Ann. Inst. Statist. Math.
Vol. 53, No. 1, 125–141 (2001)

ESTIMATING FUNCTIONS FOR NONLINEAR TIME SERIES MODELS

S. AJAY CHANDRA AND MASANOBU TANIGUCHI

Department of Mathematical Science, Graduate School of Engineering Science, Osaka University, Machikaneyama-cho 1-3, Toyonaka, Osaka 560-8531, Japan,
e-mail: chandra@sigmath.es.osaka-u.ac.jp; taniguti@sigmath.es.osaka-u.ac.jp

(Received April 10, 2000; revised August 17, 2000)

Abstract. This paper discusses the problem of estimation for two classes of nonlinear models, namely random coefficient autoregressive (RCA) and autoregressive conditional heteroskedasticity (ARCH) models. For the RCA model, first assuming that the nuisance parameters are known we construct an estimator for parameters of interest based on Godambe's asymptotically optimal estimating function. Then, using the conditional least squares (CLS) estimator given by Tjøstheim (1986, *Stochastic Process. Appl.*, **21**, 251–273) and classical moment estimators for the nuisance parameters, we propose an estimated version of this estimator. These results are extended to the case of vector parameter. Next, we turn to discuss the problem of estimating the ARCH model with unknown parameter vector. We construct an estimator for parameters of interest based on Godambe's optimal estimator allowing that a part of the estimator depends on unknown parameters. Then, substituting the CLS estimators for the unknown parameters, the estimated version is proposed. Comparisons between the CLS and estimated optimal estimator of the RCA model and between the CLS and estimated version of the ARCH model are given via simulation studies.

Key words and phrases: Nonlinear time series models, random coefficient autoregressive models, autoregressive conditional heteroskedasticity models, conditional least squares estimator, estimating function, classical moment estimator, asymptotic optimality.

1. Introduction

In the last two decades, wide classes of nonlinear time series models have been studied, for example, Engle (1982), Nicholls and Quinn (1982), Tjøstheim (1986), Tong (1990), Pötscher and Prucha (1997), and Hafner (1998). One of these classes which has received a considerable amount of attention is that of random coefficient models. These are important in the engineering and econometrics literature since many data sets in fields such as hydrology, metrology and biology exhibit occasional sharp spikes, which cannot be sufficiently explained by classical linear time series models. Such features arise when the coefficients of the model considered have random characteristics. This situation led to a consideration of random coefficient autoregressive (RCA) models. For these models, Nicholls and Quinn (1982) developed a rigorous statistical theory which covers the case of vector-valued autoregressive models when the random coefficients and error process are Gaussian and mutually independent. Feigin and Tweedie (1985) studied the stationarity, ergodicity and finiteness of moments for these models.

Another important class of nonlinear time series models, introduced by Engle (1982) is the class of autoregressive conditional heteroskedasticity (ARCH) models for conditional variances which proved to be extremely useful in analyzing economic time series. Since their introduction, ARCH type models have become perhaps the most popular and extensively studied financial econometric models. The literature on the subject is so vast that we will restrict ourselves to directing the reader to fairly comprehensive reviews by Bollerslev *et al.* (1992) and Shephard (1996). A detailed treatment of ARCH models at a textbook level is also given by Gouriéroux (1997).

In the estimation of nonlinear time series models, Tjøstheim (1986) proposed a conditional least squares (CLS) estimator, and elucidated its asymptotics. Recently, Hwang and Basawa (1998) discussed the estimation problems based on such and weighted least squares estimators for RCA models in cases of known and unknown nuisance parameters and, established their asymptotics in which the random coefficients are permitted to be correlated with the error process. They also studied their asymptotic behaviour through simulation. As another method, Godambe (1960, 1985) developed the theory of estimating function for stochastic models, and introduced the concept of asymptotically optimal estimating function (see also Heyde (1997)). In econometrics, generalized method of moments (GMM) estimation developed by Hansen (1982), is widely used in the theory and applications (see also Newey and McFadden (1994) and Wooldridge (1994)). The GMM estimation and Godambe's estimating function method are essentially the same. But these two methods have been developed independently.

In this paper, we consider the problem of estimation for RCA and ARCH models. For the RCA model, first we construct Godambe's optimal estimator for the parameter of interest assuming that the nuisance parameters are known. Then, using the CLS estimator and classical moment estimators for the nuisance parameters we propose the estimated version of Godambe's optimal estimator. Such results are extended to a vector case. Next, we turn to discuss the problem of estimating the ARCH model with unknown parameter vector. We construct an estimator based on Godambe's optimal estimator. Then, an estimated version of this estimator is proposed via the CLS estimator. These new estimators are expressed in closed forms. We study the asymptotic behaviour of the estimators. Also, we compare, via simulation, the mean square error of the estimators in various situations.

2. Conditional least squares estimation

In this section, we present Tjøstheim's (1986) results which were obtained by reformulating and extending the arguments of Klimko and Nelson (1978) to nonlinear time series.

Let $\{\boldsymbol{X}_t; t = 0, \pm 1, \dots\}$ be a stochastic process taking values in $\boldsymbol{R}^p$ and defined on a probability space $(\Omega, \mathcal{F}, P)$. Here, $\{\boldsymbol{X}_t\}$ is possibly a strictly stationary ergodic nonlinear time series. In addition, suppose that $E\{\|\boldsymbol{X}_t\|^2\} < \infty$ so that $\{\boldsymbol{X}_t\}$ is second order stationary, where $\|\cdot\|$ denotes the Euclidean norm. We assume that observations $(\boldsymbol{X}_1, \dots, \boldsymbol{X}_n)$ are available. The probability distribution of $(\boldsymbol{X}_1, \dots, \boldsymbol{X}_n)$ is specified by unknown parameter $\boldsymbol{\theta}^0 = (\theta_1^0, \dots, \theta_q^0)' \in \Theta \subset \boldsymbol{R}^q$. Then consider a general real-valued penalty function $Q_n(\boldsymbol{\theta}) = Q_n(\boldsymbol{X}_1, \dots, \boldsymbol{X}_n; \boldsymbol{\theta})$ depending on the observations and on a parameter vector $\boldsymbol{\theta} \in \Theta$.

We specify the penalty function. Let $\mathcal{F}_t(m)$ be the σ-field generated by $\{\boldsymbol{X}_s; t-m \leq s \leq t\}$, where m is an appropriate integer. If $\{\boldsymbol{X}_t\}$ is a nonlinear autoregressive model of order k, we can take $m = k$. Let $\tilde{\boldsymbol{X}}_{t|t-1}(\boldsymbol{\theta}) = E_\theta\{\boldsymbol{X}_t \mid \mathcal{F}_{t-1}(m)\}$ be an optimal one-step least squares predictor of $\boldsymbol{X}_t$ based on $\boldsymbol{X}_{t-1}, \ldots, \boldsymbol{X}_{t-m}$.

Consider the penalty function

$$Q_n(\boldsymbol{\theta}) = \sum_{t=m+1}^{n} \{\boldsymbol{X}_t - \tilde{\boldsymbol{X}}_{t|t-1}(\boldsymbol{\theta})\}'\{\boldsymbol{X}_t - \tilde{\boldsymbol{X}}_{t|t-1}(\boldsymbol{\theta})\}.$$

The CLS estimator $\hat{\boldsymbol{\theta}}_n^{(CL)}$ of $\boldsymbol{\theta}^0$ is defined by $\hat{\boldsymbol{\theta}}_n^{(CL)} = \arg\min_{\theta\in\Theta} Q_n(\boldsymbol{\theta})$.

Then, under some regularity conditions of Cramer type, Tjøstheim ((1986), 254–256) showed that

$$\hat{\boldsymbol{\theta}}_n^{(CL)} \longrightarrow^{a.s.} \boldsymbol{\theta}^0, \quad \text{and} \quad \sqrt{n}(\hat{\boldsymbol{\theta}}_n^{(CL)} - \boldsymbol{\theta}^0) \rightarrow^{d} \mathcal{N}(\boldsymbol{0}, \boldsymbol{U}^{-1}\boldsymbol{R}\boldsymbol{U}^{-1}), \quad \text{as } n \rightarrow \infty,$$

where

$$\boldsymbol{U} = E\left\{\frac{\partial}{\partial\boldsymbol{\theta}}\tilde{\boldsymbol{X}}'_{t|t-1}(\boldsymbol{\theta}^0)\frac{\partial}{\partial\boldsymbol{\theta}}\tilde{\boldsymbol{X}}_{t|t-1}(\boldsymbol{\theta}^0)\right\}, \quad \text{and}$$

$$\boldsymbol{R} = E\left\{\frac{\partial}{\partial\boldsymbol{\theta}}\tilde{\boldsymbol{X}}'_{t|t-1}(\boldsymbol{\theta}^0)\{\boldsymbol{X}_t - \tilde{\boldsymbol{X}}_{t|t-1}(\boldsymbol{\theta}^0)\}\{\boldsymbol{X}_t - \tilde{\boldsymbol{X}}_{t|t-1}(\boldsymbol{\theta}^0)\}'\frac{\partial}{\partial\boldsymbol{\theta}}\tilde{\boldsymbol{X}}_{t|t-1}(\boldsymbol{\theta}^0)\right\}.$$

The CLS estimation can be applied to a wide class of nonlinear time series. However, $\hat{\boldsymbol{\theta}}_n^{(CL)}$ in general is not asymptotically efficient. Thus, we next discuss an asymptotically efficient estimator proposed by Godambe (1985).

3. Estimating function approach

Godambe (1985) developed the theory of estimating function for stochastic models, and introduced the concept of asymptotically optimal estimating function. The estimating function approach in the context of parametric models provides a logical frame for estimation of parameter(s) of interest in the presence of nuisance parameter(s).

Let $\boldsymbol{X}^{(n)} = (X_1, \ldots, X_n)'$ be a collection of random variables forming a stochastic process. Let $\boldsymbol{F}$ be a class of probability distributions F on $\boldsymbol{R}^n$ and $\theta = \theta(F)$, $F \in \boldsymbol{F}$ be a real parameter. We assume that the probability distribution of $\boldsymbol{X}^{(n)}$ belongs to $\boldsymbol{F}$. A real function g of $\boldsymbol{X}^{(n)}$ and θ, satisfying certain regularity conditions, is called a regular unbiased estimating function if

$$E_F[g(\boldsymbol{X}^{(n)}; \theta(F))] = 0, \qquad F \in \boldsymbol{F}. \tag{3.1}$$

Among all regular unbiased estimating functions g, g^* is said to be optimum if

$$\mathrm{E}_F[g^2(\boldsymbol{X}^{(n)}; \theta(F))] \Big/ \left\{\mathrm{E}_F\left[\frac{\partial}{\partial\theta}g(\boldsymbol{X}^{(n)}; \theta(F))\right]\right\}^2$$

is minimized for all $F \in \boldsymbol{F}$ at $g = g^*$. Here, the partial derivative is evaluated at $\theta = \theta(F)$. Henceforth, we denote by $\mathcal{F}_t$ the σ-field generated by $\{X_s; s \leq t\}$.

Let $\mathcal{L}$ be the class of estimating functions g of the form

$$g = \sum_{t=1}^{n} h_t a_{t-1} \tag{3.2}$$

where h_t is a function of $X_1, \dots, X_t$ and θ satisfying

$$E_F[h_t \mid \mathcal{F}_{t-1}] = 0, \qquad (t = 1, \dots, n) \tag{3.3}$$

and a_{t-1} is a function of $X_1, \dots, X_{t-1}$ and θ, for $t = 1, \dots, n$.

An example of h_t is,

$$h_t = X_t - E_F[X_t \mid \mathcal{F}_{t-1}], \tag{3.4}$$

which is the residual between X_t and its best predictor $E_F[X_t \mid \mathcal{F}_{t-1}]$ based on $X_1, \dots, X_{t-1}$.

Note that (3.3) implies that for all $F \in \boldsymbol{F}$, $E_F(h_t h_s) = 0$, $t \neq s$.

3.1 *Estimating functions for a vector $\boldsymbol{\theta}$*

In this subsection we extend the above arguments to the case of vector estimating functions.

Suppose that $\{\boldsymbol{X}_t\}$ is an r-dimensional stochastic process whose distribution depends on a parameter $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)' \in \Theta \subset \boldsymbol{R}^p$ and denote by $\mathcal{F}_t$ the σ-field generated by $\{\boldsymbol{X}_s; s \leq t\}$.

Let $\mathcal{G}$ be the class of estimating vector functions $\boldsymbol{G}(\boldsymbol{\theta})$ of the form

$$\boldsymbol{G}(\boldsymbol{\theta}) = \sum_{t=1}^{n} \boldsymbol{a}_{t-1} \boldsymbol{h}_t$$

where $\boldsymbol{h}_t = \boldsymbol{X}_t - E[\boldsymbol{X}_t \mid \mathcal{F}_{t-1}]$ is an r-dimensional vector version of (3.4) and $\boldsymbol{a}_{t-1}$ is a $p \times r$ matrix depending on $\boldsymbol{X}_1, \dots, \boldsymbol{X}_{t-1}$ and $\boldsymbol{\theta}$, $1 \leq t \leq n$.

Let $\hat{\boldsymbol{\theta}}_G$ be an estimator of $\boldsymbol{\theta}$ using the vector form of the estimating function (3.1), i.e., $\boldsymbol{G}(\hat{\boldsymbol{\theta}}_G) = \boldsymbol{0}$.

By the mean-value theorem, we have

$$\boldsymbol{G}(\hat{\boldsymbol{\theta}}_G) - \boldsymbol{G}(\boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{G}(\boldsymbol{\theta}^*)(\hat{\boldsymbol{\theta}}_G - \boldsymbol{\theta})$$

where $\|\hat{\boldsymbol{\theta}}_G - \boldsymbol{\theta}^*\| \leq \|\hat{\boldsymbol{\theta}}_G - \boldsymbol{\theta}\|$. If $\hat{\boldsymbol{\theta}}_G \to^p \boldsymbol{\theta}$, then $\boldsymbol{\theta}^* \to^p \boldsymbol{\theta}$, hence,

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_G - \boldsymbol{\theta}) \approx -\left[\frac{1}{n} \frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{G}(\boldsymbol{\theta})\right]^{-1} \frac{1}{\sqrt{n}} \boldsymbol{G}(\boldsymbol{\theta}), \tag{3.5}$$

ignoring lower order terms. Here, the first factor on the right side of (3.5) is assumed to be non-singular for all $\boldsymbol{\theta} \in \Theta$. A good estimating function should have the difference $(\hat{\boldsymbol{\theta}}_G - \boldsymbol{\theta})$ small. Notice that the asymptotic variance of $\sqrt{n}(\hat{\boldsymbol{\theta}}_G - \boldsymbol{\theta})$ is given by

$$\left[\frac{1}{n} E \frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{G}(\boldsymbol{\theta})\right]^{-1} \frac{E\{\boldsymbol{G}(\boldsymbol{\theta})\boldsymbol{G}'(\boldsymbol{\theta})\}}{n} \left(\left[\frac{1}{n} E \frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{G}(\boldsymbol{\theta})\right]^{-1}\right)' \tag{3.6}$$

where the middle factor of (3.6) is assumed to be positive definite for all $\boldsymbol{\theta} \in \Theta$.

An estimating function $\boldsymbol{G}^*(\boldsymbol{\theta})$ is said to be optimal in $\mathcal{G}$ if the quantity

$$E\left\{\left[E \frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{G}(\boldsymbol{\theta})\right]^{-1} \boldsymbol{G}(\boldsymbol{\theta})\boldsymbol{G}'(\boldsymbol{\theta}) \left(\left[E \frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{G}(\boldsymbol{\theta})\right]^{-1}\right)'\right\} \tag{3.7}$$

is minimized at $\boldsymbol{G}(\boldsymbol{\theta}) = \boldsymbol{G}^*(\boldsymbol{\theta})$ in the sense of matrix order. We assume that $\boldsymbol{h}_t$ and $\boldsymbol{a}_{t-1}$ are differentiable w.r.t. $\boldsymbol{\theta}$ for $1 \le t \le n$. The following theorem is an extension of the result of Godambe (1985) to the case of vector. Since the proof is similar, we omit the details.

THEOREM 3.1. *In the class $\mathcal{G}$ of estimating functions $\boldsymbol{G}(\boldsymbol{\theta})$, the function $\boldsymbol{G}^*(\boldsymbol{\theta})$ minimizing* (3.7) *is given by $\boldsymbol{G}^*(\boldsymbol{\theta}) = \sum_{t=1}^n \boldsymbol{a}_{t-1}^* \boldsymbol{h}_t$ where $\boldsymbol{a}_{t-1}^* = [(E(\partial \boldsymbol{h}_t/\partial\boldsymbol{\theta}) \mid \mathcal{F}_{t-1})']$ $[(E(\boldsymbol{h}_t\boldsymbol{h}_t') \mid \mathcal{F}_{t-1})^{-1}]$.*

4. Random coefficient autoregressive models

In this section, we discuss an estimation procedure for an important class of nonlinear time series models. A time series $\{X_t\}$ is said to follow a random coefficient autoregressive (RCA) model of order k, if X_t satisfies

$$X_t = \sum_{i=1}^{k} (\theta_i + b_i(t)) X_{t-i} + \varepsilon_t. \tag{4.1}$$

For this model the following conditions are imposed.

ASSUMPTION 1. (i) $\theta_i, i = 1, \ldots, k$, are parameters to be estimated;
(ii) $\{\varepsilon_t\}$ is a sequence of i.i.d. $(0, \sigma_\varepsilon^2)$ random variables with $0 < \sigma_\varepsilon^2 < \infty$;
(iii) $\{\boldsymbol{b}_t = (b_1(t), \ldots, b_k(t))'\}$ is a sequence of i.i.d. random vectors with zero mean and covariance matrix $E(\boldsymbol{b}_t\boldsymbol{b}_t') = \boldsymbol{\Sigma}_b$;
(iv) $\{\boldsymbol{b}_t\}$ and $\{\varepsilon_t\}$ are possibly correlated i.e., $\mathrm{Cov}(\boldsymbol{b}_t, \varepsilon_t) = \rho\boldsymbol{\sigma}_b$, $\boldsymbol{\sigma}_b = (\sigma_{b_1}, \ldots, \sigma_{b_k})'$ and ρ is a constant with $|\rho| < \sigma_\varepsilon$;
(v) $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)'$ and $\boldsymbol{\Sigma}_b$ satisfy $\lambda(\boldsymbol{\theta\theta}' + \boldsymbol{\Sigma}_b) < 1$ where $\lambda(A)$ denotes the maximum eigenvalue of A in modulus.

The condition (v) is a sufficient condition for $\{X_t\}$ to be stationary and ergodic (see Feigin and Tweedie, (1985)).

4.1 *Scalar case*

In this subsection, we describe an estimation procedure for the scalar RCA models. Given an observed stretch $X_1, \ldots, X_n$, we shall estimate the parameter $\boldsymbol{\theta}$ of interest using Theorem 3.1 and assuming that the parameters $(\rho, \sigma_\varepsilon^2, \boldsymbol{\Sigma}_b)$ are known. The true value of $\boldsymbol{\theta}$ is denoted by $\boldsymbol{\theta}^0$. Define $\boldsymbol{X}_{t-1} = (X_{t-1}, \ldots, X_{t-k})'$ and rewrite (4.1) as

$$X_t = \boldsymbol{X}_{t-1}'\boldsymbol{\theta} + w_t, \quad w_t = \boldsymbol{X}_{t-1}'\boldsymbol{b}_t + \varepsilon_t.$$

Since $\{X_t\}$ is generated by $\{\varepsilon_t\}$ and $\{\boldsymbol{b}_t\}$, we observe $\mathcal{F}_t \subset \sigma\{(\boldsymbol{b}_t, \varepsilon_t), (\boldsymbol{b}_{t-1}, \varepsilon_{t-1}), \ldots\}$.

Noting that $\boldsymbol{b}_t$ and ε_t are independent of $\{(\boldsymbol{b}_{t-1}, \varepsilon_{t-1}), (\boldsymbol{b}_{t-2}, \varepsilon_{t-2}), \ldots,\}$ and $\boldsymbol{X}_{t-1} \in \mathcal{F}_{t-1}$, it is seen that

$$E[w_t \mid \mathcal{F}_{t-1}] = 0, \tag{4.2}$$

hence, from (3.4) we can set $h_t = X_t - \boldsymbol{X}_{t-1}'\boldsymbol{\theta}$. Differentiating h_t with respect to $\boldsymbol{\theta}$ and taking its expectation yield $E[(\partial h_t/\partial\boldsymbol{\theta}) \mid \mathcal{F}_{t-1}] = -\boldsymbol{X}_{t-1}$. Further,

$$E[h_t^2 \mid \mathcal{F}_{t-1}] = \sigma_\varepsilon^2 + 2Z_t + \boldsymbol{X}_{t-1}'\boldsymbol{\Sigma}_b\boldsymbol{X}_{t-1}, \qquad Z_t = \rho\boldsymbol{\sigma}_b'\boldsymbol{X}_{t-1}.$$

Let C_1, C_2 and C_3 be matrices such that the product $C_1C_2C_3$ is well defined. Applying the formula

$$\text{vec}(C_1C_2C_3) = (C_3' \otimes C_1)\text{vec}C_2 = (C_3' \otimes C_1)\boldsymbol{D}'\text{vech}C_2, \tag{4.3}$$

where $\otimes$ denotes tensor product and $\boldsymbol{D}$ is a constant matrix of zeros and ones (for the definition of $\boldsymbol{D}$, vec$(\cdot)$ and vech$(\cdot)$; see e.g., Nicholls and Quinn (1982), 11–13), we obtain

$$\begin{aligned}\boldsymbol{X}_{t-1}'\boldsymbol{\Sigma}_b\boldsymbol{X}_{t-1} &= (\boldsymbol{X}_{t-1}' \otimes \boldsymbol{X}_{t-1}')\,\text{vec}\,\boldsymbol{\Sigma}_b = \{\text{vec}(\boldsymbol{X}_{t-1}\boldsymbol{X}_{t-1}')\}'\boldsymbol{D}_k'\,\text{vech}\,\boldsymbol{\Sigma}_b \\ &= \{\text{vec}(\boldsymbol{X}_{t-1}\boldsymbol{X}_{t-1}')\}'\boldsymbol{D}_k'\boldsymbol{R},\end{aligned}$$

and

$$E[h_t^2 \mid \mathcal{F}_{t-1}] = \sigma_\varepsilon^2 + 2Z_t + \{\text{vec}(\boldsymbol{X}_{t-1}\boldsymbol{X}_{t-1}')\}'\boldsymbol{D}_k'\boldsymbol{R} \equiv \Psi_t$$

where $\boldsymbol{D}_k$ is a constant $(k(k+1)/2) \times k^2$ matrix corresponding to $\boldsymbol{D}$ in (4.3), and $\boldsymbol{R} = \text{vech}\,\boldsymbol{\Sigma}_b$ (see e.g., Nicholls and Quinn (1982)).

Hence, from Theorem 3.1, the asymptotically optimal estimator (OE) (in Godambe's sense) for $\boldsymbol{\theta}$ is given by

$$\hat{\boldsymbol{\theta}}_n^{(OE)} = \left(\sum_{t=k+1}^{n} \boldsymbol{X}_{t-1}\boldsymbol{X}_{t-1}'/\Psi_t\right)^{-1}\left(\sum_{t=k+1}^{n} \boldsymbol{X}_{t-1}X_t/\Psi_t\right). \tag{4.4}$$

Remark 1. Note that if $(\boldsymbol{b}_t, \varepsilon_t)$ are jointly normal, and $(\rho, \sigma_\varepsilon^2, \boldsymbol{\Sigma}_b)$ are assumed to be known, $\hat{\boldsymbol{\theta}}_n^{(OE)}$ is the maximum likelihood estimator of $\boldsymbol{\theta}$.

In the first-order RCA model $X_t = (\theta + b_t)X_{t-1} + \varepsilon_t$, it is easy to see that the OE of θ is given by

$$\hat{\theta}_n^{(OE)} = \sum_{t=2}^{n} a_{t-1}^* X_t \Big/ \sum_{t=2}^{n} a_{t-1}^* X_{t-1} \tag{4.5}$$

where $a_{t-1}^* = -X_{t-1}/\{\sigma_\varepsilon^2 + 2\rho' X_{t-1} + \sigma_b^2 X_{t-1}^2\}$ with $\rho' = \rho\sigma_b$.

Remark 2. If $\rho' = 0$, the (OE)* of θ becomes

$$\hat{\theta}_n^{(OE)^*} = \sum_{t=2}^{n} a_{t-1}^{**} X_t \Big/ \sum_{t=2}^{n} a_{t-1}^{**} X_{t-1}$$

where $a_{t-1}^{**} = -X_{t-1}/(\sigma_\varepsilon^2 + \sigma_b^2 X_{t-1}^2)$ and this estimator is formally identical to (2.4) of Thavaneswaran and Abraham (1988).

Next, we establish the asymptotics of (4.4) by employing the ergodic theorem and Billingsley's theorem ((1961), 788–792) for martingales.

THEOREM 4.1. *Suppose that the process* $\{X_t\}$ *given by* (4.1) *satisfies Assumption* 1 *and* $E(X_t^2) < \infty$. *If* $E\{\boldsymbol{X}_{t-1}\boldsymbol{X}_{t-1}'/\Psi_t\}$ *is a positive definite matrix with bounded elements, then*:

(i) *there exists a sequence of estimators* $\{\hat{\boldsymbol{\theta}}_n^{(OE)}\}$ *such that* $\hat{\boldsymbol{\theta}}_n^{(OE)} \xrightarrow{a.s.} \boldsymbol{\theta}^0$.

(ii) $\sqrt{n}(\hat{\boldsymbol{\theta}}_n^{(OE)} - \boldsymbol{\theta}^0) \rightarrow^d \mathcal{N}(\mathbf{0}, [E\{\boldsymbol{X}_{t-1}\boldsymbol{X}'_{t-1}/\Psi_t\}]^{-1})$.

PROOF. From (4.4), we have

$$\hat{\boldsymbol{\theta}}_n^{(OE)} - \boldsymbol{\theta}^0 = \left\{(n-k)^{-1} \sum_{t=k+1}^{n} \boldsymbol{X}_{t-1}\boldsymbol{X}'_{t-1}/\Psi_t\right\}^{-1} (n-k)^{-1} \sum_{t=k+1}^{n} \boldsymbol{X}_{t-1}w_t/\Psi_t.$$

Since $\{X_t\}$, $\{\boldsymbol{b}_t\}$ and $\{\varepsilon_t\}$ are strictly stationary and ergodic, so are $\{\boldsymbol{X}_{t-1}\boldsymbol{X}'_{t-1}/\Psi_t\}$ and $\{\boldsymbol{X}_{t-1}w_t/\Psi_t\}$. Furthermore, $E\{\boldsymbol{X}_{t-1}\boldsymbol{X}'_{t-1}/\Psi_t\}$ is finite, and from (4.2) $E\{\boldsymbol{X}_{t-1}w_t/\Psi_t\} = \mathbf{0}$. The second moments of $\{\boldsymbol{b}_t\}$ and $\{\varepsilon_t\}$ are assumed to exist, and thus by the ergodic theorem

$$(n-k)^{-1} \sum_{t=k+1}^{n} \boldsymbol{X}_{t-1}\boldsymbol{X}'_{t-1}/\Psi_t \longrightarrow^{a.s.} E(\boldsymbol{X}_{t-1}\boldsymbol{X}'_{t-1}/\Psi_t), \quad \text{and}$$

$$(n-k)^{-1} \sum_{t=k+1}^{n} \boldsymbol{X}_{t-1}w_t/\Psi_t \longrightarrow^{a.s.} E\{\boldsymbol{X}_{t-1}w_t/\Psi_t\} = \mathbf{0},$$

which imply $\hat{\boldsymbol{\theta}}_n^{(OE)} \longrightarrow^{a.s.} \boldsymbol{\theta}^0$.

Next, the joint asymptotic normality of the estimators $\{\hat{\boldsymbol{\theta}}_n^{(OE)}\}$ may be verified by using the Cramer-Wold device and Billingsley's theorem (1961) for martingales. For any given $\boldsymbol{c} = (c_1, \ldots, c_k)' \neq 0$, it is easily seen that the distribution of

$$(n-k)^{-1/2} \sum_{t=k+1}^{n} (\boldsymbol{c}'\boldsymbol{X}_{t-1}/\Psi_t)w_t$$

converges to the normal distribution with mean zero and variance

$$E\{(\boldsymbol{c}'\boldsymbol{X}_{t-1}w_t/\Psi_t)^2\} = \boldsymbol{c}'E\{\boldsymbol{X}_{t-1}\boldsymbol{X}'_{t-1}/\Psi_t\}\boldsymbol{c} \quad \text{as} \quad n \to \infty.$$

Hence, by Slutsky's theorem, $\sqrt{n}(\hat{\boldsymbol{\theta}}_n^{(OE)} - \boldsymbol{\theta}^0) \rightarrow^d \mathcal{N}(\mathbf{0}, [\mathrm{E}\{\boldsymbol{X}_{t-1}\boldsymbol{X}'_{t-1}/\Psi_t\}]^{-1})$. □

Remark 3. The assumption that $E\{\boldsymbol{X}_{t-1}\boldsymbol{X}'_{t-1}/\Psi_t\}$ is positive definite is not restrictive. If the following two natural conditions:

(i) the probability mass of the distribution of the random vector $\boldsymbol{X}_{t-1}$ does not reduce to a lower dimensional space (less than k),

(ii) $\{X_t\}$ is a nondeterministic process,

are satisfied, then the assumption holds.

4.2 *Vector case*

In the previous subsection, we described the estimation procedure for the scalar RCA models in the case when nuisance parameters are known. This section generalizes such estimation procedure of more general RCA models to a multivariate case.

Let $\{\boldsymbol{X}_t\}$ be a p-vector time series generated by

$$\boldsymbol{X}_t = \sum_{i=1}^{k} (\beta_i + B_i(t))\boldsymbol{X}_{t-i} + \boldsymbol{\varepsilon}_t, \tag{4.6}$$

which satisfies the following assumption.

ASSUMPTION 2. (i) $\beta_i, i = 1, \ldots, k$, is a $p \times p$ constant matrix to be estimated;
(ii) $\{\varepsilon_t\}$ is a sequence of i.i.d. p-vector random variables with zero mean and covariance matrix $\mathbf{\Gamma}$;
(iii) $\{\boldsymbol{B}_t = (\text{vec}'(B_1(t)), \ldots, \text{vec}'(B_k(t)))'\}$ is a sequence of i.i.d. $q \times 1$ $(q = kp^2)$ random vectors with mean zero and covariance matrix $E(\boldsymbol{B}_t\boldsymbol{B}_t') = \boldsymbol{\Omega}$;
(iv) $\{\varepsilon_t\}$ and $\{\boldsymbol{B}_t\}$ are possibly correlated i.e., $\text{Cov}(\varepsilon_t, \boldsymbol{B}_t) = \rho\boldsymbol{\sigma}_B$ where $\boldsymbol{\sigma}_B$ is a $p \times q$ matrix and ρ is a constant with $|\rho| < \|\mathbf{\Gamma}\|^{1/2}$;
(v) $\boldsymbol{\beta} = (\text{vec}'(\beta_1), \ldots, \text{vec}'(\beta_k))'$ and $\boldsymbol{\Omega}$ satisfy $\lambda(\boldsymbol{\beta}\boldsymbol{\beta}' + \boldsymbol{\Omega}) < 1$ so that $\{\boldsymbol{X}_t\}$ is stationary and ergodic.

Based on an observed stretch $(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$ from $\{\boldsymbol{X}_t\}$, we shall estimate the parameter $\boldsymbol{\beta}$ of interest using Theorem 3.1 and assuming the parameters $(\rho, \mathbf{\Gamma}, \boldsymbol{\Omega})$ are known. The true value of $\boldsymbol{\beta}$ is denoted by $\boldsymbol{\beta}^0$. Define a $kp \times 1$ random vector by $\boldsymbol{Y}_{t-1} = (\boldsymbol{X}_{t-1}', \ldots, \boldsymbol{X}_{t-k}')'$. Applying the formula in (4.3), we can rewrite (4.6) as $\boldsymbol{X}_t = (\boldsymbol{Y}_{t-1}' \otimes \boldsymbol{I}_p)\boldsymbol{\beta} + \boldsymbol{w}_t$, where $\boldsymbol{w}_t = (\boldsymbol{Y}_{t-1}' \otimes \boldsymbol{I}_p)\boldsymbol{B}_t + \varepsilon_t$.

By analogy with the arguments set out for the univariate case, the asymptotically OE estimator (in Godambe's sense) for $\boldsymbol{\beta}$ is given by

$$\hat{\boldsymbol{\beta}}_n^{(OE)} = \left(\sum_{t=k+1}^{n} (\boldsymbol{Y}_{t-1} \otimes \boldsymbol{I}_p)\boldsymbol{\Upsilon}_t^{-1}(\boldsymbol{Y}_{t-1}' \otimes \boldsymbol{I}_p)\right)^{-1} \left(\sum_{t=k+1}^{n} (\boldsymbol{Y}_{t-1} \otimes \boldsymbol{I}_p)\boldsymbol{\Upsilon}_t^{-1}\boldsymbol{X}_t\right)$$

where $\boldsymbol{\Upsilon}_t = \boldsymbol{V}_t'\boldsymbol{\omega} + 2\boldsymbol{Z}_t + \mathbf{\Gamma}$, with $\mathbf{V}_t = \boldsymbol{D}_q[\text{vec}(\boldsymbol{Y}_{t-1}\boldsymbol{Y}_{t-1}') \otimes \boldsymbol{I}_{p^2}]$, $\boldsymbol{\omega} = \text{vech}\,\boldsymbol{\Omega}$ and $\boldsymbol{Z}_t = \rho(\boldsymbol{Y}_{t-1}' \otimes \boldsymbol{I}_p)\boldsymbol{\sigma}_B'$, and $\boldsymbol{D}_q$ is a constant $(q(q+1)/2) \times q^2$ matrix corresponding to $\boldsymbol{D}$ in (4.3).

We can establish the asymptotic properties of $\hat{\boldsymbol{\beta}}_n^{(OE)}$ via the ergodic theorem and Billingsley's theorem (1961) for martingales similarly as in Theorem 4.1.

THEOREM 4.2. *Suppose that the process* $\{\boldsymbol{X}_t\}$ *given by* (4.6) *satisfies Assumption* 2 *and* $E\{\|\boldsymbol{X}_t\|^2\} < \infty$. *If* $E[(\boldsymbol{Y}_{t-1} \otimes \boldsymbol{I}_p)\boldsymbol{\Upsilon}_t^{-1}(\boldsymbol{Y}_{t-1}' \otimes \boldsymbol{I}_p)]$ *is a positive definite matrix with bounded elements, then*:
(i) *there exists a sequence of estimators* $\{\hat{\boldsymbol{\beta}}_n^{(OE)}\}$ *such that* $\hat{\boldsymbol{\beta}}_n^{(OE)} \xrightarrow{a.s.} \boldsymbol{\beta}^0$.
(ii) $\sqrt{n}(\hat{\boldsymbol{\beta}}_n^{(OE)} - \boldsymbol{\beta}^0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \{\mathrm{E}[(\boldsymbol{Y}_{t-1} \otimes \boldsymbol{I}_p)\boldsymbol{\Upsilon}_t^{-1}(\boldsymbol{Y}_{t-1}' \otimes \boldsymbol{I}_p)]\}^{-1})$.

5. Proposed estimation procedure

In Section 4, we constructed the asymptotically optimal estimators $\hat{\boldsymbol{\theta}}_n^{(OE)}$ and $\hat{\boldsymbol{\beta}}_n^{(OE)}$ under the assumption that the nuisance parameters are known. However, it is often the case that the nuisance parameters are unknown.

5.1 *Scalar case*

In this subsection, we consider the case when the nuisance parameters of (4.1) are unknown. Estimating the nuisance parameters by the classical method of moments and CLS estimator we propose an estimated version of the OE estimator.

For simplicity, we consider the model

$$X_t = (\theta + b_t)X_{t-1} + \varepsilon_t \tag{5.1}$$

where θ is an unknown parameter, and $\{b_t\}$ and $\{\varepsilon_t\}$ are sequences of i.i.d. random variables with zero mean and unknown variances $0 < \sigma_b^2 < \infty$ and $0 < \sigma_\varepsilon^2 < \infty$, respectively. In addition, $\{b_t\}$ and $\{\varepsilon_t\}$ are possibly correlated i.e., $\mathrm{Cov}(b_t, \varepsilon_t) = \rho\sigma_b$ where ρ is an unknown constant with $|\rho| < \sigma_\varepsilon$.

For (5.1), Nicholls and Quinn (1982) obtained the fourth moment condition, namely $\theta^4 + 6\theta^2\sigma_b^2 + 3\sigma_b^4 < 1$, in the case when independent sequences $\{\varepsilon_t\}$ and $\{b_t\}$ are Gaussian. Our target is to estimate the unknown parameters $(\theta, \rho, \sigma_b^2, \sigma_\varepsilon^2)$ on the basis of observed stretch $X_1, \ldots, X_n$.

Write

$$R(h) = E[X_t X_{t-h}], \quad h = 0, \pm 1, \ldots, \quad \text{and}$$
$$R(s_1, s_2) = E[X_t X_{t-s_1} X_{t-s_2}], \quad s_1, s_2 = 0, \pm 1, \ldots.$$

Natural moment estimators of $R(h)$ and $R(s_1, s_2)$ are, respectively, given by

$$\hat{R}(h) = \frac{1}{n}\sum_{t=1}^{n-|h|} X_t X_{t-|h|}, \quad \text{and} \quad \hat{R}(s_1, s_2) = \frac{1}{n}\sum_{t=1}^{n-\gamma} X_t X_{t-s_1} X_{t-s_2}, \quad \gamma = \max(0, s_1, s_2).$$

Under the condition $\theta^2 + \sigma_b^2 < 1$, it can be shown that

$$R(0) = E[X_t^2] = \sigma_\varepsilon^2 / \{1 - (\theta^2 + \sigma_b^2)\}, \tag{5.2}$$
$$R(0,1) = E[X_t^2 X_{t-1}] = (\theta^2 + \sigma_b^2) R(0,0) + 2\rho' R(0), \quad \text{and} \tag{5.3}$$
$$R(0,2) = E[X_t^2 X_{t-2}] = (\theta^2 + \sigma_b^2) R(0,1) + 2\rho' R(1). \tag{5.4}$$

For (5.1), the CLS estimator of θ is given by

$$\hat{\theta}_n^{(CL)} = \sum_{t=2}^{n} X_t X_{t-1} \Big/ \sum_{t=2}^{n} X_{t-1}^2. \tag{5.5}$$

If we substitute $\hat{R}(\cdot)$, $\hat{R}(\cdot,\cdot)$ and $\hat{\theta}_n^{(CL)}$ into the corresponding quantities in (5.2)–(5.4), we can get the moment estimators of σ_b^2, σ_ε^2 and ρ. We explain the procedure below.

Estimation of σ_b^2 and σ_ε^2 requires elimination of the third term from (5.3) and (5.4). To do this, multiply both sides of (5.3) by $R(1)/R(0)$, then

$$R(1)R(0,1)/R(0) = [(\theta^2 + \sigma_b^2) R(0,0)/R(0) + 2\rho'] R(1). \tag{5.6}$$

Subtracting (5.6) from (5.4), and substituting $\hat{\theta}_n^{(CL)}$, $\hat{R}(h)$ and $\hat{R}(s_1, s_2)$ for the corresponding quantities, and solving for σ_b^2, we obtain the estimator $\hat{\sigma}_b^2$:

$$\hat{\sigma}_b^2 = \hat{\delta} - (\hat{\theta}_n^{(CL)})^2 \tag{5.7}$$

where $\hat{\delta} = (\hat{b} - \hat{d})/(\hat{a} - \hat{c})$, with $\hat{a} = \hat{R}(0)\hat{R}(0,1)$, $\hat{b} = \hat{R}(0)\hat{R}(0,2)$, $\hat{c} = \hat{R}(1)\hat{R}(0,0)$ and $\hat{d} = \hat{R}(1)\hat{R}(0,1)$. Using (5.7) and the respective estimators, we get from (5.2) and (5.3) the estimators of σ_ε^2 and ρ:

$$\hat{\sigma}_\varepsilon^2 = (1 - \hat{\delta})\hat{R}(0), \quad \hat{\rho} = \hat{\zeta}^{-1} \hat{R}(0,0)[\hat{R}(0,1) - \hat{\delta}]$$

where $\hat{\zeta} = 2\hat{\sigma}_b \hat{R}(0)$.

Recalling $\hat{\theta}_n^{(OE)}$ of (4.5) in the case that the nuisance parameters are known, we can propose the estimated optimal estimator (EOE) of θ in the closed form:

$$(5.8)\ \hat{\hat{\theta}}_n^{(EOE)} = \sum_{t=2}^{n} \frac{X_t X_{t-1}}{\hat{\Pi}_t} \Bigg/ \sum_{t=2}^{n} \frac{X_{t-1}^2}{\hat{\Pi}_t}, \qquad \text{with} \quad \hat{\Pi}_t = \hat{\sigma}_\varepsilon^2 + 2\hat{\rho}' X_{t-1} + \hat{\sigma}_b^2 X_{t-1}^2.$$

If $\rho' = 0$, the estimator becomes simple. In fact, from (5.3), we can estimate σ_b^2 by $\hat{\sigma}_b^2 = \hat{\phi} - (\hat{\theta}_n^{(CL)})^2$ where $\hat{\phi} = \hat{R}(0,1)/\hat{R}(0,0)$. Thus, from (5.2) the estimator of σ_ε^2 is given by $\hat{\sigma}^2\varepsilon = (1-\hat{\phi})\hat{R}(0)$. Hence, if $\rho' = 0$, the estimated optimal estimator (EOE)* in the closed form is given by

$$(5.9) \qquad \hat{\hat{\theta}}_n^{(EOE)^*} = \sum_{t=2}^{n} \frac{X_t X_{t-1}}{\hat{\Phi}_t} \Bigg/ \sum_{t=2}^{n} \frac{X_{t-1}^2}{\hat{\Phi}_t}, \qquad \text{with} \quad \hat{\Phi}_t = \hat{\sigma}_\varepsilon^2 + \hat{\sigma}_b^2 X_{t-1}^2.$$

5.2 *Vector case*

In the previous subsection, we obtained explicit expressions of the estimators in (5.8) and (5.9). This section proposes an asymptotically optimal procedure based on the previous estimation procedure for the model $X_t = \boldsymbol{X}_{t-1}'(\boldsymbol{\theta} + \boldsymbol{b}_t) + \varepsilon_t$. Here, $\{\boldsymbol{b}_t\}$ and $\{\varepsilon_t\}$ are supposed to be Gaussian and mutually independent.

The second-order moments of $\{X_t\}$ is given by

$$(5.10) \qquad E[X_t^2] = \mathrm{tr}[\boldsymbol{\Gamma}(\boldsymbol{\theta}\boldsymbol{\theta}' + \boldsymbol{\Sigma}_b)] + \sigma_\varepsilon^2$$

where $\boldsymbol{\Gamma} = E[\boldsymbol{X}_{t-1}\boldsymbol{X}_{t-1}']$. $\boldsymbol{\Gamma}$ and $E[X_t^2]$ are estimable by the method of moments.

Since $\boldsymbol{\Sigma}_b$ is symmetric, we need only $l = k(k+1)/2$ dimension vector denoted by $\mathrm{vech}\boldsymbol{\Sigma}_b$. Next, it is easily seen that

$$\begin{aligned} E[\boldsymbol{R}_l X_t^2] &= E[\boldsymbol{R}_l\{\boldsymbol{X}_{t-1}'(\boldsymbol{\theta} + \boldsymbol{b}_t) + \varepsilon_t\}\{(\boldsymbol{\theta}' + \boldsymbol{b}_t')\boldsymbol{X}_{t-1} + \varepsilon_t\}] \\ &= E[\mathrm{vec}\{\boldsymbol{R}_l\boldsymbol{X}_{t-1}'\boldsymbol{\theta}\boldsymbol{\theta}'\boldsymbol{X}_{t-1}\}] + E[\mathrm{vec}\{\boldsymbol{R}_l\boldsymbol{X}_{t-1}'\boldsymbol{\Sigma}_b\boldsymbol{X}_{t-1}\}] \end{aligned}$$

where $\boldsymbol{R}_l = (X_{t-1}, \ldots, X_{t-l})'$. Noting the formula in (4.3), we can write

$$\begin{aligned} (5.11) \quad & \underbrace{E[\boldsymbol{R}_l X_t^2]}_{(a)} \\ &= \underbrace{E[(\boldsymbol{X}_{t-1}' \otimes \boldsymbol{R}_l\boldsymbol{X}_{t-1}')]}_{(b)} \boldsymbol{D}_1' \underbrace{\mathrm{vech}(\boldsymbol{\theta}\boldsymbol{\theta}')}_{(c)} + \underbrace{E[(\boldsymbol{X}_{t-1}' \otimes \boldsymbol{R}_l\boldsymbol{X}_{t-1}')]}_{(d)} \boldsymbol{D}_2'\mathrm{vech}\boldsymbol{\Sigma}_b \end{aligned}$$

where $\{(a),(b),(d)\}$ and (c) are estimable by the method of moments and CLS estimator respectively, and $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$ are constant $l \times k^2$ matrices corresponding to $\boldsymbol{D}$ in (4.3). Suppose that an observed stretch $X_1, \ldots, X_n$ is available. Then, from (5.11)

$$\begin{aligned} (5.12) \qquad \mathrm{vech}\hat{\boldsymbol{\Sigma}}_b = {} & [E[(\boldsymbol{X}_{t-1}' \widehat{\otimes \boldsymbol{R}_l}\boldsymbol{X}_{t-1}')]\boldsymbol{D}_2']^{-1} \\ & \times [E[\widehat{\boldsymbol{R}_l X_t^2}] - E[(\boldsymbol{X}_{t-1}' \widehat{\otimes \boldsymbol{R}_l}\boldsymbol{X}_{t-1}')]\boldsymbol{D}_1'\mathrm{vech}(\hat{\boldsymbol{\theta}}_{(CL)}\hat{\boldsymbol{\theta}}_{(CL)}')] \end{aligned}$$

where

$$\hat{\boldsymbol{\theta}}_n^{(CL)} = \left(\sum_{t=2}^{n} \boldsymbol{X}_{t-1}\boldsymbol{X}_{t-1}'\right)^{-1} \left(\sum_{t=2}^{n} \boldsymbol{X}_{t-1}X_t\right),$$

and $\widehat{[\cdot]}$'s are moment estimators of $[\cdot]$. Substituting (5.12) into (5.10) yields

$$\hat{\sigma}_\varepsilon^2 = \widehat{E[X_t^2]} - \mathrm{tr}[\hat{\boldsymbol{\Gamma}}(\hat{\boldsymbol{\theta}}_{(CL)}\hat{\boldsymbol{\theta}}'_{(CL)} + \hat{\boldsymbol{\Sigma}}_b)].$$

Thus, we propose the EOE (in Godame's sense) of $\boldsymbol{\theta}$ in the closed form:

$$\hat{\boldsymbol{\theta}}_n^{(EOE)} = \left(\sum_{t=k+1}^{n} \boldsymbol{X}_{t-1}\boldsymbol{X}'_{t-1}/\hat{\Psi}_t^*\right)^{-1}\left(\sum_{t=k+1}^{n} \boldsymbol{X}_{t-1}X_t/\hat{\Psi}_t^*\right)$$

where $\hat{\Psi}_t^* = \hat{\sigma}_\varepsilon^2 + \boldsymbol{K}'_t \mathrm{vech}\hat{\boldsymbol{\Sigma}}_b$ with $\boldsymbol{K}_t = \boldsymbol{D}_2\mathrm{vec}\{(\boldsymbol{X}_{t-1}\boldsymbol{X}'_{t-1})\}$.

6. Estimation for autoregressive conditional heteroskedastic models

In this section, we discuss the problem of estimating another important class of nonlinear time series models which is considered to be extremely useful in financial engineering. A time series $\{X_t\}$ is said to follow an ARCH process of order m introduced by Engle (1982), if X_t satisfies

$$X_t = \varepsilon_t\sqrt{U_t} \tag{6.1}$$

where U_t evolves according to $U_t = \alpha_0 + \sum_{i=1}^{m} \alpha_i X_{t-i}^2$.

For this model, we impose the following conditions.

ASSUMPTION 3. (i) The parameter $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \ldots, \alpha_m)'$ is to be estimated where $\alpha_0 > 0$ and $\alpha_i \geq 0$, $i = 1, \ldots, m$ being the constraints on $\boldsymbol{\alpha}$ to ensure that the variance remains positive;

(ii) $\{\varepsilon_t\}$ is a sequence of i.i.d. $(0,1)$ random variables with fourth-order cumulant κ_4, and ε_t is independent of X_s, $s < t$;

(iii) $\sum_{i=1}^{m} \alpha_i < 1$ so that $\{X_t\}$ is strictly stationary and ergodic.

The model in (6.1) can be rewritten in an alternative form that is easy to compare with an autoregressive representation:

$$Y_t = \boldsymbol{\alpha}'\boldsymbol{Y}_{t-1} + \eta_t, \qquad U_t = \boldsymbol{\alpha}'\boldsymbol{Y}_{t-1} \tag{6.2}$$

where $Y_t = X_t^2$, $\boldsymbol{Y}_{t-1} = (1, Y_{t-1}, \ldots, Y_{t-m})'$ and $\eta_t = U_t(\varepsilon_t^2 - 1)$. Henceforth, denote by $\mathcal{F}_t$ the σ-field generated by $\{Y_t, Y_{t-1}, \ldots\}$. We note that the disturbance term, η_t, in (6.2) is a martingale difference since $E[\eta_t \mid \mathcal{F}_{t-1}] = 0$. Thus, from (3.4) we can set

$$h_t = Y_t - E[Y_t \mid \mathcal{F}_{t-1}] = Y_t - \boldsymbol{\alpha}'\boldsymbol{Y}_{t-1}. \tag{6.3}$$

Then,

$$E[(\partial h_t/\partial\boldsymbol{\alpha}) \mid \mathcal{F}_{t-1}] = -\boldsymbol{Y}_{t-1}, \qquad \text{and} \qquad E[h_t^2 \mid \mathcal{F}_{t-1}] = (\kappa_4 + 2)U_t^2. \tag{6.4}$$

Now, we shall consider the OE estimator for the parameter $\boldsymbol{\alpha}$ of interest in the following two cases I and II. The true value of $\boldsymbol{\alpha}$ is denoted by $\boldsymbol{\alpha}^0$. Suppose that an observed stretch $Y_1, \ldots, Y_n$ from $\{Y_t\}$ is available.

Case I. Estimator based on U_t.

By virtue of Theorem 3.1, (6.3) and (6.4), the OE (in Godame's sense) for $\boldsymbol{\alpha}$ is given by

$$\hat{\boldsymbol{\alpha}}_n^{(OE)} = \left(\sum_{t=m+1}^{n} \boldsymbol{Y}_{t-1}\boldsymbol{Y}'_{t-1}/U_t^2\right)^{-1}\left(\sum_{t=m+1}^{n} \boldsymbol{Y}_{t-1}Y_t/U_t^2\right). \tag{6.5}$$

The following theorem establishes the asymptotic properties of $\hat{\boldsymbol{\alpha}}_n^{(OE)}$.

THEOREM 6.1. *Suppose that the process $\{Y_t\}$ given by (6.2) satisfies Assumption 3 and $E(Y_t^2) < \infty$. If $E\{\boldsymbol{Y}_{t-1}\boldsymbol{Y}'_{t-1}/U_t^2\}$ is a positive definite matrix with bounded elements, then*:

(i) *there exists a sequence of estimators $\{\hat{\boldsymbol{\alpha}}_n^{(OE)}\}$ such that $\hat{\boldsymbol{\alpha}}_n^{(OE)} \longrightarrow^{a.s.} \boldsymbol{\alpha}^0$.*

(ii) $\sqrt{n}(\hat{\boldsymbol{\alpha}}_n^{(OE)} - \boldsymbol{\alpha}^0) \rightarrow^d \mathcal{N}(\mathbf{0}, (\kappa_4 + 2)[E\{\boldsymbol{Y}_{t-1}\boldsymbol{Y}'_{t-1}/U_t^2\}]^{-1})$.

PROOF. Note that

$$\hat{\boldsymbol{\alpha}}_n^{(OE)} - \boldsymbol{\alpha}^0 = \left\{\sum_{t=m+1}^{n} \boldsymbol{Y}_{t-1}\boldsymbol{Y}'_{t-1}/U_t^2\right\}^{-1} \sum_{t=m+1}^{n} \boldsymbol{Y}_{t-1}\eta_t/U_t^2. \tag{6.6}$$

Consider the first factor on the right side of (6.6). As in Section 4, by the ergodic theorem

$$(n-m)^{-1} \sum_{t=m+1}^{n} \boldsymbol{Y}_{t-1}\boldsymbol{Y}'_{t-1}/U_t^2 \longrightarrow^{a.s.} E(\boldsymbol{Y}_{t-1}\boldsymbol{Y}'_{t-1}/U_t^2).$$

Since $\boldsymbol{Y}_{t-1}$ and η_t are strictly stationary and ergodic, so is $\boldsymbol{Y}_{t-1}\eta_t/U_t^2$. Hence, by the ergodic theorem we obtain

$$(n-m)^{-1} \sum_{t=m+1}^{n} \boldsymbol{Y}_{t-1}\eta_t/U_t^2 \longrightarrow^{a.s.} E\{(\boldsymbol{Y}_{t-1}\eta_t/U_t^2)\} = \mathbf{0},$$

which implies $\hat{\boldsymbol{\alpha}}_n^{(OE)} \longrightarrow^{a.s.} \boldsymbol{\alpha}^0$.

Next turn to derive the asymptotic distribution of $\hat{\boldsymbol{\alpha}}_n^{(OE)}$. For any given vector $\boldsymbol{c} = (c_1, \ldots, c_{m+1})' \neq 0$, noting the central limit theorem for stationary ergodic martingale difference (see Billingsley (1961)), and Cramer-Wold device, we observe that

$$(n-m)^{-1/2} \sum_{t=m+1}^{n} \{\boldsymbol{c}'\boldsymbol{Y}_{t-1}/U_t^2\}\eta_t \rightarrow^d \mathcal{N}(\mathbf{0}, (\kappa_4 + 2)\boldsymbol{c}'E\{\boldsymbol{Y}_{t-1}\boldsymbol{Y}'_{t-1}/U_t^2\}\boldsymbol{c}) \quad \text{as} \quad n \to \infty.$$

Hence, by Slutsky's theorem

$$\sqrt{n}(\hat{\boldsymbol{\alpha}}_n^{(OE)} - \boldsymbol{\alpha}^0) \rightarrow^d \mathcal{N}[\mathbf{0}, (\kappa_4 + 2)\{E(\boldsymbol{Y}_{t-1}\boldsymbol{Y}'_{t-1}/U_t^2)\}^{-1}]. \qquad \square$$

Since $\hat{\boldsymbol{\alpha}}_n^{(OE)}$ depends on U_t through the unknown parameter $\boldsymbol{\alpha}$ namely, $\hat{\boldsymbol{\alpha}}_n^{(OE)} = \hat{\boldsymbol{\alpha}}_n^{(OE)}(\boldsymbol{\alpha})$, we cannot use as an estimator of $\boldsymbol{\alpha}$. So we consider an estimated version of it, $\hat{\hat{\boldsymbol{\alpha}}}_n^{(EOE)} = \hat{\boldsymbol{\alpha}}_n^{(OE)}(\hat{\boldsymbol{\alpha}}_n^{(CL)})$ where $\hat{\boldsymbol{\alpha}}_n^{(CL)}$ is the CLS estimator of $\boldsymbol{\alpha}$.

Case II. Estimated version of $\hat{\boldsymbol{\alpha}}_n^{(OE)}$.

For (6.2), the CLS estimator $\hat{\boldsymbol{\alpha}}_n^{(CL)}$ of $\boldsymbol{\alpha}$ is given by

$$\hat{\boldsymbol{\alpha}}_n^{(CL)} = \left(\sum_{t=2}^{n} \boldsymbol{Y}_{t-1}\boldsymbol{Y}'_{t-1}\right)^{-1} \left(\sum_{t=2}^{n} \boldsymbol{Y}_{t-1}Y_t\right). \tag{6.7}$$

We note that the asymptotic properties of (6.7) follow readily from Tjøstheim's (1986) results.

Now, we can propose the EOE (in Godambe's sense) of $\boldsymbol{\alpha}$ in the closed form:

$$\hat{\boldsymbol{\alpha}}_n^{(EOE)} = \left(\sum_{t=m+1}^{n} \boldsymbol{Y}_{t-1}\boldsymbol{Y}'_{t-1}/\hat{U}_t^2\right)^{-1} \left(\sum_{t=m+1}^{n} \boldsymbol{Y}_{t-1}Y_t/\hat{U}_t^2\right), \tag{6.8}$$

where $\hat{U}_t = \hat{\boldsymbol{\alpha}}'_{(CL)}\boldsymbol{Y}_{t-1}$. Recall (6.7) that $\hat{\boldsymbol{\alpha}}_n^{(CL)} \longrightarrow^{a.s.} \boldsymbol{\alpha}^0$, as $n \to \infty$. Hence, we understand that $\hat{U}_t$ behaves like U_t, for $t = m+1, \ldots, n$.

To establish the limiting distribution of (6.8), we impose the following condition.

ASSUMPTION 4. $EY_t^4 < \infty$,

A sufficient condition for Assumption 4 is given by Engle ((1982), Theorem 1) in the case of $m = 1$.

THEOREM 6.2. *Under the Assumptions* 3 *and* 4, *we have, as* $n \to \infty$,

$$\sqrt{n}(\hat{\boldsymbol{\alpha}}_n^{(EOE)} - \boldsymbol{\alpha}^0) \to^d \mathcal{N}(\mathbf{0}, (\kappa_4 + 2)[E\{\boldsymbol{Y}_{t-1}\boldsymbol{Y}'_{t-1}/U_t^2\}]^{-1}).$$

PROOF. If we show that, as $n \to \infty$,

$$(n-m)^{-1} \sum_{t=m+1}^{n} (\boldsymbol{Y}_{t-1}\boldsymbol{Y}'_{t-1})\{(\hat{\boldsymbol{\alpha}}_n^{(CL)})'\boldsymbol{Y}_{t-1}\boldsymbol{Y}'_{t-1}\hat{\boldsymbol{\alpha}}_n^{(CL)}\}^{-1} - (n-m)^{-1} \sum_{t=m+1}^{n} (\boldsymbol{Y}_{t-1}\boldsymbol{Y}'_{t-1})\{\boldsymbol{\alpha}^{0\prime}\boldsymbol{Y}_{t-1}\boldsymbol{Y}'_{t-1}\boldsymbol{\alpha}^0\}^{-1} = o_p(1), \tag{6.9}$$

$$(n-m)^{-1/2} \sum_{t=m+1}^{n} (\boldsymbol{Y}_{t-1}\eta_t)\{(\hat{\boldsymbol{\alpha}}_n^{(CL)})'\boldsymbol{Y}_{t-1}\boldsymbol{Y}'_{t-1}\hat{\boldsymbol{\alpha}}_n^{(CL)}\}^{-1} - (n-m)^{-1/2} \sum_{t=m+1}^{n} (\boldsymbol{Y}_{t-1}\eta_t)\{\boldsymbol{\alpha}^{0\prime}\boldsymbol{Y}_{t-1}\boldsymbol{Y}'_{t-1}\boldsymbol{\alpha}^0\}^{-1} = o_p(1), \tag{6.10}$$

the proof is reduced to that of Theorem 6.1. Since the proof of (6.9) is similar to that of (6.10) we prove (6.10) only. Recall $\hat{\boldsymbol{\alpha}}_n^{(CL)} \longrightarrow^{a.s.} \boldsymbol{\alpha}^0$, hence, for any $\epsilon > 0$ there exists N_ϵ such that $\|\hat{\boldsymbol{\alpha}}_n^{(CL)} - \boldsymbol{\alpha}^0\| < \epsilon$, for all $n \geq N_\epsilon$, with probability one. Expanding the left hand side of (6.10) in Taylor expansion with respect to $\hat{\boldsymbol{\alpha}}_n^{(CL)}$ at $\boldsymbol{\alpha}^0$, we can see that it is dominated by

$$(6.11)\left|\frac{1}{n}\sum_{t=m+1}^{n} \boldsymbol{Y}_{t-1}\eta_t\bar{\boldsymbol{\alpha}}_\epsilon'(\boldsymbol{Y}_{t-1}\boldsymbol{Y}_{t-1}')\{\bar{\boldsymbol{\alpha}}_\epsilon'\boldsymbol{Y}_{t-1}\boldsymbol{Y}_{t-1}'\bar{\boldsymbol{\alpha}}_\epsilon\}^{-2}\right| \times O_p\{\sqrt{n}(\hat{\boldsymbol{\alpha}}_n^{(CL)}-\boldsymbol{\alpha}^0)\},$$

where $\bar{\boldsymbol{\alpha}}_\epsilon$ is nonrandom and lies in the ball $\{\boldsymbol{\alpha} : \|\boldsymbol{\alpha}-\boldsymbol{\alpha}^0\| \le \epsilon\}$. It is easily shown that the term in the absolute value of (6.11) converges to zero a.s. Hence we get the desired result. □

In the first-order ARCH model

$$(6.12)\qquad y_t = \boldsymbol{\alpha}'\mathbf{y}_{t-1} + \eta_t, \qquad u_t = \boldsymbol{\alpha}'\mathbf{y}_{t-1},$$

it is easy to see that the EOE of $\boldsymbol{\alpha}$ is given by

$$(6.13)\qquad \hat{\boldsymbol{\alpha}}_n^{(EOE)} = \left(\sum_{t=2}^{n}\mathbf{y}_{t-1}\mathbf{y}_{t-1}'/\hat{u}_t^2\right)^{-1}\left(\sum_{t=2}^{n}\mathbf{y}_{t-1}y_t/\hat{u}_t^2\right),$$

where $\boldsymbol{\alpha} = (\alpha_0, \alpha_1)'$, $\mathbf{y}_{t-1} = (1, y_{t-1})'$, $\hat{u}_t = \hat{\boldsymbol{\alpha}}_{(CL)}'\mathbf{y}_{t-1}$,

$$(6.14)\qquad \hat{\boldsymbol{\alpha}}_n^{(CL)} = \left(\sum_{t=2}^{n}\mathbf{y}_{t-1}\mathbf{y}_{t-1}'\right)^{-1}\left(\sum_{t=2}^{n}\mathbf{y}_{t-1}y_t\right).$$

Remark 4. Combined with the conditional least squares and classical moment estimators we applied the estimation method proposed by Godambe (1985) to the RCA and ARCH models. As pointed out by an anonymous referee, the score function in (3.2) is given by the GMM formulation (e.g., Hansen (1982)) in the discrete time series framework. As we said in Introduction, the estimation method by Godambe and GMM method are essentially the same. However, these two methods have been developed independently in mathematical statistics and econometrics, respectively.

7. Simulation study

Some simulations are performed to give some ideas about the asymptotic efficiency of the proposed estimators and algorithm in Subsection 5.1 and Section 6 through the models (5.1) and (6.12) respectively, for several sets of data of different sizes. We consider the simulation of these models separately.

Example 1. (RCA model) Consider the model

$$X_t = (\theta + b_t)X_{t-1} + \varepsilon_t, \qquad X_t = 0 \quad \text{for } t < 0.$$

Let $\{\boldsymbol{\xi}_t = (\varepsilon_t^0, b_t^0)'\}$ be a sequence of i.i.d. random vectors distributed as

$$\mathcal{N}\left[\begin{pmatrix}0\\0\end{pmatrix}, \boldsymbol{\Omega} = \begin{pmatrix}1 & 0\\ 0 & \sigma_{b^0}^2\end{pmatrix}\right].$$

Table 1.

$\theta = 0.2, 0.6$				
α	β	$\sigma_{\varepsilon^0}^2$	$\sigma_{b^0}^2$	ρ
0.2	0.3	1	0.16	0.1740
			0.36	0.1360

Table 2.

$\theta = 0.2, 0.6$				
α	β	$\sigma_{\varepsilon^0}^2$	$\sigma_{b^0}^2$	ρ
−0.3	0.6	1	0.16	−0.5220
			0.36	−0.4080

Table 3. MSE of the estimators and average estimates of ρ.

Replication=100, $\theta = 0.2$, $\sigma_{\varepsilon^0}^2 = 1$				
Estimator	ρ	$n = 200$	$n = 500$	$n = 1000$
$\hat{\rho}$	0.1360	0.1100	0.1151	0.1290
	0.1740	0.1342	0.1437	0.1650
$\hat{\theta}_n^{(CL)}$	0.1360	0.0280	0.0095	0.0091
	0.1740	0.0251	0.0092	0.0089
$\hat{\hat{\theta}}_n^{(EOE)*}$	0.1360	0.0272	0.0088	0.0086
	0.1740	0.0265	0.0086	0.0082
$\hat{\hat{\theta}}_n^{(EOE)}$	0.1360	0.0252	0.0081	0.0084
	0.1740	0.0245	0.0080	0.0082
Replication=100, $\theta = 0.6$, $\sigma_{\varepsilon^0}^2 = 1$				
Estimator	ρ	$n = 200$	$n = 500$	$n = 1000$
$\hat{\rho}$	0.1360	0.1080	0.1124	0.1265
	0.1740	0.1275	0.1399	0.1640
$\hat{\theta}_n^{(CL)}$	0.1360	0.0310	0.0096	0.0094
	0.1740	0.0271	0.0093	0.0090
$\hat{\hat{\theta}}_n^{(EOE)*}$	0.1360	0.0285	0.0090	0.0088
	0.1740	0.0278	0.0087	0.0085
$\hat{\hat{\theta}}_n^{(EOE)}$	0.1360	0.0263	0.0084	0.0085
	0.1740	0.0258	0.0080	0.0083

Then, define $\boldsymbol{\eta}_t = (\varepsilon_t, b_t)'$ by $\boldsymbol{A}'\boldsymbol{\xi}_t$ where $\boldsymbol{A} = \begin{pmatrix} \alpha & \beta \\ \beta & \alpha \end{pmatrix}$, which implies that the covariance matrix of $\boldsymbol{\eta}_t$ is $\boldsymbol{A}'\boldsymbol{\Omega}\boldsymbol{A}$. Thus, $\rho = \alpha\beta(1 + \sigma_{b^0}^2)/\sigma_{b^0}$.

In this example, we consider two cases of Tables 1 and 2: Here, (α, β) and $(\theta, \sigma_{b^0}^2)$ are chosen so that $|\rho| < \sigma_{\varepsilon^0}$ and $\theta^2 + \sigma_{b^0}^2 < 1$ respectively.

For each case, the process is simulated for the sample sizes of 200, 500 and 1000, of which hundred replications are performed. The sample mean square error (MSE) of the estimators (5.5), (5.8), (5.9) and average estimates of ρ are given in each case. These results are summarized in Tables 3 and 4. It is seen from these tables that $\hat{\hat{\theta}}_n^{(EOE)}$ is more efficient than the other estimators, and that the efficiency of $\hat{\hat{\theta}}_n^{(EOE)}$ increases as the sample size n increases, or θ and ρ decrease. The results are found to be satisfactory and agree well with the theoretical results.

Table 4. MSE of the estimators and average estimates of ρ.

Replication=100, $\theta = 0.2$, $\sigma_{\varepsilon 0}^2 = 1$				
Estimator	ρ	$n = 200$	$n = 500$	$n = 1000$
$\hat{\rho}$	−0.5220	−0.4685	−0.4751	−0.4931
	−0.4080	−0.3552	−0.3615	−0.3782
$\hat{\theta}_n^{(CL)}$	−0.5220	0.0292	0.0094	0.0090
	−0.4080	0.0286	0.0091	0.0088
$\hat{\theta}_n^{(EOE)*}$	−0.5220	0.0289	0.0092	0.0087
	−0.4080	0.0273	0.0089	0.0086
$\hat{\theta}_n^{(EOE)}$	−0.5220	0.0274	0.0090	0.0085
	−0.4080	0.0268	0.0088	0.0084
Replication=100, $\theta = 0.6$, $\sigma_{\varepsilon 0}^2 = 1$				
Estimator	ρ	$n = 200$	$n = 500$	$n = 1000$
$\hat{\rho}$	−0.5220	−0.4354	−0.4542	−0.4851
	−0.4080	−0.3340	−0.3610	−0.3725
$\hat{\theta}_n^{(CL)}$	−0.5220	0.0312	0.0095	0.0093
	−0.4080	0.0295	0.0092	0.0091
$\hat{\theta}_n^{(EOE)*}$	−0.5220	0.0290	0.0092	0.0089
	−0.4080	0.0289	0.0091	0.0088
$\hat{\theta}_n^{(EOE)}$	−0.5220	0.0271	0.0092	0.0087
	−0.4080	0.0268	0.0090	0.0086

Table 5. MSE of the estimators.

Replication=100, $\alpha_0 = 1$, $\alpha = 0.8$, $\sigma_\varepsilon^2 = 1$			
Estimator	$n = 200$	$n = 500$	$n = 1000$
$\hat{\boldsymbol{\alpha}}_n^{(CL)}$	0.0410	0.0082	0.0051
$\hat{\boldsymbol{\alpha}}_n^{(EOE)}$	0.0382	0.0071	0.0049
Replication=100, $\alpha_0 = 1$, $\alpha = 0.2$, $\sigma_\varepsilon^2 = 1$			
Estimator	$n = 200$	$n = 500$	$n = 1000$
$\hat{\boldsymbol{\alpha}}_n^{(CL)}$	0.0390	0.0076	0.0048
$\hat{\boldsymbol{\alpha}}_n^{(EOE)}$	0.0350	0.0072	0.0047

Table 6. MSE of the estimators.

Replication=100, $\alpha_0 = 20$, $\alpha = 0.8$, $\sigma_\varepsilon^2 = 1$			
Estimator	$n = 200$	$n = 500$	$n = 1000$
$\hat{\boldsymbol{\alpha}}_n^{(CL)}$	0.0443	0.0086	0.0053
$\hat{\boldsymbol{\alpha}}_n^{(EOE)}$	0.0412	0.0079	0.0051
Replication=100,$\alpha_0 = 20$, $\alpha = 0.2$, $\sigma_\varepsilon^2 = 1$			
Estimator	$n = 200$	$n = 500$	$n = 1000$
$\hat{\boldsymbol{\alpha}}_n^{(CL)}$	0.0420	0.0078	0.0050
$\hat{\boldsymbol{\alpha}}_n^{(EOE)}$	0.0381	0.0074	0.0048

Example 2. (ARCH model) In this example we consider the model

$$X_t = \varepsilon_t \sqrt{U_t}, \quad U_t = \alpha_0 + \alpha X_{t-1}^2, \quad X_t = 0, \text{ for } t < 0 \tag{7.1}$$

where $\{\varepsilon_t\}$ is a sequence of i.i.d. (0,1) random variables, $\alpha_0 > 0$ and $0 \le \alpha < 1$.

For $\alpha_0 = (1, 20)$ and $\alpha = (0.2, 0.8)$, we generated realizations from (7.1) with $n = 200, 500, 1000$. Each simulation was replicated hundred times and the sample MSE was calculated for the estimators in (6.13) and (6.14). Tables 5 and 6 summarize these results. It can be seen that $\hat{\boldsymbol{\alpha}}_n^{(EOE)}$ is more efficient than $\hat{\boldsymbol{\alpha}}_n^{(CL)}$. The efficiency of the former estimator increases against its counterpart as α_0 or α decreases. These results confirm the theoretical results given in Section 6.

Acknowledgements

The authors gratefully acknowledge two anonymous referees for careful reading and comments, which substantially improve the initial version of the paper. We also thank Professor S. Shirahata for his support and encouragement.

REFERENCES

Billingsley, P. (1961). The Lindeberg-Lévy theorem for martingales, *Proc. Amer. Math. Soc.*, **12**, 788–792.

Bollerslev, T., Chou, R. Y. and Kroner, K. F. (1992). ARCH modeling in finance: A review of the theory and empirical evidence, *J. Econometrics*, **52**, 5–59.

Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of U. K. inflation, *Econometrica*, **50**, 987–1007.

Feigin, P. D. and Tweedie, R. L. (1985). Random coefficient autoregressive processes: A Markov chain analysis of stationarity and finiteness of moments, *J. Time Ser. Anal.*, **6**, 1–14.

Godambe, V. P. (1960). An optimum property of regular maximum likelihood equation, *Ann. Math. Statist.*, **31**, 1208–11.

Godambe, V. P. (1985). The foundations of finite sample estimation in stochastic processes, *Biometrika*, **72**, 419–28.

Gouriéroux, C. (1997). *ARCH Models and Financial Applications*, Springer, New York.

Hafner, C. M. (1998). *Nonlinear Time Series Analysis with Applications to Foreign Exchange Rate Volatility*, Physica, Heidelberg.

Hansen, L. P. (1982). Large sample properties of generalised method of moments estimators, *Econometrica*, **50**, 1029–1054.

Heyde, C. C. (1997). *Quasi-likelihood and Its Application: A General Approach to Optimal Parameter Estimation*, Springer, New York.

Hwang, S. Y. and Basawa, I. V. (1998). Parameter estimation for generalised random coefficient autoregressive processes, *J. Statist. Plann. Inference*, **68**, 323–337.

Klimko, L. A. and Nelson, P. I. (1978). On conditional least squares estimation for stochastic processes, *Ann. Statist.*, **6**, 629–642.

Newey, W. K. and McFadden, D. L. (1994). Large sample estimation and hypothesis testing, *Handbook of Econometrics*, Vol. 4 (eds. R. F. Engle and D. L. McFadden), 2111–2245, Elsevier, New York.

Nicholls, D. F. and Quinn, B. G. (1982). *Random Coefficient Autoregressive Models: An Introduction*, Lecture Notes in Statist., Vol. 11, Springer, New York.

Pötscher, B. M. and Prucha, I. R. (1997). *Dynamic Nonlinear Econometric Models: Asymptotic Theory*, Springer, Berlin.

Shephard, N. (1996). Statistical aspects of ARCH and stochastic volatility, (eds. D. R. Cox, D. V. Hinkley and O. E. Barndorff-Nielson), *Time Series Models in Econometrics, Finance and Other Fields*, 1–55, Chapman & Hall, London.

Thavaneswaran, A. and Abraham, B. (1988). Estimation for nonlinear time series models using estimating equations, *J. Time Ser. Anal.*, **13**, 99–108.

Tjøstheim, D. (1986). Estimation in nonlinear time series models, *Stochastic. Process Appl.*, **21**, 251–273.

Tong, H. (1990). *Nonlinear Time Series: A Dynamical System Approach*, Oxford University Press, Oxford.

Wooldridge, J. M. (1994). Estimating and inference for dependent processes, *Handbook of Econometrics*, Vol. 4 (eds. R. F. Engle and D. L. McFadden), 2639–2738, Elsevier, New York.