Ann. Inst. Statist. Math.
Vol. 57, No. 4, 733–745 (2005)

A NEW INSTRUMENTAL VARIABLE ESTIMATION FOR DIFFUSION PROCESSES*

BEONG SOO SO

Department of Statistics, Ewha Womans University, Seoul 120-750, Korea

(Received June 21, 2004; revised September 13, 2004)

Abstract. We consider the problem of parametric inference from continuous sample paths of the diffusion processes $\{x(t)\}$ generated by the system of possibly non-stationary and/or nonlinear Ito stochastic differential equations. We propose a new instrumental variable estimator of the parameter whose pivotal statistic has a Gaussian distribution for all possible values of parameter. The new estimator enables us to construct exact level-α confidence intervals and tests for the parameter in the possibly non-stationary and/or nonlinear diffusion processes. Applications to several non-stationary and/or nonlinear diffusion processes are considered as examples.

Key words and phrases: Non-stationary nonlinear diffusion, instrumental variable estimator.

1. Introduction

We consider the vector-valued continuous-time stochastic process $\{x(t)\}$ in a state space $S \subset R^k$, generated by the following system of possibly non-stationary and/or nonlinear Ito stochastic differential equations;

$$dx(t) = a(t, x(t))dt + b(t, x(t))\theta dt + \sigma(t, x(t))dW(t), \qquad 0 \le t \le T, \tag{1.1}$$

where $x(t) = (x_1(t), \dots, x_k(t))'$, $a(t,x)$, $b(t,x)$, $\sigma(t,x)$ are $k\times 1$, $k\times 1$, $k\times p$, $k\times k$ matrix-valued functions respectively, $\theta = (\theta_1, \dots, \theta_p)'$ is an unknown vector of p parameters of interest with $p \le k$, $W(t) = (W_1(t), \dots, W_k(t))'$ is a vector of k independent standard Brownian motion processes.

Recently, the use of continuous-time processes described by Ito type stochastic differential equations has become very popular in financial economics. For example, in capital asset pricing models, many of the financial time series such as stock prices, exchange rates, and interest rates are usually assumed to satisfy stochastic differential equations of the type (1.1). See Karlin and Taylor (1981) for a large number of applications of diffusion processes for various stochastic modelling and Hull (1999) and James and Webber (2000) for more specific applications in various capital asset pricing models.

In this paper, we consider the problem of statistical inference on the parameter θ of the model (1.1) from the continuous sample path $\{x(t); t \in [0, T]\}$ of the process up to time T. Statistical inference for continuous-time diffusion type processes is, of course, not new and has been the subject of extensive research effort in the mathematics and statistics literatures. See Basawa and Prakasa Rao ((1980), Chapter 9), Prakasa

*This work was supported by Korea Research Foundation Grant (KRF-2001-015-DP0057).

Rao ((1999), Chapter 2) and Kutoyants (2003) for a recent survey and the extensive bibliography of this vast and still growing literature.

However, most of these works are focused on the asymptotic properties of the classical estimators such as least squares estimator (LSE) or maximum likelihood estimator (MLE) as $T \to \infty$. For the special model of the type (1.1) with $\sigma(t, x) = I_k \sigma$ for $\sigma > 0$, both conditional MLE and LSE of the unknown parameter θ are the same and is given by

$$\hat{\theta}_o = \left[\int_0^T b' b(t, x(t)) dt \right]^{-1} \int_0^T b'(t, x(t)) [dx(t) - a(t, x(t)) dt].$$

In spite of the asymptotic optimality theory developed for this estimator as in Prakasa Rao ((1999), Chapter 2), standard inference procedures based on LSE or MLE have several drawbacks from the applications point of view.

First, most of standard asymptotic theory depends heavily on the stringent assumption of the stationarity of the underlying process $\{x(t)\}$ which severely limits the domain of the permissible parameters. This restriction excludes from consideration many interesting non-stationary models, such as random walk and cointegration, commonly encountered in the current economics and finance literatures.

Secondly, even for the stationary model, most of the optimality results are based on the asymptotic distribution of the MLE as $T \to \infty$. Thus the finite sample properties of the MLE-based procedures are largely unknown especially for non-stationary and/or nonlinear processes and there is no simple explicit guideline on the value T for the validity of the asymptotic results. Both of these features may diminish the practical utility of the LSE-based methods in analyzing real data. Therefore we need alternative methods which can overcome these problems of the LSE-based methods and enable us to make a simple and flexible inference on the parameter θ which is valid not only for stationary processes with a large sample size but also for possibly non-stationary and/or nonlinear processes with a small sample size.

In this paper, we propose a new estimator of θ based on the special instrumental variables and establish important finite sample properties of the estimator such as the median-unbiasedness and the normality of the corresponding pivotal quantity for possibly non-stationary and/or nonlinear diffusion processes. Then we develop exact level-α confidence intervals and tests of the hypotheses on θ in non-stationary and/or nonlinear diffusion models.

In the sequel, in order to simplify the notation, we will assume that $\sigma(t, x) = I_k \sigma$ for fixed $\sigma > 0$ without loss of generality unless otherwise stated. See Section 4 for the extensions to the heteroscedastic models.

Our new estimator is motivated by the following moment condition

$$E \left[\int_0^T K(t, x(t)) dW(t) \right] = 0, \qquad T \geq 0 \tag{1.2}$$

for any F_t-measurable $p \times k$ matrix-valued function $K(t, x) = [k_1, \ldots, k_p]'$ of p *instrumental variables* vectors where F_t is the σ-field generated by $\{x(s); s \in [0, t]\}$ for $t \geq 0$.

Now the sample analog of (1.2) is given by

$$\int_0^T K(t, x(t)) [dx(t) - a(t, x(t)) dt - b(t, x(t)) \theta dt] = 0$$

which in turn motivates the definition of the following estimator.

DEFINITION 1.1. (Instrumental variables estimator) For the matrix of instrumental variables $K(t, x(t))$, the estimator defined by

$$(1.3) \qquad \hat{\theta}_c = \left[\int_0^T K(t, x(t))b(t, x(t))dt\right]^{-1} \int_0^T K(t, x(t))[dx(t) - a(t, x(t))dt]$$

is called the *instrumental variables estimator* (IVE) of θ.

By the appropriate choice of the matrix $K(t, x(t))$, we can generate a variety of IVEs of θ. For example LSE $\hat{\theta}_o$ corresponds to the choice of the instrumental variables $K(t, x(t)) = b'(t, x(t))$.

As an alternative to LSE, we will consider the special type of IVE $\hat{\theta}_c$ based on the p-*orthonormal* instrumental variables $K^*(t, x(t)) = b^{*\prime}(t, x(t))$ where $b^*(t, x) = [b_1^*, \ldots, b_p^*]$ is the matrix of the p-orthonormal vectors $\{b_r^*\}_{r=1}^p$ in R^k constructed from the p-column vectors $\{b_r\}_{r=1}^p$ of the matrix $b(t, x)$ of rank p by the Gram-Schmidt orthogonalization process:

$$b_1^* = b_1/|b_1|,$$
$$b_2^* = (b_2 - (b_2, b_1^*)b_1^*)/|b_2 - (b_2, b_1^*)b_1^*|,$$
$$\ldots$$
$$b_p^* = \left(b_p - \sum_{i=1}^{p-1}(b_p, b_i^*)b_i^*\right) \Big/ \left|b_p - \sum_{i=1}^{p-1}(b_p, b_i^*)b_i^*\right|,$$

where $(a, b) = \sum_{i=1}^k a_i b_i$ is an inner product of the vectors a, b in R^k and $|x|^2 = (x, x) = \sum_{i=1}^k x_i^2$ is the square of the Euclidean norm of the vector x.

In view of formal similarity to the sign-based IVE first proposed by Cauchy (1836) in the context of linear regression model, we will call the new IVE Cauchy estimator.

DEFINITION 1.2. (Cauchy estimator) The IVE based on the matrix of orthonormal instrumental variables $b^*(t, x(t))$ is called a *Cauchy estimator*.

See also So and Shin (1999) for the similar definition of the Cauchy estimator of the discrete-time stochastic processes.

From the definition of the estimator $\hat{\theta}_c$, the vector of pivotal quantities τ_c based on $\hat{\theta}_c$ is given by

$$(1.4) \qquad \tau_c(\theta) = \left[\int_0^T b^{*\prime}b(t, x(t))dt\right](\hat{\theta}_c - \theta)/\sigma T^{1/2} = \int_0^T b^{*\prime}(t, x(t))dW(t)/T^{1/2}.$$

In the next section, we will establish the distribution of the pivotal quantity τ_c.

Following this introduction, in Section 2, we prove a key Lemma for deriving the distribution of the proposed pivotal statistic and then we construct exact level-α confidence regions and tests of the hypotheses on the parameter θ. In Section 3, applications to several non-stationary and/or non-linear diffusion type processes are considered with some encouraging simulation results. Section 4 concludes with some discussions on extensions to other models.

2. Finite sample properties of the estimator

In the sequel, $A \sim B$ denotes that A and B have the same distribution, $|a|^2 = \mathrm{tr}(a'a)$ denotes the square of the Euclidean norm of the matrix a, $N_k(\mu, \sigma^2)$ denotes the k-variate normal distribution with a mean vector μ and a covariance matrix σ^2.

In order to ensure the existence and uniqueness of a solution to the system of stochastic differential equations (1.1), we require

A: There exists some constant $K > 0$ such that the matrix-valued functions $a(t,x)$, $b(t,x)$, and $\sigma(t,x)$ satisfy the following conditions for all $x, y \in S \subset R^k$ and $s, t \in [0,T]$:

a) $|a(t,x) - a(t,y)| + |b(t,x) - b(t,y)| + |\sigma(t,x) - \sigma(t,y)| \leq K|x-y|$,
b) $|a(s,x) - a(t,x)| + |b(s,x) - b(t,x)| + |\sigma(s,x) - \sigma(t,x)| \leq K|s-t|$,
c) $|a(t,x)|^2 + |b(t,x)|^2 + |\sigma(t,x)|^2 \leq K^2(1+|x|^2)$.

Now we are ready to prove the following lemma which will be useful in establishing finite sample properties of the new estimator.

LEMMA 2.1. *Let $v(t) = [v_1(t), \ldots, v_p(t)]$ be a $k \times p$-matrix of non-anticipating stochastic processes adapted to the σ-fields F_t satisfying the following orthogonality condition:*

B: *$v(t)'v(t) = I_p$ for any $t \in [0,T]$.*

If we let

$$M_t = \int_0^t v'(s)dW(s), \qquad t \geq 0,$$

then

a) *$\{M_t, F_t\}_{t=0}^T$ is a martingale,*
b) *$\{M_t\}_{t=0}^T \sim \{(W_1(t), \ldots, W_p(t))\}_{t=0}^T$,*
c) *$M_T \sim N_p(0, TI_p)$.*

PROOF. By the definition of stochastic integral with respect to Wiener Process and the orthogonality of $v(t)$, M_t is a continuous local martingale with $\langle M_t, M_t' \rangle = I_p t$. Then Levy characterization theorem of Revuz and Yor ((1999), p. 150) completes the proof.

Note that Lemma 2.1-a), b) imply that the process $\{M_t\}_{t=0}^\infty$ is a Brownian motion in R^p. Lemma 2.1-c) enables us to establish the distribution of the pivotal quantity $\tau_c(\theta)$ of (1.4) for any fixed parameter θ and for any possibly non-stationary and/or nonlinear diffusion models of the type (1.1) as is to be shown in Theorem 2.1.

THEOREM 2.1. (Normality of the pivot τ_c) *Consider model (1.1) satisfying conditions **A** with $\sigma(t,x) = I_k\sigma$ for $\sigma > 0$. Let $b(t, x(t))$ be a $k \times p$-matrix of rank p. Then, for any θ, we have*

$$\tau_c(\theta) \sim N_p(0, I_p),$$

and in particular

$$\tau_{c,p}(\theta) = \left[\int_0^T |b_p^\perp(t,x(t))|dt\right](\hat{\theta}_{c,p} - \theta_p)/\sigma T^{1/2} = \int_0^T b_p^{*\prime}(t,x(t))dW(t)/T^{1/2} \sim N(0,1),$$

where $b_p^\perp = b_p - \sum_{i=1}^{p-1}(b_p, b_i^*)b_i^*$ *and*

$$\hat{\theta}_{c,p} = \left[\int_0^T |b_p^\perp(t,x(t))|dt\right]^{-1} \int_0^T b_p^{*\prime}(t,x(t))[dx(t) - a(t,x(t))dt].$$

PROOF. First part follows immediately from Lemma 2.1 with the choice $v(t) = b^*(t,x(t))$. As for the second part, we note that $(b_i^*, b_j) = 0$, for $i > j$ and $(b_r^*, b_r) = |b_r - \sum_{i=1}^{r-1}(b_r, b_i^*)b_i^*|$, $r = 1,\ldots,p$ imply that the normalizing matrix in $\tau_c(\theta)$ is the upper triangular matrix of order p with positive diagonal elements $\{|b_r - \sum_{i=1}^{r-1}(b_r, b_i^*)b_i^*|\}_{r=1}^p$. This completes the proof.

Using the results of Theorem 2.1, we can construct the exact level-α simultaneous confidence region for $\theta \in R^p$ and the confidence intervals for the individual parameters θ_r, $r = 1,\ldots,p$ by the appropriate change of order of θ_r.

Exact level-α simultaneous confidence region for θ:

$$R_c : (\theta - \hat{\theta}_c)' U_T' U_T (\theta - \hat{\theta}_c) \leq \chi_\alpha^2(p),$$

where $U_T = [\int_0^T b^{*\prime} b(t,x(t))dt]/\sigma T^{1/2}$, and $\chi_\alpha^2(p)$ is the upper α-th quantile of the χ^2-distribution with the degrees freedom p.

Exact level-α confidence interval for θ_r:

$$I_c : \hat{\theta}_{c,r} \pm \left[\int_0^T |b_r^\perp(t,x(t))|dt\right]^{-1} \sigma T^{1/2} z_{\alpha/2},$$

where $b_r^\perp = b_r - [b_{-r}' b_{-r}]^{-1}(b_{-r}' b_r)$ with $b_{-r} = [b_1,\ldots,b_{r-1},b_{r+1},\ldots,b_p]$,

$$\hat{\theta}_{c,r} = \left[\int_0^T |b_r^\perp(t,x(t))|dt\right]^{-1} \int_0^T b_r^{\perp *\prime}(t,x(t))[dx(t) - a(t,x(t))dt],$$

$b_r^{\perp *} = b_r^\perp / |b_r^\perp|$ and z_α is the upper α-th quantile of the standard normal distribution.

We also note that the corresponding LSE-based approximate confidence intervals are given by

Approximate level-α confidence interval for θ_r:

$$I_o : \hat{\theta}_{o,r} \pm D_{rr}^{1/2} \sigma z_{\alpha/2}$$

for the stationary process where D_{rr} is the r-th diagonal element of the matrix $D = [\int_0^T b'b(t,x(t))dt]^{-1}$ of order p. This follows directly from the asymptotic distribution of the pivotal statistics

$$\tau_o = \left[\int_0^T b'b(x(t))dt\right]^{1/2}(\hat{\theta}_o - \theta) = \left[\int_0^T b'b(t,x(t))dt\right]^{-1/2}\sigma\int_0^T b'(t,x(t))dW(t)$$
$$\Rightarrow N_p(0, I_p\sigma^2)$$

as $T \to \infty$ which follows from the martingale central limit theorem for stochastic integrals due to Kutoyants (1975).

Remark 1. (Median-unbiasedness of $\hat{\theta}_{c,r}$) We note that Theorem 2.1 implies

$$P_\theta(\hat{\theta}_{c,r} \le \theta_r) = 1/2 = P_\theta(\hat{\theta}_{c,r} \ge \theta_r)$$

for any θ_r in R. However this is not true for MLE $\hat{\theta}_o$ which has a non-negligible bias typically. See Theorem 17.2 of Liptser and Shiryayev ((1999), p. 225) for the characterization of the bias and the mean squared error of the MLE in the special class of univariate Ito processes of the type (1.1). This bias is responsible for the possible power loss of the MLE-based test and the corresponding distortion of the coverage probability of the MLE-based confidence intervals for the near non-stationary model. See the simulation results of Tables 1, 2 and 3 below for the possible distortion of the empirical coverage probabilities of the 90-% confidence intervals based on LSE.

Remark 2. (Asymptotic relative efficiency of $\hat{\theta}_c$ and $\hat{\theta}_o$) Let $\{x(t)\}$ be a stationary process in R^k with $a(x), b(x)$ independent of t. Then for a real-valued parameter θ, asymptotic relative efficiency (ARE) of the Cauchy estimator $\hat{\theta}_c$ with respect to the MLE $\hat{\theta}_o$ as $T \to \infty$ is given by

$$ARE(\hat{\theta}_c; \hat{\theta}_o) = [E|b(x(t))|]^2/E[|b(x(t))|^2]$$

from the ergodic theorem for the stationary process $\{b(x(t))\}$. For example, k-variate stationary Ornstein-Uhlenbeck process is defined by

$$dx(t) = \theta x(t)dt + \sigma dW(t),$$

where $\theta < 0$ and $x(0) \sim N_k(0, I_k\sigma^2/2|\theta|)$. Then we have

$$ARE(\hat{\theta}_c; \hat{\theta}_o) = [E|x(t)|]^2/E[|x(t)|^2] = 2\Gamma^2((k+1)/2)/k\Gamma^2(k/2) \ge 2/\pi = 0.637$$

for any $\theta < 0$ which is tabulated for selected values of k in Table 0.

We note that this result is partially supported by the simulation results for mad (mean absolute deviation) ratio $ARE = (.59/.67)^2 = .775$ of Table 3 for $k = 2$ with

Table 0. $dx(t) = \theta x(t)dt + \sigma dW(t)$, $\theta < 0$, $x(0) \sim N_k(0, I_k\sigma^2/2|\theta|)$.

k	1	2	3	4	8	∞
ARE	$2/\pi = .637$	$\pi/4 = .785$	$8/3\pi = .849$	$9\pi/32 = .884$	.940	1.0

$T = 9$, $\theta = -5$. However for non-stationary processes near $\theta = 0$ with a small sample size T, ARE is not relevant and the Cauchy estimator actually outperforms the LSE in terms of smaller mad as is shown in the simulation results of Tables 1 and 3 for $k = 1, 2$ below. We also note that although the confidence intervals I_c based on the Cauchy estimator $\hat{\theta}_c$ have the exact coverage probability $1 - \alpha$ for any finite T, those based on LSE $\hat{\theta}_o$ may suffer from considerable distortion of the coverage probability for non-stationary process as shown in Tables 1 and 2.

Now it is straightforward to construct the level-α critical regions of the tests of hypotheses on θ by inverting the pivotal quantity of the corresponding confidence regions.

Exact level-α tests for θ_r: We reject the null hypothesis $H_o : \theta_r = \theta_{ro}$ in favor of the alternative hypothesis $H_1 : \theta_r \neq \theta_{r0}$ if

$$|\tau_{c,r}(\theta_{ro})| \geq z_{\alpha/2}.$$

Similarly critical region for the exact level-α test of the null hypothesis $H_o : \theta_r = \theta_{ro}$ against the one-sided alternative $H_1 : \theta_r < \theta_{ro}$ is given by

$$\tau_{c,r}(\theta_{ro}) \leq -z_\alpha,$$

where $\tau_{c,r}(\theta_r) = [\int_0^T |b_r^\perp(t, x(t))|dt](\hat{\theta}_{c,r} - \theta_r)/\sigma T^{1/2} \sim N(0, 1)$.

Remark 3. (Alternative IVE $\hat{\theta}_c$) If we choose the instrumental variables $K^{**}(t, x(t)) = [b_1^{\perp *}, \ldots, b_p^{\perp *}]'$ where $b_r^{\perp *} = b_r^\perp / |b_r^\perp|$, $r = 1, \ldots, p$, then we have an alternative IVE

$$\hat{\theta}_{c,r} = \left[\int_0^T |b_r^\perp(t, x(t))|dt\right]^{-1} \int_0^T b_r^{\perp *\prime}(t, x(t))[dx(t) - a(t, x(t))dt]$$

of θ_r and the corresponding pivotal quantity

$$\tau_{c,r}(\theta_r) = \left[\int_0^T |b_r^\perp|(t, x(t))|dt\right] (\hat{\theta}_{c,r} - \theta_r)/\sigma T^{1/2} \sim N(0, 1)$$

with the Gaussian marginal distribution for each $r = 1, \ldots, p$. However the joint normality of $\hat{\theta}_c \in R^p$ does not hold for this estimator.

Next we consider applications of our results to the statistical inference for some non-stationary and/or nonlinear diffusion processes.

3. Applications

In the Monte-Carlo simulations of Section 3, we use the simple Euler approximation based on the discrete-time stochastic difference equation;

$$\Delta x_t = a(x_t)\Delta t + b(x_t)\theta\Delta t + \sigma(\Delta t)^{1/2}e_t, \qquad t = 0, \ldots, N, \tag{3.1}$$

where $\Delta t = T/N$, $x_t = x(t\Delta t)$, $\Delta x_t = x_{t+1} - x_t$, and e_t is a sequence of the independent standard normal random variables. This procedure seems sensible because the sample paths of the discrete approximation (3.1) and the corresponding estimators based on the approximations

$$T^{-1}\int_0^T a(x_t)dt \cong \sum_{t=1}^{N} a(x_t)/N, \qquad \int_0^T b(x_t)dx_t \cong \sum_{t=0}^{N-1} b(x_t)\Delta x_t$$

converge to those of the continuous-time diffusion process $x(t)$ as $\Delta t \to 0$. See Kloeden *et al.* (1994) for an extensive discussion on numerical approximation methods of simulation of the solution of a stochastic differential equation.

Example 1. (Ornstein-Uhlenbeck process in R^k) We consider the k-dimensional vector Gaussian process $x(t)$ driven by the linear stochastic differential equations;

$$dx(t) = \theta x(t)dt + \sigma dW(t), \qquad \theta \in R, \qquad t \in [0,T]$$

which has an unique solution

$$x(t) = x_0 e^{\theta t} + \sigma \int_0^t e^{\theta(t-s)}dW(s), \qquad t \geq 0.$$

For the stationary case $\theta < 0$, it is well known that MLE

$$\hat{\theta}_o = \int_0^T (x(t), dx(t))/\int_0^T |x(t)|^2 dt$$

has an asymptotic normal distribution as $T \to \infty$. See Basawa and Prakasa Rao ((1980), Chapter 9), Liptser and Shiryayev ((1999), Chapter 17) and Kutoyants (2003) for details. However, for the important non-stationary process with $\theta = 0$, asymptotic distribution of the MLE is not normal and is given by

$$\tau_o = \sigma^{-1}\left[\int_0^T |x(t)|^2 dt\right]^{1/2} (\hat{\theta}_o - \theta) \Rightarrow \int_0^1 W'(t)dW(t)/\left(\int_0^1 |W(t)|^2 dt\right)^{1/2}.$$

One of the undesirable consequences of the non-normality of the asymptotic distribution of MLE for θ near zero is the distortion of the coverage probability of the naive confidence intervals based on MLE as is shown in the simulation results for the empirical coverage probabilities of the LSE-based 90-% confidence intervals of Tables 1 and 3 with $k = 1, 2$ respectively.

Another difficulty of the MLE-based procedure is apparent in testing for the random walk hypothesis $H_0 : \theta = 0$ against the stationary alternative $H_1 : \theta < 0$. We need a separate table for the lower α-th quantile of the asymptotic distribution of τ_o for each k as is given by Dickey *et al.* (1984) for selected values of k in the discrete-time processes. On the other hand, the pivotal quantity based on the Cauchy estimator has a Gaussian distribution for any finite T and k regardless of the value of θ. This result greatly facilitates the solution of the difficult problem of construction of the valid confidence intervals and the corresponding tests of the parameter θ in the non-stationary process. Incidentally, we note that alternative form of the Cauchy estimator is given by Ito formula

$$\begin{aligned}\hat{\theta}_c &= \int_0^T (x^*(t), dx(t))/\int_0^T |x(t)|dt \\ &= \left[|x(T)| - |x(0)| - (1/2)\sigma^2 \int_0^T \Delta|x(t)|dt\right] / \int_0^T |x(t)|dt,\end{aligned}$$

where $x^* = x/|x|$, $\Delta|x| = (k-1)/|x|$ for $k > 1$ and for $k = 1$ it is formally defined as the Dirac delta function $2\delta(x)$ and $(1/2)\int_0^T \Delta|x|(t)dt = L(0,T)$ is the local time at $x = 0$ of the univariate Ornstein-Uhlenbeck process $x(t)$ from $t = 0$ to $t = T$.

Example 2. (Nonlinear diffusion process) As an example of simple nonlinear and possibly non-stationary diffusion process, we consider the diffusion process

$$dx(t) = \theta x(t)dt/[1 + \mathrm{sign}(x(t))/2] + dW(t). \tag{3.2}$$

Even for this simple process, the nature of the asymptotic distribution of the MLE of θ as $T \to \infty$ is largely unknown especially for non-stationary case. But the inference based on the Cauchy estimator is greatly simplified by the normality of the corresponding pivot $\tau_c(\theta)$ even for non-stationary processes with a small time span T.

In order to investigate finite sample properties of the Cauchy estimator, we consider the univariate Ornstein-Uhlenbeck process and the nonlinear process (3.2). Tables 1, 2 summarize the simulation results for the bias, standard deviation (s.d.), and mean absolute deviation (mad) of the LSE and the Cauchy estimator respectively with the configuration $\Delta t = 0.01$ and the number of replications = 10,000. In general, Cauchy estimator seems to have a smaller bias but a larger variance than LSE and has a smaller mad than LSE for non-stationary models near $\theta = 0$. As for the empirical coverage

Table 1. $T = 1$, $k = 1$, $x(0) = 0$, $dx(t) = \theta x(t)dt + dW(t)$.

	LSE				Cauchy			
θ	$E(\hat{\theta}_o)$	s.d.	mad	Pr.	$E(\hat{\theta}_c)$	s.d.	mad	Pr.
1.0	−.58	2.90	1.97	.865	−.098	3.07	1.83	.900
0.1	−1.63	3.07	2.21	.882	−1.01	3.28	2.11	.899
0.0	−1.75	3.15	2.25	.879	−1.14	3.34	2.18	.902
−0.1	−1.93	3.19	2.32	.881	−1.32	3.47	2.26	.898
−1.0	−2.86	3.38	2.47	.890	−2.14	3.75	2.53	.898
−5.0	−6.86	4.25	3.21	.899	−6.12	4.99	3.69	.900

Note: number of replication = 10,000; $dt = 0.01$; $\sigma = 1$; nominal coverage probability = 0.90.

Table 2. $T = 4$, $k = 1$, $x(0) = 0$, $dx(t) = \theta x(t)dt/[1 + \mathrm{sign}(x(t))/2] + dW(t)$.

	LSE				Cauchy			
θ	$E(\hat{\theta}_o)$	s.d.	mad	Pr.	$E(\hat{\theta}_c)$	s.d.	mad	Pr.
2.0	1.99	.15	.02	.898	1.98	.15	.03	.899
1.0	0.90	.40	.13	.869	.93	.37	.13	.898
0.1	−.27	.69	.46	.868	−.12	.71	.45	.899
0.0	−.38	.70	.49	.879	−.22	.73	.47	.903
−0.1	−.51	.75	.52	.883	−.33	.79	.51	.907
−1.0	−1.43	.94	.71	.897	−1.25	1.12	.82	.896
−5.0	−5.38	1.57	1.23	.899	−5.19	1.98	1.56	.898

Note: number of replication = 10,000; $dt = 0.01$; $\sigma = 1$; nominal coverage probability = 0.90.

probability (Pr.) of the confidence intervals with a 90-% nominal coverage probability, LSE-based intervals suffer from considerable distortion of the coverage probability for non-stationary models near $\theta = 0$ in contrast to the stable behavior of the corresponding intervals based on Cauchy estimator.

Example 3. (Complex-valued Ito process in R^2) As an interesting example of model (1.1), we consider the complex-valued process $x(t) = x_1(t) + ix_2(t)$ in R^2 defined by

$$dx(t) = a(t,x)dt + \theta b(t,x(t))dt + dW(t),$$

where $i^2 = -1$, $a(t,x), b(t,x)$ are fixed complex-valued functions of $t \in R$, $x \in R^2$, $\theta = \theta_1 + i\theta_2$ is a complex parameter and $W(t) = W_1(t) + iW_2(t)$ is a Brownian motion in the complex plane. We can verify that the Cauchy estimator of θ is given by

$$\hat{\theta}_c = \left[\int_0^T |b(t,x(t))|dt\right]^{-1} \int_0^T \mathrm{sign}(\bar{b}(t,x(t)))[dx(t) - a(t,x(t))dt],$$

where $|x|^2 = x_1^2 + x_2^2$, $\mathrm{sign}(x) = x/|x|$, $\bar{x} = x_1 - ix_2$ is the complex conjugate of x. Moreover the pivotal statistics $\tau_c(\theta)$ of the Cauchy estimator has the bivariate normal distribution for any $T > 0$;

$$\tau_c(\theta) = \int_0^T |b(t,x(t))|dt(\hat{\theta}_c - \theta)/T^{1/2} \sim N_2(0, I_2).$$

As a specific application to geophysical problem, Arato *et al.* (1962) employed the simple linear model;

$$dx(t) = \theta x(t)dt + dW(t), \qquad x(0) = x_o$$

in order to model the random oscillation of the instantaneous axis of the rotation of the earth. They used MLE for the estimation of the complex parameter θ under the stationarity assumption $\theta_1 < 0$ and studied asymptotic properties of the estimator as $T \to \infty$. See Liptser and Shiryayev ((1999), Section 17.4) for more details. However their results are not applicable to small T or to non-stationary case $\theta_1 \geq 0$. On the other hand, our procedures based on the Cauchy estimator are valid not only for small T but also for non-stationary models. Table 3 summarizes the simulation results for performance of the two estimators of the complex parameter $\theta = \theta_1 + i\theta_2$, with $\theta_2 = 0$.

Table 3. $T = 9$, $k = 2$, $x(0) = 0$, $dx(t) = \theta x(t)dt + dW(t)$.

	LSE				Cauchy			
θ	$E(\hat{\theta}_o)$	s.d.	mad	Pr.	$E(\hat{\theta}_c)$	s.d.	mad	Pr.
0.1	.02	.17	.11	.856	.05	.17	.11	.896
0.0	−.09	.19	.14	.880	−.06	.20	.13	.898
−0.1	−.20	.22	.16	.890	−.16	.23	.16	.899
−1.0	−1.11	.38	.29	.905	−1.06	.42	.32	.901
−5.0	−5.10	.75	.59	.902	−5.05	.85	.67	.903

Note: number of replication = 10,000; $dt = 0.01$; $\sigma = 1$; nominal coverage probability = 0.90.

Table 4. $k = 1$, $x(0) = 0$, $dx(t) = \theta x(t)dt + dW(t)$.

	$T = 1$		$T = 4$	
θ	τ_o	τ_c	τ_o	τ_c
0.0	.050	.050	.051	.050
−0.1	.050	.051	.056	.062
−0.5	.057	.065	.119	.114
−1.0	.073	.079	.236	.200
−1.5	.094	.092	.399	.313
−2.0	.115	.107	.588	.426
−5.0	.320	.261	.998	.908

Note: nominal level = 0.05; number of replication = 10,000; $dt = 0.01$; $\sigma = 1$.

It is clear that Cauchy estimator outperforms LSE in the neighborhood of $\theta = 0$ in terms of the smaller mad and the more stable coverage probability.

Example 4. (Test for random walk in R^k) For the important special case $a(t, x) = 0$ with $p = 1$, we may consider the general problem of test of random walk hypothesis $H_0 : \theta = 0$ against possibly nonlinear stationary alternative hypotheses $H_1 : \theta < 0$ of the type (1.1). It is straightforward to construct the one-sided normal test from the pivotal quantity (1.4) of the Cauchy estimator. However, for the MLE-based test, we need a separate table for the quantiles of the finite sample or asymptotic null distribution of the quantity

$$\int_0^T b'(W(t))dW(t) / \left(\int_0^T |b(W(t))|^2 dt \right)^{1/2}$$

which can be obtained through Monte-Carlo experiments on a case-by-case basis. Furthermore, we point out the intrinsic difficulty in constructing reliable MLE-based confidence interval of the parameter θ near zero due to the inevitable non-normality of the corresponding t-statistic.

Table 4 summarizes limited simulation results for the powers of the level-0.05 random walk test $H_0 : \theta = 0$ against the stationary alternative $H_1 : \theta < 0$ of the univariate Ornstein-Uhlenbeck process with $b(x_t) = x_t$. It shows the local power advantage of the normal test τ_c based on the Cauchy estimator over that τ_o based on the LSE for small samples $T = 1, 4$ respectively.

4. Summary and further extensions

In this paper, we proposed a new IVE of the parameter of the Ito type diffusion processes and developed exact level-α confidence intervals and tests for the parameters which are valid not only for stationary but also for possibly nonlinear and/or non-stationary processes.

Our approach to the parameter estimation is based on the special orthogonal instrumental variables of unit length and has a good efficiency for non-stationary model as well as a modest relative efficiency with respect to MLE for the stationary model. Furthermore it has several desirable small sample properties such as median-unbiasedness even for the possibly nonlinear and non-stationary processes. This is especially important for applications in which the total time span T of the data set is small, since it is precisely in such cases that the inferences must depend critically upon finite sample properties of the estimator.

We note that the desirable finite sample properties of the tests and confidence intervals based on the new IVE can be extended to other class of possibly non-stationary and/or nonlinear continuous-time diffusion type processes. We may consider extension to heteroscedastic diffusion processes of the type

$$dx(t) = a(t, x(t))dt + b(t, x(t))\theta dt + \sigma(t, x(t))dW(t), \qquad t \geq 0,$$

where $\sigma(t, x(t))$ is a known nonsingular matrix of order k. It is straightforward to extend main results of this paper to this case if we modify the definition of the new estimator $\hat{\theta}_c$ and the corresponding pivotal quantity τ_c in (1.3) and (1.4) according to the scheme

$$\sigma^{-1}dx(t) = \sigma^{-1}a(t, x(t))dt + \sigma^{-1}b(t, x(t))\theta dt + dW(t).$$

As an interesting class of stochastic processes, we mention the process generated by the stochastic differential equations with time delay

$$dx(t) = a(x(t-r))dt + b(x(t-r))\theta dt + \sigma(x(t-r))dW(t), \qquad t \geq 0$$

for some $r > 0$. For example, for the special linear model with time delay, Gushchin and Küchler (1999) develop a sophisticated asymptotic likelihood theory for MLE with 11 different limit distributions depending on the value of the parameters. Meanwhile our result provides a simple alternative Gaussian finite sample procedure with possible loss of efficiency but without any stationarity and/or linearity conditions.

In view of the nice finite sample properties and the possible applications in nonstationary processes, a potentially fruitful direction for further investigation will be the extension of the finite sample results to general multi-parameter diffusion processes with $p > k$ and possibly for a discretely observed data.

References

Arato, M., Kolmogorov, A. N. and Sinai, Ya. G. (1962). Evaluation of the parameters of a complex stationary Gauss-Markov process, *Soviet Mathematics Doklady*, **3**, 1368–1371.

Basawa, I. V. and Prakasa Rao, B. L. S. (1980). *Statistical Inference for Stochastic Processes*, Academic Press, London.

Cauchy, A. L. (1836). On a new formula for solving the problem of interpolation in a manner applicable to physical investigations, *Philosophical Magazine*, **8**, 459–486.

Dickey, D. A., Hasza, D. P. and Fuller, W. A. (1984). Testing for unit roots in seasonal time series, *Journal of the American Statistical Association*, **79**, 355–377.

Gushchin, A. and Küchler, U. (1999). Asymptotic inference for a linear stochastic differential equation with time delay, *Bernoulli*, **5**(6), 1059–1098.

Hull, J. C. (1999). *Options, Futures and Other Derivative Securities*, Prentice-Hall, New York.

James, J. and Webber, N. (2000). *Interest Rate Modelling*, Wiley, New York.

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

Kloeden, P. E., Platen, E. and Schurz, H. (1994). *Numerical Solution of Stochastic Differential Equations*, Springer, Berlin.

Kutoyants, Yu. A. (1975). On the hypothesis testing problem and asymptotic normality of stochastic integrals, *Theory of Probability and Its Applications*, **20**, 376–384.

Kutoyants, Yu. A. (2003). *Statistical Inference for Ergodic Diffusion Processes*, Springer, New York.

Liptser, R. S. and Shiryayev, A. N. (1999). *Statistics of Random Processes II*; *Applications*, 2nd ed., Springer, New York.

Prakasa Rao, B. L. S. (1999). *Statistical Inference for Diffusion Type Processes*, Oxford University Press, New York.

Revuz, D. and Yor, M. (1999). *Continuous Martingales and Brownian Motion*, 3rd ed., Springer, New York.

So, B. S. and Shin, D. W. (1999). Cauchy estimators for autoregressive processes with applications to unit root tests and confidence intervals, *Econometric Theory*, **15**, 165–176.