IMES DISCUSSION PAPER SERIES

# A New Approach to the Estimation of Stochastic Differential Equations with an Application to the Japanese Interest Rates

Atsuyuki KOGURE

**Discussion Paper No. 97-E-8**

INSTITUTE FOR MONETARY AND ECONOMIC STUDIES

BANK OF JAPAN

C.P.O BOX 203 TOKYO

100-91 JAPAN

**IMES Discussion Paper Series 97-E-8**

**September 1997**

# A New Approach to the Estimation of Stochastic Differential Equations with an Application to the Japanese Interest Rates

**Atsuyuki KOGURE***

**Abstract**

In this paper we propose a new approach to the estimation of the stochastic differential equations (SDE's) from discretely sampled data and apply the proposed method to examine the volatility of the interest rate on the Japanese call money rate data. For the estimation of the SDE's from discretely sampled data, the traditional approach in the econometric literature is to use a discretization of the original continuous-time model. This approach requires the assumption that the sampling intervals tend to 0, as the sample size increases, in order to insure the consistency of the estimators. In many applications in finance, however, the sampling intervals are not sufficiently small. We propose an estimation procedure which does not require the restrictive assumption. We derive a set of moment equations from the original continuous-time model and use them as the orthogonality conditions for a generalized method of moments (GMM) estimation procedure. The GMM estimates thus obtained are shown to provide the consistency and the asymptotic normality, as the sample size increases, with the sampling intervals being kept fixed. The proposed method is applied to an interest rate model on the Japanese call money rate data and is compared with the traditional approach.

**Keywords:** Call money rate data; Continuous time model; Discretely sampled data; Generalized method of moments; Volatility

**JEL classification:** C13; G12

***** Department of Economics, Faculty of Law and Economics, Chiba University, 1-33 Yayoi-cho, Inage-ku, Chiba-shi 263, Japan

This material is based upon work supported by the Bank of Japan. I am grateful to an anonymous referee for helpful suggestions.

# 1 Introduction

In many fields of finance the use of stochastic differential equations (SDE's) has become an indispensable ingredient to model the dynamics of a variety of random phenomena such as the behavior of asset prices. In real applications the parameters of the equation are unknown and need to be estimated. In most cases we have at our disposal only discretely sampled data on the equation and then it is a common practice to use the discretization of the original continuous-time model for the estimation.

For example, in a comparative study of various single-factor term structure models nested in an SDE of the form

$$dX_t = (\alpha + \beta X_t)dt + \sigma X_t^{\gamma} dW_t, \tag{1.1}$$

Chan, Karolyi, Longstaff and Sanders(1992a, 1992b) used the following discrete-time model

$$X_{t_i} - X_{t_{i-1}} = \alpha + \beta X_{t_{i-1}} + \varepsilon_{t_i}, \quad 1 \leq i \leq n \tag{1.2}$$

$$\varepsilon_{t_i} | X_{t_{i-1}} \sim N\left(0, \sigma^2 X_{t_{i-1}}^{2\gamma}\right)$$

where $t_i$'s are sampling points when the process is observed.[1] They apply the generalized method of moments (GMM) estimation to the discrete-time model (1.2) to estimate the parameters of the original continuous-time model (1.1).

The GMM estimates thus obtained are known to be consistent and asymptotically normally-distributed under some regularity conditions; see Hansen (1982). However this is under the assumption that the data are generated from the discrete-time model (1.2). Thus we cannot expect that the GMM estimators converge to the true parameter values of the continuous-time model (1.1), as $n \to \infty$. The same is true if the maximum likelihood (ML) estimation method is used for the GMM estimation. Theoretically this point may be defended by assuming that

$$\max\{(t_i - t_{i-1}), 1 \leq i \leq n\} \to 0 \tag{1.3}$$

as $n \to \infty$. Under this assumption the process defined by (1.2) converges weakly to the process defined by (1.1) as $n \to \infty$ and thus the GMM estimators, as well as the ML estimators, are consistent.

[1] This type of discretization scheme, known as the Euler's method, seems popular in finance. Various schemes such as Heun and Runge-Kutta methods have been proposed in the other fields. Recently Ozaki (1992) introduced a new discretization scheme which may yield computationally stable discrete-time models.

But it should be noted that in many applications the sampling points $t_i$'s are fixed and $\max\{(t_i - t_{i-1}), 1 \leq i \leq n\}$ is not sufficiently small. Then the discrete-time model (1.2) may misrepresent the true nature of model (1.1).

Statistical inference for SDE's has been relatively well developed under the assumption (1.3); see Praksa Rao (1988) for a survey in the area. However, there seem to be very few studies which address the estimation problem of SDE's without assuming (1.3) A rare exceptions is provided by Lo (1988) in which an ML estimation is considered based on the exact conditional probability density of $X_{t_i}$ given the level of $X_{t_{i-1}}$ which is obtained as a solution of the Kolmogorov's forward equation. However, the Lo's method seems limited mainly to the theoretical issue because even a simple equation such as (1.1) cannot provide explicit solutions aside from a few exceptional cases such that $\gamma = 0$. [2]

The main purpose of the present study is to present an alternative estimation procedure for SDE's that is free from the discretizations such as (1.2) and does not require (1.3). We derive a set of moment equations from the original continuous-time model and use them as the orthogonality conditions for a GMM procedure. The GMM estimates thus obtained are shown to converge to the true parameter values of the continuous-time model without (1.3).

In Section 2 we present a model and a lemma to derive a set of moment equations. In Section 3 we propose a GMM estimation procedure and present some asymptotic results. In Section 4 we apply the proposed method to estimate the parameter $\gamma$ of (1.1) on the Japanese call money rate data and compare the results with those based on the discretization. In Section 5 we give some brief concluding remarks.

## 2 ORTHOGONALITY CONDITIONS

In this section we derive orthogonality conditions from a given SDE to be used for the GMM estimation.

Let $\{X_t,\ t \geq 0\}$ be a real-valued process satisfying the stochastic differential equation

$$dX_t = a(X_t,\ \theta)dt + b(X_t,\ \theta)dW_t, \quad t \geq 0, \tag{2.1}$$

where $\{W_t,\ t \geq 0\}$ is a standard Brownian motion and $\theta$ represents a vector of unknown parameters

[2] When $\gamma = 0.5$, the solution is explicitly given as a noncentral chi-square density (Cox Ingersoll and Ross, 1985, p.391). In this case, however, it seems difficult to apply the Lo's method to the actual data since the noncentral chi-square density involves the modified Bessel function.

in a compact subset $\Theta$ of the $p$-dimensional Euclidean space. We suppose that $a(X_t,\ \theta)$ and $b(X_t,\ \theta)$ are known real-valued continuous functions on $R \times \Theta$

We are concerned with the estimation problem about $\theta$ when $\{X_t,\ t \geq 0\}$ is observed at equidistant sampling points $\{t_i,\ 1 \leq i \leq n\}$ where $t_{i+1} - t_i = \Delta$ for all $i \geq 1$. Unlike the traditional approach, we do not assume that $\Delta \to 0$.

Henceforth we shall make the following assumptions.

**Assumption 2.1**

*(i) The initial random variable $X$ is such that $E(X^2) < \infty$.*

*(ii) For each $\theta$ there is a strictly increasing continuous function $\rho$ with positive constant $K$ such that for each $x, y$*

$$|b(x,\theta) - b(y,\theta)| \leq \rho(|x-y|)$$

*where $\rho(0) = 0$ and for all $\epsilon > 0$*

$$\int_0^\epsilon \rho^{-2}(u)du = \infty.$$

*(iii) For each $\theta$ there is a strictly increasing and concave function $\kappa$ with $|a(x) - a(y)| \leq \kappa(|x-y|)$ where $\kappa(0) = 0$ and for all $\epsilon > 0$*

$$\int_0^\epsilon \kappa^{-1}(u)du = \infty.$$

Under Assumptions 2.1 equation (2.1) has a unique (weak) solution with probability 1 and the solution is a continuous local martingale (see Durrett, 1996, Theorem (5.3.3) and §8.5).

**Assumption 2.2** *For all $t > 0$ $X_t$ has the same distribution as the initial random variable $X$.*

This assumption implies that $\{X_t,\ t \geq 0\}$ is stationary. Now we have the following lemma:

**Lemma 2.1** *Let Assumptions 2.1 and 2.2 be satisfied. Let $h$ be a twice continuously differentiable function on $\mathbb{R}^1$ into $\mathbb{R}^1$ such that $E[|h(X_t)|] < \infty$. Then*

$$E[h'(X_t)a(X_t,\theta) + (1/2)h''(X_t)b^2(X_t,\theta)] = 0. \tag{2.2.}$$

*Proof.* By Ito formula (e.g. Durrett, 1996, Theorem 2.7.1) $Y_t \equiv h(X_t)$ satisfies integral of the form

$$Y_t - Y_0 = \int_0^t \{h'(X_s)a(X_s,\theta) + (1/2)h''(X_s)b(X_s,\theta)ds + \int_0^t h''(X_s)b^2(X_s,\theta)dW_s.$$

Taking expectations we have

$$(2.3) \qquad E[\int_0^t \{h'(X_s)a(X_s,\theta) + (1/2)h''(X_s)b^2(X_s,\theta)\}ds]$$

$$= E[\int_0^t h''(X_s)b^2(X_s,\theta)dW_s] - E[Y_t - Y_0] = 0.$$

By Fubini's Theorem

$$(2.4) \qquad E[\int_0^t \{h'(X_s)a(X_s,\theta) + (1/2)h''(X_s)b^2(X_s,\theta)ds]$$

$$= \int_0^t E[\{h'(X_s)a(X_s,\theta) + (1/2)h''(X_s)b^2(X_s,\theta)]ds$$

$$= tE[\{h'(X_0)a(X_0,\theta) + (1/2)h''(X_0)b^2(X_0,\theta)].$$

Combining (2.3) with (2.4) we have the desired conclusion. □

# 3 THE ESTIMATION

The sample version of equation (2.2) is

$$(3.1) \qquad (1/n)\sum_{i=1}^{n}[h'(X_{t_i})a(X_{t_i},\theta) + (1/2)h''(X_{t_i})b(X_{t_i},\theta)] = 0.$$

We propose to use (3.1) as a device for generating a set of orthogonality conditions for the GMM estimation.

For any integer $r \geq p$ let $h_1,\ h_2,\ ,\ldots,\ h_r$ be a set of twice continuously differentiable functions. For each $1 \leq i \leq n$ write $X_i \equiv X_{t_i}$ and put

$$M(X_i,\ \theta) = \begin{pmatrix} h_1'(X_i)a(X_i,\theta) + \frac{1}{2}h_1''(X_i)b(X_i,\theta) \\ h_2'(X_i)a(X_i,\theta) + \frac{1}{2}h_2''(X_i)b(X_i,\theta) \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ h_r'(X_i)a(X_i,\theta) + \frac{1}{2}h_r''(X_i)b(X_i,\theta) \end{pmatrix}.$$

Let $\theta^*$ denote the true value of $\theta$. Then by Lemma 2.1 we have a set of population moment equations

$$E[M(X_i,\theta^*)] = 0.$$

Define the sample moment as

$$M_n(\theta) \equiv \frac{1}{n}\sum_{i=1}^{n} M(X_i,\ \theta).$$

Then the minimizer $\hat{\theta}$ of

$$Q_n(\theta) \equiv M_n(\theta)' M_n(\theta).$$

is a member of a class of GMM estimators of $\theta^*$.

Note that under the conditions assumed in Section 2 $X_t$ is a continuous stationary Markov process. Then for each $t \geq 0$ define the transition operator $H_t$ of $X$ by

$$H_t f(x) = E[f(X_t)|X_0 = x], \quad -\infty < x < \infty$$

for any bounded Borel measurable function $f$ on $\mathbb{R}^1$. Now define

$$|H_t|_2 = \sup_{\{f:E[f(X)]=0\}} \frac{E[(H_t f)^2]}{E[f^2]}^{1/2}.$$

Then $\{X_t,\ t \geq 0\}$ is said to satisfy the condition $G_2(s, \alpha)$ if there is $s > 0$ such that $|H_s|_2 \leq \alpha$ for some $0 < \alpha < 1$ and $s > 0$ (Banon, 1978).

To derive the asymptotic properties of $\hat{\theta}$ we make the additional assumption on $\{X_t,\ t \geq 0\}$:

**Assumption 3.1** *$\{X_t,\ t \geq 0\}$ satisfies the condition $G_2(s, \alpha)$ for some $0 < \alpha < 1$ and $s > 0$.*

Note that the $G_2$ condition implies that $\{X_{t_i}, i \geq 1\}$ is the exponentially strong-mixing condition (e.g. Rosenblatt, 1971, p.200). Thus, under Assumptions 2.1-2.3 and 3.1, $\{X_i, 1 \leq i \leq n\}$ can be regarded as observations on a discrete stationary process satisfying the exponentially strong-mixing condition.

Conditions for the consistency and asymptotic normality of $\hat{\theta}$ are presented in the following two theorems. These theorems imply that it is possible to estimate the parameters of SDE's without appealing to the discretization. The proofs of the theorems are given in Appendix.

**Theorem 3.1** *Let Assumptions 2.1, 2.2 and 3.1 be satisfied and assume in addition that:*

*(i) for each $1 \leq j \leq r$ and for each $x \in R$ $m_j(x, \theta)$ is continuous on $\Theta$ where $m_j(x, \theta)$ is the $j$th element of $M(x, \theta)$;*

*(ii) for each $j$*

$$E[sup_{\theta \in \Theta} |m_j(X_1, \theta)|] < \infty;$$

*(iii) $\bar{M}(\theta) \equiv E[M(X_1, \theta)]$ is continuous on $\Theta$.*

*(iv) $\bar{M}(\theta)$ has a unique $0$ at $\theta = \theta^*$.*

*Then* $\hat{\theta} \to \theta^*$ *in pr. as* $n \to \infty$.

**Theorem 3.2** *Let the conditions of Theorem 3.1 be satisfied and assume in addition that:*

*(i)* $\Theta$ *is convex and* $\theta^*$ *is an interior point of* $\Theta$

*(ii)* $(\partial/\partial\theta)Q_n(\theta)$ *and* $(\partial/\partial\theta)(\partial/\partial\theta')Q_n(\theta)$ *are well defined as vector and matrix, respectively, of random functions on* $\Theta$*;*

*(iii) for* $1 \leq j \leq r$ *and for* $1 \leq k_1, k_2 \leq p$

$$E[sup_{\theta\in\Theta}|(\partial/\partial\theta_{k_1})m_j(X_1,\theta)|] < \infty$$

*and*

$$E[sup_{\theta\in\Theta}|(\partial/\partial\theta_{k_1})(\partial/\partial\theta_{k_2})m_j(X_1,\theta)|] < \infty;$$

*(iv) for* $1 \leq j \leq r$ *and for* $1 \leq k_1, k_2 \leq p$

$$E[(\partial/\partial\theta_{k_1})m_j(X_1,\theta)] = (\partial/\partial\theta_{k_1})E[m_j(X_1,\theta)]$$

*and*

$$E[(\partial/\partial\theta_{k_1})(\partial/\partial\theta_{k_2})m_j(X_1,\theta)] = (\partial/\partial\theta_{k_1})(\partial/\partial\theta_{k_2})E[m_j(X_1,\theta)];$$

*(v) The matrix* $A \equiv (\partial/\partial\theta)(\partial/\partial\theta')Q(\theta^*)$ *is non-singular.*

*Then*

$$\Sigma \equiv \lim_{n\to\infty} Var\left((1/n)\sum_{i=1}^{n}\left(\sum_{j=1}^{r}E[m_j(\theta^*)](\partial/\partial\theta)m_j(X_i,\theta)\right)\right)$$

*converges, and if*

*(vi)* $\Sigma$ *is non-singular,*

*then* $\sqrt{n}(\hat{\theta}-\theta^*) \to N(0, 4A^{-1}\Sigma A^{-1})$ *in distribution as* $n \to \infty$.

# 4 AN APPLICATION TO THE INTEREST RATES MODEL

In this section we apply the proposed method to estimate the parameter $\gamma$ of the stochastic differential equation (1.1) on the Japanese call money rate data. We also compare the results with those based on the discretization.

Let $X_t$ represent the interest rate at time $t$. Then the parameter $\gamma$ is the elasticity of the volatility to the the level of $X_t$. Many alternative term structure models are nested in (1.1) and the $\gamma$ is the most important parameter to distinguish among the alternative models. When $\gamma$ is 0, equation (1.1) represents a model of constant volatility which is used in Vacisek (1977). When $\gamma$ is greater than 0, model (1.1) becomes a model of varying volatility. The case where $\gamma$ is 0.5 corresponds to the square root process of Cox, Ingersoll and Ross (1985). The case where $\gamma$ is 1 corresponds to the process which appears in Brennan and Schwartz (1980).

## 4.1 THE IDENTIFIABILITY

With orthogonality conditions such as (3.1) Condition (iv) of Theorem 3.1 is not satisfied for the parameter vector $\theta \equiv (\alpha, \beta, \sigma^2, \gamma)$ unless some value of $\alpha, \beta$ and $\sigma$ is known. However our method is still applicable if our main interest is in the estimation of $\gamma$. Now define the transformed parameters

$$\mu \equiv \alpha/\sigma^2, \nu \equiv \beta/\sigma^2.$$

and let $h$ be a twice continuously differentiable function. Then by Lemma 2.1 we have

$$E[h'(X_1)(\mu + \nu X_1) + (1/2)h''(X_1)X_1^{2\gamma}] = 0.$$

It is a moot point what functional forms are appropriate for $h$. One idea would be to choose $h$ so that the asymptotic variance matrix is minimized in some sense. Here we simply choose it to be a class of polynomials, which correspond to using ordinary moments of $X_1$.

Putting, for $j = 1$, 2 and 3

$$h_j(x) = x^j/j$$

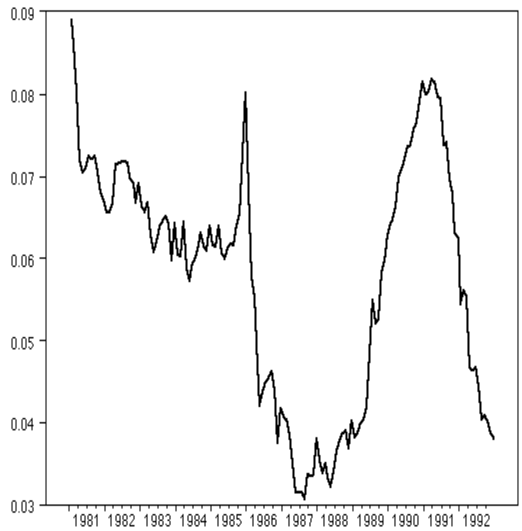
we have

$$\begin{cases} \mu + E[X_1]\nu = 0 \\ E[X_1]\mu + E[X_1^2]\gamma + (1/2)E[X_1^{2\gamma}] = 0 \\ E[X_1^2]\mu + E[X_1^3]\gamma + E[X_1^{2\gamma+1}] = 0. \end{cases}$$

By eliminating $\mu$ and $\nu$ from the above set of equations we obtain

$$Var(X_1)E[X_1^{2\gamma+1}] - (1/2)Cov(X_1, X_1^3)E[X_1^{2\gamma}] = 0. \tag{4.1}$$

Figure 1: Monthly Call Money Rates from January 1981 through December 1992




Thus a sufficient condition for $(\mu, \nu, \gamma)$ to be identified is that equation (4.1) has a unique zero. It seems difficult to check whether this condition holds in general. However, for the cases in which $\gamma$ is 0 and $\gamma$ is 0.5, two exceptions where the marginal distribution of $\{X(t),\ t \geq 0\}$ is known, a simple algebra shows that this condition is satisfied in both case.

### 4.2 THE DATA

Our data is a monthly call money rate series in the Japanese market, which runs from from January 1981 to December 1992, yielding 144 values. It is collected from *Bank of Japan Monthly Statistics*. It is the rate for the overnight repayment and has often been used as a typical short-term interest rate in the past empirical studies of the Japanese economy.

Figure 1 plots the data and Table 1 and 2 summarize its characteristics.

Table 1: Basic Statistics of Monthly Call Money Rates from January 1981 through December 1992

| $n$ | Mean | Std. Dev. | Minimum | Maximum | Skewness | Kurtosis |
|---|---|---|---|---|---|---|
| 144 | .05786 | .01494 | .03090 | .08908 | -0.24191 | -1.07237 |

Table 2: Autocorrelations of Monthly Call Money Rates from January 1981 through December 1992

| $\rho_1$ | $\rho_2$ | $\rho_3$ | $\rho_4$ | $\rho_5$ | $\rho_6$ | $\rho_7$ | $\rho_8$ | $\rho_9$ | $\rho_{10}$ | $\rho_{11}$ | $\rho_{12}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| .96 | .91 | .86 | .81 | .77 | .73 | .68 | .63 | .58 | .53 | .47 | .42 |

### 4.3 ESTIMATION RESULTS

In addition to $h_1, h_2, h_3$ used in Section 4.1 we tried $h_4, h_5, h_6, h_7, h_8$ where for $j = 4, 5, 6, 7, 8$

$$h_j(x) \equiv \frac{x^{j-2\gamma}}{(j-2\gamma)\sigma^2}.$$

The results are summarized in Table 3. The values in the parentheses are asymptotic t-statistics.

We also applied the ML and GMM methods used in Chan, Karolyi, Longstaff and Sanders (1992a, 1992b) which adopted the orthogonality conditions derived from the discrete-time model (1.2). The results are summarized in Table 4 with asymptotic t-statistics in parentheses. It must be noted that these t-statistics are calculated under the asymptotic situation where $n \to \infty$ while $\Delta$ is fixed. Thus some reservation is needed to interpret these values since the proper asymptotic set-up in this case should be such that $\Delta \to 0$ as $n \to \infty$.

Table 3: Estimates by the proposed method

| orthogonality conditions | $\gamma$ | $\mu$ | $\nu$ |
|---|---|---|---|
| case 1 | 0.5075 | 9.0907 | -177.9801 |
| | (184264.48) | (17.86) | (-21.61) |
| case 2 | 0.5187 | 8.7447 | -162.9475 |
| | (3586874.07) | (16.23) | (-19.95) |
| case 3 | 0.535 | 13.6757 | -238.28 |
| | (16094558.10) | (13.73) | (-17.62) |

## 5 CONCLUDING REMARKS

The GMM estimation proposed in this paper provides us with an important alternative to existing methods for estimating the stochastic differential equations when the sampling intervals are not

Table 4: Estimates based on the discrete-time model

| | $\gamma$ | $\mu$ | $\nu$ | $\alpha$ | $\beta$ | $\sigma^2$ |
|---|---|---|---|---|---|---|
| ML | 0.5878 | 4.6589 | -102.5972 | 0.01534 | -0.3377 | 0.0033 |
| | (5.53) | | | (2.59) | (-3.07) | (1.64) |
| GMM | 0.5026 | 8.8237 | -188.3432 | 0.0179 | -0.3823 | 0.0020 |
| | (1.60) | | | (0.34) | (-0.46) | (0.55) |

sufficiently small. We have shown that the consistency and the asymptotic normality can be established without the traditional assumption that the sampling intervals tend to 0.

We have also applied the proposed method to estimate the volatility parameter of the interest rate model on the Japanese call money rate data and compared the results with those based on the discretization.

The proposed method has a limitation that all the information resulting from its orthogonality conditions is that contained in the marginal distribution of the SDE. Since the correspondence between the SDE and its marginal distribution and is not unique in general (Ozaki, 1985), this may cause some difficulty in identifying the parameters. However, as indicated in Section 4, it is still possible to identify some of the important parameters and thus to apply the proposed method.

# APPENDIX

## PROOF OF THEOREM 3.1

We use the following fundamental theorem taken from Bierens (1994, page 64, Theorem 4.2.1).

**Theorem A.1 (Bierens)**

*Let $((Q_n(\theta))$ be a sequence of random functions on a compact set $\Theta \subset R^m$ such that for a continuous real function $Q(\theta)$ on $\Theta$,*

$$\sup_{\theta\in\Theta} |Q_n(\theta) - \bar{Q}(\theta)| = o_p(1). \tag{A.1}$$

*Let $\theta_n$ be any random vector in $\Theta$ satisfying*

$$Q_n(\theta_n) = \inf_{\theta\in\Theta} Q_n(\theta)$$

*and let $\theta_0$ be a unique point in $\Theta$ such that*

$$Q(\theta_0) = \inf_{\theta \in \Theta} Q(\theta).$$

*Then $\theta_n \to \theta_0$ in probability.*$\square$

To obtain the desired conclusion, define

$$\bar{Q}(\theta) \equiv \bar{M}(\theta)'\bar{M}(\theta).$$

By conditions (iii) and (iv) of Theorem 3.1 $\bar{Q}(\theta)$ is continuous on $\Theta$ and has a unique minimum at $\theta = \theta^*$. Thus it remains to show (A.1). Now put

$$L \equiv max_{1\le j\le r} E[sup_{\theta\in\Theta}|m_j(X_1,\theta)|].$$

Then we have

$$\begin{aligned}
&\bar{Q}(\theta) - Q_n(\theta) \\
&\quad = (M_n(\theta) - \bar{M}(\theta))'(M_n(\theta) - \bar{M}(\theta)) + 2(M_n(\theta) - \bar{M}(\theta))'\bar{M}(\theta)) \\
&\quad = ((M_n(\theta) - \bar{M}(\theta)))'((M_n(\theta) - \bar{M}(\theta))) + 2((M_n(\theta) - \bar{M}(\theta)))'\bar{M}(\theta) \\
&\quad \le r^2\{max_{1\le j\le r} sup_{\theta\in\Theta}|(1/n)\sum_{i=1}^{n}(m_j(X_i,\theta) - E[m_j(X_i,\theta)])|\}^2 \\
&\quad\quad + 2r^2L\{max_{1\le j\le r} sup_{\theta\in\Theta}|(1/n)\sum_{i=1}^{n}(m_j(X_i,\theta) - E[m_j(X_i,\theta)])|\}
\end{aligned}$$

Thus it suffices to show that for each $1 \le j \le r$

$$sup_{\theta\in\Theta}|(1/n)\sum_{i=1}^{n}(m_j(X_i,\theta) - E[m_j(X_i,\theta)])| = o_p(1),$$

which follows from Theorem 6.3.4 of Bierens (1994) in view of Assumption 3.1 and conditions (i)-(iv) of Theorem 3.1. $\square$

## PROOF OF THEOREM 3.2

We use the following fundamental theorem taken from Bierens (1994, page 65, Theorem 4.2.2).

**Theorem A.2 (Bierens)**

*Let the conditions of Theorem A.1 be satisfied and assume in addition that*

*(i) $\Theta$ is convex and $\theta_0$ is an interior point of $\Theta$;*

*(ii) $(\partial/\partial\theta')Q_n(\theta)$ and $(\partial/\partial\theta)(\partial/\partial\theta')Q_n(\theta)$ are well defined as vector and matrix, respectively of random functions on $\Theta$;*

*(iii)* $\sqrt{n}(\partial/\partial\theta')Q_n(\theta_0) \to N_m(0, A_1)$ *where* $A_1$ *is a positive semi-definite* $m \times m$ *matrix;*

*(iv) for* $k_1, k_2 = 1, 2, \ldots, m$,

$$(A.2) \qquad \sup_{\theta} |(\partial/\partial\theta_{k_1})(\partial/\partial\theta_{k_2})Q_n(\theta) - (\partial/\partial\theta_{k_1})(\partial/\partial\theta_{k_2})Q(\theta)| = o_p(1),$$

*where the sup is taken on a neighborhood of* $\theta_0$ *and the limit function involved is continuous at* $\theta_0$*;*

*(v) the matrix* $A_2 = (\partial/\partial\theta)(\partial/\partial\theta')Q(\theta_0)$ *is non-singular.*

*Then* $\sqrt{n}(\theta_n - \theta_0) \to N_m\left(0, A_2^{-1}A_1A_2^{-1}\right)$. □

First, notice that, for $1 \leq k_1, k_2 \leq p$, (A.2) follows from the same argument as in Proof of Theorem 3.1 using conditions (iii) and (iv) of Theorem 3.2. Thus we only need to show that

$$\sqrt{n}(\partial/\partial\theta)Q_n(\theta^*) \to N(0, 4\Sigma) \text{ in distribution.}$$

To see this let $d$ be an arbitrary $p$-dimensional vector of constant elements and let

$$Z_i \equiv \sum_{j=1}^{r} E[m_j(\theta^*)](\partial/\partial)m_j(X_i, \theta^*).$$

Then the asymptotic distribution of

$$d'(\partial/\partial\theta)Q_n(\theta^*) = 2\sum_{j=1}^{r}(1/n)\sum_{i=1}^{n} m_j(X_i, \theta^*)(1/n)\sum_{i=1}^{n} d'(\partial/\partial\theta)m_j(x_i, \theta^*)$$

is equivalent to that of $(2/n)\sum_{i=1}^{n} d'Z_i$ since $(1/n)\sum_{i=1}^{n} m_j(X_i, \theta) = E[m_j(\theta)] + o_p(1)$ for each $j$. Therefore it suffices to show that

$$(A.3) \qquad (1/\sqrt{n})\sum_{i=1}^{n} d'Z_i \to N(0, d'\Sigma d) \text{ in distribution}$$

for each $d \neq 0$.

Notice that $\{d'Z_i\}$satisfies the conditions of Theorem 18.5.2 of Ibragimov and Linnik (1971) and that $E(d'Z_i) = 0$. Thus

$$\begin{aligned} \sigma^2(d) &\equiv lim_{n\to\infty} Var\left((1/n)\sum_{i=1}^{n} d'Z_i\right) \\ &= d' lim_{n\to\infty} Var\left((1/n)\sum_{i=1}^{n} Z_i\right) d \end{aligned}$$

converges for each $d \neq 0$, and so, $\Sigma = lim_{n\to\infty} Var\left((1/n)\sum_{i=1}^{n} Z_i\right)$ converges. Then it follows from (iv) that $\sigma^2(d) = d'\Sigma d > 0$ for any $d \neq 0$, which implies (A.3). □

## References


Banon, G. (1974): "Nonparametric Identification for Diffusion Processes," *SIAM J. of Control Optim.*, 16, 380-395.

Bierens, H. J. (1994): *Topics in Advanced Econometrics.* Cambridge: Cambridge University Press.

Brennan, M. J., and E. S. Schwartz (1980): "Analyzing Convertible Bonds," *J. of Financial and Quantitative Analysis*, 15, 907-929.

Chan, K. C., G.A. Karolyi, F. A. Longstaff, and A. B. Sanders. (1992a): "The Volatility of Japanese Interest Rates: A Comparison of Alternative Term Structure Models," *Pacific-Basin Capital Markets Research*, 3, 119-136.

Chan, K. C., G.A. Karolyi, F. A. Longstaff, and A. B. Sanders. (1992b): "An Empirical Comparison of Alternative Models of the Short-term Interest Rate," *J. of Finance*, 47, 1209-1227.

Cox, J. C., J. E. Ingersoll, and S. A. Ross (1985): "A Theory of the Term Structure of Interest Rates," *Econometrica*, 53, 385-407.

Durrett, E. (1996): *Stochastic Calculus: A Practical Introduction.* Boca Raton, Florida: CRS Press

Hansen, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029-1054.

Ibragimov, I. A. and Yu. V. Linnik (1971): *Independent and Stationary Sequences of Random Variables.* Groningen: Wolters-Noordhoff Publishing.

Lo, A. W. (1988): "Maximum Likelihood Estimation of Generalized Itô Processes with Discretely Sampled Data," *Econometric Theory*, 4, 231-247.

Ozaki, T. (1985): "Non-linear Time Series Models and Dynamical Systems," In em Handbook of Statistics, 5 (Edited by E. J. Hannan, et al), 25-83, Amsterdam:North-Holland.

Ozaki, T. (1992): "A Bridge between Nonlinear Time Series Models and Nonlinear Stochastic Dynamical Systems: A Local Linearization Approach," *Statistica Sinica*, 2, 113-135.

Praksa Rao, B. L. S. (1988): "Statistical Inference from Sampled Data for Stochastic Processes" In *Contemporary Mathematics*, 80 (Edited by N. U. Prabhu), 249-284, American Mathematical Society.

Rosenblatt, M. (1971): *Markov Processes, Structure and Asymptotic Behavior.* New York: Springer-Verlag.

Vasicek, O. (1979): "An Equilibrium Characterization of the Term Structure," *J. of Financial Economics*, 5,177-188.