$, and

$$\begin{aligned}&V_{n,i}^\alpha(\sigma;k_0,k_1)\\&=\begin{cases}\displaystyle\frac{1}{b(X_{t_{i-1}^n},\sigma)^\alpha}\Big\{1+\sum_{j=1}^{k_1}h_n^jd_j^\alpha(X_{t_{i-1}^n},\sigma)\Big\}\Bigg[\frac{1}{\sqrt{1+\alpha}}\\ \displaystyle-\Big(1+\frac{1}{\alpha}\Big)\exp\Big(-\frac{\alpha}{2}\frac{(X_{t_i^n}-r_{k_0}(h_n,X_{t_{i-1}^n},\sigma))^2}{b(X_{t_{i-1}^n},\sigma)^2h_n}\Big\{1+\sum_{j=1}^{k_1}h_n^jd_j(X_{t_{i-1}^n},\sigma)\Big\}\Big)\Bigg], & \alpha>0,\\ \displaystyle\frac{(X_{t_i^n}-r_{k_0}(h_n,X_{t_{i-1}^n},\sigma))^2}{b(X_{t_{i-1}^n},\sigma)^2h_n}\Big\{1+\sum_{j=1}^{k_1}h_n^jd_j(X_{t_{i-1}^n},\sigma)\Big\}\\ \displaystyle\qquad+\log b(X_{t_{i-1}^n},\sigma)^2+\sum_{i=1}^{k_1}h_n^je_j(X_{t_{i-1}^n},\sigma), & \alpha=0.\end{cases}\end{aligned}$$

Remark 1. By letting $k_0 = k_1 = k$, the MDPD estimator with $\alpha = 0$ is the estimator of Kessler (1997). In the case of $k_0 = 1$ and $k_0 = 0$, the weak consistency of the estimator was examined in Song et al. (2007).

Remark 2. According to our simulation study, the estimator with large k_1 produces more precise estimation results when the sampling interval is not sufficiently short, as in low-frequency sampling cases (see Section 3).

To establish the consistency and asymptotic normality of the MDPD estimator in (2.5), we need some conditions.

A1. There exists a constant C such that, for any x and y,

$$|a(x) - a(y)| + |b(x, \sigma_0) - b(y, \sigma_0)| \leq C|x - y|,$$

where σ_0 denotes the true parameter of σ.

A2. The process X from (2.1) is ergodic with invariant measure μ_0 such that $\int x^k d\mu_0(x) < \infty$ for all $k \geq 0$.

A3. $\sup_t E|X_t|^k < \infty$ for all $k \geq 0$.

A4. $\inf_{x,\sigma} b(x, \sigma)^2 > 0$.

A5. If $\mu_0\{x : b(x, \sigma)^2 = b(x, \sigma_0)^2\} = 1$, then $\sigma = \sigma_0$.

A6. For a positive integer k, the functions a and b are continuously differentiable with respect to x up to order k, and all those derivatives belong to $\mathcal{P} := \{f : |f| \leq C(1+|x|)^C \text{ for some } C\}$, where C does not depend on the parameter.

(i) The function b is differentiable with respect to σ and its derivative belongs to $\mathcal{P}$.

(ii) All the x-derivatives of b up to order k are twice differentiable with respect to σ; b is three times differentiable with respect to σ. Moreover, all those derivatives belong to $\mathcal{P}$.

Theorem 1. *Suppose that* **A1**–**A6(i)** *with* $2k$ *hold, and* $nh_n^p \to 0$ *for some* $p > 1$. *Then, for any* $k_0 \in \{1, \ldots, k\}$ *and* $k_1 \in \{0\} \cup \mathbb{N}$, $\hat{\sigma}_{\alpha,n}(k_0, k_1)$ *converges almost surely to* σ_0 *for each* $\alpha \geq 0$.

Theorem 2. *Suppose that* **A1**–**A5** *and* **A6(ii)** *with* $2k$ *hold, and* σ *is in the interior of* Θ. *For any* $k_0 \in \{1, \ldots, k\}$, $k_1 \in \{0\} \cup \mathbb{N}$ *and each* $\alpha \geq 0$, *if* $nh_n^2 \to 0$, *then*

$$\sqrt{n}(\hat{\sigma}_{\alpha,n}(k_0, k_1) - \sigma_0) \longrightarrow N(0, \Sigma_\alpha) \quad \textit{in distribution},$$

where

$$\Sigma_\alpha = \mathcal{K}(\alpha)\int \frac{(\partial_\sigma b(x,\sigma_0))^2}{b(x,\sigma_0)^{2\alpha+2}}\,d\mu_0(x)\Big\{\int \frac{(\partial_\sigma b(x,\sigma_0))^2}{b(x,\sigma_0)^{\alpha+2}}\,d\mu_0(x)\Big\}^{-2},$$

$$\mathcal{K}(\alpha) = \frac{(1+\alpha)^3}{(2+\alpha^2)^2}\Big\{2\frac{(1+\alpha)^2(1+2\alpha^2)}{(1+2\alpha)^2\sqrt{1+2\alpha}} - \frac{\alpha^2}{1+\alpha}\Big\}.$$

Remark 3. In case $\alpha = 0$ and $k_0 = k_1 = k$, this asymptotic normality also holds when $nh_n^{2k+1} \to 0$. For more details, we refer to Theorem 1 in Kessler (1997).

Remark 4. Choosing an optimal α is an important issue. Several studies of the problem have been made (see, e.g., Warwick (2005), Fujisawa and Eguchi (2006), and Durio and Isaia (2011)). Conventionally, a small α is recommended because a large α may lead to a big loss in efficiency when the portion of outliers is not as large as speculated. It is also noteworthy that the ML estimator and the MDPD estimator are likely to produce similar values in estimates when data are not contaminated or the extent of the contamination is not severe. As a rule of thumb, a small α may be preferred when the MDPD estimates are similar to the ML estimate, whereas a relatively large α should be selected in the cases where the differences are large. In this regard, Durio and Isaia (2011) defined a similarity measure and implemented the idea by using a bootstrap test on the similarity.

We can also consider a robust estimator for diffusion processes with an unknown drift parameter, $dX_t = a(X_t,\theta)dt + b(X_t,\sigma)dW_t$. In this case, the contrast function can be defined by replacing $m(X_{t^n_{i-1}},\sigma)$ and $v(X_{t^n_{i-1}},\sigma)$ in (2.3) with

$$\mathbf{r}_{k_0}(h_n, X_{t^n_{i-1}}, \theta, \sigma) := x + \sum_{i=1}^{k_0} \frac{h_n^i}{i!} L^i_{\theta,\sigma}\, X_{t^n_{i-1}},$$

$$b(X_{t^n_{i-1}},\sigma)^2\big\{1 + \bar{\Gamma}_{k_0+1}(h_n, X_{t^n_{i-1}}, \theta, \sigma)\big\},$$

respectively, where

$$L_{\theta,\sigma} g(x) = a(x,\theta)\partial_x g(x) + \frac{1}{2} b(x,\sigma)^2 \partial_x^2 g(x) \quad \text{for } g \in C^2(\mathbb{R})$$

and $\bar{\Gamma}_{k_0+1}(h_n, X_{t^n_{i-1}}, \theta, \sigma)$, say $\bar{\mathbf{\Gamma}}_{k_0+1}$, is the one given in Kessler (1997). Let $V^\alpha_{n,i}(\theta,\sigma;k_0,k_1)$ be the counterpart of $V^\alpha_{n,i}(\sigma;k_0,k_1)$ in (2.5) by substituting d_j, d_j^α, and e_j with the coefficients of h_n^j in the Taylor expansions of $\big(1+\bar{\mathbf{\Gamma}}_{k_0+1}\big)^{-1}$, $\big(1+\bar{\mathbf{\Gamma}}_{k_0+1}\big)^{-\alpha/2}$, and $\log\big(1+\bar{\mathbf{\Gamma}}_{k_0+1}\big)$, respectively. Then, the MDPD estimator is

$$\begin{pmatrix} \hat{\theta}_{\alpha,n}(k_0,k_1) \\ \hat{\sigma}_{\alpha,n}(k_0,k_1) \end{pmatrix} = \underset{(\theta,\sigma)\in\Theta}{\operatorname{argmin}}\ \frac{1}{n}\sum_{i=1}^n V^\alpha_{n,i}(\theta,\sigma;k_0,k_1)\,. \tag{2.6}$$

To establish the consistency of this estimator, the assumptions **A1** and **A6**(i) need replacement by the following.

A1′. There exists a constant C such that for any x, y,

$$|a(x,\theta_0) - a(y,\theta_0)| + |b(x,\sigma_0) - b(y,\sigma_0)| \le C|x-y|\,.$$

A6′(i). The functions a and b are differentiable with respect to θ and σ, respectively, and their derivatives belong to $\mathcal{P}$.

Theorem 3. *Suppose that* **A1′**–**A6′(i)** *with* $2k$ *hold, and* $nh_n^p \to 0$ *for some* $p > 1$. *Then, for any* $k_0 \in \{1, \ldots, k\}$ *and* $k_1 \in \{0\} \cup \mathbb{N}$, $\hat{\sigma}_{\alpha,n}(k_0, k_1)$ *at (2.6) converges almost surely to* σ_0 *for each* $\alpha \ge 0$.

Remark 5. For the consistency of $\hat{\theta}_{\alpha,n}(k_0, k_1)$, we follow the arguments on pages 217-218 of Kessler (1997). Since Θ is compact and $\hat{\sigma}_{\alpha,n}(k_0, k_1)$ converges to σ_0, we can take a subsequence $\{(\hat{\theta}_{\alpha,n_k}, \hat{\sigma}_{\alpha,n_k})\}$ converging to $(\theta_\infty, \sigma_0)$ for some θ_∞. Then, it can be shown that for $\alpha > 0$,

$$\begin{aligned}
&\frac{1}{n_k h_{n_k}} \sum_{i=1}^{n_k} \left\{ V^\alpha_{n_k,i}(\hat{\theta}_{\alpha,n_k}, \hat{\sigma}_{\alpha,n_k}; k_0, k_1) - V^\alpha_{n_k,i}(\theta_0, \hat{\sigma}_{\alpha,n_k}; k_0, k_1) \right\} \\
&\xrightarrow{a.s.} \frac{1}{2\sqrt{1+\alpha}} \int \frac{(a(x,\theta_\infty) - a(x,\theta_0))^2}{b(x,\sigma_0)^{\alpha+2}} d\mu_0(x) - \frac{1}{2\sqrt{1+\alpha}} R(\theta_\infty),
\end{aligned}$$

where

$$\begin{aligned}
&R(\theta_\infty) \\
&= \int \left\{ \frac{\mathbf{d_1}(x,\theta_\infty,\sigma_0) - \mathbf{d_1}(x,\theta_0,\sigma_0)}{b(x,\sigma_0)^\alpha} + \frac{4+\alpha}{1+\alpha} \frac{a(x,\theta_\infty) - a(x,\theta_0)}{b(x,\sigma_0)^{\alpha+2}} b(x,\sigma_0)\partial_x b(x,\sigma_0) \right\} \\
&\qquad d\mu_0(x)
\end{aligned}$$

and $\mathbf{d_1}$ is the coefficient of h_n in the Taylor expansion of $\left(1 + \bar{\mathbf{\Gamma}}_{k_0+1}\right)^{-1}$. Unlike the MDPD estimator with $\alpha = 0$ (see Lemma 3 in Kessler (1997)), the troublesome term $R(\theta_\infty)$ appears in the limit. One of the sufficient conditions for the consistency of $\hat{\theta}_{\alpha,n}(k_0, k_1)$ is that the integrand in $R(\theta_\infty)$ is zero for μ_0-a.s.. For instance, we can choose the case where the expansions of $\left(1 + \bar{\mathbf{\Gamma}}_{k_0+1}\right)^{-1}$ and $\left(1 + \bar{\mathbf{\Gamma}}_{k_0+1}\right)^{-\alpha/2}$ are not under consideration and $\partial_x b(x,\sigma) = 0$. Such examples can be found in Lee and Song (2013).

3. Simulation Study

In the present simulation, we compared the performance of the MDPD estimator (MDPDE) with $\alpha > 0$ and $\alpha = 0$, the Quasi-ML estimator (QMLE) that includes Kessler (1997) estimator. Toward this, we considered the stochastic differential equation

$$dX_t = -X_t dt + \left(1 + \frac{\sigma}{1 + X_t^2}\right) dW_t, \quad X_0 = 0, \tag{3.1}$$

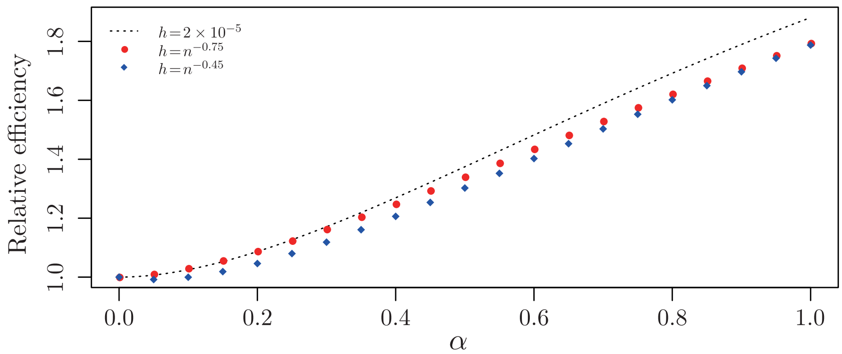

Figure 1. Relative efficiency of the MDPDE with α ($n = 10^6$).

with the true parameter $\sigma_0 = 1$. The path of X was generated via the Milstein scheme with the generating interval of $h = 2 \times 10^{-5}$, and the sample $\{X_{o,t_i^n}\}_{i=1}^n$ was observed with the sampling intervals of $h_n = n^{-0.75}$ and $n^{-0.45}$. In this setting, the QMLE with $k_0 = k_1 = 1$ is Kessler's estimator. The sample size under consideration was 1,000. When $n =$1,000, $h_n = 1,000^{-0.75} \approx 1.5/250$ (resp. $1,000^{-0.45} \approx 11/250$) corresponds to the interval of 1.5 (resp. 11) trading days in financial applications, and thus, it can describe a high (resp. low)-frequency case. The comparison was based on the sample mean squared error (MSE) and

$$d_R := \frac{\text{MSE of MDPDE with } \alpha}{\text{MSE of QMLE}}.$$

We examined the case where the data were not contaminated by outliers. Based on 1,000 repetitions, the mean, standard deviation (SD), MSE of the estimates, and d_R were calculated for $k_0 = 1, 2$ and $k_1 = 0, 1$. The estimation results are presented in Table 1. Here the MDPDE with α close to 0 performs similarly to the QMLE. The estimators with $k_1 = 1$ show better performance than the estimator with $k_1 = 0$ when $h_n = 1,000^{-0.45}$, which suggests that a larger k_1 is recommended in a low-frequency sampling case. k_0 had little effect on the estimation in our simulations. The QMLE outperforms the MDPDE when $h_n = 1,000^{-0.75}$, whereas the MDPDE with small α shows a slightly better performance in the case of $h_n = 1,000^{-0.45}$. A possible explanation for this might be that the transition distribution of $X_{t_i^n}|X_{t_{i-1}^n}$ is not sufficiently close to the normal distribution due to the long sampling interval. It is, nevertheless, expected that the QMLE would outperform the MDPDE as the sample size increases. The point is that the performance of the MDPDE with α close to 0 is not poor, and the efficiency of the MDPDE decreases with an increase in α.

We can observe this in two figures. The dotted line in Figure 1 displays relative efficiency according to α, the ratio of asymptotic variance of the MDPDE

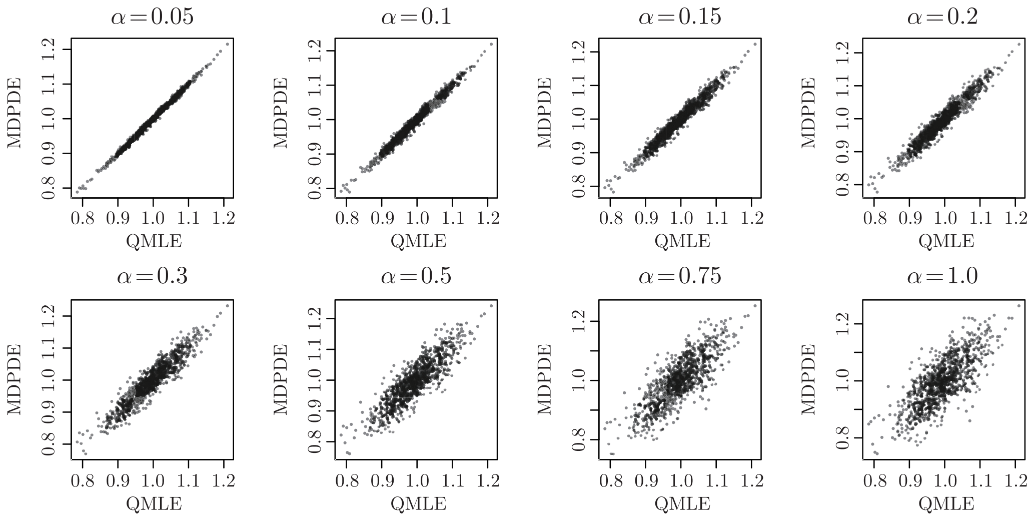

Figure 2. Scatter plots of QMLE and MDPDE when data are not contaminated ($h_n = 1,000^{-0.75}$).

with α to that of the QMLE. We used Lemma 3 in supplementary material to calculate the integrations in the variance, where the sample of size $n = 20/h$ was generated and the sampling interval was set to the generating interval h. It can be seen that the relative efficiency increases up to about 1.9, and around $\alpha = 0.2$, an efficiency of about 1.1 is produced. In this figure, the circles and lozenges represent the ratio of the sample variance of the MDPD estimates to that of the QML estimates for $h_n = n^{-0.75}$ and $n^{-0.45}$, respectively. The lozenges lie below the circles, indicating that the MDPDE is more efficient when the sampling interval is larger. Figure 2 depicts the scatter plots of pairs of QML and MDPD estimates when $h_n = n^{-0.75}$, and confirms the results. The plots for the sampling interval of $h_n = n^{-0.45}$ are similar to those for $h_n = n^{-0.75}$ and are not reported.

We addressed the case in which outliers are involved in the observations. Here, it was assumed that the sample $\{X_{o,t_i^n}\}_{i=0}^n$ from (3.1) was contaminated by outliers $\{X_{c,t_i^n}\}_{i=0}^n \sim$ i.i.d. $N(0, \sigma_V^2)$, and that the observed random variables followed the scheme $X_{t_i^n} = X_{o,t_i^n} + p_i X_{c,t_i^n}$, where $\{p_i\}_{i=0}^n$ are i.i.d. Bernoulli random variables with success probability p; $\{p_i\}$, $\{X_{o,t_i^n}\}$, and $\{X_{c,t_i^n}\}$ were assumed to be independent. We considered the cases of $\sigma_V^2 = 0.5, 1$, $p = 0.01, 0.03$ for $h_n = n^{-0.75}$, and $p = 0.03, 0.05$ for $h_n = n^{-0.45}$. The mean, standard deviation, MSE, and d_R based on $\{X_{t_i^n}\}_{i=0}^n$ were calculated from 1,000 repetitions. Tables 2−5 demonstrate that the MDPDE performs much better than the QMLE when the data are severely contaminated, that is, p or σ_V^2 increases, and the scatter plots in Figure 3 also show the strong robustness of the MDPDE well. Here, it is important to note that the QMLE is more damaged by outliers when

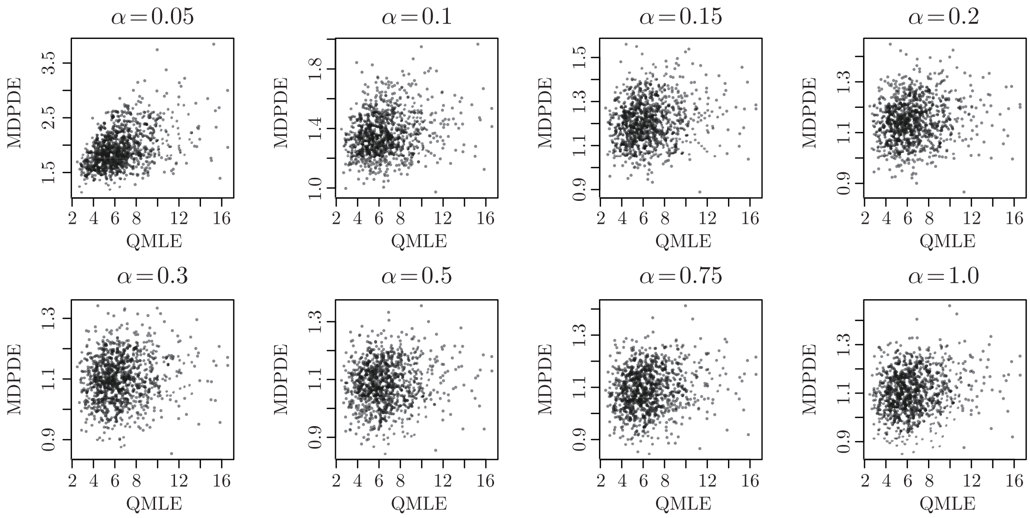

Figure 3. Scatter plots of QMLE and MDPDE when data are contaminated ($p = 0.03, \sigma_V^2 = 1$, and $h_n = 1,000^{-0.75}$).

$h_n = n^{-0.75}$ than when $h_n = n^{-0.45}$, mainly due to the fact that the normal likelihood is easily damaged because of its short tail. This indicates that when estimating diffusion models without jump-components in high-frequency sampling cases, particular attention should be paid to the existence of outliers or jumps. In such cases, the MDPDE can be a promising estimator for the dispersion parameter.

In addition, we dealt with a diffusion process with an unknown drift parameter

$$dX_t = -\theta X_t dt + \left(1 + \frac{\sigma}{1 + X_t^2}\right) dW_t, \quad X_0 = 0\,.$$

We set the true parameter $(\theta_0, \sigma_0) = (3, 1)$. $\sigma_V^2 = 0, 0.5, 1$, $p = 0.03$, $k_0 = 1$, $k_1 = 0$, and $h_n = n^{-0.75}$ were considered. Here, $\sigma_V^2 = 0$ represents the case of no contamination. We used the MDPDE at (2.6) to estimate the parameter (θ, σ). Estimation results are summarized in Tables 6 and 7. It can be seen that the MDPDE for σ has strong robust properties and performs similarly to the QMLE when the data are not contaminated. Thus, the MDPDE can be a good option when diffusion model has an unknown drift parameter. Concerning θ, while its MDPDE also seems to be robust against outliers, the consistency of the MDPD estimator does not seem to hold. It appears that a decreasing trend exists in the sample mean of estimates as α increases. This phenomenon is still observed when the sample size is large (see Table 7). Since the shape of the contrast function of the MDPDE with a small α is similar to that of the QMLE, the estimator with a small α, for example α lying in [0.05, 0.2], could be used to estimate θ.

Table 1. Mean, SD, MSE, and d_R when data are not contaminated ($h_n = 1,000^{-\delta}$).

δ	k_0	k_1		*MDPDE*								
				QMLE	0.05	0.1	0.15	0.2	0.3	0.5	0.75	1.0
0.75	1	0	Mean	0.992	0.992	0.992	0.992	0.993	0.993	0.994	0.996	0.997
			SD	0.064	0.064	0.065	0.065	0.066	0.069	0.074	0.080	0.085
			MSE	0.004	0.004	0.004	0.004	0.004	0.005	0.005	0.006	0.007
			d_R	1.000	1.009	1.027	1.052	1.082	1.154	1.325	1.554	1.766
	1	1	Mean	0.999	0.999	0.999	1.000	1.000	1.000	1.001	1.003	1.003
			SD	0.064	0.064	0.065	0.066	0.067	0.069	0.074	0.080	0.085
			MSE	0.004	0.004	0.004	0.004	0.004	0.005	0.005	0.006	0.007
			d_R	1.000	1.010	1.029	1.055	1.087	1.161	1.339	1.574	1.792
	2	0	Mean	0.992	0.992	0.992	0.992	0.993	0.993	0.994	0.996	0.997
			SD	0.064	0.064	0.065	0.065	0.066	0.069	0.074	0.080	0.085
			MSE	0.004	0.004	0.004	0.004	0.004	0.005	0.005	0.006	0.007
			d_R	1.000	1.009	1.027	1.052	1.082	1.154	1.325	1.553	1.766
	2	1	Mean	0.999	0.999	0.999	1.000	1.000	1.000	1.001	1.003	1.003
			SD	0.064	0.064	0.065	0.066	0.067	0.069	0.074	0.080	0.085
			MSE	0.004	0.004	0.004	0.004	0.004	0.005	0.005	0.006	0.007
			d_R	1.000	1.010	1.029	1.055	1.087	1.161	1.338	1.574	1.792
0.45	1	0	Mean	0.946	0.946	0.946	0.946	0.947	0.949	0.953	0.957	0.960
			SD	0.059	0.059	0.059	0.060	0.060	0.062	0.067	0.073	0.079
			MSE	0.006	0.006	0.006	0.006	0.006	0.007	0.007	0.007	0.008
			d_R	1.000	0.998	0.998	1.000	1.004	1.015	1.053	1.125	1.212
	1	1	Mean	1.001	1.001	1.001	1.001	1.001	1.003	1.005	1.008	1.009
			SD	0.061	0.061	0.061	0.061	0.062	0.064	0.069	0.075	0.081
			MSE	0.004	0.004	0.004	0.004	0.004	0.004	0.005	0.006	0.007
			d_R	1.000	0.991	0.998	1.016	1.042	1.113	1.295	1.542	1.768
	2	0	Mean	0.946	0.945	0.946	0.946	0.947	0.948	0.952	0.956	0.959
			SD	0.059	0.059	0.059	0.060	0.060	0.062	0.067	0.073	0.079
			MSE	0.006	0.006	0.006	0.006	0.006	0.007	0.007	0.007	0.008
			d_R	1.000	0.999	1.000	1.004	1.008	1.020	1.060	1.134	1.220
	2	1	Mean	1.001	1.001	1.001	1.001	1.001	1.002	1.004	1.007	1.008
			SD	0.061	0.061	0.061	0.062	0.062	0.064	0.069	0.075	0.081
			MSE	0.004	0.004	0.004	0.004	0.004	0.004	0.005	0.006	0.007
			d_R	1.000	0.991	0.998	1.015	1.042	1.112	1.292	1.537	1.762

However, as mentioned in Remark 5, the MDPDE for θ in (2.6) could be a biased estimator. Thus, one should be careful when using the MDPDE for estimating the drift parameter.

4. Conclusion

This paper presents a robust estimator for the dispersion parameter in discretely observed diffusion processes. To construct the contrast function, we ap-

Table 2. Estimation results when data are contaminated ($p = 0.01, h_n = 1,000^{-0.75}$).

σ_V^2	k_0	k_1		MDPDE								
				QMLE	0.05	0.1	0.15	0.2	0.3	0.5	0.75	1.0
0.5	1	0	Mean	1.900	1.218	1.108	1.069	1.050	1.034	1.026	1.028	1.032
			SD	0.539	0.119	0.081	0.072	0.069	0.069	0.072	0.078	0.084
			MSE	1.101	0.062	0.018	0.010	0.007	0.006	0.006	0.007	0.008
			d_R	1.000	0.056	0.017	0.009	0.007	0.005	0.005	0.006	0.007
	2	1	Mean	1.914	1.226	1.115	1.076	1.057	1.041	1.034	1.035	1.039
			SD	0.544	0.119	0.081	0.072	0.070	0.069	0.072	0.078	0.084
			MSE	1.132	0.066	0.020	0.011	0.008	0.006	0.006	0.007	0.009
			d_R	1.000	0.058	0.018	0.010	0.007	0.006	0.006	0.007	0.008
1.0	1	0	Mean	2.868	1.225	1.095	1.058	1.042	1.028	1.023	1.027	1.032
			SD	1.242	0.123	0.080	0.073	0.071	0.072	0.077	0.083	0.089
			MSE	5.032	0.066	0.015	0.009	0.007	0.006	0.006	0.008	0.009
			d_R	1.000	0.013	0.003	0.002	0.001	0.001	0.001	0.002	0.002
	2	1	Mean	2.889	1.233	1.103	1.065	1.049	1.036	1.030	1.033	1.039
			SD	1.246	0.123	0.080	0.073	0.071	0.072	0.077	0.083	0.089
			MSE	5.121	0.070	0.017	0.010	0.007	0.006	0.007	0.008	0.009
			d_R	1.000	0.014	0.003	0.002	0.001	0.001	0.001	0.002	0.002

Table 3. Estimation results when data are contaminated ($p = 0.03, h_n = 1,000^{-0.75}$).

σ_V^2	k_0	k_1		MDPDE								
				QMLE	0.05	0.1	0.15	0.2	0.3	0.5	0.75	1.0
0.5	1	0	Mean	3.658	1.840	1.411	1.257	1.186	1.126	1.096	1.097	1.108
			SD	1.033	0.287	0.153	0.110	0.094	0.084	0.083	0.088	0.094
			MSE	8.132	0.788	0.192	0.078	0.044	0.023	0.016	0.017	0.021
			d_R	1.000	0.097	0.024	0.010	0.005	0.003	0.002	0.002	0.003
	2	1	Mean	3.686	1.853	1.420	1.265	1.194	1.134	1.103	1.104	1.115
			SD	1.033	0.288	0.153	0.110	0.094	0.084	0.083	0.088	0.094
			MSE	8.281	0.811	0.200	0.082	0.047	0.025	0.018	0.019	0.022
			d_R	1.000	0.098	0.024	0.010	0.006	0.003	0.002	0.002	0.003
1.0	1	0	Mean	6.625	1.945	1.361	1.208	1.147	1.100	1.082	1.091	1.107
			SD	2.324	0.354	0.148	0.102	0.087	0.078	0.079	0.084	0.090
			MSE	37.04	1.019	0.152	0.054	0.029	0.016	0.013	0.015	0.020
			d_R	1.000	0.028	0.004	0.001	0.001	0.000	0.000	0.000	0.001
	2	1	Mean	6.707	1.960	1.370	1.216	1.155	1.107	1.090	1.098	1.114
			SD	2.370	0.357	0.148	0.102	0.087	0.079	0.079	0.084	0.090
			MSE	38.19	1.049	0.159	0.057	0.032	0.018	0.014	0.017	0.021
			d_R	1.000	0.027	0.004	0.001	0.001	0.000	0.000	0.000	0.001

proximate the transition density of the diffusion process to the Gaussian density by using the approach of Kessler (1997) and then adopt the MDPD estimation

Table 4. Estimation results when data are contaminated ($p = 0.03, h_n = 1,000^{-0.45}$).

σ_V^2	k_0	k_1		*MDPDE*								
				QMLE	0.05	0.1	0.15	0.2	0.3	0.5	0.75	1.0
0.5	1	0	Mean	1.270	1.187	1.141	1.111	1.091	1.067	1.047	1.042	1.044
			SD	0.125	0.095	0.084	0.079	0.077	0.075	0.077	0.083	0.088
			MSE	0.088	0.044	0.027	0.019	0.014	0.010	0.008	0.009	0.010
			d_R	1.000	0.498	0.305	0.211	0.161	0.114	0.093	0.098	0.111
	2	1	Mean	1.343	1.255	1.206	1.174	1.152	1.126	1.102	1.095	1.095
			SD	0.132	0.100	0.088	0.083	0.080	0.078	0.080	0.085	0.090
			MSE	0.135	0.075	0.050	0.037	0.030	0.022	0.017	0.016	0.017
			d_R	1.000	0.555	0.370	0.274	0.219	0.161	0.125	0.120	0.127
1.0	1	0	Mean	1.584	1.326	1.216	1.157	1.121	1.081	1.053	1.049	1.053
			SD	0.224	0.126	0.099	0.088	0.082	0.078	0.079	0.084	0.090
			MSE	0.391	0.122	0.057	0.032	0.021	0.013	0.009	0.009	0.011
			d_R	1.000	0.312	0.145	0.083	0.055	0.033	0.023	0.024	0.028
	2	1	Mean	1.679	1.403	1.286	1.222	1.183	1.140	1.109	1.102	1.104
			SD	0.238	0.133	0.104	0.092	0.086	0.081	0.082	0.086	0.092
			MSE	0.517	0.180	0.092	0.058	0.041	0.026	0.019	0.018	0.019
			d_R	1.000	0.348	0.179	0.112	0.079	0.051	0.036	0.034	0.037

Table 5. Estimation results when data are contaminated ($p = 0.05, h_n = 1,000^{-0.45}$).

σ_V^2	k_0	k_1		*MDPDE*								
				QMLE	0.05	0.1	0.15	0.2	0.3	0.5	0.75	1.0
0.5	1	0	Mean	1.496	1.363	1.285	1.234	1.199	1.154	1.116	1.103	1.104
			SD	0.161	0.119	0.102	0.093	0.088	0.083	0.083	0.086	0.091
			MSE	0.272	0.146	0.092	0.063	0.047	0.031	0.020	0.018	0.019
			d_R	1.000	0.535	0.336	0.233	0.173	0.113	0.074	0.066	0.070
	2	1	Mean	1.583	1.442	1.358	1.304	1.266	1.218	1.174	1.158	1.157
			SD	0.169	0.125	0.106	0.096	0.091	0.086	0.085	0.088	0.093
			MSE	0.369	0.210	0.140	0.102	0.079	0.055	0.038	0.033	0.033
			d_R	1.000	0.570	0.379	0.275	0.214	0.148	0.102	0.089	0.090
1.0	1	0	Mean	2.039	1.616	1.428	1.323	1.258	1.186	1.132	1.119	1.124
			SD	0.302	0.167	0.127	0.108	0.098	0.089	0.086	0.089	0.094
			MSE	1.171	0.407	0.199	0.116	0.076	0.042	0.025	0.022	0.024
			d_R	1.000	0.348	0.170	0.099	0.065	0.036	0.021	0.019	0.021
	2	1	Mean	2.166	1.714	1.511	1.398	1.327	1.249	1.190	1.174	1.177
			SD	0.322	0.177	0.133	0.112	0.102	0.092	0.089	0.091	0.096
			MSE	1.463	0.541	0.279	0.171	0.117	0.071	0.044	0.039	0.040
			d_R	1.000	0.370	0.191	0.117	0.080	0.048	0.030	0.026	0.028

method of Basu et al. (1998). According to simulations, the proposed estimator possesses strong robustness against outliers while still having a high efficiency

Table 6. Mean, SD, MSE, and d_R ($p = 0.03, h_n = 1,000^{-0.75}$).

σ_V^2			*MDPDE*								
			QMLE	0.05	0.1	0.15	0.2	0.3	0.5	0.75	1.0
0	θ	Mean	3.266	3.204	3.148	3.097	3.050	2.969	2.842	2.731	2.655
		SD	0.966	0.955	0.948	0.945	0.943	0.948	0.973	1.025	1.088
		MSE	1.004	0.954	0.921	0.902	0.893	0.899	0.972	1.122	1.302
		d_R	1.000	0.950	0.918	0.898	0.889	0.896	0.969	1.118	1.297
	σ	Mean	0.982	0.983	0.983	0.983	0.983	0.984	0.984	0.984	0.984
		SD	0.051	0.052	0.052	0.053	0.053	0.055	0.060	0.066	0.070
		MSE	0.003	0.003	0.003	0.003	0.003	0.003	0.004	0.005	0.005
		d_R	1.000	1.001	1.013	1.033	1.061	1.131	1.306	1.539	1.748
0.5	θ	Mean	12.62	4.922	3.908	3.561	3.381	3.182	2.979	2.842	2.760
		SD	5.742	1.675	1.291	1.178	1.124	1.076	1.058	1.085	1.134
		MSE	125.6	6.500	2.491	1.702	1.410	1.192	1.120	1.202	1.343
		d_R	1.000	0.052	0.020	0.014	0.011	0.009	0.009	0.010	0.011
	σ	Mean	2.868	1.653	1.329	1.207	1.150	1.099	1.072	1.071	1.079
		SD	0.641	0.210	0.117	0.087	0.076	0.069	0.068	0.073	0.078
		MSE	3.900	0.470	0.122	0.051	0.028	0.015	0.010	0.010	0.012
		d_R	1.000	0.121	0.031	0.013	0.007	0.004	0.003	0.003	0.003
1.0		Mean	28.37	4.984	3.791	3.473	3.316	3.141	2.958	2.830	2.751
		SD	12.002	1.728	1.191	1.086	1.045	1.015	1.015	1.052	1.104
		MSE	787.9	6.923	2.045	1.402	1.191	1.049	1.032	1.135	1.281
		d_R	1.000	0.009	0.003	0.002	0.002	0.001	0.001	0.001	0.002
	σ	Mean	4.968	1.724	1.288	1.168	1.118	1.077	1.060	1.065	1.076
		SD	1.413	0.265	0.117	0.084	0.073	0.066	0.067	0.072	0.077
		MSE	17.74	0.594	0.097	0.035	0.019	0.010	0.008	0.009	0.012
		d_R	1.000	0.033	0.005	0.002	0.001	0.001	0.000	0.001	0.001

Table 7. Estimation results when data are not contaminated ($h_n = n^{-0.75}$).

n			*MDPDE*								
			QMLE	0.05	0.1	0.15	0.2	0.3	0.5	0.75	1.0
5,000	θ	Mean	3.194	3.134	3.080	3.031	2.986	2.909	2.789	2.686	2.615
		SD	0.781	0.773	0.768	0.765	0.764	0.767	0.787	0.825	0.872
		MSE	0.647	0.615	0.596	0.586	0.584	0.597	0.663	0.779	0.909
	σ	Mean	0.994	0.994	0.994	0.994	0.994	0.994	0.994	0.995	0.995
		SD	0.024	0.024	0.024	0.024	0.025	0.026	0.028	0.031	0.033
		MSE	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
10,000	θ	Mean	3.160	3.100	3.047	2.998	2.954	2.878	2.762	2.660	2.591
		SD	0.703	0.696	0.692	0.690	0.689	0.694	0.716	0.758	0.810
		MSE	0.520	0.495	0.481	0.476	0.477	0.496	0.569	0.690	0.823
	σ	Mean	0.997	0.997	0.997	0.997	0.997	0.997	0.997	0.997	0.997
		SD	0.017	0.017	0.017	0.017	0.018	0.018	0.020	0.021	0.023
		MSE	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001

relative to the QML estimator. In particular, the QML estimator is observed to be severely damaged by outliers in a high-frequency sampling case, whereas our estimator shows much better performance. This indicates that in the presence of outliers or jumps, particular attention should be paid when estimating diffusion models without jump components. In such situations, the MDPD estimator can be a good alternative.

Although we deal with the univariate diffusion processes, the estimation method can be generalized to the multivariate cases. Other approximation techniques, such as Aït-Sahalia (2002), can be employed in the construction of the MDPD estimator. In this case, the estimator is anticipated to be more efficient when data are observed at a low frequency or a sampling interval is fixed. Robust estimation of the drift parameter is also of great interest because this parameter plays its own role, for example, in the calculation of the value-at-risk (VaR). Although the MDPD estimator for the drift parameter has been proposed, we cannot guarantee its consistency except for some special cases. It is considered that some bias reduction techniques could be applied. We leave these issues as possible topics of future research.

Supplementary Materials

The online supplementary material contains some technical lemmas and the proofs of Theorems 1, 2, and 3.

Acknowledgements

The author would like to thank the referees and an associate editor for carefully examining the paper and providing valuable comments to improve its quality. This work was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT and Future Planning (NRF-2013R1A1A1008032).

References

Aït-Sahalia, Y. (2002). Maximum-likelihood estimation of discretely-sampled diffusions: a closed-form approximation approach. *Econometrica* **70**, 223-262.

Aït-Sahalia, Y. (2008). Closed-form likelihood expansions for multivariate diffusions. *Ann. Statist.* **36**, 906-937.

Basu, A., Harris, I. R., Hjort, N. L. and Jones, M. C. (1998). Robust and efficient estimation by minimizing a density power divergence. *Biometrika* **85**, 549-559.

Beskos, A., Papaspiliopoulos, O. and Roberts, G. (2009). Monte Carlo maximum likelihood estimation for discretely observed diffusion processes. *Ann. Statist.* **37**, 223-245.

Cichocki, A. and Amari, S. (2010). Families of alpha- beta- and gamma- divergences: Flexible and robust measures of similarities. *Entropy* **12**, 1532-1568.

Dacunha-Castelle, D. and Florens-Zmirou, D. (1986). Estimation of the coefficient of a diffusion from discrete observations. *Stochastics* **19**, 263-284.

Durham, G. and Gallant, R. (2002). Numerical techniques for maximum likelihood estimation of continuous time diffusion processes. *J. Busi. Econom. Statist.* **20**, 297-316.

Durio, A. and Isaia, E. D. (2011). The minimum density power divergence approach in building robust regression models. *Informatica* **2**, 43-56.

Friedman, A. (1975). *Stochastic Differential Equations and Applications*. Academic Press, New York.

Fujisawa, H. and Eguchi, S. (2006). Robust estimation in the normal mixture model. *J. Statist. Plann. Inference* **136**, 3989- 4011.

Juárez, S. F. and Schucany, W. R. (2004). Robust and efficient estimation for the generalized pareto distribution. *Extremes* **7**, 237-251.

Kessler, M. (1997). Estimation of an ergodic diffusion from discrete observations. *Scand. J. Statist.* **24**, 211-229.

Kim, B. and Lee, S. (2013). Robust estimation for copula parameter in SCOMDY models. *J. Time Ser. Anal.* **34**, 302-314.

Kleppe, T. S., Yu, J. and Skaug, H. J. (2014). Maximum likelihood estimation of partially observed diffusion models. *J. Econom.* **180**, 73-80.

Kloeden, P. E. and Platen, E. (1999). *Numerical Solution of Stochastic Differential Equations*. Springer-Verlag, New York.

Kutoyants, Y. (2004). *Statistical Inference for Ergodic Diffusion Processes*. Springer, New York.

Liesenfeld, R. and Richard, J.-F. (2006). Classical and Bayesian analysis of univariate and multivariate stochastic volatility models. *Econom. Rev.* **25**, 335-360.

Lee, S. and Song, J. (2013). Minimum density power divergence estimator for diffusion processes. *Ann. Inst. Statist. Math.* **65**, 213-236.

Pardo, L. (2006). *Statistical Inference Based on Divergence Measures*, Chapman and Hall/CRC.

Pedersen, A. (1995). A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations. *Scand. J. Statist.* **22**, 55-71.

Phillips, P. C. B. and Yu, J. (2009). Maximum likelihood and Gaussian estimation of continuous time models in finance. *Handbook of Financial Time Series*. Springer, 497-530.

Prakasa Rao, B. L. S. (1999). *Statistical Inference for Diffusion Type Processes*. Arnold, London.

Richard, J.-F. and Zhang, W. (2007). Efficient high-dimensional importance sampling. *J. Econom.* **127**, 1385-1411.

Shephard, N. and Pitt, M. K. (1997). Likelihood analysis of non-Gaussian measurement time series. *Biometrika* **84**, 653-667.

Song, J., Lee, S., Na, O., and Kim, H. (2007). Minimum density power divergence estimator for diffusion parameter in discretely observed diffusion processes. *Korean Commun. Statist.* **14**, 267-280.

Warwick, J. (2005). A data-based method for selecting tuning parameters in minimum distance estimators. *Comput. Statist. Data Anal.* **48**, 571-585.

Yoshida, N. (1992). Estimation for diffusion processes from discrete observations. *J. Multivariate Anal.* **41**, 220-242.

Department of Computer Science and Statistics, Jeju National University, Jeju, 63243, Korea.

E-mail: jmsong@jejunu.ac.kr

(Received May 2014; accepted January 2016)