QBS
Vol. 36, No. 2, pp. 105-110 (2017)
https://doi.org/10.22283/qbs.2017.36.2.105
pISSN 2288-1344 eISSN 2508-7185 **Review Article**

Quantitative Bio-Science
Official Journal of the Institute of Natural Science at Keimyung University
Available Online at https://qbs.kmu.ac.kr:442/

Statistical Issues on Instrumental Scores for Non-likelihood Stochastic Models

Sun Young Hwang[1,*], Tae Yoon Kim[2]

[1]Department of Statistics, Sookmyung Women's University, Seoul 14310, Korea
[2]Department of Statistics, Keimyung University, Daegu 42601, Korea
(Received October 23, 2017; Revised November 7, 2017; Accepted November 12, 2017)

ABSTRACT

For the data exhibiting a dependency structure, the exact likelihood is rarely available to researchers mainly because of unobserved initial values and unknown innovation distributions. It is the case in practice to assume a tractable score for the data for the sake of easy analysis. The adopted tractable score is referred to as the instrumental score in order to discriminate from the true Fisher's score. In this review paper, various existing inferential methodologies in stochastic models (e.g., conditional least squares, pseudo likelihood, quasi-likelihood, quasi-maximum likelihood, Godambe's linear scores) are reviewed under a unified framework of the instrumental scores. Applications to bifurcating auto-regressions in the context of cell lineage studies are discussed.

Key words : Instrumental score, Quasi-likelihood, Unknown likelihood

1. Introduction and Preliminaries

We consider the stationary stochastic models $\{X_t, t\geq 1\}$ adapted to the increasing sequence of sigma fields $\{F_t, t\geq 1\}$. Introduce the conditional mean μ_t and the prediction error ϵ_t defined by

$$\mu_t = E(X_t|F_{t-1}) \text{ and } X_t - \mu_t = \epsilon_t. \tag{1}$$

To accommodate conditionally heteroscedastic models, let h_t denote the conditional variance and the standardized version $\{e_t, t\geq 1\}$ of $\{X_t, t\geq 1\}$ is referred to as the "innovation" process associated with the stochastic model $\{X_t\}$. Specifically,

$$e_t = \frac{X_t-\mu_t}{\sqrt{h_t}} = \frac{\epsilon_t}{\sqrt{h_t}} \text{ where } h_t = Var(X_t|F_{t-1}) \tag{2}$$

where F_0 denotes the trivial sigma field. First note that μ_t and h_t are F_{t-1} measurable and thus the innovation $\{e_t, t\geq 1\}$ constitutes a zero mean and unit variance sequence of (uncorrelated) martingale differences. It is customary to model $\{e_t, t\geq 1\}$ as iid (independent and identically distributed) with mean zero and variance one. Now, so called the parameter (vector) of interest comes into the stochastic model. Let θ denote $(k\times 1)$ vector parameter taking values in Θ which is an open subset of the k-dimensional Euclidean space.

Typically, θ indexes μ_t and h_t in such a way that $\mu_t(\theta)= E(X_t|F_{t-1})$ and $h_t(\theta)=Var(X_t|F_{t-1})$ and therefore θ is called a parameter (vector) of interest. It will be assumed that the functional form of $\mu_t(\theta)$ and $h_t(\theta)$ is specified. For instance, the formulation

$$\mu_t(\theta)=\phi X_{t-1} \text{ and } h_t(\theta)=\alpha_0+\alpha_1\epsilon_{t-1}^2+\beta_1 h_{t-1}(\theta)$$

is named as AR(1)-GARCH(1,1) model where the parameter of interest is given by $\theta=(\phi,\alpha_0,\alpha_1,\beta_1)^T$ with $k=4$. To be more specific, the AR(1)-GARCH(1,1) leads to

$$X_t-\phi X_{t-1}=\epsilon_t$$

* Correspondence should be addressed to Dr. Sun Young Hwang, Department of Statistics, Sookmyung Women's University, Seoul 14310, Korea. Tel: +82-2-710-9622, Fax: +82-2-710-9283, E-mail: shwang@sookmyung.ac.kr

$$\epsilon_t = \sqrt{h_t} e_t$$
$$h_t = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 h_{t-1}. \tag{3}$$

Various stochastic models in the literature can be employed by making appropriate choices of $\mu_t(\theta)$ and $h_t(\theta)$. See for instance Hwang et al. [1].

We move to discussions on the likelihood of the data $\{X_1, X_2, \ldots, X_n\}$. When the likelihood is correctly specified, the maximum likelihood (ML) method is readily available to estimate parameters. Recall that first order derivative of the true log-likelihood is referred to as Fisher's score and variance of the Fisher's score is named as Fisher information. For the data exhibiting a dependency structure, however, exact likelihood is rarely available to researchers mainly due to unobserved initial values and unknown innovation distributions. From now on, we are interested in so called non-likelihood case in the sense that exact likelihood is not specified although it does exist. The initial value problem happens commonly in stochastic models. The likelihood of a stochastic model is usually conditionally on initial values. For the ARMA(1,1) given by

$$X_t = \phi X_{t-1} + \epsilon_t - \delta \epsilon_{t-1}$$

initial value is (X_0, ϵ_0) and therefore likelihood of the data $X_1, \ldots, X_n$ is obtained conditionally on the unobservable random quantity (X_0, ϵ_0). In other words, the likelihood for ARMA(1,1) is not an exact one but a conditional one. The non-likelihood case arises because the innovation distribution (and likelihood) may not be known and stochastic models are defined only through a first few conditional moments without requiring specification of the likelihood. Further, the true innovation distribution (and likelihood) may be too complicated to use for practical purposes. To deal with non-likelihood cases, it is usual to assume a tractable score for the data for the sake of easy analysis. The adopted tractable score is referred to as the instrumental score in order to discriminate from the true Fisher's score. The term "instrumental" is used to indicate that it acts as an alternative to the true (but unknown) Fisher's score. Based on the authors' recent publications, in this paper, various existing inferential methodologies in stochastic models (e.g., conditional least squares, pseudo likelihood, quasi-likelihood, quasi-maximum likelihood, Godambe's linear scores) are reviewed under a unified framework of the instrumental scores. Applications to bifurcating auto-regressions in the context of cell lineage studies are discussed. Most of the results in the paper are taken and adapted from the authors' recent publications listed in the reference.

2. Instrumental Scores for Non-likelihood Stochastic Models

We begin with the Fisher's score $S_n(\theta)$. Let $p_t(\theta)$ denote the (true) conditional density of X_t given F_{t-1}. The maximum likelihood (ML) method is based on the Fisher's score $S_n(\theta)$ defined by

$$S_n(\theta) = \sum_{t=1}^{n} \frac{\partial \ln p_t(\theta)}{\partial \theta} : (k \times 1) \text{ vector} \tag{4}$$

where $p_1(\theta)$ is unconditional density of X_1. The Fisher information matrix $\Gamma_n(\theta)$ is defined by the covariance matrix of $S_n(\theta)$, that is,

$$\Gamma_n(\theta) = Var(S_n(\theta)) = E[S_n(\theta) S_n^T(\theta)] \tag{5}$$

where "T" denotes "transpose". Under some regularity conditions (cf. Hwang and Basawa [2]), $S_n(\theta)$ is asymptotically normal with mean zero vector and covariance matrix $\Gamma_n(\theta)$. Extending $S_n(\theta)$ toward non-likelihood cases, consider the following $(k \times 1)$ instrumental score $U_n(\theta)$ defined by

$$U_n(\theta) = \sum_{t=1}^{n} u_t(\theta) : (k \times 1) \text{ vector} \tag{6}$$

where $\{u_t(\theta)\}$ is a sequence of martingale differences i.e., $E(u_t(\theta)|F_{t-1}) = 0$. Here, F_t denote the σ-field generated by $X_t, X_{t-1}, \ldots, X_1$. We collect all the instrumental scores into $U = \{U_n(\theta)\}$. It is obvious that $S_n(\theta)$ is a member of U. Some (important) special members of U follow.

Conditional Least Squares (CLS): With the conditional mean $\mu_t(\theta)$, consider the instrumental score $U_n(\theta) = \sum_{t=1}^{n} u_t(\theta)$ with

$$u_t(\theta) = \left(\frac{\partial \mu_t(\theta)}{\partial \theta}\right) \cdot (X_t - \mu_t(\theta)) : (k \times 1) \tag{7}$$

which is referred to as a CLS-score.

Generalized Least Squares (GLS): Let $g_t(\theta)$ denote a discrepancy between X_t and $\mu_t(\theta)$ such as $g_t(\theta) = X_t - \mu_t(\theta)$. Generalized least squares (GLS) based on $g_t(\theta)$ is obtainedby minimizing

$$\sum_{t=1}^{n} g_t^T(\theta) [Var(g_t(\theta)|F_{t-1})]^{-1} g_t(\theta).$$

It is shown that the GLS score is given by (refer to Hwang [3])

$$u_t(\theta)=\sum_{t=1}^{n}\frac{\partial g_t^T(\theta)}{\partial\theta}[Var(g_t(\theta)|F_{t-1})]^{-1}g_t(\theta) : (k\times 1) \quad (8)$$

Maximum Likelihood (ML): Maximum Likelihood score refers to the choice

$$u_t(\theta)=\frac{\partial\, log\;\, p_t(\theta)}{\partial\theta}$$

where $p_t(\theta)$ is the conditional density of X_t given F_{t-1}.

Quasi-Likelihood (QL): Without requiring the knowledge of the likelihood, QL method assumes only first two conditional moments: $\mu_t(\theta)=E(X_t|F_{t-1})$ and $h_t(\theta)=Var(X_t|F_{t-1})$. A quasi-likelihood (QL) score based on the martingale difference $X_t-\mu_t(\theta)$ is given by $U_n(\theta)=\sum_{t=1}^{n}u_t(\theta)$ where

$$u_t(\theta)=\frac{\partial\mu_t(\theta)}{\partial\theta}\cdot\frac{(X_t-\mu_t(\theta))}{h_t(\theta)} \;:\; (k\times 1) \quad (9)$$

The QL score based on the martingale difference vector

$$g_t(\theta)=\begin{pmatrix} X_t-\mu_t(\theta) \\ (X_t-\mu_t(\theta))^2-h_t(\theta)\end{pmatrix} \quad (10)$$

is given by

$$U_n(\theta)=\sum_{t=1}^{n}\left(\frac{\partial\mu_t(\theta)}{\partial\theta},\frac{\partial h_t(\theta)}{\partial\theta}\right)V_t^{-1}(\theta)g_t(\theta) \quad (11)$$

where $V_t(\theta)$: the (2×2) conditional covariance matrix of $g_t(\theta)$, viz.,

$$V_t(\theta)=E[g_t(\theta)g_t^T(\theta)|F_{t-1}]=\begin{pmatrix} h_t(\theta) & \mu_{3t}(\theta) \\ \mu_{3t}(\theta) & \mu_{4t}(\theta)-h_t^2(\theta)\end{pmatrix} \quad (12)$$

with $\mu_{3t}(\theta)=E[(X_t-\mu_t(\theta))^3|F_{t-1}]$ and $\mu_{4t}(\theta)=E[(X_t-\mu_t(\theta))^4|F_{t-1}]$. See Godambe [4], Hwang and Basawa [3] and Hwang et al. [5].

Pseudo Likelihood (PL): A tractable likelihood is called a pseudo-likelihood (PL) which may be a falsely specified likelihood. The PL estimator is the zero of the pseudo-likelihood score which is given by $U_n(\theta)=\sum_{t=1}^{n}\frac{\partial\, ln\; f_t(\theta)}{\partial\theta}$. Here $f_t(\theta)$ is a pseudo-conditional density (of X_t given F_{t-1}) which may differ from the true (but unknown) conditional density $p_t(\theta)$ appearing in the ML. Some pseudo-conditional densities are useful in describing fatter tails than normal density (see, for instance, Tsay [6, Ch.3]). The standardized t-distribution with r degrees of freedom $(r>2)$ is given by

$$f(e_t)=\frac{\Gamma[(r+1)/2]}{\Gamma(r/2)\sqrt{(r-2)\pi}}\left(1+\frac{e_t^2}{(r-2)}\right)^{-(r+1)/2}.$$

The generalized error distribution (GED) with parameter $v>0$ is of the form

$$f(e_t)=\frac{v\, exp\left(-\frac{1}{2}|x/\lambda|^v\right)}{\lambda 2^{(1+1/v)}\Gamma(1/v)}$$

with $\lambda^2=2^{(-2/v)}\Gamma(1/v)\Gamma(3/v)$.

Quasi Maximum Likelihood (QML): Quasi-maximum likelihood (QML) score is obtained by assuming that the likelihood is Gaussian for simplicity. QML estimator is the zero of the QML score. QML-estimator is consistent and asymptotically normal under some regularity conditions (cf., Hwang et al. [1]).

Godambe's Optimum Score (GOS): Godambe [4] considered the following "linear" scores $G_n(\theta)$ defined by

$$G_n(\theta)=\sum_{t=1}^{n}w_{t-1}(\theta)a_t(\theta) \quad (13)$$

where $a_t(\theta)$ is a fixed martingale difference vector of dimension d and $w_{t-1}(\theta)$ is $(k\times d)$ weight matrix whose components are F_{t-1} measurable. The Godambe's class of "linear" instrumental scores is generated by varying the "coefficients" $w_{t-1}(\theta)$ while $a_t(\theta)$ being fixed in (13). We shall refer to the resulting Godambe's class as $L=\{G_n(\theta)\}$ which is clearly a subset of $U=\{U_n(\theta)\}$. The Godambe's optimum score (GOS) which will be denoted by $G_n^O(\theta)$ within Godambe's class L is formulated by (see Hwang [3])

$$\text{GOS}: G_n^O(\theta)=\sum_{t=1}^{n}w_{t-1}^O(\theta)a_t(\theta) \quad (14)$$

where F_{t-1} measurable $(k\times d)$ matrix $w_{t-1}^O(\theta)$ is given by

$$w_{t-1}^O(\theta)=E_{t-1}[\partial a_t^T(\theta)/\partial\theta]Var_{t-1}^{-1}(a_t(\theta)). \quad (15)$$

where and in what follows E_{t-1} and Var_{t-1} denote conditional expectation and conditional variance given F_{t-1}. The GOS in (14) can be viewed as a generalization of the QL score defined in (11). Specifically, if one chooses $g_t(\theta)$ in (10) for $a_t(\theta)$ in (13), then the GOS $G_n^O(\theta)$ given in (14) reduces to the QL score in (11). Consequently, GOS extends the scope of the QL methodology. Refer to, for instance, Hwang and Basawa [7] and Hwang [3].

3. Efficiency Issue of Instrumental Scores and Quasi-information

Recall the class $U=\{U_n(\theta)\}$ of all the instrumental scores in which the Fisher's score $S_n(\theta)$ is a special member. The Fisher information matrix $\Gamma_n(\theta)$ is defined as the covariance matrix of $S_n(\theta)$. See (4) and (5). The MLE of $\hat{\theta}_{ML}$ is obtained by solving $S_n(\theta)=0$. Via ergodic theorem, the limit matrix of $n^{-1}\Gamma_n(\theta)$ exists and is denoted by

$$\Gamma(\theta)=plim[n^{-1}\Gamma_n(\theta)]. \tag{16}$$

Under some regularity conditions (see Theorem 3.2 of Hwang and Basawa [2] for details), it can be verified that

(i) $\sqrt{n}(\hat{\theta}_{ML}-\theta)\xrightarrow{d}N(0,\Gamma^{-1})$ (17)

and

(ii) Consider any instrumental score in $U=\{U_n(\theta)\}$ and let $\hat{\theta}_n$ be obtained from solving $U_n(\theta)=0$. Then, there exits a constant matrix $A(\theta)$ such that

$$\sqrt{n}(\hat{\theta}_n-\theta)\xrightarrow{d}N(0,A^{-1}) \tag{18}$$

where $A^{-1}-\Gamma^{-1}$ is a non-negative definite matrix, and $\xrightarrow{d}$ denotes "convergence in distribution".

Consequently, one can conclude

Proposition 1: The ML estimator $\hat{\theta}_{ML}$ is asymptotically efficient in the sense of having the "smallest" covariance matrix among all the estimators $\hat{\theta}_n$ within $U=\{U_n(\theta)\}$.

Although it is the best among $U=\{U_n(\theta)\}$, $\hat{\theta}_{ML}$ can not be implemented because of the non-likelihood case. To argue the second best score, we need to restrict $U=\{U_n(\theta)\}$ to the Godambe's class $L=\{G_n(\theta)\}$ of "linear" instrumental scores. See (13). It is obvious that $L\subset U$. The Godambe's optimum score (GOS) $G_n^O(\theta)$ is given by (14) and (15). Analogously to the Fisher information, we define the quasi-information as the covariance matrix of $G_n^O(\theta)$, that is

$$\begin{aligned}Q_{G_n^O}(\theta)&=Var(G_n^O(\theta))\\&=E[E_{t-1}[\partial a_t^T(\theta)/\partial\theta]Var_{t-1}^{-1}(a_t(\theta))E_{t-1}[\partial a_t(\theta)/\partial\theta^T]].\end{aligned} \tag{19}$$

Define the limiting average quasi-information as

$$Q(\theta)=plim[n^{-1}Q_{G_n^O}(\theta)]. \tag{20}$$

Let $\hat{\theta}_n^O$ be the solution of $G_n^O(\theta)=0$. Under some regularity conditions (see Theorem 3.3 of Hwang and Basawa [2] and Lemma 1 of Hwang [3] for details), we are able to state

(i) $\sqrt{n}(\hat{\theta}_n^O-\theta)\xrightarrow{d}N(0,Q(\theta)^{-1})$ (21)

and

(ii) Consider any instrumental score in $L=\{G_n(\theta)\}$ and let $\hat{\theta}_n$ be obtained from solving $G_n(\theta)=0$. Then, for some constant matrix $B(\theta)$,

$$\sqrt{n}(\hat{\theta}_n-\theta)\xrightarrow{d}N(0,B^{-1}) \tag{22}$$

where $B^{-1}-Q(\theta)^{-1}$ is a non-negative definite matrix.

Consequently, analogous to Proposition 1, within the restricted class $L=\{G_n(\theta)\}$,

Proposition 2: The GOS estimator $\hat{\theta}_n^O$ is asymptotically efficient in the sense of having the "smallest" covariance matrix among all the estimators $\hat{\theta}_n$ within $L=\{G_n(\theta)\}$.

Combining Propositions 1 and 2, GOS score $G_n^O(\theta)$ acts as the ML score $S_n(\theta)$ within the restricted class $L\subset U$ (see for instance Hwang and Kim [8]). The information of $G_n^O(\theta)$ is referred to as "quasi-information" (i.e., feasible second best information) which is asymptotically smaller than the Fisher information.

4. Nuisance Parameter Issue on Instrumental Scores

As noted in Hwang and Kim [8] and Hwang [3], although GOS $G_n^O(\theta)$ defined in (14) and (15) enjoys a certain asymptotic efficiency as presented in Proposition 2, it has a drawback in terms of a practical calculation. The weight in (15)

$$w_{t-1}^O(\theta)=E_{t-1}[\partial a_t^T(\theta)/\partial\theta]Var_{t-1}^{-1}(a_t(\theta)) \tag{23}$$

may depend on (unknown) nuisance parameters η arising from the true (but unknown) distribution and therefore $G_n^O(\theta)=0$ may produce an estimate $\hat{\theta}_n^O$ involving unknown nuisance parameter η. However, this problem can be resolved (often in time series) via replacing the nuisance parameter η by QML estimate $\hat{\eta}$ (See Section 2, QML). For instance, revisit the QL score in (11) and (12). From the relationship (2) between the innovation process $\{e_t\}$ and observation process $\{X_t\}$, we have $X_t=\mu_t+\sqrt{h_t}e_t$. Thus, the QL score $U_n(\theta)$ involves the nuisance parameter $\eta=(Ee_t^3,Ee_t^4)^T$. Refer to Hwang [3] for various examples of GOS $G_n^O(\theta)$ involving nuisance parameter η. One can replace η by QML residuals without affecting asymptotic efficiency of $\hat{\theta}_n^O$.

Proposition 3: Under some regularity conditions in Theorem 2 of Hwang [3], one can solve $G_n^O(\theta)=0$ to get $\hat{\theta}_n^O$ via feasible steps as follows. (i) Step 1: Identify nuisance η parameter appearing in the GOS $G_n^O(\theta)$. (ii) Step 2: Fit the stochastic model via QML obtained from Gaussian likelihood and calculate QML residual $\{\hat{e}_t\}$. (iii) Step 3: Estimate η by $\hat{\eta}_{QML}$ based on the QML residual $\{\hat{e}_t\}$. (iv) Step 4: Replace η by $\hat{\eta}_{QML}$ in the GOS $G_n^O(\theta)$ and solve $G_n^O(\theta)=0$ to get $\hat{\theta}_n^O$.

5. Non-stationarity Issue on Instrumental Scores

Asymptotic efficiency issues stated in Propositions 1 and 2 are based on the assumption that the stochastic model is stationary. Extensions to non-stationary models are made in the literature. Refer to, e.g., Hwang et al. [5], Hwang [9] and Hwang and Kim [8] where various non-stationary processes including non-stationary ARCH, explosive AR(1), non-stationary random coefficient model and Branching Markov process are discussed in the context of instrumental scores. The main idea to deal with a non-stationary model is to use random norm in stead of constant norm ($\sqrt{n}$) to get limit distributions.

Consider the class L and the GOS $G_n^O(\theta)=\sum_{t=1}^n w_{t-1}^O(\theta)$ $a_t(\theta)\in L$ for which the sum of conditional covariance matrices $V_n^O(\theta)$ is defined by

$$V_n^O(\theta)=\sum_{t=1}^n w_{t-1}^O(\theta)w_{t-1}^{O\ T}(\theta)E_{t-1}a_t^2(\theta) : (k\times k). \quad (24)$$

Let non-random and non-singular $(k\times k)$ matrix $C(\theta)$ be such that

$$C(\theta)=plim[V_n^{-1/2}(\theta)[\partial G_n(\theta)/\partial\theta^T]V_n^{-1/2}(\theta)]. \quad (25)$$

Let $\hat{\theta}_n^O$ be the solution of $G_n^O(\theta)=0$ for the non-stationary process. Under some regularity conditions as in Theorem 1 of Hwang and Kim [8],

$$V_n^{O1/2}(\theta)(\hat{\theta}_n^O-\theta)\xrightarrow{d}N(0,C(\theta)^{-1}(C(\theta)^{-1})^T) \quad (26)$$

where $C(\theta)$ is defined in (25). Consequently, $\hat{\theta}_n^O$ is asymptotically normal with mean zero and covariance matrix $J_n(\theta)=V_n^{O-1/2}(\theta)C(\theta)^{-1}(C(\theta)^{-1})^TV_n^{O-1/2}(\theta)$. Due to Hwang and Kim [8], one can continue to obtain

Proposition 4: Consider the non-stationary stochastic model. The GOS estimator $\hat{\theta}_n^O$ is efficient in the sense of having the asymptotically "smallest" covariance matrix within $L=\{G_n(\theta)\}$.

6. Applications to Bifurcating Auto-regressions

Cowan and Staudte [10] suggested bifurcating autoregression (BAR) to analyze data from cell lineage studies. The stochastic model BAR suits the bifurcating data such as blood pressure, cholesterol level and protein content of a cell. Let X_t represent observation on individual t. In a bifurcating model, (mother) t produces two sisters $(2t,2t+1)$. Denote $t(1)$ denote mother of the individual t. For instance with $t=5$, $5(1)=2$. Refer to Hwang and Kim [11] and Hwang [12] and references therein for various bifurcating data structures. Consider the following heteroscedastic BAR given by

$$X_t=\mu_{t(1)}+\sqrt{h_{t(1)}}\cdot e_t,\ i\geq 2 \quad (27)$$

where the innovation $\{e_t\}$ is iid with mean zero and variance unity, and its distribution is unspecified. Here, $\mu_{t(1)}=E(X_t|X_{t(1)})$ and $h_{t(1)}=Var(X_t|X_{t(1)})$. Based on the discrepancy vector $a_t(\theta)$ in GOS and hence in QLin (10)

$$a_t(\theta)=\begin{pmatrix} X_t-\mu_{t(1)}(\theta) \\ (X_t-\mu_{t(1)}(\theta))^2-h_{t(1)}(\theta)\end{pmatrix} \quad (28)$$

the GOS $G_n^O(\theta)$ is formulated by (see (11) and (12))

$$G_n^O(\theta)=\sum_{t=1}^n\left(\frac{\partial\mu_{t(1)}(\theta)}{\partial\theta},\frac{\partial h_{t(1)}(\theta)}{\partial\theta}\right)V_{t(1)}^{-1}(\theta)a_t(\theta) \quad (29)$$

with

$$V_{t(1)}(\theta)=E[a_t(\theta)a_t^T(\theta)|X_{t(1)}]. \quad (30)$$

The corresponding quasi-information $G_n^O(\theta)$ is given by

$$G_n^O(\theta)=E\left[\left(\frac{\partial\mu_{t(1)}(\theta)}{\partial\theta},\frac{\partial h_{t(1)}(\theta)}{\partial\theta}\right)V_{t(1)}^{-1}(\theta)\left(\frac{\partial\mu_{t(1)}(\theta)}{\partial\theta},\frac{\partial h_{t(1)}(\theta)}{\partial\theta}\right)^T\right] \quad (31)$$

which attains the maximum information matrix within the class of linear scores L (see Proposition 2). As an alternative instrumental score, one may adopt weighted least squares (WLS) minimizing $\sum_{t=1}^n\left(\frac{X_t-\mu_{t(1)}(\theta)}{\sqrt{h_{t(1)}(\theta)}}\right)^2$. If WLS score belongs to L, then $G_n^O(\theta)$ is better than WLS score and therefore $\hat{\theta}_n^O$ has the smaller asymptotic variance than that of $\hat{\theta}_{WLS}$.

Acknowledgements

This research was supported by Basic Science Research Program through the NRF of Korea funded by the Ministry of

Education (2015-057031). T.Y. Kim's research was supported by Basic Science Research Program through the NRF of Korea funded by the Ministry of Education (NRF 2016R1D1A1B 03934375).

References

1. Hwang SY, Choi MS, Yeo IK. Quasilikelihood and quasi-maximum likelihood for GARCH-type processes: estimating function approach. J Korean Stat Soc 2014;43:631-641.
2. Hwang SY, Basawa IV. Martingale estimating functions for stochastic processes: a review toward a unifying tool. In: Lahiri S et al., editors. Contemporary developments in statistical theory. Switzerland: Springer; 2014. p. 9-28.
3. Hwang SY. Feasible optimum Godambe scores for a semi-parametric GARCH time series. J Korean Stat Soc 2017;46: 104-112.
4. Godambe VP. The foundation of finite sample estimation in stochastic processes. Biometrika 1985;72:419-428.
5. Hwang SY, Basawa IV, Choi MS, Lee SD. Non-ergodic martingale estimating functions and related asymptotics. Statistics 2014;48:487-507.
6. Tsay RS. Analysis of financial time series. 3rd ed. New York: Wiley; 2010.
7. Hwang SY, Basawa IV. Godambe estimating functions and asymptotic optimal inference. Stat Probabil Lett 2011;81:1121-1127.
8. Hwang SY, Kim TY. Non-stationary quasilikelihood and asymptotic optimality. J Korean Stat Soc 2014;43:475-482.
9. Hwang SY. Some characterizations of non-ergodic estimating functions for stochastic processes. J Korean Stat Soc 2015;44: 661-667.
10. Cowan R, Staudte RG. The bifurcating autoregression model in cell lineage studies. Biometrics 1986;42:769-783.
11. Hwang SY, Kim TY. Threshold approaches to asymmetric models for heteroscedastic time series. QBS 2016;35:37-40.
12. Hwang SY. An overview on models for tree-indexed time series. QBS 2011;30:9-11.