Chin. Ann. Math.
26B:3(2005),335–346.

ITERATIVE QUASI-LIKELIHOOD FOR SEEMINGLY UNRELATED REGRESSION SYSTEMS****

LIN Lu* FAN Yunzheng** DU Jun*** YUAN Yuan***

Abstract

In the seemingly unrelated regression systems, the existing quasi-likelihood is always involved in the difficult problem of calculating inverse of a high order matrix specially for large systems. To avoid this problem, the new quasi-likelihood proposed in this paper is based mainly on a linearly iterative process of some unbiased estimating functions. Some finite sample properties and asymptotic behaviours for this new quasi-likelihood are investigated. These results show that the new quasi-likelihood for parameter of interest is E-sufficient, iteratively efficient and approximately efficient. Some examples are given to illustrate the theoretical results.

Keywords Seemingly unrelated regression system, Quasi-likelihood, Unbiased estimating function

2000 MR Subject Classification 62J02, 62F12

§ 1 . Introduction

The seemingly unrelated regression (SUR) system is useful in many fields such as econometrics, industry, biology and so on. The inference about SUR system has received much attention. Serivastava and Dwivedi [16] presented a brief survey of the developments in the parameter estimation of SUR system. The works on the topic proposed by Zellner [20], Revankar [13], Wang [17] and so on focus mainly on the seemingly unrelated linear regression systems and the methods seem not to be extended to the seemingly unrelated nonlinear regression systems. Lin [10] introduced a Bayes quasi-likelihood to the seemingly unrelated nonlinear regression systems, but the consistency of parameter estimation depends on the prior distribution and then the properties will be bad when the chosen prior is wrong. So the investigation for SUR system is still a challenge.

In this paper the main goal is to estimate the parameters of SUR systems through some estimating functions. Generally, an unbiased estimating function $g(\theta, Y, x)$ is defined to be a function of the data (Y, x) and parameter θ having zero mean for all θ. In other words,

Manuscript received April 26, 2004. Received October 25, 2004.

*School of Mathematics and System Sciences, Shandong University, Jinan 250100, Shandong, China. **E-mail:** linlu@sdu.edu.cn

**School of Mathematics and System Sciences, Shandong University, Jinan 250100, Shandong, China; Department of Mathematics, Nantong University, Nantong 226007, Jiangsu, China.

***School of Mathematical Sciences, Nankai University, Tianjin 300071, China.

****Project supported by the National Natural Science Foundation of China (No.10371059, No.10171051).

$E\{g(\theta, Y, x)\} = 0$ for all θ. One purpose of an estimating function is to produce an estimate $\hat{\theta}$ of the parameter from data (Y, x), the estimate being obtained as a root of the equation $g(\theta, Y, x) = 0$. Consequently, if the parameter θ is p-dimensional, it is necessary at least that the range of g is p-dimensional with nonsingular derivative matrix.

Wedderburn [18], in his attempt to establish a method of inference based exclusively on the first two moments of the sample, introduced the quasi-score function. Suppose that the $n \times 1$ random variable $Y = (Y_1, \cdots, Y_n)'$ has mean $\mu(\theta)$ and covariance matrix $\sigma^2 V(\theta)$. Both are known functions of the p-dimensional parameter θ, and $V(\theta)$ is a positive definite matrix. In this case, a special unbiased estimating function, called the quasi-score function of θ, is defined as

$$q(\theta) = \sigma^{-2} \dot{\mu}'(\theta) V^{-1}(\theta) e(\theta, Y), \tag{1.1}$$

where $\dot{\mu}$ is an $n \times p$ matrix with components $\partial \mu_i / \partial \theta_j$ and $\operatorname{rank}\{\dot{\mu}(\theta)\} = p$, $e(\theta, Y) = Y - \mu(\theta)$. The quasi-score function is a linear unbiased function based only on the first two moments of the observations. It is well known that the quasi-score function is the optimum in the class of linear unbiased estimating functions and, consequently, the quasi-likelihood estimator $\hat{\theta}$ obtained from the quasi-score function is the optimum in the class of the estimators obtained from the linear unbiased estimating functions under some criterions (see [5, 8]).

In this paper we mainly consider the following SUR system

$$\begin{cases} E(Y_i) = \mu_i(\theta_i, x_i), \ \operatorname{Var}(Y_i) = \sigma_i^2 V_i(\theta_i, x_i) & \text{for } i = 1, \cdots, m, \\ \operatorname{Cov}(Y_i, Y_j) = \sigma_{ij} V_{ij}(\theta_i, \theta_j, x_i, x_j) & \text{for } i \neq j, \ i, j = 1, \cdots, m, \end{cases} \tag{1.2}$$

where response variables $Y_i = (Y_{i1}, \cdots, Y_{in})'$ depend on non-random covariates $x_i = (x_{i1}, \cdots, x_{in})'$ through known regression functions $\mu_i(\theta_i, x_i) = (h_i(\theta_i, x_{i1}), \cdots, h_i(\theta_i, x_{in}))'$, $\theta_i = (\theta_{i1}, \cdots, \theta_{ip_i})' \in \Theta_i$ are vectors of unknown parameters, $V_i(\theta_i, x_i)$ are positive definite matrices and their components are known functions of θ_i and x_i, similarly, the components of $V_{ij}(\theta_i, \theta_j, x_i, x_j)$ are known functions of θ_i, θ_j, x_i and x_j. This model covers commonly used models such as seemingly unrelated nonlinear regression system, seemingly unrelated linear regression system, seemingly unrelated generalized linear regression system and so on. The character of Model (1.2) is that each regression equation has its own parameter and variables, and any two of response variables are related to each other. Without loss of generality, we assume, in this paper, the main goal is to introduce an estimating function for parameter θ_1 in the first regression equation and then the method for estimating the other parameters follows.

According to the definition of the quasi-score function (1.1), the quasi-score function for full parameter $\theta = (\theta_1, \cdots, \theta_m)'$ in Model (1.2) has the form of

$$q(\theta) = \begin{bmatrix} \dot{\mu}_1' & 0 & \cdots & 0 \\ 0 & \dot{\mu}_2' & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & \dot{\mu}_m' \end{bmatrix} \begin{bmatrix} \sigma_1^2 V_1 & \sigma_{12} V_{12} & \cdots & \sigma_{1m} V_{1m} \\ \sigma_{21} V_{21} & \sigma_2^2 V_2 & \cdots & \sigma_{2m} V_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{m1} V_{m1} & \sigma_{m2} V_{m2} & \cdots & \sigma_m^2 V_m \end{bmatrix}^{-1} \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_m \end{bmatrix} \tag{1.3}$$

with $e_i = Y_i - \mu_i(\theta_i, x_i)$. This score function is always involved in the difficult situation of calculating inverse of covariance matrix specially for the case with large m. Even if

$m = 2$, the simplest SUR system, the calculating problem is also difficult because the order of the covariance matrix is a large number $2n$. At the same time, the profile quasi-likelihood, designed for estimating the 1-dimensional parameter of interest, is also involved in calculating the inverse of covariance matrix (see [9]). So the existing quasi-likelihood is not applicable to the case of SUR systems and then a new design is desired.

To avoid the above problem, a new quasi-likelihood is proposed in this paper, which is based mainly on a linearly iterative process. From this iterative process, the quasi-score function for parameter θ_1 is constructed as such an unbiased estimating function depending mainly on the first regression equation and secondarily on the related information from the other regression equations. The remainder of the article is organized as follows. In Section 2, we define a quasi-score function for parameter θ_1 in SUR system with $m = 2$. This score function is an optimum linear combination of the first two regression equations. In Section 3, we study finite sample properties of this quasi-score function and the asymptotic behaviours of the quasi-likelihood estimate obtained from this quasi-score function. These results show that the new quasi-score function for parameter of interest is E-sufficient, iteratively efficient and approximately efficient in a class of linear unbiased estimating functions. The purpose of Section 4 is to develop the method of this quasi-score to multi-SUR systems and generalized SUR systems. In Section 5, some examples are given to illustrate our results. The proofs of the lemmas and the theorems are all presented in Section 6.

§ 2. Quasi-score Function

For convenience, in this section, we first consider a particular situation in which there are only two seemingly unrelated regression equations, i.e., $m = 2$ in Model (1.2). According to the definition of quasi-score function presented as (1.1), a quasi-score function of θ_1 based only on the first regression equation should be defined as

$$q_1(\theta_1) = \sigma_1^{-2}\dot{\mu}_1'(\theta_1)V_1^{-1}(\theta_1)e_1(\theta_1, Y_1).$$

If we add the related information from the second equation to the score function above, a class of estimating functions based on the first two equations can be constructed as

$$q_{12} = \sigma_1^{-2}\dot{\mu}_1'(\theta_1)V_1^{-1}(\theta_1)e_1(\theta_1, Y_1) + Ae_2(\theta_2, Y_2), \tag{2.1}$$

where A is an arbitrary non-random matrix. This is a simple linear iterative process from $q_1(\theta_1)$ to an unbiased estimating function $q_{12}(\theta_1, \theta_2)$. According to the concept of optimum estimating function in [2, 5, 9], an optimum choice of A should maximize $E(q_{12}q_{12}')$. It can be verified immediately that the optimum choice of A is

$$A^* = -\frac{\sigma_{12}}{\sigma_1^2\sigma_2^2}\dot{\mu}_1'(\theta_1)V_1^{-1}(\theta_1)V_{12}(\theta_1, \theta_2)V_2^{-1}(\theta_2).$$

Therefore, in the class of estimating functions defined by (2.1) the optimum choice is

$$q_{12}^*(\theta_1, \theta_2) = \frac{1}{\sigma_1^2}\dot{\mu}_1'(\theta_1)V_1^{-1}(\theta_1)e_1(\theta_1, Y_1)$$

$$- \frac{\sigma_{12}}{\sigma_1^2\sigma_2^2}\dot{\mu}_1'(\theta_1)V_1^{-1}(\theta_1)V_{12}(\theta_1,\theta_2)V_2^{-1}(\theta_2)e_2(\theta_2,Y_2).$$

On the other hand, the estimating function $q_{12}^*(\theta_1,\theta_2)$ depends also on parameter θ_2. To eliminate the influence from θ_2, we need the quasi-score function related only to the second regression equation

$$q_2(\theta_2) = \sigma_2^{-2}\dot{\mu}_2'(\theta_2)V_2^{-1}(\theta_2)e_2(\theta_2,Y_2).$$

By summing up the two estimating functions $q_{12}^*(\theta_1,\theta_2)$ and $q_2(\theta_2)$, if σ_{12} and σ_2^2 are known, a new quasi-score function for θ_1 is defined as

$$\begin{cases} q_{12}^*(\theta_1,\theta_2) = \dfrac{1}{\sigma_1^2}\dot{\mu}_1'(\theta_1)V_1^{-1}(\theta_1)e_1(\theta_1,Y_1) \\ \qquad\qquad\qquad -\dfrac{\sigma_{12}}{\sigma_1^2\sigma_2^2}\dot{\mu}_1'(\theta_1)V_1^{-1}(\theta_1)V_{12}(\theta_1,\theta_2)V_2^{-1}(\theta_2)e_2(\theta_2,Y_2), \\ q_2(\theta_2) = \dfrac{1}{\sigma_2^2}\dot{\mu}_2'(\theta_2)V_2^{-1}(\theta_2)e_2(\theta_2,Y_2). \end{cases} \tag{2.2}$$

Note that the quasi-score function (2.2) is designed only for estimating parameter θ_1. Then the practical process to estimate θ_1 has two steps. The first step is to get an estimator $\hat{\theta}_2$ of θ_2 from the quasi-score function $q_2(\theta_2)$, and secondly, after θ_2 being replaced by $\hat{\theta}_2$ in $q_{12}^*(\theta_1,\theta_2)$, we obtain the quasi-score function $q_{12}^*(\theta_1,\hat{\theta}_2)$ for θ_1. This iterative process shows that the quasi-score function $q_{12}^*(\theta_1,\hat{\theta}_2)$ depends mainly on the first regression equation and secondarily on the second regression equation.

We call $q_{12}^*(\theta_1,\hat{\theta}_2)$ the iterative quasi-score function for θ_1. If $\hat{\theta}_{12}$ satisfies

$$q_{12}^*(\theta_1,\hat{\theta}_2)|_{\theta_1=\hat{\theta}_{12}} = 0,$$

then we call $\hat{\theta}_{12}$ the iterative quasi-likelihood estimator (IQLE) of θ_1.

§ 3. Main Properties

As mentioned in the previous section, the two estimating functions $q_{12}^*(\theta_1,\theta_2)$ and $q_2(\theta_2)$ play different roles in parameter estimation. According to the theory of estimating functions in [5, 15], in order to estimate the parameter of interest θ_1, it is necessary that the estimating function for parameter θ_1 is E-sufficient. For this purpose, we consider the class of linear unbiased estimating functions

$$\Psi^{(2)} = \{A_1(\theta_1,\theta_2)e_1(\theta_1,Y_1) + A_2(\theta_1,\theta_2)e_2(\theta_2,Y_2) : A_1, A_2 \text{ are arbitrary } p_1\times n \text{ matrices}\}$$

and assume that the order of the differentiation and integration can be interchanged. Following the theory of estimating functions in [5, 15], an estimating function $\psi(\theta_1,\theta_2)$ is said to be E-ancillary for parameter θ_1 if

$$E\Big(\frac{\partial\psi(\theta_1,\theta_2)}{\partial\theta_1}\Big) = 0 \qquad \text{for all } \ \theta_1\in\Theta_1,\ \theta_2\in\Theta_2,$$

where Θ_i is the parameter space of θ_i.

Lemma 3.1. *Let* $\mathcal{A}_1^{(2)} \subset \Psi^{(2)}$ *be the class of E-ancillary functions for parameter* θ_1 *and* $\partial\mu_1(\theta_1)/\partial\theta_{11}|_{\theta_1=\theta_1^*}, \cdots, \partial\mu_1(\theta_1)/\partial\theta_{1p_1}|_{\theta_1=\theta_1^*}$ *be linearly independent for some* $\theta_1^* \in \Theta_1$. *Then*

$$\mathcal{A}_1^{(2)} = \{A_2(\theta_1,\theta_2)e_2(\theta_2,Y_2) : A_2 \textit{ is an arbitrary } p_1 \times n \textit{ matrix}\}.$$

From Lemma 3.1, we have the following theorem, which describes the small sample property of $q_{12}^*(\theta_1,\theta_2)$.

Theorem 3.1. *Under the conditions of Lemma* 3.1, $E(q_{12}^*(\theta_1,\theta_2)\psi'(\theta_1,\theta_2)) = 0$ *for all* $\psi(\theta_1,\theta_2) \in \mathcal{A}_1^{(2)}$, $\theta_1 \in \Theta_1$ *and* $\theta_2 \in \Theta_2$.

Remark 3.1. Theorem 3.1 shows that $q_{12}^*(\theta_1,\theta_2)$ is orthogonal to $\mathcal{A}_1^{(2)}$ under the inner product defined by $\langle a,b\rangle_\theta = E(ab')$. Thus $q_{12}^*(\theta_1,\theta_2)$ is E-sufficient for θ_1 in $\Psi^{(2)}$ (see [5, 15]). Since $q_2(\theta_2)$ is unrelated to θ_1, it is E-ancillary for θ_1 and, consequently, $q_{12}^*(\theta_1,\theta_2)$ is orthogonal to $q_2(\theta_2)$. This orthogonality plays a key role in the efficiency and normality for estimating θ_1 (for details, see the proof of Theorem 3.2).

Denote $\dot\mu_1' V_1^{-1}$ by $(s_{it}^{(1)})$ and $\dot\mu_1' V_1^{-1} V_{12} V_2^{-1}$ by $(s_{it}^{(2)})$, $i = 1,\cdots,p_1, t = 1,\cdots,n$. Let $n \times p_1 \times p_1$ array matrices $W^{(k)} = (W_{tij}^{(k)}) = (\partial s_{it}^{(k)}/\partial\theta_{1j})$, $Q^{(2)} = (Q_{tij}^{(2)}) = (\partial s_{it}^{(2)}/\partial\theta_{2j})$ and vectors $W_{ij}^{(k)} = (W_{1ij}^{(k)},\cdots,W_{nij}^{(k)})'$, $Q_{ij}^{(2)} = (Q_{1ij}^{(2)},\cdots,Q_{nij}^{(2)})'$, $k = 1,2$.

In order to obtain the consistency and asymptotic normality for IQLE $\hat\theta_{12}$, we need the following regularity conditions.

(A) The rank of $\partial\mu_1(\theta_1)/\partial\theta_1|_{\theta_1=\theta_1^0}$ is p_1, where θ_i^0 stands for the true value of θ_i in Model (1.2). In a neighborhood of $(\theta_1^{0\prime},\theta_2^{0\prime})'$, $s_{it}^{(k)}$, $W_{tij}^{(k)}$ and $\partial W_{tij}^{(k)}/\partial\theta_{1r}$ are continuous and bounded.

(B) In a neighborhood of $(\theta_1^{0\prime},\theta_2^{0\prime})'$,

$$\begin{aligned}
&\lim_{n\to\infty}\frac{1}{n}\dot\mu_k' V_k^{-1}\dot\mu_k = \Delta_k, \qquad k = 1,2,\\
&\lim_{n\to\infty}\frac{1}{n}\dot\mu_1' V_1^{-1}V_{12}V_2^{-1}\dot\mu_2 = \Delta_{12}, \quad \lim_{n\to\infty}\frac{1}{n}\dot\mu_1' V_1^{-1}V_{12}V_2^{-1}V_{21}V_1^{-1}\dot\mu_1 = \Lambda_1,\\
&\lim_{n\to\infty}\frac{1}{n}\big(W_{ij}^{(k)}\big)' V_k W_{ij}^{(k)} = F_{kij},\\
&\lim_{n\to\infty}\frac{1}{n}\big(Q_{ij}^{(2)}\big)' V_2 Q_{ij}^{(2)} = G_{2ij}, \qquad k = 1,2, \quad i,j = 1,\cdots,p_1.
\end{aligned}$$

These conditions are standard in the theory of quasi-likelihood (see [9]). Under the conditions above, we have the following lemma and theorem.

Lemma 3.2. *If Conditions* (A) *and* (B) *hold, then IQLE of* θ_1 *satisfies*

$$\hat\theta_{12} - \theta_1^0 = O_P(n^{-1/2}) \qquad \textit{as} \;\; n\to\infty.$$

From Lemma 3.1, we get the following theorem.

Theorem 3.2. *If the conditions of Lemma* 3.1, (A) *and* (B) *hold, then IQLE of* θ_1 *satisfies*

$$\sqrt{n}(\hat\theta_{12} - \theta_1^0) \xrightarrow{\mathcal{D}} N(0, \Upsilon_{12}(\theta_1^0,\theta_2^0)) \qquad \textit{as} \;\; n\to\infty,$$

where

$$\Upsilon_{12}(\theta_1^0,\theta_2^0)=\sigma_1^2\Delta_1^{-1}(\theta_1^0)-\frac{\sigma_{12}^2}{\sigma_2^2}\Delta_1^{-1}(\theta_1^0)(\Lambda_1(\theta_1^0,\theta_2^0)-\Lambda_2(\theta_1^0,\theta_2^0))\Delta_1^{-1}(\theta_1^0),$$

$\Lambda_2=\Delta_{12}\Delta_2^{-1}\Delta_{21}$, $\Delta_{21}=\Delta_{12}'$, *and* Δ_k, Δ_{12} *and* Λ_1 *are presented as in Condition* (B).

Remark 3.2. Let $\hat\theta_1$ be the quasi-likelihood estimator of θ_1 obtained only from quasi-score function

$$q_1(\theta_1)=\sigma_1^{-2}\dot\mu_1'(\theta_1)V_1^{-1}(\theta_1)e_1(\theta_1,Y_1).$$

Then, under the above regularity conditions (see [11, 12]), we have

$$\sqrt{n}(\hat\theta_1-\theta_1^0)\xrightarrow{\mathcal{D}}N(0,\sigma_1^2\Delta_1^{-1}(\theta_1^0)).$$

Furthermore, we can verify that $\Lambda_1\ge\Lambda_2$ and then $\Upsilon_{12}\le\sigma_1^2\Delta_1^{-1}$. It shows that IQLE $\hat\theta_{12}$ is asymptotically more efficient than the original quasi-likelihood estimator $\hat\theta_1$, and the contribution quantity from the related information of the second regression equation is

$$\frac{\sigma_{12}^2}{\sigma_2^2}\Delta_1^{-1}(\theta_1^0)(\Lambda_1(\theta_1^0,\theta_2^0)-\Lambda_2(\theta_1^0,\theta_2^0))\Delta_1^{-1}(\theta_1^0)\ge 0.$$

On the other hand, if we use the quasi-score function (1.3) of full parameters to estimate θ_1 and the estimator is denoted by $\hat\theta_1^*$, then we can prove that the asymptotic covariance of $\hat\theta_1^*$ is Υ_{12}^* satisfying

$$\Upsilon_{12}^{*-1}=\frac{1}{\sigma_1^2}\Delta_1+\frac{\sigma_{12}^2}{\sigma_1^4\sigma_2^2}\Lambda_1+\frac{\sigma_{12}^4}{\sigma_1^4\sigma_2^4}R^*,$$

and

$$R^*\le\Lambda_1\le\Delta_1.$$

From Theorem 3.2 we get the similar result

$$\Upsilon_{12}^{-1}=\frac{1}{\sigma_1^2}\Delta_1+\frac{\sigma_{12}^2}{\sigma_1^4\sigma_2^2}\Lambda_1+\frac{\sigma_{12}^4}{\sigma_1^4\sigma_2^4}R,$$

where $R\le\Lambda_1\le\Delta_1$. The results above show that Υ_{12}^{-1} and Υ_{12}^{*-1} are equal in the two main parts and as a result $\hat\theta_{12}$ and $\hat\theta_1^*$, approximately, have the same efficiency.

From the assumption in which σ_2^2 and σ_{12} are known, we get the definition and the results above. When they are unknown, we can use their estimators, $\hat\sigma_2^2$ and $\hat\sigma_{12}$ say, to replace them. The methods to estimate σ_2^2 and σ_{12} are similar to those in [13] and [17]. Since these estimators are consistent, the results such as the asymptotic normality presented in Theorem 3.2 all hold (see [1]). So, in what follows, we always assume that σ_2^2 and σ_{12} or σ_i^2 and σ_{ij} in multi-SUR systems are known.

§ 4 . Extension

We now extend the method as proposed in the previous sections to general multi-SUR systems with m regression equations. Assume that we have had an iterative quasi-score function based on the first to k-th regressions equations denoted by $q_{1\cdots k}^*$. Adding the information from $(k+1)$-th regression equation to $q^*_{1\cdots k}$, we get $q^*_{1\cdots k(k+1)} = q^*_{1\cdots k}+A^*e_{k+1}$ for a suitable matrix A^*, by which $q^*_{1\cdots k(k+1)}$ obtains its maximum covariance. Since $q^*_{1\cdots k(k+1)}$ is also dependent on $\theta_2, \cdots, \theta_{k+1}$, we need the ancillary quasi-score functions defined by $q_i = \sigma_i^{-2}\dot{\mu}_i' V_i^{-1} e_i$ for $i = 2, \cdots, k+1$ to eliminate the influence from the parameters $\theta_2, \cdots, \theta_{k+1}$. Then the iterative quasi-score function for θ_1 in Model (1.2) is defined by

$$\begin{cases} q^*_{12\cdots m} = \dfrac{1}{\sigma_1^2}\dot{\mu}_1' V_1^{-1} e_1 - \displaystyle\sum_{i=2}^{m} \frac{\sigma_{1i}}{\sigma_1^2\sigma_i^2}\dot{\mu}_1' V_1^{-1} V_{1i} V_i^{-1} e_i \\ \qquad + \displaystyle\sum_{2\le i<j\le m} \frac{\sigma_{1i}\sigma_{ij}}{\sigma_1^2\sigma_i^2\sigma_j^2}\dot{\mu}_1' V_1^{-1} V_{1i} V_i^{-1} V_{ij} V_j^{-1} e_j \\ \qquad +(-1)^l \displaystyle\sum_{2\le i_1<\cdots<i_l\le m} \frac{\sigma_{1i_1}\cdots\sigma_{i_{l-1}i_l}}{\sigma_1^2\sigma_{i_1}^2\cdots\sigma_{i_l}^2}\dot{\mu}_1' V_1^{-1} V_{1i_1} V_{i_1}^{-1} V_{i_1i_2} V_{i_2}^{-1}\cdots V_{i_l}^{-1} e_{i_l} \\ \qquad +\cdots \\ \qquad +(-1)^m \dfrac{\sigma_{12}\cdots\sigma_{(m-1)m}}{\sigma_1^2\sigma_2^2\cdots\sigma_m^2}\dot{\mu}_1' V_1^{-1} V_{12} V_2^{-1} V_{23} V_3^{-1}\cdots V_m^{-1} e_m, \\ q_i = \dfrac{1}{\sigma_i^2}\dot{\mu}_i' V_i^{-1} e_i, \qquad i = 2, \cdots, m. \end{cases} \tag{4.1}$$

We can verify that the score functions defined above have orthogonality as

$$E(q^*_{12\cdots(k+1)}(\theta_1,\theta_2,\cdots,\theta_k,\theta_{k+1})(A_{(k+1)}e_{(k+1)})') = 0 \qquad \text{for} \quad A_{(k+1)}e_{(k+1)} \in \mathcal{A}_1^{(k+1)},$$

where $\mathcal{A}_1^{(k+1)} = \{A_{(k+1)}e_{(k+1)} : A_{(k+1)}(\theta_1,\theta_{k+1}) \ \text{ is an arbitrary matrix}\} \subset \Psi^{(k+1)}$ is the class of E-ancillary functions for θ_1 and

$$\Psi^{(k+1)} = \{A_1e_1 + A_{(k+1)}e_{(k+1)} : A_1(\theta_1,\theta_{k+1}) : A_{(k+1)}(\theta_1,\theta_{k+1}) \ \text{ are arbitrary matrices}\}$$

is the class of linear unbiased estimating functions. At the same time, q_i, $i = 2, \cdots, m$, are independent of θ_1. Thus these properties make the score functions $q_i, i = 2, \cdots, m$ to be E-ancillary and $q^*_{12\cdots m}$ to be E-sufficient for parameter θ_1.

From the equations $q_i = 0$ we get the estimators $\hat{\theta}_i$ of θ_i for $i = 2, \cdots, m$. And then by putting them into $q^*_{12\cdots m}$, we obtain the iterative quasi-score function $q^*_{12\cdots m}(\theta_1, \hat{\theta}_2, \cdots, \hat{\theta}_m)$ for θ_1. The root of the equation $q^*_{12\cdots m}(\theta_1, \hat{\theta}_2, \cdots, \hat{\theta}_m) = 0$ is denoted by $\hat{\theta}_{1\cdots m}$. We call it IQLE of θ_1. It can be verified that, under some regularity conditions similar to (A) and (B) above,

$$\sqrt{n}(\hat{\theta}_{1\cdots m} - \theta_1^0) \xrightarrow{\mathcal{D}} N(0, \Upsilon_{1\cdots m}(\theta_1^0, \cdots, \theta_m^0)) \qquad \text{as} \ \ n\to\infty$$

and the asymptotic covariance satisfies that $\Upsilon_{1\cdots(m-1)} \ge \Upsilon_{1\cdots m}$ for any m. It shows that $\hat{\theta}_{1\cdots m}$ is iteratively efficient.

We now introduce some examples to illustrate our theoretical results.

Example 4.1. Consider the following seemingly unrelated linear regression system

$$\begin{cases} E(Y_i) = X_i\theta_i, \quad \mathrm{Var}(Y_i) = \sigma_i^2 I, \qquad i = 1, 2, \\ \mathrm{Cov}(Y_1, Y_2) = \sigma_{12} I, \end{cases}$$

where $X_i = (x_{i1}, \cdots, x_{in})'$ are $n \times p_i$ design matrices. It follows from the formula (2.2) that the IQLE of θ_1 is

$$\hat{\theta}_{12} = \hat{\theta}_1 - \frac{\sigma_{12}}{\sigma_2^2}(X_1'X_1)^{-1}X_1'(I - P_{X_2})Y_2,$$

where $\hat{\theta}_1 = (X_1'X_1)^{-1}X_1'Y_1$ and $P_{X_2} = X_2(X_2'X_2)^{-1}X_2'$. We can verify that $E(\hat{\theta}_{12}) = \theta_1$ and $\hat{\theta}_{12}$ has minimum covariance in the class of linear unbiased estimations

$$\{\hat{\theta}_1 + Ae_2(\hat{\theta}_2, Y_2) : A \text{ is an arbitrary } p_1 \times n \text{ matrix}\},$$

where $\hat{\theta}_2 = (X_2'X_2)^{-1}X_2'Y_2$. Furthermore, $\hat{\theta}_{12}$ is a linear function of Y_i and

$$\mathrm{Var}(\hat{\theta}_{12}) = \sigma_1^2(X_1'X_1)^{-1} - \frac{\sigma_{12}^2}{\sigma_2^2}(X_1'X_1)^{-1}(X_1'X_1 - X_1'P_{X_2}X_1)(X_1'X_1)^{-1}.$$

Thus, it is clear from the central limit theorem that

$$\sqrt{n}(\hat{\theta}_{12} - \theta_1^0) \xrightarrow{\mathcal{D}} N(0, \Upsilon_{12}),$$

where

$$\Upsilon_{12} = \sigma_1^2\Delta_1^{-1} - \frac{\sigma_{12}^2}{\sigma_2^2}\Delta_1^{-1}(\Delta_1 - \Lambda_2)\Delta_1^{-1},$$
$$\Delta_1 = \lim_{n\to\infty}\frac{1}{n}X_1'X_1, \qquad \Lambda_2 = \lim_{n\to\infty}\frac{1}{n}X_1'P_{X_2}X_1.$$

This just illustrates the results in Theorem 3.2 and Remark 3.2.

Generally, if $i = 1, \cdots, m$ in the linear model above, then the IQLE of θ_1 is

$$\begin{aligned}\hat{\theta}_{12\cdots m} = \hat{\theta}_1 &- \sum_{i=2}^{m}\frac{\sigma_{1i}}{\sigma_i^2}(X_1'X_1)^{-1}X_1'N_iY_i \\ &+ \sum_{2\le i<j\le m}\frac{\sigma_{1i}\sigma_{ij}}{\sigma_i^2\sigma_j^2}(X_1'X_1)^{-1}X_1'N_iN_jY_j \\ &+ (-1)^l\sum_{2\le i_1<\cdots<i_l\le m}\frac{\sigma_{1i_1}\cdots\sigma_{i_{l-1}i_l}}{\sigma_{i_1}^2\cdots\sigma_{i_l}^2}(X_1'X_1)^{-1}X_1'N_{i_1}\cdots N_{i_l}Y_{i_l} \\ &+ \cdots \\ &+ (-1)^m\frac{\sigma_{12}\cdots\sigma_{(m-1)m}}{\sigma_2^2\cdots\sigma_m^2}(X_1'X_1)^{-1}X_1'N_2\cdots N_mY_m,\end{aligned}$$

where $N_i = I - P_{X_i}$. For all m, we have $E(\hat{\theta}_{12\cdots m}) = \theta_1$ and $\mathrm{Var}(\hat{\theta}_{12\cdots m}) \le \mathrm{Var}(\hat{\theta}_{12\cdots(m-1)})$ (see [17]). It illustrates the iterative efficiency.

Example 4.2. In this example, we assume that $Y_{i1}, \cdots, Y_{in}$ are independently (but not necessary identically) distributed, the density function of Y_{ij} has the form of $f_{ij}(|y_{ij} - \theta_i|)$ and function $f_{ij}(y)$ has symmetic center $y = \theta_i$, $i = 1, 2$. We first consider the following unbiased estimating functions

$$g(y_{ij}, \theta_i) = \mathrm{sign}(y_{ij} - \theta_i). \tag{4.2}$$

Let

$$\mathrm{Var}(g(Y_{ij},\theta_i))=\sigma_i^2\lambda_j,\qquad \mathrm{Cov}(g(Y_{1j},\theta_1),g(Y_{2j},\theta_2))=\sigma_{12}\kappa_j$$

and

$$\mathrm{Cov}(g(Y_{1j_1},\theta_1),g(Y_{2j_2},\theta_2))=0,\qquad j_1\neq j_2.$$

According to common definition of derivative, however, $\partial g/\partial\theta_i$ can not be defined on some points. In this case, following Godambe and Thompson [3], we define

$$E\Big(\frac{\partial g(Y_{ij},\theta_i)}{\partial\theta_i}\Big)=\lim_{\varepsilon\to0}E(g(Y_{ij},\theta_i+\varepsilon)-g(Y_{ij},\theta_i))/\varepsilon.$$

From this definition, it can be verified that

$$-\frac{1}{2}E\Big(\frac{\partial g(Y_{ij},\theta_i)}{\partial\theta_i}\Big)=\lim_{\varepsilon\to0}P(\theta_i<Y_{ij}<\theta_i+\varepsilon)/\varepsilon=f_{ij}(0),$$

where $f_{ij}(0)$ is the density of Y_{ij} valued at $y_{ij}=\theta_i$, i.e., $f_{ij}(0)=f_{ij}(|y_{ij}-\theta_i|)|_{y_{ij}=\theta_i}$. Then we get the quasi-score function for θ_2:

$$q_2(\theta_2)=\sigma_2^{-2}\sum_{j=1}^n\lambda_j^{-1}f_{2j}(0)g(Y_{2j},\theta_2).$$

Denote the root of the equation $q_2(\theta_2)=0$ by $\hat\theta_2$. We now assume that $Y_{1k_1}\le Y_{1k_2}\le\cdots\le Y_{1k_n}$, $Y_{2j_1}\le Y_{2j_2}\le\cdots\le Y_{2j_n}$, and m satisfies

$$-\sum_{s=1}^m\lambda_{j_s}^{-1}f_{2j_s}(0)+\sum_{s=m+1}^n\lambda_{j_s}^{-1}f_{2j_s}(0)\ge0\ \text{ and }\ -\sum_{s=1}^{m+1}\lambda_{j_s}^{-1}f_{2j_s}(0)+\sum_{s=m+2}^n\lambda_{j_s}^{-1}f_{2j_s}(0)\le0.$$

We can verify that $\hat\theta_2\doteq\frac12(Y_{2j_m}+Y_{2j_{(m+1)}})$, the m-th ordering statistic of $Y_{21},\cdots,Y_{2n}$. Thus it follows that the iterative quasi-score function for θ_1 is

$$\begin{aligned}q_{12}^*(\theta_1,\hat\theta_2)&=\sigma_1^{-2}\sum_{j=1}^n\lambda_j^{-1}f_{1j}(0)g(Y_{1j},\theta_1)\\&\quad-\sigma_{12}\sigma_1^{-2}\sigma_2^{-2}\Big(-\sum_{s=1}^m\lambda_{j_s}^{-2}\kappa_{j_s}f_{1j_s}(0)+\sum_{s=m+1}^n\lambda_{j_s}^{-2}\kappa_{j_s}f_{1j_s}(0)\Big).\end{aligned}$$

By the equation $q_{12}^*(\theta_1)=0$ we see that the IQLE of θ_1 is $\hat\theta_{12}\doteq\frac12(Y_{1k_l}+Y_{1k_{(l+1)}})$, where l satisfies

$$-\sum_{s=1}^l\lambda_{k_s}^{-1}f_{1k_s}(0)+\sum_{s=l+1}^n\lambda_{k_s}^{-1}f_{1k_s}(0)\ge\frac{\sigma_{12}}{\sigma_2^2}\Big(-\sum_{s=1}^m\lambda_{j_s}^{-2}\kappa_{j_s}f_{1j_s}(0)+\sum_{s=m+1}^n\lambda_{j_s}^{-2}\kappa_{j_s}f_{1j_s}(0)\Big),$$

$$-\sum_{s=1}^{l+1}\lambda_{k_s}^{-1}f_{1k_s}(0)+\sum_{s=l+2}^n\lambda_{k_s}^{-1}f_{1k_s}(0)\le\frac{\sigma_{12}}{\sigma_2^2}\Big(-\sum_{s=1}^m\lambda_{j_s}^{-2}\kappa_{j_s}f_{1j_s}(0)+\sum_{s=m+1}^n\lambda_{j_s}^{-2}\kappa_{j_s}f_{1j_s}(0)\Big).$$

Obviously, $\hat\theta_{12}$ is the l-th ordering statistic of $Y_{11},\cdots,Y_{1n}$. In this case the asymptotic variance of $\sqrt{n}\,\hat\theta_{12}$ is

$$\Upsilon_{12}=\sigma_1^2\Delta_1^{-1}-\frac{\sigma_{12}^2}{\sigma_2^2}\Delta_1^{-1}(\Delta_1-\Delta_{12}\Delta_1^{-1}\Delta_{12})\Delta_1^{-1},\tag{4.3}$$

where

$$\Delta_1 = \lim_{n\to\infty} \frac{4}{n}\sum_{j=1}^{n} \lambda_j^{-1} f_{1j}^2(0), \qquad \Delta_{12} = \lim_{n\to\infty} \frac{4}{n}\sum_{s=1}^{n} \lambda_j^{-2}\kappa_j f_{1j}(0) f_{2j}(0).$$

On the other hand, if unbiased function is chosen as

$$g(y_{ij}, \theta_i) = y_{ij} - \theta_i, \tag{4.4}$$

then, from Example 4.1, we can see that the IQLE of θ_1 is $\hat\theta_{12}^* = \overline{Y}_1 = \frac{1}{n}\sum\limits_{j=1}^{n} Y_{1j}$, which is unrelated to $Y_{21}, \cdots, Y_{2n}$. Then the asymptotic covariance of $\sqrt{n}\,\hat\theta_{12}^*$ is

$$\Upsilon_{12}^* = \lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} \sigma_1^2 \lambda_j.$$

It follows from (4.3) and the result above that, if the correlation between $Y_{11}, \cdots, Y_{1n}$ and $Y_{21}, \cdots, Y_{2n}$ is large enough and $Y_{21}, \cdots, Y_{2n}$ are dispersive enough, i.e., $\frac{|\sigma_{12}|}{(\sigma_1^2\sigma_2^2)}$ and σ_2^2 are large enough, then $\Upsilon_{12} \le \Upsilon_{12}^*$. In this case we should choose (4.2) as unbiased estimating function. Otherwise, we should choose (4.4) as unbiased estimating function.

§ 5 . Proofs

Proof of Lemma 3.1. By the definition of E-ancillary we can verify the lemma immediately.

Proof of Theorem 3.1. By the definitions of $q_{12}^*(\theta_1, \theta_2)$ and Model (1.2) and Lemma 3.1 we can verify the theorem immediately.

Proof of Lemma 3.2. According to the theory of quasi-likelihood (see [12, 13]), we have

$$\hat\theta_2 - \theta_2^0 = O_P(n^{-1/2}). \tag{5.1}$$

By the definition of IQLE of θ_1, we have $q_{12}^*(\theta_1, \hat\theta_2)|_{\theta_1=\hat\theta_{12}} = 0$. If it is expanded at $(\theta_1^0, \theta_2^0)'$ by Taylor expansion, we get

$$\begin{aligned} 0 = {} & \dot\mu_1'(\theta_1^0)V_1^{-1}(\theta_1^0)e_1(\theta_1^0) - \frac{\sigma_{12}}{\sigma_2^2}\dot\mu_1'(\theta_1^0)V_1^{-1}(\theta_1^0)V_{12}(\theta_1^0,\theta_2^0)V_2^{-1}(\theta_2^0)e_2(\theta_2^0) \\ & + \Big\{ -\dot\mu_1'(\theta_{1c}^0)V_1^{-1}(\theta_{1c}^0)\dot\mu_1(\theta_{1c}^0) + [e_1'(\theta_{1c})][W^{(1)}(\theta_{1c},\theta_{2c})] \\ & - \frac{\sigma_{12}}{\sigma_2^2}[e_2'(\theta_{2c})][W^{(2)}(\theta_{1c},\theta_{2c})]\Big\}(\hat\theta_{12} - \theta_1^0) \\ & - \frac{\sigma_{12}}{\sigma_2^2}\{-\dot\mu_1'(\theta_{1c}^0)V_1^{-1}(\theta_{1c}^0)V_{12}(\theta_{1c}^0,\theta_{2c}^0)V_2^{-1}(\theta_{2c}^0)\dot\mu_2 \\ & + [e_2'(\theta_{2c})][Q^{(2)}(\theta_{1c},\theta_{2c})]\}(\hat\theta_2 - \theta_2^0), \end{aligned} \tag{5.2}$$

where $\|\theta_{1c} - \theta_1^0\| < \|\hat\theta_{12} - \theta_1^0\|$, $\|\theta_{2c} - \theta_1^0\| < \|\hat\theta_2 - \theta_2^0\|$ and $[\,\cdot\,][\,\cdot\,]$ stands for multiplying a matrix by an array matrix (see [19]). Now we calculate the orders of convergence of every terms in Equation (5.2). From Condition (B), it is clear that

$$E(\dot\mu_1'(\theta_1^0)V_1^{-1}(\theta_1^0)e_1(\theta_1^0)) = 0, \qquad \mathrm{Var}(\dot\mu_1'(\theta_1^0)V_1^{-1}(\theta_1^0)e_1(\theta_1^0)) = O(n).$$

As a result,

$$\dot{\mu}_1'(\theta_1^0)V_1^{-1}(\theta_1^0)e_1(\theta_1^0) = O_P(n^{\frac{1}{2}}), \tag{5.3}$$

$$E(e_1'(\theta_1^0)W_{ij}^{(1)}) = 0, \qquad \mathrm{Var}(e_1'(\theta_1^0)W_{ij}^{(1)}) = O(n),$$

and then

$$[e_1'(\theta_{1c})][W^{(1)}(\theta_{1c},\theta_{2c})] = [e_1'(\theta_1^0)][W^{(1)}(\theta_1^0,\theta_2^0)] + C_n = O_P(n^{\frac{1}{2}}), \tag{5.4}$$

where C_n is uniformly bounded by a constant for $n = 1, 2, \cdots$. Similarly,

$$\dot{\mu}_1'(\theta_1^0)V_1^{-1}(\theta_1^0)V_{12}(\theta_1^0,\theta_2^0)V_2^{-1}(\theta_2^0)e_2(\theta_2^0) = O_P(n^{\frac{1}{2}}) \tag{5.5}$$

and

$$[e_2'(\theta_{2c})][W^{(2)}(\theta_{1c},\theta_{2c})] = O_P(n^{\frac{1}{2}}), \qquad [e_2'(\theta_{2c})][Q^{(2)}(\theta_{1c},\theta_{2c})] = O_P(n^{\frac{1}{2}}). \tag{5.6}$$

From (5.1) and (5.3)–(5.6), by comparing with the orders of convergence of every terms in Equation (5.2), we get $\hat{\theta}_{12} - \theta_1^0 = O_P(n^{-1/2})$. The proof is completed.

Proof of Theorem 3.2. It is clear from the equation $q_{12}^*(\theta_1,\hat{\theta}_2)|_{\theta_1=\hat{\theta}_{12}} = 0$, Lemma 3.2 and Condition (A) that we have the following Taylor expression

$$\begin{aligned}
0 &= \dot{\mu}_1'(\theta_1^0)V_1^{-1}(\theta_1^0)e_1(\theta_1^0) - \frac{\sigma_{12}}{\sigma_2^2}\dot{\mu}_1'(\theta_1^0)V_1^{-1}(\theta_1^0)V_{12}(\theta_1^0,\theta_2^0)V_2^{-1}(\theta_2^0)e_2(\theta_2^0)\\
&\quad + \Big\{ - \dot{\mu}_1'(\theta_1^0)V_1^{-1}(\theta_1^0)\dot{\mu}_1 + [e_1'(\theta_1^0)][W^{(1)}(\theta_1^0,\theta_2^0)] - \frac{\sigma_{12}}{\sigma_2^2}[e_2'(\theta_2^0)][W^{(2)}(\theta_1^0,\theta_2^0)]\Big\}(\hat{\theta}_{12} - \theta_1^0)\\
&\quad - \frac{\sigma_{12}}{\sigma_2^2}\{-\dot{\mu}_1'(\theta_1^0)V_1^{-1}(\theta_1^0)V_{12}(\theta_1^0,\theta_2^0)V_2^{-1}(\theta_2^0)\dot{\mu}_2 + [e_2'(\theta_2^0)][Q^{(2)}(\theta_1^0,\theta_2^0)]\}(\hat{\theta}_2 - \theta_2^0)\\
&\quad + o_P(|\hat{\theta}_{12} - \theta_1^0|) + o_P(|\hat{\theta}_2 - \theta_2^0|).
\end{aligned}$$

By Lemma 3.1, Condition (B), (5.3)–(5.6) and the expression above, we get

$$\begin{aligned}
\sqrt{n}(\hat{\theta}_{12} - \theta_1^0) &= \Big\{\frac{1}{n}\dot{\mu}_1'(\theta_1^0)V_1^{-1}(\theta_1^0)\dot{\mu}_1(\theta_1^0)\Big\}^{-1} \times \Big\{\frac{1}{\sqrt{n}}\dot{\mu}_1'(\theta_1^0)V_1^{-1}(\theta_1^0)e_1(\theta_1^0)\\
&\quad + \frac{1}{\sqrt{n}}([e_1'(\theta_1^0)][W^{(1)}(\theta_1^0,\theta_2^0)] - \frac{\sigma_{12}}{\sigma_2^2}[e_2'(\theta_2^0)][W^{(2)}(\theta_1^0,\theta_2^0)])(\hat{\theta}_{12} - \theta_1^0)\\
&\quad - \frac{\sigma_{12}}{\sigma_2^2}\frac{1}{\sqrt{n}}\dot{\mu}_1'(\theta_1^0)V_1^{-1}(\theta_1^0)V_{12}(\theta_1^0,\theta_2^0)V_2^{-1}(\theta_1^0)e_2(\theta_2^0)\\
&\quad + \frac{\sigma_{12}}{\sigma_2^2}\frac{1}{n}\dot{\mu}_1'(\theta_1^0)V_1^{-1}(\theta_1^0)V_{12}(\theta_1^0,\theta_2^0)V_2^{-1}(\theta_1^0)\dot{\mu}_2(\theta_2^0)\sqrt{n}(\hat{\theta}_2 - \theta_2^0)\\
&\quad - \frac{\sigma_{12}}{\sigma_2^2}\frac{1}{\sqrt{n}}[e_2'(\theta_2^0)][Q^{(2)}(\theta_1^0,\theta_2^0)](\hat{\theta}_2 - \theta_2^0)\\
&\quad + \frac{1}{\sqrt{n}}o_P(|\hat{\theta}_{12} - \theta_1^0|) + \frac{1}{\sqrt{n}}o_P(|\hat{\theta}_2 - \theta_2^0|)\Big\}\\
&= \Big\{\frac{1}{n}\dot{\mu}_1'(\theta_1^0)V_1^{-1}(\theta_1^0)\dot{\mu}_1(\theta_1^0)\Big\}^{-1}\Big\{\frac{\sigma_1^2}{\sqrt{n}}q_{12}^*(\theta_1^0,\theta_2^0) + \frac{\sigma_{12}}{\sigma_2^2}\frac{1}{n}\dot{\mu}_1'(\theta_1^0)\\
&\quad \cdot V_1^{-1}(\theta_1^0)V_{12}(\theta_1^0,\theta_2^0)V_2^{-1}(\theta_2^0)\dot{\mu}_2(\theta_2^0)\sqrt{n}(\hat{\theta}_2 - \theta_2^0)\Big\} + o_P(1).
\end{aligned} \tag{5.7}$$

On the other hand, from the equation $q_2(\hat{\theta}_2) = 0$ and the method above, we get

$$\sqrt{n}(\hat{\theta}_2 - \theta_2^0) = \left\{ \frac{1}{n} \dot{\mu}_2'(\theta_2^0) V_2^{-1}(\theta_2^0) \dot{\mu}_2(\theta_2^0) \right\}^{-1} \frac{\sigma_2^2}{\sqrt{n}} q_2(\theta_2^0) + o_P(1). \tag{5.8}$$

Lemma 3.1, (5.7) and (5.8) imply that the conclusion in the theorem holds. The proof is completed.

References

[1] Chen, K. & Heyde, C. C., On asymptotic optimality of estimating functions, *JSPI.*, **48**(1995), 101–112.

[2] Godambe, V. P. & Heyde, C. D., Quasi-likelihood and optimal estimation, *Interal Statist Rev.*, **55**(1987), 231–244.

[3] Godambe, V. P. & Thompson, M. E., Robust estimation through estimating equation, *Biometrika*, **71**(1984), 115–125.

[4] Godambe, V. P. & Thompson, M. E., An extension of quasi-likelihood estimation, *J. Statist. Plann. Inference*, **22**(1989), 137–152.

[5] Heyde, C. C., Quasi-likelihood and Its Application, Springer-Verlag, New York, 1997.

[6] Jung, S., Quasi-likelihood for median regression models, *JASA.*, **91**(1996), 251–257.

[7] Li, B. & McCullagh, P., Potential functions and conservative estimating functions, *Ann. Statist.*, **22**(1994), 340–356.

[8] Lin, L., Some properties for quasi likelihood estimate in nonlinear regression models (in Chinese), *Acta Mathematicae Applicatae Sinica*, **22**(1999), 307–310.

[9] Lin, L., Maximum information and optimum estimating function, *Chin. Ann. Math.*, **24B:**3(2003), 349–358.

[10] Lin, L. & Zhang, R. C., Profile quasi-likelihood, *Statistics and Probability Letters*, **56**(2002), 147–154.

[11] Lin, L., Bayes quasi-likelihood with auxiliary information in seeming unrelated nonlinear regression systems (in Chinese), *Acta Mathematica Sinica*, **45**(2002), 1227–1234.

[12] McCullagh, P., Quasi-likelihood functions, *Ann. of Statist.*, **11**(1983), 59–67.

[13] McCullagh, P. & Nelder, J. A., Generalized Linear Models, 2nd edition, Chmpman and Hall, London, 1989.

[14] Revankar, N. S., Some finite sample results in the context of seemingly unrelated regression equations, *JASA.*, **69**(1974), 187–190.

[15] Small, C. G. & Mcleish, D. L., Hilbert Space Methods in Probability and Statistical Inference, John Wiley and Sons. Inc., 1994.

[16] Srivastava, V. K. & Dwivedi, Estimation of seeming unrelated regression equation, *J. Econometrika*, **10**(1978), 15–32.

[17] Wang, S. G., A new estimate of regression coefficients in seeming unrelated regression system, *Science in China, Series A,* **32**(1989), 808–816.

[18] Wedderburn, R. W. M., Quasi-likelihood functions, generalized linear models, and the Guess-Newton method, *Biometrika*, **61**(1974), 439–447.

[19] Wei, B. C., Exponential Family Nonlinear Models, Springer-Verlag, Singapore, 1998.

[20] Zellner, A., An efficient method of estimating seeming unrelated regression test for aggregation bias, *JASA.*, **57**(1962), 348–368.