The Korean Communications in Statistics **Vol. 14** No. 2, 2007, pp. 267–280

Minimum Density Power Divergence Estimator for Diffusion Parameter in Discretely Observed Diffusion Processes*

Junmo Song[1], Sangyeol Lee[2], Okyoung Na[3] and Hyojung Kim[4]

Abstract

In this paper, we consider the robust estimation for diffusion processes when the sample is observed discretely. As a robust estimator, we consider the minimizing density power divergence estimator (MDPDE) proposed by Basu *et al.* (1998). It is shown that the MDPDE for diffusion process is weakly consistent. A simulation study demonstrates the robustness of the MDPDE.

Keywords: ergodic diffusion processes; Brownian motion; minimum density power divergence estimator; robustness.

1. Introduction

The diffusion process has long been popular in modeling stochastic phenomena occurring in various fields such as finance, engineering, physical and medical sciences. The statistical inference for diffusion processes is either on the basis of a continuous sample during a specific period, or a sample discretely observed from diffusion processes. Kutoyants (2004) is a representative reference for the former case. With regard to the latter case, we refer to Dacunha-Castelle and Florens-Zmirou (1986), Florens-Zmirou (1989), Yoshida (1992), Kessler (1997, 2000), Kessler and Sørensen (1999). Some basic results are well summarized in Prakasa Rao (1999).

---

* This research was supported by Korean Research Foundation Grant (2003-070-C00008).

1) Graduate student, Department of Statistics, Seoul National University, Seoul 151-742, Korea.

2) Professor, Department of Statistics, Seoul National University, Seoul 151-742, Korea.

3) Post doctor, Department of Statistics, Seoul National University, Seoul 151-742, Korea. Correspondence : okyoung0@hanmail.net

4) Post doctor, Department of Statistics, Seoul National University, Seoul 151-742, Korea.

In various statistical parametric models, it is well known that the estimators based on Gaussian likelihood are severely influenced by outliers or extreme values. Naturally, one can conjecture that the similar situations happen in the estimation procedure for diffusion processes. Actually, according to our simulation study, Kessler's (1997) estimator based on the Gaussian approximation method is damaged by outliers. The purpose of this paper is to designate a robust estimator diffusion processes. For constructing a robust estimator, we adopt the idea of Basu *et al.* (1998) (BHHJ for abbreviation) which introduces an estimation procedure to minimize a density-based divergence measures, called density power divergences. Compared to other existing density-based divergence methods, such as Beran (1977), Tamura and Boos (1986) and Simpson (1987), which use the Hellinger distance, and Basu and Lindsay (1994) and Cao *et al.* (1995), their method is known to have merit of not requiring any smoothing methods. Moreover, BHHJ demonstrated that the minimum density power divergence estimator (MDPDE for abbreviation) possesses strong robust properties with little loss in asymptotic efficiency relative to the maximum likelihood estimator (MLE). Therefore, it can be viewed as a good alternative in terms of efficiency and robustness.

The organization of this paper is as follows. In Section 2, we introduce the construction of the robust estimator using the BHHJ's procedure and present the weak consistency of the proposed estimator. The result is an extension of the weak consistency result in Lee and Song (2006). In Section 3, we perform a simulation study and compare the proposed estimator with Kessler's estimator. The proof for the main result in Section 2 is provided in Section 4.

2. Main Result

For a family of parametric distributions $\{F_\theta : \theta \in \Theta \subset \mathbb{R}^m\}$ possessing densities $\{f_\theta\}$ and for a distribution G with density g, BHHJ defined the minimum density power divergence functional $T_\alpha(\cdot)$ by

$$d_\alpha\left(g, f_{T_\alpha(g)}\right) = \min_{\theta\in\Theta} d_\alpha(g, f_\theta),$$

where

$$d_\alpha(g,f) := \begin{cases} \displaystyle\int\left\{f^{1+\alpha}(z) - \left(1+\frac{1}{\alpha}\right)g(z)f^\alpha(z) + \frac{1}{\alpha}g^{1+\alpha}(z)\right\}dz & , \alpha > 0, \\ \displaystyle\int g(z)\left(\log g(z) - \log f(z)\right)dz & , \alpha = 0. \end{cases} \tag{2.1}$$

Given random sample $X_1, \ldots, X_n$ with unknown density g, the MDPDE is defined by

$$\hat{\theta}_{\alpha,n} = \operatorname*{argmin}_{\theta \in \Theta} D_{\alpha,n}(\theta), \tag{2.2}$$

where $D_{\alpha,n}(\theta) = n^{-1} \sum_{t=1}^{n} V_\alpha(\theta; X_t)$ and

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

BHHJ showed that $\hat{\theta}_{\alpha,n}$ is weakly consistent for $T_\alpha(\cdot)$ and asymptotically normal and demonstrated that the estimator has strong robust properties against outliers.

This approach can be extended to regression models. Let $\{f_\theta(y|x)\}$ be a parametric family of regression models indexed by the parameter $\theta \in \Theta \subset \mathbb{R}^m$ and let $g(y|x)$ be the true density for Y given $X = x$. Substituting f_θ and g in (2.1) by $f_\theta(\cdot|x)$ and $g(\cdot|x)$ respectively, a family of the x-conditional versions of density power divergences is obtained as follows:

$$d_\alpha\left(g(\cdot|x), f_\theta(\cdot|x)\right) := \begin{cases} \int \left\{ f_\theta^{1+\alpha}(y|x) - \left(1 + \frac{1}{\alpha}\right) g(z) f_\theta^\alpha(y|x) + \frac{1}{\alpha} g^{1+\alpha}(y|x) \right\} dy, & \alpha > 0, \\ \int g(y|x) \left(\log g(y|x) - \log f_\theta(y|x)\right) dy, & \alpha = 0. \end{cases}$$

Similarly to (2.2), given data pairs $\{(X_t, Y_t) : 1 \le t \le n\}$, the MDPDE for regression model is defined by

$$\hat{\theta}_{\alpha,n} = \operatorname*{argmin}_{\theta \in \Theta} \begin{cases} \sum_{t=1}^{n} \int f_\theta^{1+\alpha}(y|X_t)dz - \left(1 + \frac{1}{\alpha}\right) \sum_{t=1}^{n} f_\theta^\alpha(Y_t|X_t), & \alpha > 0, \\ -\sum_{t=1}^{n} \log f_\theta(Y_t|X_t), & \alpha = 0. \end{cases} \tag{2.3}$$

Now, we apply this method to the parameter estimation for diffusion processes. Let us consider the stochastic differential equation

$$dX_t = a(X_t, \sigma_0)dW_t + b(X_t)dt, \quad X_0 = x_0, \ t \ge 0, \tag{2.4}$$

where $\sigma_0 \in \Theta$, a compact subset of $\mathbb{R}^p$, a and b are real valued functions and $\{W_t : t \geq 0\}$ is a standard Brownian motion. It is assumed that a and b are smooth enough to ensure the uniqueness in law of the solution to (2.4). We denote this law by P_0.

Suppose that $X_{t_i^n}, 1 \leq i \leq n$, are discretely observed from (2.4), where $t_i^n = ih_n$, $h_n \to 0$ and $nh_n \to \infty$. From (2.4), we have that

$$X_{t_i^n} = X_{t_{i-1}^n} + a(X_{t_{i-1}^n}, \sigma_0) Z_{n,i} \sqrt{h_n} + b(X_{t_{i-1}^n}) h_n + \Delta_{n,i},$$

where $Z_{n,i} = (W_{t_i^n} - W_{t_{i-1}^n})/\sqrt{h_n}$ and

$$\Delta_{n,i} = \int_{t_{i-1}^n}^{t_i^n} \left\{ a(X_s, \sigma_0) - a(X_{t_{i-1}^n}, \sigma_0) \right\} dW_s + \int_{t_{i-1}^n}^{t_i^n} \left\{ b(X_s) - b(X_{t_{i-1}^n}) \right\} ds.$$

Note that $Z_{n,1}, \ldots, Z_{n,n}$ are *i.i.d.* $N(0,1)$ and $|\Delta_{n,i}| = O_{P_0}(h_n)$ (*cf.* Lemma 4.1). Ignoring $\Delta_{n,i}$, we can see that for large n, $X_{t_i^n} | G_{i-1}^n, 1 \leq i \leq n$, behave like independent r.v.'s following $N(X_{t_{i-1}^n} + b(X_{t_{i-1}^n}) h_n, a(X_{t_{i-1}^n}, \sigma_0)^2 h_n)$, where G_i^n denotes the sigma field generated by $\{W_s : s \leq t_i^n\}$. Hence, viewing the observations as regression data pairs $\{(X_{t_{i-1}^n}, X_{t_i^n}) : 1 \leq i \leq n\}$ and applying (2.3) to them (in this case, the family of parametric distributions is the normal distributions), we can define the MDPDE for (2.4) as

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

where

$$V_{\alpha,n,i}(\sigma) = \begin{cases} \left(\frac{1}{a(X_{t_{i-1}^n}, \sigma)^2} \right)^{\frac{\alpha}{2}} \left[\frac{1}{\sqrt{1+\alpha}} - \left(1 + \frac{1}{\alpha}\right) \exp\left\{ -\frac{\alpha}{2} \frac{(X_{t_i^n} - X_{t_{i-1}^n} - b(X_{t_{i-1}^n}) h_n)^2}{a(X_{t_{i-1}^n}, \sigma)^2 h_n} \right\} \right] & , \alpha > 0, \\ \frac{(X_{t_i^n} - X_{t_{i-1}^n} - b(X_{t_{i-1}^n}) h_n)^2}{a(X_{t_{i-1}^n}, \sigma)^2 h_n} + \log a(X_{t_{i-1}^n}, \sigma)^2 & , \alpha = 0. \end{cases}$$

Remark 2.1 The contrast function, $l_{p,n}(\sigma)$, of Kessler (1997, p. 227) asymptotically coincides with $\sum_{i=1}^n V_{0,n,i}(\sigma)$ since by Lemmas 8–10 of Kessler (1997),

$$\sup_{\sigma \in \Theta} \left| \frac{1}{n} l_{p,n}(\sigma) - \frac{1}{n} \sum_{i=1}^n V_{0,n,i}(\sigma) \right| \leq \frac{h_n}{n} \sum_{i=1}^n C \left(1 + |X_{t_{i-1}^n}|\right)^C$$

$$
\begin{aligned}
&+\frac{1}{n}\sum_{i=1}^{n} C\left(1+|X_{t_{i-1}^n}|\right)^C \left\{X_{t_i^n} - X_{t_{i-1}^n} - b(X_{t_{i-1}^n})h_n\right\} \\
&+\frac{1}{n}\sum_{i=1}^{n} C\left(1+|X_{t_{i-1}^n}|\right)^C \left\{X_{t_i^n} - r_{k_0}(h_n, X_{t_{i-1}^n}), \sigma)\right\}^2 \\
&= O_{P_0}(h_n),
\end{aligned}
$$

where k_0 and r_{k_0} are the those given in (3.5) of Kessler (1997).

Below we establish weak consistency of the MDPDE in (2.5). For this task, we set $\mathcal{F} = \{f(x,\sigma) : |f| \le C(1+|x|)^C \text{ for some } C\}$, where C does not depend on σ, and assume the conditions as follows:

(A1) There exists a constant C_1 such that for any x and y,

$$|a(x,\sigma_0) - a(y,\sigma_0)| + |b(x) - b(y)| \le C_1|x-y|.$$

(A2) The process X from (2.4) is ergodic with its invariant measure μ_0 such that $\int x^k d\mu_0(x) < \infty$, for all $k \ge 0$.

(A3) $\sup_t E_0|X_t|^k < \infty$, for all $k \ge 0$.

(A4) The function a is differentiable with respect to x and σ; a and the derivatives belong to $\mathcal{F}$.

(A5) $\inf_{x,\sigma} a(x,\sigma)^2 > 0$.

(A6) If $\mu_0\{x : a(x,\sigma)^2 = a(x,\sigma_0)^2\} = 1$, then $\sigma = \sigma_0$.

Here is the main result of this paper.

Theorem 2.1 *Assume that (A1)–(A6) hold. For each $\alpha \ge 0$, if $h_n \to 0$, $nh_n \to \infty$ and $nh_n^q \to 0$ for some $q > 1$, then $\hat{\sigma}_{\alpha,n}$ converges weakly to σ_0.*

3. Simulation Study

In this section, we compare the performance of the MDPDE and Kessler's estimator (KE) for the diffusion process:

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

because the KE is a typical estimator based on the Gaussian approximation method. In our simulation, the case $\sigma_0 = 1$ is considered and the path of X is generated via the Euler scheme with generating interval 10^{-4}. The minimizer is obtained on the parameter space $\Theta = [0, 20]$. The sample $\{X_{o,t_i^n}\}_{i=1}^n$ is observed discretely with sampling interval $h_n = n^{-0.75}$. Comparison is based on the mean squared error (MSE) and

$$d_R := \frac{\text{MSE of MDPDE}}{\text{MSE of KE}}.$$

First, we handle the case that the observations are not contaminated by outliers. Based on 1000 repetitions, the mean, standard deviation (SD), MSE of the estimates and d_R are calculated for $n = 500, 800, 1000$. The results which are presented in Table 3.1 show that the KE outperforms the MDPDE and the MDPDE with α close to 0 performs similarly to the KE. It is also seen that the performance of the MDPDE with α not close to 0, say $\alpha > 0.1$, is not bad.

Next, we examine the case that outliers are involved in the data. Here, we consider the situation that the sample $\{X_{o,t_i^n}\}_{i=1}^n$ from (3.1) is contaminated by the outliers $X_{c,t_i^n} \overset{i.i.d.}{\sim} N(0, \sigma_V^2)$ and the observed r.v.'s follow the scheme $X_{t_i^n} = X_{o,t_i^n} + p_i X_{c,t_i^n}$, where p_i are *i.i.d.* Bernoulli r.v.'s with success probability p. It is assumed that $\{p_i\}$, $\{X_{o,t_i^n}\}$ and $\{X_{c,t_i^n}\}$ are all independent. The mean, SD, MSE of the estimates and d_R based on $\{X_{t_i^n}\}$ are calculated out of 1000 repetitions for $n = 1000$.

Table 3.1: The Mean, SD, MSE and d_R without outliers

		n=500				n=800				n=1000			
		MEAN	SD	MSE	d_R	MEAN	SD	MSE	d_R	MEAN	SD	MSE	d_R
KE		0.9945	0.0849	0.0072	1.0000	0.9941	0.0695	0.0049	1.0000	0.9993	0.0618	0.0038	1.0000
	0	**0.9825**	**0.0844**	**0.0074**	**1.0254**	**0.9856**	**0.0692**	**0.0050**	**1.0268**	**0.9922**	**0.0616**	**0.0039**	**1.0086**
	0.01	0.9826	0.0844	0.0074	1.0258	0.9856	0.0692	0.0050	1.0271	0.9922	0.0616	0.0039	1.0092
	0.02	0.9826	0.0844	0.0074	1.0268	0.9856	0.0693	0.0050	1.0280	0.9922	0.0617	0.0039	1.0105
	0.03	0.9827	0.0845	0.0074	1.0284	0.9856	0.0693	0.0050	1.0295	0.9922	0.0617	0.0039	1.0122
	0.04	0.9828	0.0846	0.0075	1.0305	0.9857	0.0694	0.0050	1.0314	0.9922	0.0618	0.0039	1.0146
	0.05	0.9828	0.0847	0.0075	1.0332	0.9857	0.0695	0.0050	1.0338	0.9922	0.0619	0.0039	1.0174
	0.06	0.9829	0.0849	0.0075	1.0363	0.9857	0.0696	0.0050	1.0367	0.9922	0.0620	0.0039	1.0207
M	0.07	0.9829	0.0851	0.0075	1.0399	0.9857	0.0697	0.0051	1.0399	0.9922	0.0621	0.0039	1.0244
D	0.08	0.9830	0.0852	0.0076	1.0440	0.9857	0.0698	0.0051	1.0436	0.9922	0.0622	0.0039	1.0286
P	0.09	0.9831	0.0854	0.0076	1.0484	0.9857	0.0700	0.0051	1.0476	0.9923	0.0624	0.0040	1.0331
D	0.1	0.9831	0.0857	0.0076	1.0533	0.9858	0.0701	0.0051	1.0520	0.9923	0.0625	0.0040	1.0380
E	0.15	0.9835	0.0870	0.0078	1.0828	0.9859	0.0711	0.0053	1.0786	0.9923	0.0635	0.0041	1.0677
	0.2	0.9838	0.0886	0.0081	1.1199	0.9860	0.0722	0.0054	1.1118	0.9925	0.0646	0.0042	1.1044
	0.3	0.9845	0.0923	0.0088	1.2109	0.9863	0.0750	0.0058	1.1934	0.9927	0.0672	0.0046	1.1933
	0.4	0.9851	0.0965	0.0095	1.3171	0.9865	0.0781	0.0063	1.2891	0.9930	0.0701	0.0050	1.2963
	0.5	0.9857	0.1008	0.0104	1.4317	0.9868	0.0813	0.0068	1.3938	0.9933	0.0731	0.0054	1.4073
	0.6	0.9863	0.1050	0.0112	1.5496	0.9870	0.0845	0.0073	1.5033	0.9936	0.0760	0.0058	1.5217
	0.8	0.9873	0.1128	0.0129	1.7807	0.9873	0.0907	0.0084	1.7241	0.9942	0.0815	0.0067	1.7471
	1.0	0.9880	0.1195	0.0144	1.9914	0.9874	0.0962	0.0094	1.9333	0.9947	0.0863	0.0075	1.9546

Table 3.2: The Mean, SD, MSE and d_R with 0.1% outliers

		$\sigma_V^2 = 0.5$				$\sigma_V^2 = 1$				$\sigma_V^2 = 2$			
		MEAN	SD	MSE	d_R	MEAN	SD	MSE	d_R	MEAN	SD	MSE	d_R
KE		1.0940	0.1729	0.0388	1.0000	1.1899	0.3369	0.1495	1.0000	1.3863	0.7804	0.7583	1.0000
	0	1.0863	0.1720	0.0370	0.9553	1.1812	0.3341	0.1445	0.9660	1.3759	0.7781	0.7467	0.9848
	0.01	1.0518	0.1034	0.0134	0.3449	1.0746	0.1325	0.0231	0.1547	1.0823	0.1313	0.0240	0.0317
	0.02	1.0355	0.0831	0.0082	0.2109	1.0407	0.0923	0.0102	0.0680	1.0399	0.0863	0.0090	0.0119
	0.03	1.0264	0.0749	0.0063	0.1627	1.0250	0.0787	0.0068	0.0456	1.0235	0.0741	0.0060	0.0080
	0.04	1.0206	0.0708	0.0054	0.1402	1.0164	0.0728	0.0056	0.0372	1.0150	0.0694	0.0050	0.0067
	0.05	1.0166	0.0684	0.0050	0.1279	1.0110	0.0698	0.0050	0.0333	1.0098	0.0672	0.0046	0.0061
	0.06	1.0137	0.0670	0.0047	0.1206	1.0074	0.0680	0.0047	0.0313	1.0064	0.0660	0.0044	0.0058
M	0.07	1.0115	0.0661	0.0045	0.1160	1.0048	0.0669	0.0045	0.0301	1.0040	0.0653	0.0043	0.0057
D	0.08	1.0098	0.0654	0.0044	0.1130	1.0028	0.0662	0.0044	0.0294	1.0022	0.0649	0.0042	0.0056
P	0.09	1.0085	0.0650	0.0043	0.1110	1.0013	0.0658	0.0043	0.0289	1.0009	0.0647	0.0042	0.0055
D	0.1	1.0074	0.0648	0.0042	0.1097	1.0002	0.0655	0.0043	0.0287	0.9998	0.0645	0.0042	0.0055
E	**0.15**	**1.0041**	**0.0646**	**0.0042**	**0.1081**	**0.9969**	**0.0651**	**0.0043**	**0.0284**	**0.9967**	**0.0645**	**0.0042**	**0.0055**
	0.2	1.0026	0.0653	0.0043	0.1100	0.9955	0.0656	0.0043	0.0289	0.9954	0.0652	0.0043	0.0056
	0.3	1.0014	0.0674	0.0045	0.1173	0.9946	0.0675	0.0046	0.0307	0.9943	0.0671	0.0045	0.0060
	0.4	1.0011	0.0700	0.0049	0.1266	0.9946	0.0699	0.0049	0.0329	0.9940	0.0695	0.0049	0.0064
	0.5	1.0011	0.0728	0.0053	0.1367	0.9949	0.0725	0.0053	0.0353	0.9939	0.0721	0.0052	0.0069
	0.6	1.0013	0.0755	0.0057	0.1471	0.9954	0.0750	0.0056	0.0378	0.9941	0.0748	0.0056	0.0074
	0.8	1.0018	0.0805	0.0065	0.1673	0.9964	0.0797	0.0064	0.0426	0.9945	0.0798	0.0064	0.0084
	1.0	1.0023	0.0847	0.0072	0.1854	0.9973	0.0838	0.0070	0.0470	0.9950	0.0841	0.0071	0.0094

Table 3.3: The Mean, SD, MSE and d_R with 1% outliers

		$\sigma_V^2 = 0.5$				$\sigma_V^2 = 1$				$\sigma_V^2 = 2$			
		MEAN	SD	MSE	d_R	MEAN	SD	MSE	d_R	MEAN	SD	MSE	d_R
KE		1.8798	0.5554	1.0826	1.0000	2.8863	1.1982	4.9940	1.0000	5.5998	3.5021	33.423	1.0000
	0	1.8663	0.5519	1.0551	0.9746	2.8605	1.1849	4.8655	0.9743	5.5250	3.4307	32.246	0.9648
	0.01	1.5857	0.3330	0.4539	0.4193	1.9049	0.4717	1.0414	0.2085	2.2682	0.6855	2.0783	0.0622
	0.02	1.4276	0.2337	0.2374	0.2193	1.5621	0.3010	0.4066	0.0814	1.6287	0.3471	0.5157	0.0154
	0.03	1.3279	0.1791	0.1396	0.1289	1.3868	0.2133	0.1952	0.0391	1.3807	0.2149	0.1911	0.0057
	0.04	1.2610	0.1462	0.0895	0.0827	1.2846	0.1636	0.1078	0.0216	1.2603	0.1551	0.0918	0.0027
	0.05	1.2139	0.1250	0.0614	0.0567	1.2198	0.1338	0.0662	0.0133	1.1922	0.1238	0.0522	0.0016
	0.06	1.1795	0.1106	0.0445	0.0411	1.1761	0.1148	0.0442	0.0089	1.1495	0.1056	0.0335	0.0010
M	0.07	1.1536	0.1005	0.0337	0.0311	1.1451	0.1023	0.0315	0.0063	1.1207	0.0942	0.0234	0.0007
D	0.08	1.1334	0.0932	0.0265	0.0245	1.1222	0.0938	0.0237	0.0048	1.1003	0.0866	0.0176	0.0005
P	0.09	1.1175	0.0878	0.0215	0.0199	1.1048	0.0879	0.0187	0.0037	1.0851	0.0814	0.0139	0.0004
D	0.1	1.1046	0.0837	0.0179	0.0166	1.0912	0.0836	0.0153	0.0031	1.0736	0.0777	0.0115	0.0003
E	0.15	1.0660	0.0736	0.0098	0.0090	1.0532	0.0737	0.0083	0.0017	1.0426	0.0694	0.0066	0.0002
	0.2	1.0478	0.0704	0.0072	0.0067	1.0367	0.0708	0.0064	0.0013	1.0299	0.0675	0.0054	0.0002
	0.3	1.0321	0.0696	0.0059	0.0054	**1.0233**	**0.0705**	**0.0055**	**0.0011**	**1.0206**	**0.0682**	**0.0051**	**0.0002**
	0.4	**1.0265**	**0.0710**	**0.0057**	**0.0053**	1.0190	0.0723	0.0056	0.0011	1.0186	0.0707	0.0053	0.0002
	0.5	1.0247	0.0731	0.0059	0.0055	1.0180	0.0747	0.0059	0.0012	1.0192	0.0737	0.0058	0.0002
	0.6	1.0248	0.0754	0.0063	0.0058	1.0185	0.0773	0.0063	0.0013	1.0208	0.0769	0.0063	0.0002
	0.8	1.0273	0.0801	0.0072	0.0066	1.0215	0.0823	0.0072	0.0014	1.0253	0.0828	0.0075	0.0002
	1.0	1.0308	0.0845	0.0081	0.0075	1.0252	0.0866	0.0081	0.0016	1.0300	0.0880	0.0086	0.0003

Tables 3.2–3.4 summarize the results for $p = 0.001, 0.01, 0.05$ and $\sigma_V^2 = 0.5, 1, 2$. From these Tables, it can be seen that the values of MSE of KE and MDPDE with α close to 0 increase considerably as σ_V^2 increases. Meanwhile, MSE of MDPDE with $\alpha > 0.1$ does not increase. So, d_R decreases to 0 as σ_V^2 increases. The bold phased figures denote the α that gives a minimal MSE, i.e, a minimal d_R. It is observed that the producing minimal d_R increases as p increases

Table 3.4: The Mean, SD, MSE and d_R with 5% outliers

		$\sigma_V^2 = 0.5$				$\sigma_V^2 = 1$				$\sigma_V^2 = 2$			
		MEAN	SD	MSE	d_R	MEAN	SD	MSE	d_R	MEAN	SD	MSE	d_R
KE		5.2477	1.3401	19.839	1.0000	10.359	3.3003	98.475	1.0000	18.816	2.3562	322.96	1.0000
	0	5.2097	1.3588	19.568	0.9864	10.179	3.2152	94.594	0.9606	18.663	2.4759	318.12	0.9850
	0.01	4.3141	1.0206	12.025	0.6061	6.9600	1.8803	39.057	0.3966	11.932	3.3789	130.92	0.4054
	0.02	3.7174	0.8249	8.0645	0.4065	5.3883	1.3832	21.171	0.2150	7.9786	2.2855	53.925	0.1670
	0.03	3.2715	0.6848	5.6287	0.2837	4.3695	1.0883	12.538	0.1273	5.7388	1.6534	25.190	0.0780
	0.04	2.9205	0.5786	4.0232	0.2028	3.6327	0.8751	7.6972	0.0782	4.2505	1.2507	12.130	0.0376
	0.05	2.6371	0.4946	2.9247	0.1474	3.0752	0.7076	4.8070	0.0488	3.2448	0.9329	5.9095	0.0183
	0.06	2.4051	0.4260	2.1559	0.1087	2.6492	0.5723	3.0474	0.0309	2.5844	0.6734	2.9638	0.0092
M	0.07	2.2141	0.3687	1.6101	0.0812	2.3259	0.4639	1.9731	0.0200	2.1594	0.4809	1.5754	0.0049
D	0.08	2.0564	0.3208	1.2189	0.0614	2.0827	0.3789	1.3157	0.0134	1.8864	0.3528	0.9102	0.0028
P	0.09	1.9261	0.2808	0.9366	0.0472	1.8999	0.3132	0.9080	0.0092	1.7060	0.2724	0.5726	0.0018
D	0.1	1.8183	0.2478	0.7310	0.0368	1.7618	0.2636	0.6499	0.0066	1.5819	0.2212	0.3875	0.0012
E	0.15	1.4953	0.1542	0.2691	0.0136	1.4141	0.1482	0.1935	0.0020	1.3042	0.1260	0.1084	0.0003
	0.2	1.3507	0.1195	0.1372	0.0069	1.2847	0.1127	0.0938	0.0010	1.2105	0.1024	0.0548	0.0002
	0.3	1.2323	0.0970	0.0634	0.0032	1.1896	0.0924	0.0445	0.0005	1.1465	0.0896	0.0295	0.0001
	0.4	1.1895	0.0910	0.0442	0.0022	1.1600	0.0884	0.0334	0.0003	**1.1304**	**0.0873**	**0.0246**	**0.0001**
	0.5	1.1732	0.0897	0.0380	0.0019	**1.1522**	**0.0887**	**0.0310**	**0.0003**	1.1301	0.0881	0.0247	0.0001
	0.6	**1.1689**	**0.0904**	**0.0367**	**0.0018**	1.1541	0.0907	0.0320	0.0003	1.1367	0.0902	0.0268	0.0001
	0.8	1.1758	0.0940	0.0398	0.0020	1.1696	0.0963	0.0380	0.0004	1.1582	0.0956	0.0342	0.0001
	1.0	1.1904	0.0986	0.0460	0.0023	1.1902	0.1024	0.0467	0.0005	1.1825	0.1014	0.0436	0.0001

and does not receive an effect a lot in σ_V^2 change

From this result, it can be concluded that the MDPDE is more robust than the KE. Meanwhile, it can be seen that the α's yielding minimal d_R vary with the cases. This indicates that choosing an optimal α is not an easy task in actual usage. However, it may be an important issue to select an optimal α. One possible way is, as in the trimmed mean context, to choose an α to produce the smallest asymptotic variance of the MDPDE. See, for instance, Hong and Kim (2001) or Warwick and Jones (2005). Conventionally, α in [0.1, 0.2] is recommended since the MDPDE with such an α still keeps the efficiency and robustness properties. The same can be applied to our case. However, our simulation result suggests that a broader range of α's, say, in [0.15, 0.5] could be employed in construction of the MDPDE.

4. Proofs

We will provide the proof for the case of $\alpha > 0$ since the proofs of the case of $\alpha = 0$ are similar to that of $\alpha > 0$. In what follows, we denote

$$a_i(\sigma) = a(X_{t_i^n}, \sigma), b_i = b(X_{t_i^n}), Z_i = Z_{n,i}.$$

The symbol $C > 0$ denotes a universal constant. In particular, C_k is a constant depending on k.

Lemma 4.1 *Suppose that (A1) and (A3) hold. Then, for $k \geq 1$ and $t_{i-1}^n \leq t \leq t_i^n$,*

$$\max_{i \leq n} E_0 |\Delta_i|^{2k} = O(h_n^{2k}).$$

Proof: By Lemma 6 of Kessler (1997), we can have

$$E_0 \left| X_t - X_{t_{i-1}^n} \right|^{2k} \leq C_k h_n^k.$$

Then, using this and Theorem 6.3 of Friedman (1975), page 85, we have

$$\begin{aligned}
&E_0|\Delta_i|^{2k} \\
&\leq 2^{2k-1} E_0 \left\{ \int_{t_{i-1}^n}^{t_i^n} (a(X_s, \sigma_0) - a_{i-1}(\sigma_0))\, dW_s \right\}^{2k} + 2^{2k-1} E_0 \left\{ \int_{t_{i-1}^n}^{t_i^n} \left(b(X_s) - b(X_{t_{i-1}^n}) \right) ds \right\}^{2k} \\
&\leq C_k h_n^{k-1} E_0 \int_{t_{i-1}^n}^{t_i^n} \{a(X_s, \sigma_0) - a_{i-1}(\sigma_0)\}^{2k} ds + C_k h_n^{2k-1} E_0 \int_{t_{i-1}^n}^{t_i^n} \left\{ b(X_s) - b(X_{t_{i-1}^n}) \right\}^{2k} ds \\
&\leq C_k h_n^{k-1} \int_{t_{i-1}^n}^{t_i^n} E_0 |X_s - X_{t_{i-1}^n}|^{2k} ds + C_k h_n^{2k-1} \int_{t_{i-1}^n}^{t_i^n} E_0 |X_s - X_{t_{i-1}^n}|^{2k} ds \\
&\leq C_k h_n^{2k} + C_k h_n^{3k} = O(h_n^{2k}).
\end{aligned}$$

This asserts the lemma. □

Lemma 4.2 *Suppose that (A1) and (A3) hold. If $f \in \mathcal{F}$ and $nh_n^q \to 0$ for some $q > 1$. Then,*

$$\sup_{\sigma \in \Theta} \max_{i \leq n} \left| f^j(X_{t_{i-1}^n}, \sigma) Z_i^k \Delta_i^l h_n^m \right| = o(1) \quad a.s.,$$

where $j, k, l \in \{0, 1, 2, \ldots\}$ and $m > -l$.

Proof: Using (A3), Lemma 4.1 and the fact that $EZ_i^k < \infty$ for all $k \geq 0$, we have that for any $\epsilon, \kappa > 0$,

$$\begin{aligned}
&\sum_{n=1}^{\infty} P_0 \left(\max_{i \leq n} \left| (1 + |X_{t_{i-1}^n}|)^C Z_i^k \Delta_i^l h_n^m \right| > \epsilon \right) \\
&\leq \sum_{n=1}^{\infty} \frac{n}{\epsilon^\kappa} \max_{i \leq n} E_0 \left| (1 + |X_{t_{i-1}^n}|)^C Z_i^k \Delta_i^l h_n^m \right|^\kappa \\
&\leq C \sum_{n=1}^{\infty} n h_n^{(l+m)\kappa} \\
&= \sum_{n=1}^{\infty} o\left(n^{1-(l+m)\kappa/q} \right).
\end{aligned}$$

By choosing κ such that $1 - (l+m)\kappa/q \leq -2$, we establish the lemma. □

Lemma 4.3 (Kessler, 1997). *Suppose that (A1)-(A3) hold. If* $f(x,\sigma) : \mathbb{R}\times\Theta \to \mathbb{R}$ *is differentiable with respect to* x *and the derivative belongs to* $\mathcal{F}$, *then*

$$\left| \frac{1}{n}\sum_{i=1}^{n} f(X_{t_{i-1}^n},\sigma) - \int f(x,\sigma)\mu_0(dx) \right| = o_{P_0}(1).$$

If in addition, $\partial_\sigma f$ *exists and belongs to* $\mathcal{F}$, *the above convergence holds uniformly in* σ.

Lemma 4.4 *Suppose that (A1)-(A5) hold. If* $f(x,\sigma) : \mathbb{R}\times\Theta \to \mathbb{R}$ *belongs to* $\mathcal{F}$ *and is differentiable with respect to* x *and* σ *with derivatives belonging to* $\mathcal{F}$, *then*

$$\sup_{\sigma\in\Theta} \left| \frac{1}{n}\sum_{i=1}^{n} f(X_{t_{i-1}^n},\sigma) e^{-\frac{\alpha}{2}\frac{a_{i-1}(\sigma_0)^2}{a_{i-1}(\sigma)^2}Z_i^2} - \left(1+\alpha\frac{a_{i-1}(\sigma_0)^2}{a_{i-1}(\sigma)^2}\right)^{-1/2} \int f(x,\sigma)\mu_0(dx) \right|$$
$$= o_{P_0}(1).$$

Proof: By (A4) and (A5), we can readily see that $\partial_x(1+\alpha(a(x,\sigma_0)^2/a(x,\sigma)^2))^{-1/2} \in \mathcal{F}$. Let

$$h_i(\sigma) = \frac{1}{n} f(X_{t_{i-1}^n},\sigma) e^{-\frac{\alpha}{2}\frac{a_{i-1}(\sigma_0)^2}{a_{i-1}(\sigma)^2}Z_{i-1}^2}.$$

Then, by Lemma 4.3, we have that

$$\begin{aligned}\sum_{i=1}^{n} E_0\left\{h_i(\sigma)|G_{i-1}^n\right\} &= \frac{1}{n}\sum_{i=1}^{n}\left(1+\alpha\frac{a_{i-1}(\sigma_0)^2}{a_{i-1}(\sigma)^2}\right)^{-1/2} f(X_{t_{i-1}^n},\sigma) \\ &\to \int \left(1+\alpha\frac{a(x,\sigma_0)^2}{a(x,\sigma)^2}\right)^{-1/2} f(x,\sigma)\mu_0(dx)\end{aligned}$$

and

$$\sum_{i=1}^{n} E_0\left\{h_i(\sigma)^2|G_{i-1}^n\right\} \le \frac{1}{n^2}\sum_{i=1}^{n} f^2(X_{t_{i-1}^n},\sigma) = o_{P_0}(1).$$

In view of Lemma 9 of Genon-Catalot and Jacod (1993), the convergence for each σ is ensured. To establish the uniform convergence, we will prove the tightness of the sequence $\{\sum_{i=1}^n h_i(\sigma)\}$ in $C(\Theta)$, the space of continuous functions on Θ.

According to Theorem 20 of Ibragimov and Has'minskii (1981), page 378, it suffices to show that

$$E_0\left\{\left|\sum_{i=1}^n h_i(\sigma)\right|^2\right\} \le C \quad \text{for all } \sigma \tag{4.1}$$

and

$$E_0\left\{\left|\sum_{i=1}^n h_i(\sigma_1) - \sum_{i=1}^n h_i(\sigma_2)\right|^2\right\} \le C|\sigma_1-\sigma_2|^2 \quad \text{for all } \sigma_1,\sigma_2. \tag{4.2}$$

Note that (4.1) can be proved easily by using Jensen's inequality and the condition $f \in \mathcal{F}$. Further, since

$$\sup_{\sigma\in\Theta} |\partial_\sigma h_i(\sigma)| \le \frac{C}{n}(1+|X_{t_{i-1}^n}|)^C(1+\alpha Z_i^2)$$

and

$$\begin{aligned} E_0\left\{\left|\sum_{i=1}^n (h_i(\sigma_1) - h_i(\sigma_2))\right|^2\right\} &\le E_0\left\{\left|\sum_{i=1}^n \partial_\sigma h_i(\sigma^*)(\sigma_1-\sigma_2)\right|^2\right\} \\ &\le |\sigma_1-\sigma_2|^2 \frac{C}{n}\sum_{i=1}^n E_0\left\{(1+|X_{t_{i-1}^n}|)^C(1+\alpha Z_i^2)\right\} \\ &\le C|\sigma_1-\sigma_2|^2, \end{aligned}$$

where σ^* lies between σ_1 and σ_2, (4.2) is asserted. □

Lemma 4.5 *Suppose that (A1)-(A5) hold. If $h_n \to 0, nh_n \to \infty$ and $nh_n^q \to 0$ for some $q > 1$, then*

$$\sup_{\sigma\in\Theta}\left|\frac{1}{n}\sum_{i=1}^n V_{\alpha,n,i}(\sigma) - V(\sigma)\right| = o_{P_0}(1),$$

where

$$V(\sigma) := \int \left(\frac{1}{a(x,\sigma)^2}\right)^{\alpha/2}\left\{\frac{1}{\sqrt{1+\alpha}} - \left(1+\frac{1}{\alpha}\right)\left(1+\alpha\frac{a(x,\sigma_0)^2}{a(x,\sigma)^2}\right)^{-1/2}\right\}\mu_0(dx).$$

Proof: In view of Lemma 4.3, we have

$$\sup_{\sigma\in\Theta}\left|\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{a_{i-1}(\sigma)^2}\right)^{\alpha/2}-\int\left(\frac{1}{a(x,\sigma)^2}\right)^{\alpha/2}\mu_0(dx)\right|=o_{P_0}(1). \tag{4.3}$$

Let

$$K_i(\sigma):=\alpha\frac{a_{i-1}(\sigma_0)}{a_{i-1}(\sigma)^2}\frac{Z_i\Delta_i}{\sqrt{h_n}}+\frac{\alpha}{2}\frac{\Delta_i^2}{a_{i-1}(\sigma)^2h_n}.$$

Then by Lemma 4.2, $\sup_{\sigma\in\Theta}\max_{i\le n}|K_i(\sigma)|=o(1)$ a.s.. Note that

$$\begin{aligned}
&\sup_{\sigma\in\Theta}\left|\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{a_{i-1}(\sigma)^2}\right)^{\alpha/2}\left\{\exp\left(-\frac{\alpha}{2}\frac{(X_{t_i^n}-X_{t_{i-1}^n}-b_{i-1}h_n)^2}{a_{i-1}(\sigma)^2h_n}\right)-\exp\left(-\frac{\alpha}{2}\frac{a_{i-1}(\sigma_0)^2}{a_{i-1}(\sigma)^2}Z_i^2\right)\right\}\right|\\
&=\sup_{\sigma\in\Theta}\left|\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{a_{i-1}(\sigma)^2}\right)^{\alpha/2}\exp\left(-\frac{\alpha}{2}\frac{a_{i-1}(\sigma_0)^2}{a_{i-1}(\sigma)^2}Z_i^2\right)\{\exp(-K_i(\sigma))-1\}\right|\\
&\le C\sup_{\sigma\in\Theta}\max_{i\le n}|\exp(-K_i(\sigma))-1|\\
&\le C\sup_{\sigma\in\Theta}\max_{i\le n}|K_i(\sigma)|\exp\left(\sup_{\sigma\in\Theta}\max_{i\le n}|K_i(\sigma)|\right)\\
&=o(1)\quad a.s.
\end{aligned} \tag{4.4}$$

and by Lemma 4.4,

$$\begin{aligned}
\sup_{\sigma\in\Theta}\Bigg|&\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{a_{i-1}(\sigma)^2}\right)^{\alpha/2}\exp\left(-\frac{\alpha}{2}\frac{a_{i-1}(\sigma_0)^2}{a_{i-1}(\sigma)^2}Z_i^2\right)\\
&-\int\left(\frac{1}{a(x,\sigma)^2}\right)^{\alpha/2}\left(1+\alpha\frac{a(x,\sigma_0)^2}{a(x,\sigma)^2}\right)^{-1/2}\mu_0(dx)\Bigg|=o_{P_0}(1).
\end{aligned} \tag{4.5}$$

Combining (4.3)–(4.5), we assert the lemma. □

Proof of Theorem 2.1: It follows from Lemma 4.5 that for any subsequence $\{n'\}\subset\{n\}$, there exists further subsequence $\{n''\}\subset\{n'\}$ satisfying

$$\sup_{\sigma\in\Theta}\left|\frac{1}{n''}\sum_{i=1}^{n''}V_{\alpha,n'',i}(\sigma)-V(\sigma)\right|=o(1)\quad a.s..$$

Moreover, noting the fact that

$$\left(\frac{1}{x}\right)^{\alpha/2}\left\{\frac{1}{\sqrt{1+\alpha}}-\left(1+\frac{1}{\alpha}\right)\left(1+\alpha\frac{x_0}{x}\right)^{-1/2}\right\},\quad x>0,x_0>0$$

is minimized only at $x=x_0$, $V(\sigma)$ has a unique minimum at $\sigma=\sigma_0$ by (A6). Therefore, since Θ is compact, we can easily know that $\hat{\sigma}_{\alpha,n''}$ strongly converges to σ_0. This completes the proof. □

Acknowledgements

We are grateful to the Editor and referees for useful suggestions.

References

Basu, S. and Lindsay, B. G. (1994). Minimum disparity estimation for continuous models: efficiency, distributions and robustness. *Annals of the Institute of Statistical Mathematics*, **46**, 683–705.

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.

Beran, R. (1977). Minimum Hellinger distance estimates for parametric models. *The Annals of Statistics*, **5**, 445–463.

Cao, R., Cuevas, A. and Fraiman, R. (1995). Minimum distance density-based estimation. *Computational Statistics & Data Analysis*, **20**, 611–631.

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

Florens-Zmirou, D. (1989). Approximate discrete-time schemes for statistics of diffusion processes. *Statistics*, **20**, 547–557.

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

Genon-Catalot, V. and Jacod, J. (1993). On the estimation of the diffusion coefficient for multidimensional diffusion processes. *Annales Institut Henri Poincare Probabilites et Statistiques*, **29**, 119–151.

Hong, C. and Kim, Y. (2001). Automatic selection of the tuning parameter in the minimum density power divergence estimation. *Journal of the Korean Statistical Society*, **30**, 453–465.

Ibragimov, I. A. and Has'minskii, R. Z. (1981). *Statistical Estimation Asymptotic Theory.* Springer-Verlag, New York.

Kessler, M. (1997). Estimation of an ergodic diffusion from discrete observations. *Scandinavian Journal of Statistics*, **24**, 211–229.

Kessler, M. (2000). Simple and explicit estimating functions for a discretely observed diffusion process. *Scandinavian Journal of Statistics*, **27**, 65–82.

Kessler, M. and Sørensen, M. (1999). Estimating equations based on eigenfunctions for a discretely observed diffusion process. *Bernoulli*, **5**, 299–314.

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

Lee, S. and Na, O. (2005). Test for parameter change based on the estimator minimizing density-based divergence measures. *Annals of the Institute of Statistical Mathematics*, **57**, 553–573.

Lee, S. and Song, J. (2006). Minimum density power divergence estimator for diffusion processes. *submitted for publication.*

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

Masuda, H. (2005). Simple estimators for parametric Markovian trend of ergodic processes based on sampled data. *Journal of The Japan Statistical Society*, **35**, 147–170.

Simpson, D. G. (1987). Minimum Hellinger distance estimation for the analysis of count data. *Journal of the American Statistical Association*, **82**, 802–807.

Song, J. and Lee, S. (2006). Test for parameter change in discretely observed diffusion processes. *submitted for publication.*

Tamura, R. N. and Boos, D. D. (1986). Minimum Hellinger distance estimation for multivariate location and covariance. *Journal of the American Statistical Association*, **81**, 223–239.

Warwick, J. and Jones, M. C. (2005). Choosing a robustness tuning parameter. *Journal of Statistical Computation and Simulation*, **75**, 581–588.

Yoshida, N. (1992). Estimation for diffusion processes from discrete observation. *Journal of Multivariate Analysis*, **41**, 220–242.

[Received November 2006, Accepted March 2007]