Hindawi Publishing Corporation
Journal of Probability and Statistics
Volume 2011, Article ID 372512, 14 pages
doi:10.1155/2011/372512

Research Article

Joint Estimation Using Quadratic Estimating Function

Y. Liang,[1] A. Thavaneswaran,[1] and B. Abraham[2]

[1] *Department of Statistics, University of Manitoba, 338 Machray Hall, Winnipeg, MB, Canada R3T 2N2*
[2] *University of Waterloo, Waterloo, ON, Canada N2L 2G1*

Correspondence should be addressed to Y. Liang, umlian33@cc.umanitoba.ca

Received 12 January 2011; Revised 10 March 2011; Accepted 11 April 2011

Academic Editor: Ricardas Zitikis

A class of martingale estimating functions is convenient and plays an important role for inference for nonlinear time series models. However, when the information about the first four conditional moments of the observed process becomes available, the quadratic estimating functions are more informative. In this paper, a general framework for joint estimation of conditional mean and variance parameters in time series models using quadratic estimating functions is developed. Superiority of the approach is demonstrated by comparing the information associated with the optimal quadratic estimating function with the information associated with other estimating functions. The method is used to study the optimal quadratic estimating functions of the parameters of autoregressive conditional duration (ACD) models, random coefficient autoregressive (RCA) models, doubly stochastic models and regression models with ARCH errors. Closed-form expressions for the information gain are also discussed in some detail.

1. Introduction

Godambe [1] was the first to study the inference for discrete time stochastic processes using estimating function method. Thavaneswaran and Abraham [2] had studied the nonlinear time series estimation problems using linear estimating functions. Naik-Nimbalkar and Rajashi [3] and Thavaneswaran and Heyde [4] studied the filtering and prediction problems using linear estimating functions in the Bayesian context. Chandra and Taniguchi [5], Merkouris [6], and Ghahramani and Thavaneswaran [7] among others have studied the estimation problems using estimating functions. In this paper, we study the linear and quadratic *martingale* estimating functions and show that the quadratic estimating functions are more informative when the conditional mean and variance of the observed process depend on the same parameter of interest.

This paper is organized as follows. The rest of Section 1 presents the basics of estimating functions and information associated with estimating functions. Section 2 presents the general model for the multiparameter case and the form of the optimal quadratic estimating function. In Section 3, the theory is applied to four different models.

Suppose that $\{\mathbf{y}_t, t = 1,\ldots,n\}$ is a realization of a discrete time stochastic process, and its distribution depends on a vector parameter $\boldsymbol{\theta}$ belonging to an open subset $\boldsymbol{\Theta}$ of the p-dimensional Euclidean space. Let $(\Omega, \mathcal{F}, P_{\boldsymbol{\theta}})$ denote the underlying probability space, and let $\mathcal{F}_t^y$ be the σ-field generated by $\{\mathbf{y}_1,\ldots,\mathbf{y}_t, t \geq 1\}$. Let $\mathbf{h}_t = \mathbf{h}_t(\mathbf{y}_1,\ldots,\mathbf{y}_t,\boldsymbol{\theta})$, $1 \leq t \leq n$ be specified q-dimensional vectors that are martingales. We consider the class $\mathcal{M}$ of zero mean and square integrable p-dimensional martingale estimating functions of the form

$$\mathcal{M} = \left\{\mathbf{g}_n(\boldsymbol{\theta}) : \mathbf{g}_n(\boldsymbol{\theta}) = \sum_{t=1}^{n}\mathbf{a}_{t-1}\mathbf{h}_t\right\}, \tag{1.1}$$

where $\mathbf{a}_{t-1}$ are $p \times q$ matrices depending on $\mathbf{y}_1,\ldots,\mathbf{y}_{t-1}$, $1 \leq t \leq n$. The estimating functions $\mathbf{g}_n(\boldsymbol{\theta})$ are further assumed to be almost surely differentiable with respect to the components of $\boldsymbol{\theta}$ and such that $\mathrm{E}[(\partial\mathbf{g}_n(\boldsymbol{\theta})/\partial\boldsymbol{\theta}) \mid \mathcal{F}_{n-1}^y]$ and $\mathrm{E}[\mathbf{g}_n(\boldsymbol{\theta})\mathbf{g}_n(\boldsymbol{\theta})' \mid \mathcal{F}_{n-1}^y]$ are nonsingular for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and for each $n \geq 1$. The expectations are always taken with respect to $P_{\boldsymbol{\theta}}$. Estimators of $\boldsymbol{\theta}$ can be obtained by solving the estimating equation $\mathbf{g}_n(\boldsymbol{\theta}) = \mathbf{0}$. Furthermore, the $p\times p$ matrix $\mathrm{E}[\mathbf{g}_n(\boldsymbol{\theta})\mathbf{g}_n(\boldsymbol{\theta})' \mid \mathcal{F}_{n-1}^y]$ is assumed to be positive definite for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Then, in the class of all zero mean and square integrable martingale estimating functions $\mathcal{M}$, the optimal estimating function $\mathbf{g}_n^*(\boldsymbol{\theta})$ which maximizes, in the partial order of nonnegative definite matrices, the information matrix

$$\mathbf{I}_{\mathbf{g}_n}(\boldsymbol{\theta}) = \left(\mathrm{E}\left[\frac{\partial\mathbf{g}_n(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} \mid \mathcal{F}_{n-1}^y\right]\right)'\left(\mathrm{E}\left[\mathbf{g}_n(\boldsymbol{\theta})\mathbf{g}_n(\boldsymbol{\theta})' \mid \mathcal{F}_{n-1}^y\right]\right)^{-1}\left(\mathrm{E}\left[\frac{\partial\mathbf{g}_n(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} \mid \mathcal{F}_{n-1}^y\right]\right) \tag{1.2}$$

is given by

$$\mathbf{g}_n^*(\boldsymbol{\theta}) = \sum_{t=1}^{n}\mathbf{a}_{t-1}^*\mathbf{h}_t = \sum_{t=1}^{n}\left(\mathrm{E}\left[\frac{\partial\mathbf{h}_t}{\partial\boldsymbol{\theta}} \mid \mathcal{F}_{t-1}^y\right]\right)'\left(\mathrm{E}\left[\mathbf{h}_t\mathbf{h}_t' \mid \mathcal{F}_{t-1}^y\right]\right)^{-1}\mathbf{h}_t, \tag{1.3}$$

and the corresponding optimal information reduces to $\mathrm{E}[\mathbf{g}_n^*(\boldsymbol{\theta})\mathbf{g}_n^*(\boldsymbol{\theta})' \mid \mathcal{F}_{n-1}^y]$.

The function $\mathbf{g}_n^*(\boldsymbol{\theta})$ is also called the "quasi-score" and has properties similar to those of a score function in the sense that $\mathrm{E}[\mathbf{g}_n^*(\boldsymbol{\theta})] = \mathbf{0}$ and $\mathrm{E}[\mathbf{g}_n^*(\boldsymbol{\theta})\mathbf{g}_n^*(\boldsymbol{\theta})'] = -\mathrm{E}[\partial\mathbf{g}_n^*(\boldsymbol{\theta})/\partial\boldsymbol{\theta}']$. This is a more general result in the sense that for its validity, we do not need to assume that the true underlying distribution belongs to the exponential family of distributions. The maximum correlation between the optimal estimating function and the true unknown score justifies the terminology "quasi-score" for $\mathbf{g}_n^*(\boldsymbol{\theta})$. Moreover, it follows from Lindsay [8, page 916] that if we solve an unbiased estimating equation $\mathbf{g}_n(\boldsymbol{\theta}) = \mathbf{0}$ to get an estimator, then the asymptotic variance of the resulting estimator is the inverse of the information $\mathbf{I}_{\mathbf{g}_n}$. Hence, the estimator obtained from a more informative estimating equation is asymptotically more efficient.

2. General Model and Method

Consider a discrete time stochastic process $\{y_t, t = 1, 2, \ldots\}$ with conditional moments

$$\mu_t(\boldsymbol{\theta}) = \mathrm{E}\left[y_t \mid \mathcal{F}_{t-1}^y\right],$$
$$\sigma_t^2(\boldsymbol{\theta}) = \mathrm{Var}\left(y_t \mid \mathcal{F}_{t-1}^y\right),$$
$$\gamma_t(\boldsymbol{\theta}) = \frac{1}{\sigma_t^3(\boldsymbol{\theta})}\mathrm{E}\left[\left(y_t - \mu_t(\boldsymbol{\theta})\right)^3 \mid \mathcal{F}_{t-1}^y\right], \tag{2.1}$$
$$\kappa_t(\boldsymbol{\theta}) = \frac{1}{\sigma_t^4(\boldsymbol{\theta})}\mathrm{E}\left[\left(y_t - \mu_t(\boldsymbol{\theta})\right)^4 \mid \mathcal{F}_{t-1}^y\right] - 3.$$

That is, we assume that the skewness and the excess kurtosis of the standardized variable y_t do not contain any additional parameters. In order to estimate the parameter $\boldsymbol{\theta}$ based on the observations $y_1, \ldots, y_n$, we consider two classes of martingale differences $\{m_t(\boldsymbol{\theta}) = y_t - \mu_t(\boldsymbol{\theta}),\ t = 1, \ldots, n\}$ and $\{s_t(\boldsymbol{\theta}) = m_t^2(\boldsymbol{\theta}) - \sigma_t^2(\boldsymbol{\theta}),\ t = 1, \ldots, n\}$ such that

$$\langle m\rangle_t = \mathrm{E}\left[m_t^2 \mid \mathcal{F}_{t-1}^y\right] = \mathrm{E}\left[\left(y_t - \mu_t\right)^2 \mid \mathcal{F}_{t-1}^y\right] = \sigma_t^2,$$
$$\langle s\rangle_t = \mathrm{E}\left[s_t^2 \mid \mathcal{F}_{t-1}^y\right] = \mathrm{E}\left[\left(y_t - \mu_t\right)^4 + \sigma_t^4 - 2\sigma_t^2\left(y_t - \mu_t\right)^2 \mid \mathcal{F}_{t-1}^y\right] = \sigma_t^4(\kappa_t + 2), \tag{2.2}$$
$$\langle m, s\rangle_t = \mathrm{E}\left[m_t s_t \mid \mathcal{F}_{t-1}^y\right] = \mathrm{E}\left[\left(y_t - \mu_t\right)^3 - \sigma_t^2\left(y_t - \mu_t\right) \mid \mathcal{F}_{t-1}^y\right] = \sigma_t^3\gamma_t.$$

The optimal estimating functions based on the martingale differences m_t and s_t are $\mathbf{g}_M^*(\boldsymbol{\theta}) = -\sum_{t=1}^n(\partial\mu_t/\partial\boldsymbol{\theta})(m_t/\langle m\rangle_t)$ and $\mathbf{g}_S^*(\boldsymbol{\theta}) = -\sum_{t=1}^n(\partial\sigma_t^2/\partial\boldsymbol{\theta})(s_t/\langle s\rangle_t)$, respectively. Then, the information associated with $\mathbf{g}_M^*(\boldsymbol{\theta})$ and $\mathbf{g}_S^*(\boldsymbol{\theta})$ are $\mathbf{I}_{\mathbf{g}_M^*}(\boldsymbol{\theta}) = \sum_{t=1}^n(\partial\mu_t/\partial\boldsymbol{\theta})(\partial\mu_t/\partial\boldsymbol{\theta}')(1/\langle m\rangle_t)$ and $\mathbf{I}_{\mathbf{g}_S^*}(\boldsymbol{\theta}) = \sum_{t=1}^n(\partial\sigma_t^2/\partial\boldsymbol{\theta})(\partial\sigma_t^2/\partial\boldsymbol{\theta}')(1/\langle s\rangle_t)$, respectively. Crowder [9] studied the optimal quadratic estimating function with independent observations. For the discrete time stochastic process $\{y_t\}$, the following theorem provides optimality of the quadratic estimating function for the multiparameter case.

Theorem 2.1. *For the general model in* (2.1)*, in the class of all quadratic estimating functions of the form* $\mathcal{G}_Q = \{\mathbf{g}_Q(\boldsymbol{\theta}) : \mathbf{g}_Q(\boldsymbol{\theta}) = \sum_{t=1}^n(\mathbf{a}_{t-1}m_t + \mathbf{b}_{t-1}s_t)\}$,

(a) *the optimal estimating function is given by* $\mathbf{g}_Q^*(\boldsymbol{\theta}) = \sum_{t=1}^n(\mathbf{a}_{t-1}^*m_t + \mathbf{b}_{t-1}^*s_t)$*, where*

$$\mathbf{a}_{t-1}^* = \left(1 - \frac{\langle m, s\rangle_t^2}{\langle m\rangle_t\langle s\rangle_t}\right)^{-1}\left(-\frac{\partial\mu_t}{\partial\boldsymbol{\theta}}\frac{1}{\langle m\rangle_t} + \frac{\partial\sigma_t^2}{\partial\boldsymbol{\theta}}\frac{\langle m, s\rangle_t}{\langle m\rangle_t\langle s\rangle_t}\right),$$
$$\mathbf{b}_{t-1}^* = \left(1 - \frac{\langle m, s\rangle_t^2}{\langle m\rangle_t\langle s\rangle_t}\right)^{-1}\left(\frac{\partial\mu_t}{\partial\boldsymbol{\theta}}\frac{\langle m, s\rangle_t}{\langle m\rangle_t\langle s\rangle_t} - \frac{\partial\sigma_t^2}{\partial\boldsymbol{\theta}}\frac{1}{\langle s\rangle_t}\right); \tag{2.3}$$

(b) *the information* $\mathbf{I}_{g_Q^*}(\boldsymbol{\theta})$ *is given by*

$$\mathbf{I}_{\mathbf{g}_Q^*}(\boldsymbol{\theta})=\sum_{t=1}^{n}\left(1-\frac{\langle m,s\rangle_t^2}{\langle m\rangle_t\langle s\rangle_t}\right)^{-1}\left(\frac{\partial\mu_t}{\partial\boldsymbol{\theta}}\frac{\partial\mu_t}{\partial\boldsymbol{\theta}'}\frac{1}{\langle m\rangle_t}+\frac{\partial\sigma_t^2}{\partial\boldsymbol{\theta}}\frac{\partial\sigma_t^2}{\partial\boldsymbol{\theta}'}\frac{1}{\langle s\rangle_t}-\left(\frac{\partial\mu_t}{\partial\boldsymbol{\theta}}\frac{\partial\sigma_t^2}{\partial\boldsymbol{\theta}'}+\frac{\partial\sigma_t^2}{\partial\boldsymbol{\theta}}\frac{\partial\mu_t}{\partial\boldsymbol{\theta}'}\right)\frac{\langle m,s\rangle_t}{\langle m\rangle_t\langle s\rangle_t}\right); \tag{2.4}$$

(c) *the gain in information* $\mathbf{I}_{\mathbf{g}_Q^*}(\boldsymbol{\theta})-\mathbf{I}_{\mathbf{g}_M^*}(\boldsymbol{\theta})$ *is given by*

$$\sum_{t=1}^{n}\left(1-\frac{\langle m,s\rangle_t^2}{\langle m\rangle_t\langle s\rangle_t}\right)^{-1}\left(\frac{\partial\mu_t}{\partial\boldsymbol{\theta}}\frac{\partial\mu_t}{\partial\boldsymbol{\theta}'}\frac{\langle m,s\rangle_t^2}{\langle m\rangle_t^2\langle s\rangle_t}+\frac{\partial\sigma_t^2}{\partial\boldsymbol{\theta}}\frac{\partial\sigma_t^2}{\partial\boldsymbol{\theta}'}\frac{1}{\langle s\rangle_t}-\left(\frac{\partial\mu_t}{\partial\boldsymbol{\theta}}\frac{\partial\sigma_t^2}{\partial\boldsymbol{\theta}'}+\frac{\partial\sigma_t^2}{\partial\boldsymbol{\theta}}\frac{\partial\mu_t}{\partial\boldsymbol{\theta}'}\right)\frac{\langle m,s\rangle_t}{\langle m\rangle_t\langle s\rangle_t}\right); \tag{2.5}$$

(d) *the gain in information* $\mathbf{I}_{\mathbf{g}_Q^*}(\boldsymbol{\theta})-\mathbf{I}_{\mathbf{g}_S^*}(\boldsymbol{\theta})$ *is given by*

$$\begin{aligned}&\sum_{t=1}^{n}\left(1-\frac{\langle m,s\rangle_t^2}{\langle m\rangle_t\langle s\rangle_t}\right)^{-1}\left(\frac{\partial\mu_t}{\partial\boldsymbol{\theta}}\frac{\partial\mu_t}{\partial\boldsymbol{\theta}'}\frac{1}{\langle m\rangle_t}+\frac{\partial\sigma_t^2}{\partial\boldsymbol{\theta}}\frac{\partial\sigma_t^2}{\partial\boldsymbol{\theta}'}\frac{\langle m,s\rangle_t^2}{\langle m\rangle_t\langle s\rangle_t-\langle m,s\rangle_t^2}\right.\\&\qquad\left.-\left(\frac{\partial\mu_t}{\partial\boldsymbol{\theta}}\frac{\partial\sigma_t^2}{\partial\boldsymbol{\theta}'}+\frac{\partial\sigma_t^2}{\partial\boldsymbol{\theta}}\frac{\partial\mu_t}{\partial\boldsymbol{\theta}'}\right)\frac{\langle m,s\rangle_t}{\langle m\rangle_t\langle s\rangle_t}\right).\end{aligned} \tag{2.6}$$

Proof. We choose two orthogonal martingale differences m_t and $\psi_t = s_t - \sigma_t\gamma_t m_t$, where the conditional variance of ψ_t is given by $\langle\psi\rangle_t = (\langle m\rangle_t\langle s\rangle_t - \langle m,s\rangle_t^2)/\langle m\rangle_t = \sigma_t^4(\kappa_t + 2 - \gamma_t^2)$. That is, m_t and ψ_t are uncorrelated with conditional variance $\langle m\rangle_t$ and $\langle\psi\rangle_t$, respectively. Moreover, the optimal martingale estimating function and associated information based on the martingale differences ψ_t are

$$\begin{aligned}\mathbf{g}_\Psi^*(\boldsymbol{\theta}) &= \sum_{t=1}^{n}\left(\frac{\partial\mu_t}{\partial\boldsymbol{\theta}}\frac{\langle m,s\rangle_t}{\langle m\rangle_t}-\frac{\partial\sigma_t^2}{\partial\boldsymbol{\theta}}\right)\frac{\psi_t}{\langle\psi\rangle_t}\\&= \sum_{t=1}^{n}\left(1-\frac{\langle m,s\rangle_t^2}{\langle m\rangle_t\langle s\rangle_t}\right)^{-1}\\&\quad\times\left(\left(-\frac{\partial\mu_t}{\partial\boldsymbol{\theta}}\frac{\langle m,s\rangle_t^2}{\langle m\rangle_t^2\langle s\rangle_t}+\frac{\partial\sigma_t^2}{\partial\boldsymbol{\theta}}\frac{\langle m,s\rangle_t}{\langle m\rangle_t\langle s\rangle_t}\right)m_t+\left(\frac{\partial\mu_t}{\partial\boldsymbol{\theta}}\frac{\langle m,s\rangle_t}{\langle m\rangle_t\langle s\rangle_t}-\frac{\partial\sigma_t^2}{\partial\boldsymbol{\theta}}\frac{1}{\langle s\rangle_t}\right)s_t\right),\\\mathbf{I}_{\mathbf{g}_\Psi^*}(\boldsymbol{\theta}) &= \sum_{t=1}^{n}\left(\frac{\partial\mu_t}{\partial\boldsymbol{\theta}}\frac{\langle m,s\rangle_t}{\langle m\rangle_t}-\frac{\partial\sigma_t^2}{\partial\boldsymbol{\theta}}\right)\left(\frac{\partial\mu_t}{\partial\boldsymbol{\theta}'}\frac{\langle m,s\rangle_t}{\langle m\rangle_t}-\frac{\partial\sigma_t^2}{\partial\boldsymbol{\theta}'}\right)\frac{1}{\langle\psi\rangle_t}\\&= \sum_{t=1}^{n}\left(1-\frac{\langle m,s\rangle_t^2}{\langle m\rangle_t\langle s\rangle_t}\right)^{-1}\\&\quad\times\left(\frac{\partial\mu_t}{\partial\boldsymbol{\theta}}\frac{\partial\mu_t}{\partial\boldsymbol{\theta}'}\frac{\langle m,s\rangle_t^2}{\langle m\rangle_t^2\langle s\rangle_t}+\frac{\partial\sigma_t^2}{\partial\boldsymbol{\theta}}\frac{\partial\sigma_t^2}{\partial\boldsymbol{\theta}'}\frac{1}{\langle s\rangle_t}-\left(\frac{\partial\mu_t}{\partial\boldsymbol{\theta}}\frac{\partial\sigma_t^2}{\partial\boldsymbol{\theta}'}+\frac{\partial\sigma_t^2}{\partial\boldsymbol{\theta}}\frac{\partial\mu_t}{\partial\boldsymbol{\theta}'}\right)\frac{\langle m,s\rangle_t}{\langle m\rangle_t\langle s\rangle_t}\right).\end{aligned} \tag{2.7}$$

Then, the quadratic estimating function based on m_t and ψ_t becomes

$$\begin{aligned}\mathbf{g}_Q^*(\boldsymbol{\theta}) = &\sum_{t=1}^{n}\left(1-\frac{\langle m,s\rangle_t^2}{\langle m\rangle_t\langle s\rangle_t}\right)^{-1} \\ &\times\left(\left(-\frac{\partial\mu_t}{\partial\boldsymbol{\theta}}\frac{1}{\langle m\rangle_t}+\frac{\partial\sigma_t^2}{\partial\boldsymbol{\theta}}\frac{\langle m,s\rangle_t}{\langle m\rangle_t\langle s\rangle_t}\right)m_t+\left(\frac{\partial\mu_t}{\partial\boldsymbol{\theta}}\frac{\langle m,s\rangle_t}{\langle m\rangle_t\langle s\rangle_t}-\frac{\partial\sigma_t^2}{\partial\boldsymbol{\theta}}\frac{1}{\langle s\rangle_t}\right)s_t\right)\end{aligned} \tag{2.8}$$

and satisfies the sufficient condition for optimality

$$\mathrm{E}\left[\frac{\partial\mathbf{g}_Q(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\mid\mathcal{F}_{t-1}^y\right]=\mathrm{Cov}\left(\mathbf{g}_Q(\boldsymbol{\theta}),\mathbf{g}_Q^*(\boldsymbol{\theta})\mid\mathcal{F}_{t-1}^y\right)K,\quad\forall\mathbf{g}_Q(\boldsymbol{\theta})\in\mathcal{G}_Q, \tag{2.9}$$

where K is a constant matrix. Hence, $\mathbf{g}_Q^*(\boldsymbol{\theta})$ is optimal in the class $\mathcal{G}_Q$, and part (a) follows. Since m_t and ψ_t are orthogonal, the information $\mathbf{I}_{\mathbf{g}_Q^*}(\boldsymbol{\theta}) = \mathbf{I}_{\mathbf{g}_M^*}(\boldsymbol{\theta}) + \mathbf{I}_{\mathbf{g}_\Psi^*}(\boldsymbol{\theta})$ and part (b) follow. Hence, for each component θ_i, $i = 1,\dots,p$, neither $g_M^*(\theta_i)$ nor $g_S^*(\theta)$ is fully informative, that is, $I_{g_Q^*}(\theta_i) \geq I_{g_M^*}(\theta_i)$ and $I_{g_Q^*}(\theta_i) \geq I_{g_S^*}(\theta_i)$. □

Corollary 2.2. *When the conditional skewness γ and kurtosis κ are constants, the optimal quadratic estimating function and associated information, based on the martingale differences $m_t = y_t - \mu_t$ and $s_t = m_t^2 - \sigma_t^2$, are given by*

$$\begin{aligned}\mathbf{g}_Q^*(\boldsymbol{\theta}) &= \left(1-\frac{\gamma^2}{\kappa+2}\right)^{-1}\sum_{t=1}^{n}\frac{1}{\sigma_t^3}\left(\left(-\sigma_t\frac{\partial\mu_t}{\partial\boldsymbol{\theta}}+\frac{\gamma}{\kappa+2}\frac{\partial\sigma_t^2}{\partial\boldsymbol{\theta}}\right)m_t+\frac{1}{\kappa+2}\left(\gamma\frac{\partial\mu_t}{\partial\boldsymbol{\theta}}-\frac{1}{\sigma_t}\frac{\partial\sigma_t^2}{\partial\boldsymbol{\theta}}\right)s_t\right), \\ \mathbf{I}_{\mathbf{g}_Q^*}(\boldsymbol{\theta}) &= \left(1-\frac{\gamma^2}{\kappa+2}\right)^{-1}\left(\mathbf{I}_{\mathbf{g}_M^*}(\boldsymbol{\theta})+\mathbf{I}_{\mathbf{g}_S^*}(\boldsymbol{\theta})-\frac{\gamma}{\kappa+2}\sum_{t=1}^{n}\frac{1}{\sigma_t^3}\left(\frac{\partial\mu_t}{\partial\boldsymbol{\theta}}\frac{\partial\sigma_t^2}{\partial\boldsymbol{\theta}'}+\frac{\partial\sigma_t^2}{\partial\boldsymbol{\theta}}\frac{\partial\mu_t}{\partial\boldsymbol{\theta}'}\right)\right).\end{aligned} \tag{2.10}$$

3. Applications

3.1. Autoregressive Conditional Duration (ACD) Models

There is growing interest in the analysis of intraday financial data such as transaction and quote data. Such data have increasingly been made available by many stock exchanges. Unlike closing prices which are measured daily, monthly, or yearly, intra-day data or high-frequency data tend to be irregularly spaced. Furthermore, the durations between events themselves are random variables. The autoregressive conditional duration (ACD) process due to Engle and Russell [10] had been proposed to model such durations, in order to study the dynamic structure of the adjusted durations x_i, with $x_i = t_i - t_{i-1}$, where t_i is the time of the ith transaction. The crucial assumption underlying the ACD model is that the time dependence is described by a function ψ_i, where ψ_i is the conditional expectation of the adjusted duration between the $(i-1)$th and the ith trades. The basic ACD model is defined as

$$x_i = \psi_i \varepsilon_i, \qquad \psi_i = \mathrm{E}\left[x_i \mid \mathcal{F}^x_{t_i-1}\right], \tag{3.1}$$

where ε_i are the iid nonnegative random variables with density function $f(\cdot)$ and unit mean, and $\mathcal{F}^x_{t_i-1}$ is the information available at the $(i-1)$th trade. We also assume that ε_i is independent of $\mathcal{F}^x_{t-1}$. It is clear that the types of ACD models vary according to different distributions of ε_i and specifications of ψ_i. In this paper, we will discuss a specific class of models which is known as ACD (p, q) model and given by

$$x_t = \psi_t \varepsilon_t, \qquad \psi_t = \omega + \sum_{j=1}^{p} a_j x_{t-j} + \sum_{j=1}^{q} b_j \psi_{t-j}, \tag{3.2}$$

where $\omega > 0$, $a_j > 0$, $b_j > 0$, and $\sum_{j=1}^{\max(p,q)}(a_j + b_j) < 1$. We assume that ε_t's are iid nonnegative random variables with mean μ_ε, variance σ_ε^2, skewness γ_ε, and excess kurtosis κ_ε. In order to estimate the parameter vector $\boldsymbol{\theta} = (\omega, a_1, \ldots, a_p, b_1, \ldots, b_q)'$, we use the estimating function approach. For this model, the conditional moments are $\mu_t = \mu_\varepsilon \psi_t$, $\sigma_t^2 = \sigma_\varepsilon^2 \psi_t^2$, $\gamma_t = \gamma_\varepsilon$, and $\kappa_t = \kappa_\varepsilon$. Let $m_t = x_t - \mu_t$ and $s_t = m_t^2 - \sigma_t^2$ be the sequences of martingale differences such that $\langle m \rangle_t = \sigma_\varepsilon^2 \psi_t^2$, $\langle s \rangle_t = \sigma_\varepsilon^4(\kappa_\varepsilon + 2)\psi_t^4$, and $\langle m, s \rangle_t = \sigma_\varepsilon^3 \gamma_\varepsilon \psi_t^3$. The optimal estimating function and associated information based on m_t are given by $\mathbf{g}_M^*(\boldsymbol{\theta}) = -(\mu_\varepsilon/\sigma_\varepsilon^2)\sum_{t=1}^n (1/\psi_t^2)(\partial\psi_t/\partial\boldsymbol{\theta}) m_t$ and $\mathbf{I}_{\mathbf{g}_M^*}(\boldsymbol{\theta}) = (\mu_\varepsilon^2/\sigma_\varepsilon^2)\sum_{t=1}^n (1/\psi_t^2)(\partial\psi_t/\partial\boldsymbol{\theta})(\partial\psi_t/\partial\boldsymbol{\theta}')$. The optimal estimating function and the associated information based on s_t are given by $\mathbf{g}_S^*(\boldsymbol{\theta}) = -2/\sigma_\varepsilon^2(\kappa_\varepsilon + 2)\sum_{t=1}^n (1/\psi_t^3)(\partial\psi_t/\partial\boldsymbol{\theta}) s_t$ and $\mathbf{I}_{\mathbf{g}_S^*}(\boldsymbol{\theta}) = (4/(\kappa_\varepsilon + 2))\sum_{t=1}^n (1/\psi_t^2)(\partial\psi_t/\partial\boldsymbol{\theta})(\partial\psi_t/\partial\boldsymbol{\theta}')$. Then, by Corollary 2.2 that the optimal quadratic estimating function and associated information are given by

$$\begin{aligned}
\mathbf{g}_Q^*(\boldsymbol{\theta}) &= \frac{1}{\sigma_\varepsilon^2(\kappa_\varepsilon + 2 - \gamma_\varepsilon^2)} \sum_{t=1}^n \left(\frac{-\mu_\varepsilon(\kappa_\varepsilon + 2) + 2\sigma_\varepsilon\gamma_\varepsilon}{\psi_t^2} \frac{\partial\psi_t}{\partial\boldsymbol{\theta}} m_t + \frac{\mu_\varepsilon\gamma_\varepsilon - 2\sigma_\varepsilon\psi_t}{\sigma_\varepsilon\psi_t^3} \frac{\partial\psi_t}{\partial\boldsymbol{\theta}} s_t \right), \\
\mathbf{I}_{\mathbf{g}_Q^*}(\boldsymbol{\theta}) &= \left(1 - \frac{\gamma_\varepsilon^2}{\kappa_\varepsilon + 2}\right)^{-1} \left(\mathbf{I}_{\mathbf{g}_M^*}(\boldsymbol{\theta}) + \mathbf{I}_{\mathbf{g}_S^*}(\boldsymbol{\theta}) - \frac{4\mu_\varepsilon\gamma_\varepsilon}{\sigma_\varepsilon(\kappa_\varepsilon + 2)} \sum_{t=1}^n \frac{1}{\psi_t^2} \frac{\partial\psi_t}{\partial\boldsymbol{\theta}} \frac{\partial\psi_t}{\partial\boldsymbol{\theta}'} \right) \\
&= \frac{4\sigma_\varepsilon^2 + \mu_\varepsilon^2(\kappa_\varepsilon + 2) - 4\mu_\varepsilon\sigma_\varepsilon\gamma_\varepsilon}{\sigma_\varepsilon^2(\kappa_\varepsilon + 2 - \gamma_\varepsilon^2)} \sum_{t=1}^n \frac{1}{\psi_t^2} \frac{\partial\psi_t}{\partial\boldsymbol{\theta}} \frac{\partial\psi_t}{\partial\boldsymbol{\theta}'},
\end{aligned} \tag{3.3}$$

the information gain in using $\mathbf{g}_Q^*(\boldsymbol{\theta})$ over $\mathbf{g}_M^*(\boldsymbol{\theta})$ is

$$\frac{(2\sigma_\varepsilon - \mu_\varepsilon\gamma_\varepsilon)^2}{\sigma_\varepsilon^2(\kappa_\varepsilon + 2 - \gamma_\varepsilon^2)} \sum_{t=1}^n \frac{1}{\psi_t^2} \frac{\partial\psi_t}{\partial\boldsymbol{\theta}} \frac{\partial\psi_t}{\partial\boldsymbol{\theta}'}, \tag{3.4}$$

and the information gain in using $\mathbf{g}_Q^*(\boldsymbol{\theta})$ over $\mathbf{g}_S^*(\boldsymbol{\theta})$ is

$$\frac{\left(\mu_\varepsilon(\kappa_\varepsilon+2)-2\sigma_\varepsilon\gamma_\varepsilon\right)^2}{\sigma_\varepsilon^2\left(\kappa_\varepsilon+2-\gamma_\varepsilon^2\right)(\kappa_\varepsilon+2)}\sum_{t=1}^{n}\frac{1}{\psi_t^2}\frac{\partial\psi_t}{\partial\boldsymbol{\theta}}\frac{\partial\psi_t}{\partial\boldsymbol{\theta}'}, \tag{3.5}$$

which are both nonnegative definite.

When ε_t follows an exponential distribution, $\mu_\varepsilon = 1/\lambda$, $\sigma_\varepsilon^2 = 1/\lambda^2$, $\gamma_\varepsilon = 2$, and $\kappa_\varepsilon = 3$. Then, $\mathbf{I}_{\mathbf{g}_M^*}(\boldsymbol{\theta}) = \sum_{t=1}^{n}(1/\psi_t^2)(\partial\psi_t/\partial\boldsymbol{\theta})(\partial\psi_t/\partial\boldsymbol{\theta}')$, $\mathbf{I}_{\mathbf{g}_S^*}(\boldsymbol{\theta}) = (4/5)\sum_{t=1}^{n}(1/\psi_t^2)(\partial\psi_t/\partial\boldsymbol{\theta})(\partial\psi_t/\partial\boldsymbol{\theta}')$, and $\mathbf{I}_{\mathbf{g}_Q^*}(\boldsymbol{\theta}) = \sum_{t=1}^{n}(1/\psi_t^2)(\partial\psi_t/\partial\boldsymbol{\theta})(\partial\psi_t/\partial\boldsymbol{\theta}')$, and hence $\mathbf{I}_{\mathbf{g}_Q^*}(\boldsymbol{\theta}) = \mathbf{I}_{\mathbf{g}_M^*}(\boldsymbol{\theta}) > \mathbf{I}_{\mathbf{g}_S^*}(\boldsymbol{\theta})$.

3.2. *Random Coefficient Autoregressive Models*

In this section, we will investigate the properties of the quadratic estimating functions for the random coefficient autoregressive (RCA) time series which were first introduced by Nicholls and Quinn [11].

Consider the RCA model

$$y_t = (\theta + b_t)y_{t-1} + \varepsilon_t, \tag{3.6}$$

where $\{b_t\}$ and $\{\varepsilon_t\}$ are uncorrelated zero mean processes with unknown variance σ_b^2 and variance $\sigma_\varepsilon^2 = \sigma_\varepsilon^2(\theta)$ with unknown parameter θ, respectively. Further, we denote the skewness and excess kurtosis of $\{b_t\}$ by γ_b, κ_b which are known, and of $\{\varepsilon_t\}$ by $\gamma_\varepsilon(\theta)$, $\kappa_\varepsilon(\theta)$, respectively. In the model (3.6), both the parameter θ and $\beta = \sigma_b^2$ need to be estimated. Let $\boldsymbol{\theta} = (\theta,\beta)'$, we will discuss the joint estimation of θ and β. In this model, the conditional mean is $\mu_t = y_{t-1}\theta$ then and the conditional variance is $\sigma_t^2 = y_{t-1}^2\beta + \sigma_\varepsilon^2(\theta)$. The parameter θ appears simultaneously in the mean and variance. Let $m_t = y_t - \mu_t$ and $s_t = m_t^2 - \sigma_t^2$ such that $\langle m\rangle_t = y_{t-1}^2\sigma_b^2 + \sigma_\varepsilon^2$, $\langle s\rangle_t = y_{t-1}^4\sigma_b^4(\kappa_b+2) + \sigma_\varepsilon^4(\kappa_\varepsilon+2) + 4y_{t-1}^2\sigma_b^2\sigma_\varepsilon^2$, $\langle m,s\rangle_t = y_{t-1}^3\sigma_b^3\gamma_b + \sigma_\varepsilon^3\gamma_\varepsilon$. Then the conditional skewness is $\gamma_t = \langle m,s\rangle_t/\sigma_t^3$, and the conditional excess kurtosis is $\kappa_t = \langle s\rangle_t/\sigma_t^4 - 2$.

Since $\partial\mu_t/\partial\boldsymbol{\theta} = (y_{t-1},0)'$ and $\partial\sigma_t^2/\partial\boldsymbol{\theta} = (\partial\sigma_\varepsilon^2/\partial\theta, y_{t-1}^2)'$, by applying Theorem 2.1, the optimal quadratic estimating function for θ and β based on the martingale differences m_t and s_t is given by $\mathbf{g}_Q^*(\boldsymbol{\theta}) = \sum_{t=1}^{n}\mathbf{a}_{t-1}^*m_t + \mathbf{b}_{t-1}^*s_t$, where

$$\begin{aligned}
\mathbf{a}_{t-1}^* &= \left(1-\frac{\langle m,s\rangle_t^2}{\langle m\rangle_t\langle s\rangle_t}\right)^{-1}\left(\left(-\frac{y_{t-1}}{\langle m\rangle_t}+\frac{\partial\sigma_\varepsilon^2}{\partial\theta}\frac{\langle m,s\rangle_t}{\langle m\rangle_t\langle s\rangle_t}\right),\frac{y_{t-1}^2\langle m,s\rangle_t}{\langle m\rangle_t\langle s\rangle_t}\right)',\\
\mathbf{b}_{t-1}^* &= \left(1-\frac{\langle m,s\rangle_t^2}{\langle m\rangle_t\langle s\rangle_t}\right)^{-1}\left(\left(\frac{y_{t-1}\langle m,s\rangle_t}{\langle m\rangle_t\langle s\rangle_t}-\frac{\partial\sigma_\varepsilon^2}{\partial\theta}\frac{1}{\langle s\rangle_t}\right),-\frac{y_{t-1}^2}{\langle s\rangle_t}\right)'.
\end{aligned} \tag{3.7}$$

Hence, the component quadratic estimating function for θ is

$$g_Q^*(\theta) = \sum_{t=1}^{n}\left(1-\frac{\langle m,s\rangle_t^2}{\langle m\rangle_t\langle s\rangle_t}\right)^{-1} \times\left(\left(-\frac{y_{t-1}}{\langle m\rangle_t}+\frac{\partial\sigma_\varepsilon^2}{\partial\theta}\frac{\langle m,s\rangle_t}{\langle m\rangle_t\langle s\rangle_t}\right)m_t+\left(\frac{y_{t-1}\langle m,s\rangle_t}{\langle m\rangle_t\langle s\rangle_t}-\frac{\partial\sigma_\varepsilon^2}{\partial\theta}\frac{1}{\langle s\rangle_t}\right)s_t\right), \tag{3.8}$$

and the component quadratic estimating function for β is

$$g_Q^*(\beta) = \sum_{t=1}^{n}\left(1-\frac{\langle m,s\rangle_t^2}{\langle m\rangle_t\langle s\rangle_t}\right)^{-1}\left(\frac{y_{t-1}^2\langle m,s\rangle_t m_t}{\langle m\rangle_t\langle s\rangle_t}-\frac{y_{t-1}^2 s_t}{\langle s\rangle_t}\right). \tag{3.9}$$

Moreover, the information matrix of the optimal quadratic estimating function for θ and β is given by

$$\mathbf{I}_{g_Q^*}(\boldsymbol{\theta}) = \begin{pmatrix} I_{\theta\theta} & I_{\theta\beta} \\ I_{\beta\theta} & I_{\beta\beta}\end{pmatrix}, \tag{3.10}$$

where

$$I_{\theta\theta} = \sum_{t=1}^{n}\left(1-\frac{\langle m,s\rangle_t^2}{\langle m\rangle_t\langle s\rangle_t}\right)^{-1}\left(\frac{y_{t-1}^2}{\langle m\rangle_t}+\left(\frac{\partial\sigma_\varepsilon^2}{\partial\theta}\right)^2\frac{1}{\langle s\rangle_t}-2\frac{\partial\sigma_\varepsilon^2}{\partial\theta}\frac{y_{t-1}\langle m,s\rangle_t}{\langle m\rangle_t\langle s\rangle_t}\right), \tag{3.11}$$

$$I_{\theta\beta} = I_{\beta\theta} = \sum_{t=1}^{n}\left(1-\frac{\langle m,s\rangle_t^2}{\langle m\rangle_t\langle s\rangle_t}\right)^{-1}\left(\frac{\partial\sigma_\varepsilon^2}{\partial\theta}\frac{1}{\langle s\rangle_t}-\frac{y_{t-1}\langle m,s\rangle_t}{\langle m\rangle_t\langle s\rangle_t}\right)y_{t-1}^2, \tag{3.12}$$

$$I_{\beta\beta} = \sum_{t=1}^{n}\left(1-\frac{\langle m,s\rangle_t^2}{\langle m\rangle_t\langle s\rangle_t}\right)^{-1}\frac{y_{t-1}^4}{\langle s\rangle_t}. \tag{3.13}$$

In view of the parameter θ only, the conditional least squares (CLS) estimating function and the associated information are directly given by $g_{\text{CLS}}(\theta) = \sum_{t=1}^{n} y_{t-1}m_t$ and $I_{\text{CLS}}(\theta) = (\sum_{t=1}^{n} y_{t-1}^2)^2/\sum_{t=1}^{n} y_{t-1}^2\langle m\rangle_t$. The optimal martingale estimating function and the associated information based on m_t are given by $g_M^*(\theta) = -\sum_{t=1}^{n}(y_{t-1}m_t/\langle m\rangle_t)$ and $I_{g_M^*}(\theta) = \sum_{t=1}^{n}(y_{t-1}^2/\langle m\rangle_t)$. Moreover, the inequality

$$\left(\sum_{t=1}^{n} y_{t-1}^2\langle m\rangle_t\right)\left(\sum_{t=1}^{n}\frac{y_{t-1}^2}{\langle m\rangle_t}\right) \geq \left(\sum_{t=1}^{n} y_{t-1}^2\right)^2 \tag{3.14}$$

implies that $I_{\text{CLS}}(\theta) \leq I_{g_M^*}(\theta)$. Hence the optimal estimating function is more informative than the conditional least squares one. The optimal quadratic estimating function based on the martingale differences m_t and s_t is given by (3.8) and (3.11), respectively. It is obvious to see that the information of $g_Q^*(\theta)$ is larger than that of $g_M^*(\theta)$. Therefore, we can conclude that for the RCA model, $I_{\text{CLS}}(\theta) \le I_{g_M^*}(\theta) \le I_{g_Q^*}(\theta)$, and hence, the estimate obtained by solving the optimal quadratic estimating equation is more efficient than the CLS estimate and the estimate obtained by solving the optimal linear estimating equation.

3.3. Doubly Stochastic Time Series Model

Random coefficient autoregressive models we discussed in the previous section are special cases of what Tjøstheim [12] refers to as doubly stochastic time series model. In the nonlinear case, these models are given by

$$y_t = \theta_t f\left(t, \mathcal{F}_{t-1}^y\right) + \varepsilon_t, \tag{3.15}$$

where $\{\theta + b_t\}$ of (3.6) is replaced by a more general stochastic sequence $\{\theta_t\}$ and y_{t-1} is replaced by a function of the past, $\mathcal{F}_{t-1}^y$. Suppose that $\{\theta_t\}$ is a moving average sequence of the form

$$\theta_t = \theta + a_t + a_{t-1}, \tag{3.16}$$

where $\{a_t\}$ consists of square integrable independent random variables with mean zero and variance σ_a^2. We further assume that $\{\varepsilon_t\}$ and $\{a_t\}$ are independent, then $\mathrm{E}[y_t \mid \mathcal{F}_{t-1}^y]$ depends on the posterior mean $u_t = \mathrm{E}[a_t \mid \mathcal{F}_{t-1}^y]$, and variance $v_t = \mathrm{E}[(a_t - u_t)^2 \mid \mathcal{F}_{t-1}^y]$ of a_t. Under the normality assumption of $\{\varepsilon_t\}$ and $\{a_t\}$, and the initial condition $y_0 = 0$, u_t and v_t satisfy the following Kalman-like recursive algorithms (see [13, page 439]):

$$\begin{aligned}
u_t(\theta) &= \frac{\sigma_a^2 f\left(t, \mathcal{F}_{t-1}^y\right)\left(y_t - (\theta + m_{t-1}) f\left(t, \mathcal{F}_{t-1}^y\right)\right)}{\sigma_e^2(\theta) + f^2\left(t, \mathcal{F}_{t-1}^y\right)\left(\sigma_a^2 + v_{t-1}\right)}, \\
v_t(\theta) &= \sigma_a^2 - \frac{\sigma_a^4 f^2\left(t, \mathcal{F}_{t-1}^y\right)}{\sigma_e^2(\theta) + f^2\left(t, \mathcal{F}_{t-1}^y\right)\left(\sigma_a^2 + v_{t-1}\right)},
\end{aligned} \tag{3.17}$$

where $u_0 = 0$ and $v_0 = \sigma_a^2$. Hence, the conditional mean and variance of y_t are given by

$$\begin{aligned}
\mu_t(\theta) &= (\theta + u_{t-1}(\theta)) f\left(t, \mathcal{F}_{t-1}^y\right), \\
\sigma_t^2(\theta) &= \sigma_e^2(\theta) + f^2\left(t, \mathcal{F}_{t-1}^y\right)\left(\sigma_a^2 + v_{t-1}(\theta)\right),
\end{aligned} \tag{3.18}$$

which can be computed recursively.

Let $m_t = y_t - \mu_t$ and $s_t = m_t^2 - \sigma_t^2$, then $\{m_t\}$ and $\{s_t\}$ are sequences of martingale differences. We can derive that $\langle m, s \rangle_t = 0$, $\langle m \rangle_t = \sigma_e^2(\theta) + f^2(t, \mathcal{F}_{t-1}^y)(\sigma_a^2 + v_{t-1}(\theta))$, and $\langle s \rangle_t =$

$2\sigma_e^4(\theta) + 4f^2(t, \mathcal{F}_{t-1}^y)\sigma_e^2(\theta)(\sigma_a^2 + v_{t-1}(\theta)) + 2f^4(t, \mathcal{F}_{t-1}^y)(\sigma_a^2 + v_{t-1}(\theta))^2$. The optimal estimating function and associated information based on m_t are given by

$$g_M^*(\theta) = -\sum_{t=1}^{n} f\left(t, \mathcal{F}_{t-1}^y\right)\left(1 + \frac{\partial u_{t-1}(\theta)}{\partial\theta}\right)\frac{m_t}{\langle m\rangle_t},$$
$$I_{g_M^*}(\theta) = \sum_{t=1}^{n} \frac{f^2\left(t, \mathcal{F}_{t-1}^y\right)(1 + \partial u_{t-1}(\theta)/\partial\theta)^2}{\langle m\rangle_t}. \tag{3.19}$$

Then, the inequality

$$\left(\sum_{t=1}^{n} f^2\left(t, \mathcal{F}_{t-1}^y\right)\left(1 + \frac{\partial u_{t-1}(\theta)}{\partial\theta}\right)^2 \langle m\rangle_t\right)\left(\sum_{t=1}^{n} \frac{f^2\left(t, \mathcal{F}_{t-1}^y\right)(1 + \partial u_{t-1}(\theta)/\partial\theta)^2}{\langle m\rangle_t}\right)$$
$$\geq \left(\sum_{t=1}^{n} f^2\left(t, \mathcal{F}_{t-1}^y\right)\left(1 + \frac{\partial u_{t-1}(\theta)}{\partial\theta}\right)^2\right)^2 \tag{3.20}$$

implies that

$$I_{\mathrm{CLS}}(\theta) = \frac{\left(\sum_{t=1}^{n} f^2\left(t, \mathcal{F}_{t-1}^y\right)(1 + \partial u_{t-1}(\theta)/\partial\theta)^2\right)^2}{\sum_{t=1}^{n} f^2\left(t, \mathcal{F}_{t-1}^y\right)(1 + \partial u_{t-1}(\theta)/\partial\theta)^2 \langle m\rangle_t} \leq I_{g_M^*}(\theta), \tag{3.21}$$

that is, the optimal linear estimating function $g_M^*(\theta)$ is more informative than the conditional least squares estimating function $g_{\mathrm{CLS}}(\theta)$.

The optimal estimating function and the associated information based on s_t are given by

$$g_S^*(\theta) = -\sum_{t=1}^{n}\left(\frac{\partial\sigma_e^2(\theta)}{\partial\theta} + f^2\left(t, \mathcal{F}_{t-1}^y\right)\frac{\partial v_{t-1}(\theta)}{\partial\theta}\right)\frac{s_t}{\langle s\rangle_t},$$
$$I_{g_S^*}(\theta) = \sum_{t=1}^{n}\left(\frac{\partial\sigma_e^2(\theta)}{\partial\theta} + f^2\left(t, \mathcal{F}_{t-1}^y\right)\frac{\partial v_{t-1}(\theta)}{\partial\theta}\right)^2\frac{1}{\langle s\rangle_t}. \tag{3.22}$$

Hence, by Theorem 2.1, the optimal quadratic estimating function is given by

$$g_Q^*(\theta) = -\sum_{t=1}^{n}\frac{1}{\sigma_e^2(\theta) + f^2\left(t, \mathcal{F}_{t-1}^y\right)(\sigma_a^2 + v_{t-1}(\theta))}$$
$$\times\left(\left(f\left(t, \mathcal{F}_{t-1}^y\right)\left(1 + \frac{\partial u_{t-1}(\theta)}{\partial\theta}\right)\right)m_t + \frac{\partial\sigma_e^2(\theta)/\partial\theta + f^2\left(t, \mathcal{F}_{t-1}^y\right)(\partial v_{t-1}(\theta)/\partial\theta)}{\sigma_e^2(\theta) + f^2\left(t, \mathcal{F}_{t-1}^y\right)(\sigma_a^2 + v_{t-1}(\theta))}s_t\right). \tag{3.23}$$

And the associated information, $I_{g_Q^*}(\theta) = I_{g_M^*}(\theta) + I_{g_S^*}(\theta)$, is given by

$$I_{g_Q^*}(\theta) = \sum_{t=1}^{n} \frac{1}{\sigma_e^2(\theta) + f^2\left(t, \mathcal{F}_{t-1}^y\right)\left(\sigma_a^2 + v_{t-1}(\theta)\right)} \times \left(f^2\left(t, \mathcal{F}_{t-1}^y\right)\left(1 + \frac{\partial u_{t-1}(\theta)}{\partial \theta}\right)^2 + \frac{\left(\partial \sigma_e^2(\theta)/\partial\theta + f^2\left(t, \mathcal{F}_{t-1}^y\right)\left(\partial v_{t-1}(\theta)/\partial\theta\right)\right)^2}{\sigma_e^2(\theta) + f^2\left(t, \mathcal{F}_{t-1}^y\right)\left(\sigma_a^2 + v_{t-1}(\theta)\right)} \right). \tag{3.24}$$

It is obvious to see that the information of g_Q^* is larger than that of g_M^* and g_S^*, and hence, the estimate obtained by solving the optimal quadratic estimating equation is more efficient than the CLS estimate and the estimate obtained by solving the optimal linear estimating equation. Moreover, the relations

$$\begin{aligned}
\frac{\partial u_t(\theta)}{\partial\theta} &= -\frac{f^2\left(t, \mathcal{F}_{t-1}^y\right)\sigma_a^2\left(1 + \partial u_{t-1}(\theta)/\partial\theta\right)\left(\sigma_e^2(\theta) + f^2\left(t, \mathcal{F}_{t-1}^y\right)\sigma_a^2 + v_{t-1}(\theta)\right)}{\left(\sigma_e^2(\theta) + f^2\left(t, \mathcal{F}_{t-1}^y\right)\left(\sigma_a^2 + v_{t-1}(\theta)\right)\right)^2} \\
&\quad - \frac{\sigma_a^2\left(y_t - f\left(t, \mathcal{F}_{t-1}^y\right)\left(\theta + u_{t-1}(\theta)\right)\right)\left(\partial\sigma_e^2(\theta)/\partial\theta + f^2\left(t, \mathcal{F}_{t-1}^y\right)\left(\partial v_{t-1}(\theta)/\partial\theta\right)\right)}{\left(\sigma_e^2(\theta) + f^2\left(t, \mathcal{F}_{t-1}^y\right)\left(\sigma_a^2 + v_{t-1}(\theta)\right)\right)^2}, \\
\frac{\partial v_t(\theta)}{\partial\theta} &= \frac{\sigma_a^4 f^2\left(t, \mathcal{F}_{t-1}^y\right)\left(\partial\sigma_e^2(\theta)/\partial\theta + f^2\left(t, \mathcal{F}_{t-1}^y\right)\partial v_{t-1}(\theta)/\partial\theta\right)}{\left(\sigma_e^2(\theta) + f^2\left(t, \mathcal{F}_{t-1}^y\right)\left(\sigma_a^2 + v_{t-1}(\theta)\right)\right)^2}
\end{aligned} \tag{3.25}$$

can be applied to calculate the estimating functions and associated information recursively.

3.4. Regression Model with ARCH Errors

Consider a regression model with ARCH (s) errors ε_t of the form

$$y_t = \mathbf{x_t}\boldsymbol{\beta} + \varepsilon_t, \tag{3.26}$$

such that $\mathrm{E}[\varepsilon_t \mid \mathcal{F}_{t-1}^y] = 0$, and $\mathrm{Var}(\varepsilon_t \mid \mathcal{F}_{t-1}^y) = h_t = \alpha_0 + \alpha_1\varepsilon_{t-1}^2 + \cdots + \alpha_s\varepsilon_{t-s}^2$. In this model, the conditional mean is $\mu_t = \mathbf{x_t}\boldsymbol{\beta}$, the conditional variance is $\sigma_t^2 = h_t$, and the conditional skewness and excess kurtosis are assumed to be constants γ and κ, respectively. It follows form Theorem 2.1 that the optimal component quadratic estimating function for the parameter vector $\boldsymbol{\theta} = (\beta_1, \ldots, \beta_r, \alpha_0, \ldots, \alpha_s)' = (\boldsymbol{\beta}', \boldsymbol{\alpha}')'$ is

$$\mathbf{g}_Q^*(\boldsymbol{\beta}) = \frac{1}{(\kappa+2)}\left(1 - \frac{\gamma^2}{\kappa+2}\right)^{-1} \times \sum_{t=1}^{n} \frac{1}{h_t^2}\left(\left(-h_t(\kappa+2)\mathbf{x_t}' + 2h_t^{1/2}\gamma \sum_{j=1}^{s} \alpha_j \mathbf{x_t}' \varepsilon_{t-j}\right) m_t + \left(h_t^{1/2}\gamma \mathbf{x_t}' - 2\sum_{j=1}^{s} \alpha_j \mathbf{x_t}' \varepsilon_{t-j}\right) s_t\right),$$

$$\mathbf{g}_Q^*(\boldsymbol{\alpha}) = \frac{1}{(\kappa+2)}\left(1 - \frac{\gamma^2}{\kappa+2}\right)^{-1} \times \sum_{t=1}^{n} \frac{1}{h_t^2}\left(h_t^{1/2}\gamma\left(1, \varepsilon_{t-1}^2, \ldots, \varepsilon_{t-p}^2\right)' m_t - \sum_{t=1}^{n}\left(1, \varepsilon_{t-1}^2, \ldots, \varepsilon_{t-p}^2\right)' s_t\right). \tag{3.27}$$

Moreover, the information matrix for $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\alpha}')'$ is given by

$$\mathbf{I} = \left(1 - \frac{\gamma^2}{\kappa+2}\right)^{-1} \begin{pmatrix} \mathbf{I}_{\boldsymbol{\beta}\boldsymbol{\beta}} & \mathbf{I}_{\boldsymbol{\beta}\boldsymbol{\alpha}} \\ \mathbf{I}_{\boldsymbol{\alpha}\boldsymbol{\beta}} & \mathbf{I}_{\boldsymbol{\alpha}\boldsymbol{\alpha}} \end{pmatrix}, \tag{3.28}$$

where

$$\mathbf{I}_{\boldsymbol{\beta}\boldsymbol{\beta}} = \sum_{t=1}^{n}\left(\frac{\mathbf{x_t}'\mathbf{x_t}}{h_t} + \frac{4\left(1, \varepsilon_{t-1}^2, \ldots, \varepsilon_{t-s}^2\right)'\left(1, \varepsilon_{t-1}^2, \ldots, \varepsilon_{t-s}^2\right)}{h_t^2(\kappa+2)}\right),$$

$$\mathbf{I}_{\boldsymbol{\beta}\boldsymbol{\alpha}} = -\sum_{t=1}^{n} \frac{\left(h_t^{1/2}\gamma_t \mathbf{x_t}' - 2\sum_{j=1}^{s} \alpha_j \mathbf{x_t}' \varepsilon_{t-j}\right)\left(1, \varepsilon_{t-1}^2, \ldots, \varepsilon_{t-s}^2\right)}{h_t^2(\kappa+2)},$$

$$\mathbf{I}_{\boldsymbol{\alpha}\boldsymbol{\beta}} = I'_{\boldsymbol{\beta}\boldsymbol{\alpha}} = -\sum_{t=1}^{n} \frac{\left(1, \varepsilon_{t-1}^2, \ldots, \varepsilon_{t-s}^2\right)'\left(h_t^{1/2}\gamma \mathbf{x_t} - 2\sum_{j=1}^{s} \alpha_j \mathbf{x_t} \varepsilon_{t-j}\right)}{h_t^2(\kappa+2)}, \tag{3.29}$$

$$\mathbf{I}_{\boldsymbol{\alpha}\boldsymbol{\alpha}} = \sum_{t=1}^{n} \frac{\left(1, \varepsilon_{t-1}^2, \ldots, \varepsilon_{t-s}^2\right)'\left(1, \varepsilon_{t-1}^2, \ldots, \varepsilon_{t-s}^2\right)}{h_t^2(\kappa+2)}.$$

It is of interest to note that when $\{\varepsilon_t\}$ are conditionally Gaussian such that $\gamma = 0$, $\kappa = 0$,

$$E\left[\frac{\left(\sum_{j=1}^{s} \alpha_j \mathbf{x_t}' \varepsilon_{t-j}\right)\left(1, \varepsilon_{t-1}^2, \ldots, \varepsilon_{t-s}^2\right)}{h_t^2(\kappa+2)}\right] = \mathbf{0}, \tag{3.30}$$

the optimal quadratic estimating functions for $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ based on the estimating functions $m_t = y_t - \mathbf{x_t}\beta$ and $s_t = m_t^2 - h_t$, are, respectively, given by

$$\mathbf{g}_Q^*(\boldsymbol{\beta}) = -\sum_{t=1}^{n} \frac{1}{h_t^2}\left(h_t \mathbf{x_t}' m_t + \sum_{t=1}^{n}\left(\sum_{j=1}^{s} \alpha_j \mathbf{x_t}' \varepsilon_{t-j}\right) s_t \right),$$
$$\mathbf{g}_Q^*(\boldsymbol{\alpha}) = -\sum_{t=1}^{n} \frac{1}{h_t^2}\left(1, \varepsilon_{t-1}^2, \ldots, \varepsilon_{t-s}^2\right)' s_t. \tag{3.31}$$

Moreover, the information matrix for $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\alpha}')'$ in (3.28) has $\mathbf{I}_{\boldsymbol{\beta\alpha}} = \mathbf{I}_{\boldsymbol{\alpha\beta}} = \mathbf{0}$,

$$\mathbf{I}_{\boldsymbol{\beta\beta}} = \sum_{t=1}^{n} \frac{h_t \mathbf{x_t}' \mathbf{x_t} + 2\left(\sum_{j=1}^{s} \alpha_j \mathbf{x_t}' \varepsilon_{t-j}\right)\left(\sum_{j=1}^{s} \alpha_j \mathbf{x_t} \varepsilon_{t-j}\right)}{h_t^2},$$
$$\mathbf{I}_{\boldsymbol{\alpha\alpha}} = \sum_{t=1}^{n} \frac{\left(1, \varepsilon_{t-1}^2, \ldots, \varepsilon_{t-s}^2\right)'\left(1, \varepsilon_{t-1}^2, \ldots, \varepsilon_{t-s}^2\right)}{2h_t^2}. \tag{3.32}$$

4. Conclusions

In this paper, we use appropriate martingale differences and derive the general form of the optimal quadratic estimating function for the multiparameter case with dependent observations. We also show that the optimal quadratic estimating function is more informative than the estimating function used in Thavaneswaran and Abraham [2]. Following Lindsay [8], we conclude that the resulting estimates are more efficient in general. Examples based on ACD models, RCA models, doubly stochastic models, and the regression model with ARCH errors are also discussed in some detail. For RCA models and doubly stochastic models, we have shown the superiority of the approach over the CLS method.

References

[1] V. P. Godambe, "The foundations of finite sample estimation in stochastic processes," *Biometrika*, vol. 72, no. 2, pp. 419–428, 1985.

[2] A. Thavaneswaran and B. Abraham, "Estimation for nonlinear time series models using estimating equations," *Journal of Time Series Analysis*, vol. 9, no. 1, pp. 99–108, 1988.

[3] U. V. Naik-Nimbalkar and M. B. Rajarshi, "Filtering and smoothing via estimating functions," *Journal of the American Statistical Association*, vol. 90, no. 429, pp. 301–306, 1995.

[4] A. Thavaneswaran and C. C. Heyde, "Prediction via estimating functions," *Journal of Statistical Planning and Inference*, vol. 77, no. 1, pp. 89–101, 1999.

[5] S. A. Chandra and M. Taniguchi, "Estimating functions for nonlinear time series models. Nonlinear non-Gaussian models and related filtering methods," *Annals of the Institute of Statistical Mathematics*, vol. 53, no. 1, pp. 125–141, 2001.

[6] T. Merkouris, "Transform martingale estimating functions," *The Annals of Statistics*, vol. 35, no. 5, pp. 1975–2000, 2007.

[7] M. Ghahramani and A. Thavaneswaran, "Combining estimating functions for volatility," *Journal of Statistical Planning and Inference*, vol. 139, no. 4, pp. 1449–1461, 2009.

[8] B. G. Lindsay, "Using empirical partially Bayes inference for increased efficiency," *The Annals of Statistics*, vol. 13, no. 3, pp. 914–931, 1985.

[9] M. Crowder, “On linear and quadratic estimating functions,” *Biometrika*, vol. 74, no. 3, pp. 591–597, 1987.

[10] R. F. Engle and J. R. Russell, “Autoregressive conditional duration: a new model for irregularly spaced transaction data,” *Econometrica*, vol. 66, no. 5, pp. 1127–1162, 1998.

[11] D. F. Nicholls and B. G. Quinn, “The estimation of random coefficient autoregressive models. I,” *Journal of Time Series Analysis*, vol. 1, no. 1, pp. 37–46, 1980.

[12] D. Tjøstheim, “Some doubly stochastic time series models,” *Journal of Time Series Analysis*, vol. 7, no. 1, pp. 51–72, 1986.

[13] A. N. Shiryayev, *Probability*, vol. 95 of *Graduate Texts in Mathematics*, Springer, New York, NY, USA, 1984.


Advances in Operations Research
Advances in Decision Sciences
Journal of Applied Mathematics
Algebra
Journal of Probability and Statistics
The Scientific World Journal
International Journal of Differential Equations
Hindawi
Submit your manuscripts at
http://www.hindawi.com
International Journal of Combinatorics
Advances in Mathematical Physics
Journal of Complex Analysis
Journal of Mathematics
Mathematical Problems in Engineering
Abstract and Applied Analysis
Discrete Dynamics in Nature and Society
International Journal of Mathematics and Mathematical Sciences
Journal of Discrete Mathematics
Journal of Function Spaces
International Journal of Stochastic Analysis
Journal of Optimization