TAIWANESE JOURNAL OF MATHEMATICS
Vol. 8, No. 4, pp. 653-671, December 2004
This paper is available online at http://www.math.nthu.edu.tw/tjm/

# FIXED-SAMPLE OPTIMAL NON-ORTHOGONAL ESTIMATING FUNCTIONS IN THE PRESENCE OF NUISANCE PARAMETERS

Chih-Rung Chen and Lih-Chung Wang

**Abstract.** In the paper, an important necessary and sufficient condition for a commonly used non-orthogonal estimating function to be fixed-sample optimal is proposed. The class of all fixed-sample optimal non-orthogonal estimating functions is characterized under the proposed condition. A simple counterexample without any fixed-sample optimal non-orthogonal estimating function is constructed to show that the proposed condition does not necessarily hold. The usefulness and applicability of the proposed method are illustrated by two classical examples with many nuisance parameters.

## 1. Introduction

In the paper, consider the statistical space $(\Omega, \mathcal{F}, \{P_\theta : \theta \in \Theta\})$ for experiment $\mathcal{E}$, where $\Omega$ is the sample space of $\mathcal{E}$, $\mathcal{F}$ is the $\sigma$-field generated by some subsets of $\Omega$, $\theta$ ($\equiv (\psi, \lambda)$) is the parameter with parameter space $\Theta$ ($\equiv \Psi \times \Lambda$), and each $P_\theta$ is a complete probability measure on $(\Omega, \mathcal{F})$. Assume that $\psi$ is the parameter of interest in a known non-empty open subset $\Psi$ of the $p$-dimensional Euclidean space $\mathcal{R}^p$ for some $p \in \{1, 2, \ldots\}$ ($\equiv \mathcal{N}$) and that $\lambda$ is a nuisance parameter in a known or unknown non-empty set $\Lambda$. Let $Y$ be the response of $\mathcal{E}$.

To proceed the discussion, some basic notations and definitions are introduced as follows: Let $\mathcal{R}$ denote the set of all real numbers and let $\mathcal{Y}$ denote the range of $Y$. For $m, n \in \mathcal{N}$, let $\mathcal{M}_{m\times n}(\mathcal{R})$ denote the set of all $m \times n$ matrices with real-valued components.

**Definition 1.** $G$ is called an estimating function for $\psi$ if $G : \Psi \times \mathcal{Y} \mapsto \mathcal{R}^p$, i.e., $G$ is an $\mathcal{R}^p$-valued function on $\Psi \times \mathcal{Y}$.

Received March 12, 2002; accepted October 3, 2003.
Communicated by S. B. Hsu.
2000 *Mathematics Subject Classification*: 62A99.
*Key words and phrases*: Estimating function, fixed-sample optimal, Moore-Penrose inverse, non-orthogonal, nuisance parameter, parameter of interest.

**Definition 2.** An estimating function $G$ for $\psi$ is called unbiased if for each $\psi \in \Psi$, $G(\psi, Y)$ is measurable with respect to (w.r.t.) $\mathcal{F}$ such that $E_\theta(G(\psi, Y)) = 0_{p\times 1}$ for all $\lambda \in \Lambda$.

Throughout the paper, let $\mathcal{G}$ denote the class of all unbiased estimating functions $G$ for $\psi$ such that for each $\theta \in \Theta$,

(i) *$G(\psi, Y)$ is $P_\theta$-a.s. differentiable w.r.t. $\psi$ and*

(ii) *both $E_\theta(\dot{G}(\psi, Y))$ and $Cov_\theta(G(\psi, Y))$ are non-singular in $\mathcal{M}_{p\times p}(\mathcal{R})$, where $\dot{G} : \Psi \times \mathcal{Y} \mapsto \mathcal{M}_{p\times p}(\mathcal{R})$ such that $\dot{G}(\psi, y) = \partial G(\psi, y)/\partial \psi^T$ if it exists and $0_{p\times p}$ otherwise.*

**Definition 3.** For $G \in \mathcal{G}$, $G^{(s)}$ is called the standardized version of $G$ if $G^{(s)} : \Theta \times \mathcal{Y} \mapsto \mathcal{R}^p$ such that for each $\theta \in \Theta$ and $y \in \mathcal{Y}$,

$$G^{(s)}(\theta, y) = -E_\theta\left(\dot{G}^T(\psi, Y)\right) Cov_\theta^{-1}(G(\psi, Y))\, G(\psi, y). \tag{1}$$

For $G \in \mathcal{G}$, $\theta \in \Theta$, and $y \in \mathcal{Y}$, it follows from Definition 3 that $G(\psi, y) = 0_{p\times 1}$ if and only if $G^{(s)}(\theta, y) = 0_{p\times 1}$.

**Definition 4.** For $G \in \mathcal{G}$, $I_G$ is called the information of $G$ if $I_G : \Theta \mapsto \mathcal{M}_{p\times p}(\mathcal{R})$ such that for each $\theta \in \Theta$,

$$I_G(\theta) = Cov_\theta\left(G^{(s)}(\theta, Y)\right). \tag{2}$$

For $G \in \mathcal{G}$, it follows from Definitions 3 and 4 that for each $\theta \in \Theta$,

$$I_G(\theta) = E_\theta\left(\dot{G}^T(\psi, Y)\right) Cov_\theta^{-1}(G(\psi, Y))\, E_\theta\left(\dot{G}(\psi, Y)\right).$$

**Definition 5.** Let $\mathcal{G}_0$ be a non-empty subclass of $\mathcal{G}$. $G^*$ is called fixed-sample optimal within $\mathcal{G}_0$ if

(i) *$G^*$ is in $\mathcal{G}_0$ and*

(ii) *$I_{G^*}(\theta) - I_G(\theta)$ is a positive semi-definite matrix in $\mathcal{M}_{p\times p}(\mathcal{R})$ for each $G \in \mathcal{G}_0$ and $\theta \in \Theta$.*

Note that the existence of any fixed-sample optimal estimating function for $\psi$ within a non-empty subclass of $\mathcal{G}$ is by no means guaranteed. Since it is difficult to verify whether or not part (ii) in Definition 5 holds for most commonly used subclasses of $\mathcal{G}$, Heyde (1988) proposed a simple and easily verified sufficient condition for an estimating function for $\psi$ to be fixed-sample optimal within a non-empty subclass of $\mathcal{G}$ as follows:

**Lemma 1.** $($*Heyde*, $1988)$. *Let $\mathcal{G}_0$ be a non-empty subclass of $\mathcal{G}$. If $G^*$ is in $\mathcal{G}_0$ and for each $G \in \mathcal{G}_0$ and $\theta \in \Theta$,*

$$I_G(\theta) = Cov_\theta\left(G^{(s)}(\theta, Y), G^{*(s)}(\theta, Y)\right) = Cov_\theta\left(G^{*(s)}(\theta, Y), G^{(s)}(\theta, Y)\right)$$

*or*, *equivalently*,

$$\begin{aligned} &\left[E_\theta\left(\dot{G}(\psi, Y)\right)\right]^{-1} Cov_\theta(G(\psi, Y), G^*(\psi, Y)) \\ =& \left[E_\theta\left(\dot{G}^*(\psi, Y)\right)\right]^{-1} Cov_\theta(G^*(\psi, Y)), \end{aligned}$$

*then $G^*$ is a fixed-sample optimal estimating function for $\psi$ within $\mathcal{G}_0$.*

*Proof.* See page $14$ of Heyde (1997). ■

Conversely, Heyde (1988) claimed that the sufficient condition in Lemma 1 is also a necessary condition for $G^*$ to be a fixed-sample optimal estimating function for $\psi$ within a non-empty "convex" subclass of $\mathcal{G}$. However, he implicitly used the assumption given in the following lemma rather than the convexity assumption in his proof. See pages $14$ and $15$ of Heyde (1997) for detail.

**Lemma 2.** *Let $\mathcal{G}_0$ be a non-empty subclass of $\mathcal{G}$. Suppose that there exists a function $\alpha : \mathcal{G}_0 \times \mathcal{G}_0 \mapsto (0, \infty)$ such that $G_1 + \mathrm{diag}\{c_1, \ldots, c_p\}G_2 \in \mathcal{G}_0$ for all $c_1, \ldots, c_p \in (-\alpha(G_1, G_2), \alpha(G_1, G_2))$ and $G_1, G_2 \in \mathcal{G}_0$. Then the sufficient condition in Lemma 1 is also a necessary condition for $G^*$ to be a fixed-sample optimal estimating function for $\psi$ within $\mathcal{G}_0$.*

*Proof.* This lemma follows directly from a straightforward modification of the proof on pages $14$ and $15$ of Heyde (1997). ■

Note that the assumption given in Lemma 2 holds but the convexity assumption proposed by Heyde (1988) fails for most commonly used subclasses of $\mathcal{G}$, e.g., the class of all non-orthogonal estimating functions for $\psi$. See Section 2 for detail. Under the assumption given in Lemma 2, the class of all fixed-sample optimal estimating functions for $\psi$ within a non-empty subclass of $\mathcal{G}$ can be characterized as follows:

**Lemma 3.** *Let $\mathcal{G}_0$ be a non-empty subclass of $\mathcal{G}$. Suppose that the assumption given in Lemma 2 holds and that $G_0^*$ is a fixed-sample optimal estimating function for $\psi$ within $\mathcal{G}_0$. Then the following are equivalent*:

(i) *$G^*$ is a fixed-sample optimal estimating function for $\psi$ within $\mathcal{G}_0$*;

(ii) $G^*$ *is in* $\mathcal{G}_0$ *and for each* $\theta \in \Theta$,

$$G^{*(s)}(\theta, Y) = G_0^{*(s)}(\theta, Y) \quad P_\theta\text{-}a.s.;$$

(iii) $G^*$ *is in* $\mathcal{G}_0$ *and for each* $\theta \in \Theta$,

$$I_{G^*}(\theta) = I_{G_0^*}(\theta).$$

*Proof.* First of all, suppose that part (i) holds. By Definition 5, both $G^*$ and $G_0^*$ are in $\mathcal{G}_0$. By Lemma 2,

$$\begin{aligned} I_{G^*}(\theta) &= Cov_\theta\left(G^{*(s)}(\theta, Y), G_0^{*(s)}(\theta, Y)\right) \\ &= Cov_\theta\left(G_0^{*(s)}(\theta, Y), G^{*(s)}(\theta, Y)\right) \\ &= I_{G_0^*}(\theta) \end{aligned}$$

for all $\theta \in \Theta$. Thus,

$$\begin{aligned} &E_\theta\left(\left[G^{*(s)}(\theta, Y) - G_0^{*(s)}(\theta, Y)\right]\left[G^{*(s)}(\theta, Y) - G_0^{*(s)}(\theta, Y)\right]^T\right) \\ &\quad = Cov_\theta\left(G^{*(s)}(\theta, Y) - G_0^{*(s)}(\theta, Y)\right) \quad = \quad 0_{p\times p} \end{aligned}$$

for all $\theta \in \Theta$. Therefore, part (ii) holds. By Definition 4, part (iii) holds if part (ii) holds. By Definition 5, part (i) holds if part (iii) holds. Consequently, this lemma follows. ∎

The paper is organized as follows. In Section 2, an important necessary and sufficient condition for a commonly used non-orthogonal estimating function for $\psi$ to be fixed-sample optimal for $\psi$ is proposed. The class of all fixed-sample optimal non-orthogonal estimating functions for $\psi$ is characterized under the proposed condition. In Section 3, a simple counterexample without any fixed-sample optimal non-orthogonal estimating function for $\psi$ is constructed to show that the proposed condition does not necessarily hold. Finally, in Section 4, the usefulness and applicability of the proposed method are illustrated by two classical examples with many nuisance parameters.

## 2. Fixed-sample Optimal Non-orthogonal Estimating Functions

In this section, an important necessary and sufficient condition for a commonly used non-orthogonal estimating function for $\psi$ to be fixed-sample optimal for $\psi$ is proposed. The class of all fixed-sample optimal non-orthogonal estimating functions for $\psi$ is characterized under the proposed condition.

For each $\theta \in \Theta$, $i \neq i'$, and $i, i' \in \{1, \ldots, n\}$, assume that

$$E_\theta(h_i(Y)|\mathcal{F}_i) = \mu_i(\psi, Y) \quad P_\theta\text{-a.s.,} \tag{3}$$

$$Cov_\theta(h_i(Y)|\mathcal{F}_i) = \sigma^2(\theta)\, V_i(\psi, Y) \quad P_\theta\text{-a.s.,} \tag{4}$$

and

$$\begin{aligned} &E_\theta\left( [h_i(Y) - \mu_i(\psi, Y)]\, [h_{i'}(Y) - \mu_{i'}(\psi, Y)]^T \middle| \mathcal{F}_i \vee \mathcal{F}_{i'} \right) \\ &\quad = 0_{m_i \times m_{i'}} \quad P_\theta\text{-a.s.,} \end{aligned} \tag{5}$$

where $h_i : \mathcal{Y} \mapsto \mathcal{R}^{m_i}$ is a known transformation of response $Y$ for some $m_i \in \mathcal{N}$ such that $h_i(Y)$ is measurable w.r.t. $\mathcal{F}$; $\mathcal{F}_i$ is a known sub-$\sigma$-field of $\mathcal{F}$; $\mu_i : \Psi \times \mathcal{Y} \mapsto \mathcal{R}^{m_i}$ is a known function such that $\mu_i(\psi, Y)$ is measurable w.r.t. $\mathcal{F}_i$ and $P_\theta$-a.s. twice differentiable w.r.t. $\psi$ for each $\theta \in \Theta$; $\sigma^2 : \Theta \mapsto (0, \infty)$ is a known or unknown function; $V_i : \Psi \times \mathcal{Y} \mapsto \mathcal{M}_{m_i \times m_i}(\mathcal{R})$ is a known function such that $V_i(\psi, Y)$ is measurable w.r.t. $\mathcal{F}_i$ for each $\psi \in \Psi$; and $\mathcal{F}_i \vee \mathcal{F}_{i'}$ is the smallest $\sigma$-field containing both $\mathcal{F}_i$ and $\mathcal{F}_{i'}$.

For each $i \in \{1, \ldots, n\}$, let $\dot{\mu}_i : \Psi \times \mathcal{Y} \mapsto \mathcal{M}_{m_i \times p}(\mathcal{R})$ be a known function such that for each $\psi \in \Psi$ and $y \in \mathcal{Y}$, $\dot{\mu}_i(\psi, y) = \partial \mu_i(\psi, y)/\partial \psi^T$ if it exists and $0_{m_i \times p}$ otherwise. For each $i \in \{1, \ldots, n\}$, let $V_i^+ : \Psi \times \mathcal{Y} \mapsto \mathcal{M}_{m_i \times m_i}(\mathcal{R})$ be a known function such that $V_i^+(\psi, y)$ is the Moore-Penrose inverse of $V_i(\psi, y)$ for each $\psi \in \Psi$ and $y \in \mathcal{Y}$. See page 26 of Rao (1973) for the definition of the Moore-Penrose inverse. For each $\theta \in \Theta$, assume that $V_i^+(\psi, Y)$ is measurable w.r.t. $\mathcal{F}_i$ and $P_\theta$-a.s. differentiable w.r.t. $\psi$ for each $i \in \{1, \ldots, n\}$ and that $\sum_{i=1}^n E_\theta(\dot{\mu}_i^T(\psi, Y) V_i^+(\psi, Y) \dot{\mu}_i(\psi, Y))$ is non-singular in $\mathcal{M}_{p \times p}(\mathcal{R})$.

Let $\mathcal{G}_1$ denote the class of all estimating functions $G$ in $\mathcal{G}$ for $\psi$ such that for each $\psi \in \Psi$ and $y \in \mathcal{Y}$,

$$G(\psi, y) = \sum_{i=1}^n A_i(\psi, y)\, [h_i(y) - \mu_i(\psi, y)], \tag{6}$$

where $A_i : \Psi \times \mathcal{Y} \mapsto \mathcal{M}_{p \times m_i}(\mathcal{R})$ is a known function such that $A_i(\psi, Y)$ is measurable w.r.t. $\mathcal{F}_i$ and $P_\theta$-a.s. differentiable w.r.t. $\psi$ for each $\theta \in \Theta$. In the paper, $\mathcal{G}_1$ is called the class of all non-orthogonal estimating functions for $\psi$ and any estimating function $G$ in $\mathcal{G}_1$ for $\psi$ is called a non-orthogonal estimating function for $\psi$. There are two important special cases for $\mathcal{G}_1$ as follows:

**Case 1.** When $\mathcal{F}_i = \{\emptyset, \Omega\}$ for all $i \in \{1, \ldots, n\}$, $\mathcal{G}_1$ is called the class of all linear estimating functions for $\psi$ and any estimating function $G$ in $\mathcal{G}_1$ for $\psi$ is called a linear estimating function for $\psi$.

**Case 2.** When $\mathcal{F}_i \subseteq \mathcal{F}_{i+1}$ and $h_i(Y)$ is measurable w.r.t. $\mathcal{F}_{i+1}$ for each $i \in \{1, \ldots, n\}$ with $\mathcal{F}_{n+1} \equiv \mathcal{F}$, $\mathcal{G}_1$ is called the class of all martingale estimating functions for $\psi$ and any estimating function $G$ in $\mathcal{G}_1$ for $\psi$ is called a martingale estimating function for $\psi$.

A necessary and sufficient condition for an estimating function $G$ for $\psi$ satisfying equation (6) to be in $\mathcal{G}_1$ is proposed as follows:

**Lemma 4.** *Let $G$ be an estimating function for $\psi$ such that equation* $(6)$ *holds. Then the following are equivalent*:

(i) *$G$ is in $\mathcal{G}_1$;*

(ii) *both $\sum_{i=1}^n E_\theta(A_i(\psi, Y)\dot{\mu}_i(\psi, Y))$ and $\sum_{i=1}^n E_\theta(A_i(\psi, Y)V_i(\psi, Y)A_i^T(\psi, Y))$ are non-singular in $\mathcal{M}_{p\times p}(\mathcal{R})$ for each $\theta \in \Theta$.*

*Proof.* By equations (3) and (6),

$$E_\theta(G(\psi, Y)) = 0_{p\times 1}$$

and

$$E_\theta\left(\dot{G}(\psi, Y)\right) = -\sum_{i=1}^{n} E_\theta(A_i(\psi, Y)\,\dot{\mu}_i(\psi, Y)) \tag{7}$$

for all $\theta \in \Theta$. By equations (3)-(6),

$$Cov_\theta(G(\psi, Y)) = \sigma^2(\theta)\left[\sum_{i=1}^{n} E_\theta\left(A_i(\psi, Y)\,V_i(\psi, Y)\,A_i^T(\psi, Y)\right)\right] \tag{8}$$

for all $\theta \in \Theta$. Since $\sigma^2(\theta) > 0$ for all $\theta \in \Theta$, this lemma follows from the definitions of $\mathcal{G}$ and $\mathcal{G}_1$. ■

By the definition of $\mathcal{G}_1$, it is easy to see that $G \in \mathcal{G}_1$ if and only if $-G \in \mathcal{G}_1$. As $[G + (-G)]/2\ (= 0_{p\times 1}) \notin \mathcal{G}_1$, $\mathcal{G}_1$ cannot be a convex subclass of $\mathcal{G}$ and thus the convexity assumption proposed by Heyde $(1988)$ fails for $\mathcal{G}_1$. However, it is easy to verify that the assumption given in Lemma 2 holds for $\mathcal{G}_1$. Consequently, it follows from Lemma 2 that the sufficient condition in Lemma 1 is also a necessary condition for an estimating function for $\psi$ to be fixed-sample optimal within $\mathcal{G}_1$.

An important necessary and sufficient condition for an estimating function for $\psi$ to be fixed-sample optimal within $\mathcal{G}_1$ is proposed as follows:

**Lemma 5.** *Let $G^*$ be the estimating function for $\psi$ such that for each $\psi \in \Psi$ and $y \in \mathcal{Y}$,*

$$G^*(\psi, y) = \sum_{i=1}^{n} A_i^*(\psi, y)\,[h_i(y) - \mu_i(\psi, y)], \tag{9}$$

*where* $A_i^* : \Psi \times \mathcal{Y} \mapsto \mathcal{M}_{p\times m_i}(\mathcal{R})$ *is a known function such that* $A_i^*(\psi, Y)$ *is measurable w.r.t.* $\mathcal{F}_i$ *and* $P_\theta$*-a.s. differentiable w.r.t.* $\psi$ *for each* $\theta \in \Theta$*. Then the following are equivalent*:

(i) $G^*$ *is fixed-sample optimal within* $\mathcal{G}_1$;

(ii) *both* $\sum_{i=1}^n E_\theta(A_i^*(\psi,Y)\dot{\mu}_i(\psi,Y))$ *and* $\sum_{i=1}^n E_\theta(A_i^*(\psi,Y)V_i(\psi,Y)A_i^{*T}(\psi,Y))$ *are non-singular in* $\mathcal{M}_{p\times p}(\mathcal{R})$ *for each* $\theta \in \Theta$ *and*

$$\left[\sum_{i=1}^n E_\theta(A_i(\psi,Y)\,\dot{\mu}_i(\psi,Y))\right]^{-1}\left[\sum_{i=1}^n E_\theta\left(A_i(\psi,Y)\,V_i(\psi,Y)\,A_i^{*T}(\psi,Y)\right)\right]$$
$$= \left[\sum_{i=1}^n E_\theta(A_i^*(\psi,Y)\,\dot{\mu}_i(\psi,Y))\right]^{-1}\left[\sum_{i=1}^n E_\theta\left(A_i^*(\psi,Y)\,V_i(\psi,Y)\,A_i^{*T}(\psi,Y)\right)\right]$$

*for all* $G \in \mathcal{G}_1$ *satisfying equation* $(6)$ *and* $\theta \in \Theta$.

*Proof.* By equations (3)-(6) and (9),

$$Cov_\theta\left(G(\psi,Y), G^*(\psi,Y)\right) = \sigma^2(\theta)\left[\sum_{i=1}^n E_\theta\left(A_i(\psi,Y)\,V_i(\psi,Y)\,A_i^{*T}(\psi,Y)\right)\right] \tag{10}$$

for all $G \in \mathcal{G}_1$ satisfying equation (6) and $\theta \in \Theta$. Consequently, this lemma follows from Lemmas 2 and 4 and equations (7), (8), and (10). ■

A commonly used non-orthogonal estimating function for $\psi$ in the literature is given as follows:

**Lemma 6.** *Let* $C : \Psi \mapsto \mathcal{M}_{p\times p}(\mathcal{R})$ *be a known differentiable function such that* $C(\psi)$ *is non-singular for each* $\psi \in \Psi$*. Let* $G_C^*$ *be the estimating function for* $\psi$ *such that for each* $\psi \in \Psi$ *and* $y \in \mathcal{Y}$,

$$G_C^*(\psi,y) = C(\psi)\left\{\sum_{i=1}^n \dot{\mu}_i^T(\psi,y)\,V_i^+(\psi,y)\,[h_i(y) - \mu_i(\psi,y)]\right\}. \tag{11}$$

*Then* $G_C^*$ *is in* $\mathcal{G}_1$.

*Proof.* For each $i \in \{1,\ldots,n\}$, let $A_{C,i}^* : \Psi \times \mathcal{Y} \mapsto \mathcal{M}_{p\times m_i}(\mathcal{R})$ such that for each $\psi \in \Psi$ and $y \in \mathcal{Y}$,

$$A_{C,i}^*(\psi,y) = C(\psi)\,\dot{\mu}_i^T(\psi,y)\,V_i^+(\psi,y). \tag{12}$$

Observe that for each $\theta \in \Theta$,

$$\begin{aligned} &\sum_{i=1}^{n} E_\theta\left(A^*_{C,i}(\psi, Y)\, \dot{\mu}_i(\psi, Y)\right) \\ &\qquad = C(\psi)\left[\sum_{i=1}^{n} E_\theta\left(\dot{\mu}_i^T(\psi, Y)\, V_i^+(\psi, Y)\, \dot{\mu}_i(\psi, Y)\right)\right] \end{aligned} \tag{13}$$

and

$$\begin{aligned} &\sum_{i=1}^{n} E_\theta\left(A^*_{C,i}(\psi, Y)\, V_i(\psi, Y)\, A^{*T}_{C,i}(\psi, Y)\right) \\ &\qquad = C(\psi)\left[\sum_{i=1}^{n} E_\theta\left(\dot{\mu}_i^T(\psi, Y)\, V_i^+(\psi, Y)\, \dot{\mu}_i(\psi, Y)\right)\right] C^T(\psi). \end{aligned} \tag{14}$$

Since $C(\psi)$ and $\sum_{i=1}^{n} E_\theta(\dot{\mu}_i^T(\psi, Y) V_i^+(\psi, Y) \dot{\mu}_i(\psi, Y))$ are non-singular in $\mathcal{M}_{p\times p}$ $(\mathcal{R})$ for all $\theta \in \Theta$, this lemma follows from Lemma 4. ■

A simple and easily verified necessary and sufficient condition for $G^*_C$ given in Lemma 6 to be fixed-sample optimal within $\mathcal{G}_1$ is proposed as follows:

**Theorem 1.** *Let $G^*_C$ be the estimating function for $\psi$ given in Lemma 6. Then the following are equivalent*:

(i) *$G^*_C$ is fixed-sample optimal within $\mathcal{G}_1$*;

(ii) *for each $G \in \mathcal{G}_1$ satisfying equation* (6) *and $\theta \in \Theta$*,

$$\begin{aligned} &\sum_{i=1}^{n} E_\theta\left(A_i(\psi, Y)\, V_i(\psi, Y)\, V_i^+(\psi, Y)\, \dot{\mu}_i(\psi, Y)\right) \\ &\qquad = \sum_{i=1}^{n} E_\theta(A_i(\psi, Y)\, \dot{\mu}_i(\psi, Y)). \end{aligned}$$

*Proof.* For each $i \in \{1, \ldots, n\}$, let $A^*_{C,i} : \Psi \times \mathcal{Y} \mapsto \mathcal{M}_{p\times m_i}(\mathcal{R})$ such that equation (12) holds. Observe that

$$\begin{aligned} &\sum_{i=1}^{n} E_\theta\left(A_i(\psi, Y)\, V_i(\psi, Y)\, A^{*T}_{C,i}(\psi, Y)\right) \\ &\qquad = \left[\sum_{i=1}^{n} E_\theta\left(A_i(\psi, Y)\, V_i(\psi, Y)\, V_i^+(\psi, Y)\, \dot{\mu}_i(\psi, Y)\right)\right] C^T(\psi) \end{aligned} \tag{15}$$

for all $G \in \mathcal{G}_1$ satisfying equation (6) and $\theta \in \Theta$. Conseqently, this theorem follows from Lemma 5 and equations (13)-(15). ■

A very useful sufficient condition for $G_C^*$ given in Lemma 6 to be fixed-sample optimal within $\mathcal{G}_1$ is proposed as follows:

**Corollary 1.** *Let $G_C^*$ be the estimating function for $\psi$ given in Lemma 6. If the column space of $\dot{\mu}_i(\psi, Y)$ is $P_\theta$-a.s. contained in the column space of $V_i(\psi, Y)$ or, equivalently,*

$$V_i(\psi, Y)\, V_i^+(\psi, Y)\, \dot{\mu}_i(\psi, Y) = \dot{\mu}_i(\psi, Y) \quad P_\theta\text{-a.s.} \tag{16}$$

*for each $\theta \in \Theta$ and $i \in \{1, \ldots, n\}$, then $G_C^*$ is fixed-sample optimal within $\mathcal{G}_1$.*

*Proof.* By part (b) of (vi) on page 26 of Rao (1973), these two proposed conditions are equivalent. Consequently, this corollary follows directly from Theorem 1■

Another very useful sufficient condition for $G_C^*$ given in Lemma 6 to be fixed-sample optimal within $\mathcal{G}_1$ is proposed as follows:

**Corollary 2.** *Let $G_C^*$ be the estimating function for $\psi$ given in Lemma 6. If $V_i(\psi, Y)$ is $P_\theta$-a.s. non-singular in $\mathcal{M}_{m_i \times m_i}(\mathcal{R})$ for each $\theta \in \Theta$ and $i \in \{1, \ldots, n\}$, then $G_C^*$ is fixed-sample optimal within $\mathcal{G}_1$.*

*Proof.* This corollary follows directly from Corollary 1. ■

The necessary and sufficient condition given in Theorem 1 can be utilized for characterizing the class of all fixed-sample optimal estimating functions for $\psi$ within $\mathcal{G}_1$ as follows:

**Theorem 2.** *Suppose that the necessary and sufficient condition given in Theorem 1 holds. Then the following are equivalent*:

(i) *$G^*$ is a fixed-sample optimal estimating function for $\psi$ within $\mathcal{G}_1$;*

(ii) *$G^*$ is in $\mathcal{G}_1$ and for each $\theta \in \Theta$,*

$$G^{*(s)}(\theta, Y) = \sigma^{-2}(\theta) \left\{ \textstyle\sum_{i=1}^{n} \dot{\mu}_i^T(\psi, Y)\, V_i^+(\psi, Y)\, [h_i(Y) - \mu_i(\psi, Y)] \right\} \quad P_\theta\text{-a.s.};$$

(iii) *$G^*$ is in $\mathcal{G}_1$ and for each $\theta \in \Theta$,*

$$I_{G^*}(\theta) = \sigma^{-2}(\theta) \left[ \sum_{i=1}^{n} E_\theta \left( \dot{\mu}_i^T(\psi, Y)\, V_i^+(\psi, Y)\, \dot{\mu}_i(\psi, Y) \right) \right].$$

*Proof.* By Theorem 1, $G_C^*$ is fixed-sample optimal within $\mathcal{G}_1$. By equations (1), (7), (8), and (11)-(14),

$$G_C^{*(s)}(\theta, Y) = \sigma^{-2}(\theta)\left\{\sum_{i=1}^{n} \dot{\mu}_i^T(\psi, Y)\, V_i^+(\psi, Y)\, [h_i(Y) - \mu_i(\psi, Y)]\right\}$$

for all $\theta \in \Theta$. By equations (2)-(5),

$$I_{G_C^*}(\theta) = \sigma^{-2}(\theta)\left[\sum_{i=1}^{n} E_\theta\left(\dot{\mu}_i^T(\psi, Y)\, V_i^+(\psi, Y)\, \dot{\mu}_i(\psi, Y)\right)\right]$$

for all $\theta \in \Theta$. Consequently, this theorem follows from Lemma 3. ∎

## 3. A Counterexample

In this section, a simple counterexample without any fixed-sample optimal estimating function for $\psi$ within $\mathcal{G}_1$ is constructed to show that part (ii) in Theorem 1 may fail and $G_C^*$ given in Lemma 6 is not necessarily a fixed-sample optimal estimating function for $\psi$ within $\mathcal{G}_1$.

**Counterexample 1.** Let $Y$ ($\equiv (Y_1, Y_2)^T$) be the response of experiment $\mathcal{E}$ such that $Y$ is distributed as $N(\mu(\psi), \lambda V(\psi))$ for some unknown $\theta \in \Theta$, where $\theta$ ($\equiv (\psi, \lambda)$) is the parameter with parameter space $\Theta$ ($\equiv \Psi \times \Lambda$); $\psi$ is the parameter of interest in a known non-empty open subset $\Psi$ of $\mathcal{R}$; $\lambda$ is a nuisance parameter in a known or unknown non-empty subset $\Lambda$ of $(0, \infty)$; $\mu : \Psi \mapsto \mathcal{R}^2$ is a known function such that $\mu(\psi) = (\psi, \psi)^T$ for all $\psi \in \Psi$; $V : \Psi \mapsto \mathcal{M}_{2\times 2}(\mathcal{R})$ is a known function such that

$$V(\psi) = \begin{pmatrix} \cos^2(\psi) & \cos(\psi)\sin(\psi) \\ \cos(\psi)\sin(\psi) & \sin^2(\psi) \end{pmatrix} = \begin{pmatrix} \cos(\psi) \\ \sin(\psi) \end{pmatrix} (\cos(\psi), \sin(\psi))$$

for all $\psi \in \Psi$; and $N(\mu(\psi), \lambda V(\psi))$ is the bivariate normal distribution with mean vector $\mu(\psi)$ and covariance matrix $\lambda V(\psi)$. Set $n \equiv 1$, $h_1(Y) \equiv Y$, and $\mathcal{F}_1 \equiv \{\emptyset, \Omega\}$.

By equation (3), $\mu_1(\psi, Y) = \mu(\psi)$ for all $\psi \in \Psi$, which implies that $\dot{\mu}_1(\psi, Y) = (1, 1)^T$ for all $\psi \in \Psi$. By equation (4), $\sigma^2(\theta) = \lambda > 0$ and $V_1(\psi, Y) = V(\psi)$ for all $\psi \in \Psi$, which implies that $V_1^+(\psi, Y) = V(\psi)$ for all $\psi \in \Psi$. Since

$$E_\theta\left(\dot{\mu}_1^T(\psi, Y)\, V_1^+(\psi, Y)\, \dot{\mu}_1(\psi, Y)\right) = [\cos(\psi) + \sin(\psi)]^2$$

for all $\theta \in \Theta$, $E_\theta(\dot{\mu}_1^T(\psi, Y) V_1^+(\psi, Y) \dot{\mu}_1(\psi, Y)) > 0$ for all $\theta \in \Theta$ if and only if $\Psi \subseteq \mathcal{R} \backslash \{-\pi/4 + k\pi : k \in \mathcal{Z}\}$, where $\mathcal{Z}$ denotes the set of all integers.

In the following, assume that $\Psi$ is a known non-empty open subset of $\mathcal{R}\backslash\{-\pi/4+k\pi : k \in \mathcal{Z}\}$. Let $C : \Psi \mapsto \mathcal{R}\backslash\{0\}$ be a known differentiable function. Let $G_C^*$ be the estimating function for $\psi$ such that equations (11) and (12) hold. Then

$$\begin{aligned} G_C^*(\psi, Y) &= A_{C,1}^*(\psi, Y)\,[h_1(Y) - \mu_1(\psi, Y)] \\ &= C(\psi)\,\mu_1^T(\psi, Y)\,V_1^+(\psi, Y)\,[Y - \mu(\psi)] \\ &= C(\psi)\,[\cos(\psi) + \sin(\psi)]\,[\cos(\psi)\,(Y_1 - \psi) + \sin(\psi)\,(Y_2 - \psi)] \end{aligned}$$

for all $\psi \in \Psi$. By Lemma 6, $G_C^*$ is in $\mathcal{G}_1$.

Let $G^{(1)}$ be the estimating function for $\psi$ such that for each $\psi \in \Psi$ and $y \in \mathcal{Y}$,

$$G^{(1)}(\psi, y) \equiv A_1^{(1)}(\psi, y)\,[h_1(y) - \mu_1(\psi, y)] \equiv (1, 1)\,[y - \mu(\psi)].$$

Since

$$E_\theta\left(A_1^{(1)}(\psi, Y)\,\dot{\mu}_1(\psi, Y)\right) = 2 > 0$$

and

$$E_\theta\left(A_1^{(1)}(\psi, Y)\,V_1(\psi, Y)\,A_1^{(1)T}(\psi, Y)\right) = [\cos(\psi) + \sin(\psi)]^2 > 0$$

for all $\theta \in \Theta$, it follows from Lemma 4 that $G^{(1)}$ is in $\mathcal{G}_1$.

Observe that

$$\begin{aligned} &E_\theta\left(A_1^{(1)}(\psi, Y)\,V_1(\psi, Y)\,V_1^+(\psi, Y)\,\dot{\mu}_1(\psi, Y)\right) \\ &\quad = [\cos(\psi) + \sin(\psi)]^2 \\ &\quad \neq 2 \\ &\quad = E_\theta\left(A_1^{(1)}(\psi, Y)\,\dot{\mu}_1(\psi, Y)\right) \end{aligned}$$

for all $\theta \in (\Psi\backslash\{\pi/4 + k\pi : k \in \mathcal{Z}\}) \times \Lambda$. Since $(\Psi\backslash\{\pi/4 + k\pi : k \in \mathcal{Z}\}) \times \Lambda$ is a non-empty subset of $\Theta$, part (ii) in Theorem 1 fails. Consequently, it follows from Theorem 1 that $G_C^*$ is not fixed-sample optimal within $\mathcal{G}_1$ for Counterexample 1.

Moreover, there does not exist any fixed-sample optimal estimating function for $\psi$ within $\mathcal{G}_1$ for Counterexample 1, which can be shown by a contradictory argument as follows:

Suppose that there exists a fixed-sample optimal estimating function $G^*$ for $\psi$ within $\mathcal{G}_1$ such that for each $\psi \in \Psi$ and $y \in \mathcal{Y}$,

$$G^*(\psi, y) \equiv A_1^*(\psi, y)\,[h_1(y) - \mu_1(\psi, y)] \equiv (A_{11}^*(\psi), A_{12}^*(\psi))\,[y - \mu(\psi)].$$

Let $G^{(2)}$ be the estimating function for $\psi$ such that for each $\psi \in \Psi$ and $y \in \mathcal{Y}$,

$$G^{(2)}(\psi, y) \equiv A_1^{(2)}(\psi, y)\,[h_1(y) - \mu_1(\psi, y)] \equiv (\cos(\psi), \sin(\psi))\,[y - \mu(\psi)].$$

Since

$$E_\theta\left(A_1^{(2)}(\psi,Y)\,\dot{\mu}_1(\psi,Y)\right)=\cos(\psi)+\sin(\psi)\neq 0$$

and

$$E_\theta\left(A_1^{(2)}(\psi,Y)\,V_1(\psi,Y)\,A_1^{(2)T}(\psi,Y)\right)=1>0$$

for all $\theta\in\Theta$, it follows from Lemma 4 that $G^{(2)}$ is in $\mathcal{G}_1$. By Lemma 5, $A_{11}^*(\psi)\cos(\psi)+A_{12}^*(\psi)\sin(\psi)\neq 0$ for all $\psi\in\Psi$. Observe that

$$\begin{aligned}
&\left[E_\theta\left(A_1^{(1)}(\psi,Y)\,\dot{\mu}_1(\psi,Y)\right)\right]^{-1}\left[E_\theta\left(A_1^{(1)}(\psi,Y)\,V_1(\psi,Y)\,A_1^{*T}(\psi,Y)\right)\right]\\
&=\frac{[\cos(\psi)+\sin(\psi)]\,[A_{11}^*(\psi)\,\cos(\psi)+A_{12}^*(\psi)\,\sin(\psi)]}{2}\\
&\neq\frac{A_{11}^*(\psi)\,\cos(\psi)+A_{12}^*(\psi)\,\sin(\psi)}{\cos(\psi)+\sin(\psi)}\\
&=\left[E_\theta\left(A_1^{(2)}(\psi,Y)\,\dot{\mu}_1(\psi,Y)\right)\right]^{-1}\left[E_\theta\left(A_1^{(2)}(\psi,Y)\,V_1(\psi,Y)\,A_1^{*T}(\psi,Y)\right)\right]
\end{aligned}$$

for all $\theta\in(\Psi\backslash\{\pi/4+k\pi:k\in\mathcal{Z}\})\times\Lambda$. Since both $G^{(1)}$ and $G^{(2)}$ are in $\mathcal{G}_1$ and $(\Psi\backslash\{\pi/4+k\pi:k\in\mathcal{Z}\})\times\Lambda$ is a non-empty subset of $\Theta$, it follows from Lemma 5 that $G^*$ is not fixed-sample optimal within $\mathcal{G}_1$, which is a contradiction. Consequently, there does not exist any fixed-sample optimal estimating function for $\psi$ within $\mathcal{G}_1$ for Counterexample 1. ∎

In the literature, many authors claimed that $G_C^*$ given in Lemma 6 is fixed-sample optimal within $\mathcal{G}_1$ without mentioning any further condition. For example, see page 32 of Heyde (1997) when $\mathcal{G}_1$ is the class of all martingale estimating functions for $\psi$. However, by Counterexample 1, part (ii) in Theorem 1 may fail and thus $G_C^*$ is not necessarily fixed-sample optimal within $\mathcal{G}_1$. Moreover, by Counterexample 1, there does not necessarily exist any fixed-sample optimal estimating function for $\psi$ within $\mathcal{G}_1$. Consequently, part (ii) in Theorem 1 should be verified or assumed before we claim that $G_C^*$ is fixed-sample optimal within $\mathcal{G}_1$.

## 4. Applications

In this section, the usefulness and applicability of the method proposed in Section 2 are illustrated by two classical examples with many nuisance parameters.

**Example 1.** *Neyman-Scott problem* (1948). First of all, consider a special case as follows: Assume that $Y$ ($\equiv(Y_1^T,\ldots,Y_n^T)^T$) is the response of experiment $\mathcal{E}$ for some $n\in\mathcal{N}$ such that

(1) $Y_1, \ldots, Y_n$ are independent random vectors and

(2) for each $i \in \{1, \ldots, n\}$, $Y_i$ ($\equiv (Y_{i0}, Y_{i1}, \ldots, Y_{im_i})^T$) is distributed as the $(m_i+1)$-variate normal distribution with mean vector $\lambda_i 1_{(m_i+1)\times 1}$ and covariance matrix $\psi I_{m_i+1}$ for some $m_i \in \mathcal{N}$,

where $\psi$ is the parameter of interest in $(0,\infty)$ ($\equiv \Psi$), $(\lambda_1, \ldots, \lambda_n)^T$ ($\equiv \tilde{\lambda}$) is a nuisance parameter in $\mathcal{R}^n$, $(\psi, \tilde{\lambda})$ is the parameter with parameter space $\Psi \times \mathcal{R}^n$, and $I_{m_i+1}$ denotes the identity matrix of order $m_i + 1$.

Note that the maximum likelihood estimator (MLE) $\hat{\psi}_{ML}$ of $\psi$ is given by

$$
\begin{aligned}
\hat{\psi}_{ML} &= \frac{1}{n+\sum_{i=1}^n m_i} \sum_{i=1}^n \sum_{j=0}^{m_i} \left(Y_{ij} - \bar{Y}_{i\cdot}\right)^2 \\
&= \frac{1}{n+\sum_{i=1}^n m_i} \sum_{i=1}^n \left[ \sum_{j=0}^{m_i} Y_{ij}^2 - (m_i+1)\,\bar{Y}_{i\cdot}^2 \right],
\end{aligned} \tag{17}
$$

where $\bar{Y}_{i\cdot} \equiv \sum_{j=0}^{m_i} Y_{ij}/(m_i+1)$. Then $\hat{\psi}_{ML}$ is distributed as $\psi\chi^2_{\sum_{i=1}^n m_i}/(n + \sum_{i=1}^n m_i)$ with mean $\psi \sum_{i=1}^n m_i/(n+\sum_{i=1}^n m_i)$ and variance $2\psi^2 \sum_{i=1}^n m_i/(n+\sum_{i=1}^n m_i)^2 \to 0$ as $n \to \infty$, where $\chi^2_{\sum_{i=1}^n m_i}$ denotes the chi-squared distribution with $\sum_{i=1}^n m_i$ degrees of freedom. It follows from the central limit theorem (CLT) that

$$
\frac{(\sum_{i=1}^n m_i)^{1/2}}{\sqrt{2}\,\psi} \left( \frac{n+\sum_{i=1}^n m_i}{\sum_{i=1}^n m_i} \hat{\psi}_{ML} - \psi \right) \xrightarrow{d} N(0,1)
$$

as $n \to \infty$, where $N(0,1)$ denotes the standard normal distribution. Moreover, it follows from the strong law of large numbers (SLLN) that

$$
\frac{n+\sum_{i=1}^n m_i}{\sum_{i=1}^n m_i}\, \hat{\psi}_{ML} \to \psi \quad P_\theta\text{-a.s.}
$$

as $n \to \infty$. Consequently, the following are equivalent:

(i) $\sum_{i=1}^n m_i/n \to \infty$ *as* $n \to \infty$;

(ii) $\hat{\psi}_{ML}$ *is a weakly consistent estimator of* $\psi$ *as* $n \to \infty$;

(iii) $\hat{\psi}_{ML}$ *is a strongly consistent estimator of* $\psi$ *as* $n \to \infty$;

(iv) $\hat{\psi}_{ML}$ *is a mean-square-error (MSE) consistent estimator of* $\psi$ *as* $n \to \infty$, *i.e.*, $E_\theta((\hat{\psi}_{ML} - \psi)^2) \to 0$ *as* $n \to \infty$;

(v) $\hat{\psi}_{ML}$ *is an asymptotically unbiased estimator of* $\psi$ *as* $n \to \infty$, *i.e.*, $E_\theta(\hat{\psi}_{ML} - \psi) \to 0$ *as* $n \to \infty$.

In practice, $\{m_i\}_{i=1}^{\infty}$ is a bounded sequence, which implies that part (i) fails. When part (i) fails, $\hat{\psi}_{ML}$ is not a weakly consistent estimator of $\psi$ as $n \to \infty$ and thus should not be used for estimation of $\psi$. However, $(n+\sum_{i=1}^{n} m_i)\hat{\psi}_{ML}/\sum_{i=1}^{n} m_i$ is always an unbiased and weakly, MSE, and strongly consistent estimator of $\psi$ as $n \to \infty$. Consequently, $(n+\sum_{i=1}^{n} m_i)\hat{\psi}_{ML}/\sum_{i=1}^{n} m_i$ could be used for good estimation of $\psi$.

Next, consider a more general case as follows: Assume that $Y$ ($\equiv (Y_1^T, \ldots, Y_n^T)^T$) is the response of experiment $\mathcal{E}$ for some $n \in \mathcal{N}$ such that for each $i \in \{1, \ldots, n\}$, $Y_i$ ($\equiv (Y_{i0}, Y_{i1}, \ldots, Y_{im_i})^T$) has mean vector $\lambda_i 1_{(m_i+1)\times 1}$ and covariance matrix $\psi I_{m_i+1}$ for some $m_i \in \mathcal{N}$, where $\psi$ is the parameter of interest in $\Psi$ ($\equiv (0,\infty)$), $(\lambda_0, \lambda_1, \ldots, \lambda_n)$ ($\equiv \lambda$) is a nuisance parameter in $\Lambda_0 \times \mathcal{R}^n$ ($\equiv \Lambda$) for some known or unknown non-empty set $\Lambda_0$, and $(\psi, \lambda)$ ($\equiv \theta$) is the parameter with parameter space $\Psi \times \Lambda$ ($\equiv \Theta$). For example, $\Lambda_0$ is a set consisting of only one element for the special case of Example 1.

For each $i \in \{1, \ldots, n\}$, set $\bar{Y}_{i\cdot} \equiv \sum_{j=0}^{m_i} Y_{ij}/(m_i+1)$ and $h_i(Y) \equiv \sum_{j=0}^{m_i}(Y_{ij} - \bar{Y}_{i\cdot})^2$. For each $\theta \in \Theta$, $i \neq i'$, and $i, i' \in \{1, \ldots, n\}$, assume that

$$Var_\theta(h_i(Y)) = m_i\,\sigma^2(\theta) \tag{18}$$

and

$$Cov_\theta(h_i(Y), h_{i'}(Y)) = 0, \tag{19}$$

where $\sigma^2 : \Theta \mapsto (0,\infty)$ is a known or unknown function. For example, $\sigma^2 : \Theta \mapsto (0,\infty)$ is a known function such that $\sigma^2(\theta) = 2\psi^2$ for all $\theta \in \Theta$ for the special case of Example 1.

Assume that one of the following conditions holds:

(i) $\mathcal{F}_i = \{\emptyset, \Omega\}$ *for all* $i \in \{1, \ldots, n\}$.

(ii) $\mathcal{F}_i$ *is the* $\sigma$*-field generated by* $\bar{Y}_{i\cdot}$ *for each* $i \in \{1, \ldots, n\}$.

(iii) $\mathcal{F}_i$ *is the* $\sigma$*-field generated by* $\bar{Y}_{1\cdot}, \ldots, \bar{Y}_{n\cdot}$ *for each* $i \in \{1, \ldots, n\}$.

(iv) $\mathcal{F}_1 = \{\emptyset, \Omega\}$ *and* $\mathcal{F}_i$ *is the* $\sigma$*-field generated by* $Y_1, \ldots, Y_{i-1}$ *for each* $i \in \{2, \ldots, n\}$.

(v) $\mathcal{F}_1$ *is the* $\sigma$*-field generated by* $\bar{Y}_{1\cdot}$ *and* $\mathcal{F}_i$ *is the* $\sigma$*-field generated by* $Y_1, \ldots, Y_{i-1}, \bar{Y}_{i\cdot}$ *for each* $i \in \{2, \ldots, n\}$.

For each $\theta \in \Theta$ and $i, i' \in \{1, \ldots, n\}$, assume that

$$E_\theta(h_i(Y)|\mathcal{F}_i) = E_\theta(h_i(Y)) \quad P_\theta\text{-a.s.} \tag{20}$$

and

$$(21)\quad \begin{aligned} &E_\theta([h_i(Y) - E_\theta(h_i(Y))]\,[h_{i'}(Y) - E_\theta(h_{i'}(Y))]|\mathcal{F}_i \vee \mathcal{F}_{i'}) \\ &= E_\theta([h_i(Y) - E_\theta(h_i(Y))]\,[h_{i'}(Y) - E_\theta(h_{i'}(Y))]) \quad P_\theta\text{-a.s.} \end{aligned}$$

For example, equations (20) and (21) hold under any one of Conditions 1-5 for the special case of Example 1.

For each $i \in \{1, \ldots, n\}$, let $\mu_i, V_i : \Psi \times \mathcal{Y} \mapsto \mathcal{R}$ be known functions such that $\mu_i(\psi, y) = m_i\psi$ and $V_i(\psi, y) = m_i$ for all $\psi \in \Psi$ and $y \in \mathcal{Y}$. Then $\dot{\mu}_i(\psi, Y) = m_i$ and $V_i^+(\psi, Y) = 1/m_i$ for all $\psi \in \Psi$ and $i \in \{1, \ldots, n\}$. By equations (18)-(21), it is easy to verify that equations (3)-(5) hold.

Let $C : \Psi \mapsto \mathcal{R}\backslash\{0\}$ be a known differentiable function and let $G_C^*$ be the estimating function for $\psi$ given in Lemma 6. Then

$$G_C^*(\psi, Y) = C(\psi)\,[\textstyle\sum_{i=1}^n\, h_i(Y) - (\sum_{i=1}^n\, m_i)\,\psi]$$

for all $\psi \in \Psi$. By Corollary 2, $G_C^*$ is fixed-sample optimal within $\mathcal{G}_1$. Moreover, by Theorem 2, the following are equivalent:

(i) *$G^*$ is a fixed-sample optimal estimating function for $\psi$ within $\mathcal{G}_1$;*

(ii) *$G^*$ is in $\mathcal{G}_1$ and for each $\theta \in \Theta$,*

$$G^{*(s)}(\theta, Y) = \frac{\sum_{i=1}^n\, h_i(Y) - (\sum_{i=1}^n\, m_i)\,\psi}{\sigma^2(\theta)} \quad P_\theta\text{-a.s.};$$

(iii) *$G^*$ is in $\mathcal{G}_1$ and for each $\theta \in \Theta$,*

$$I_{G^*}(\theta) = \frac{\sum_{i=1}^n\, m_i}{\sigma^2(\theta)}.$$

Set

$$\hat{\psi}_{FSO} \equiv \frac{\sum_{i=1}^n\, h_i(Y)}{\sum_{i=1}^n\, m_i}.$$

Then $\hat{\psi}_{FSO} = (n+\sum_{i=1}^n m_i)\hat{\psi}_{ML}/\sum_{i=1}^n m_i$, where $\hat{\psi}_{ML}$ is given in equation (17). Since $G^*(\psi, Y)|_{\psi=\hat{\psi}_{FSO}} = 0$ $P_\theta$-a.s. for each fixed-sample optimal estimating function $G^*$ for $\psi$ within $\mathcal{G}_1$, $\hat{\psi}_{FSO}$ is called the fixed-sample optimal estimator (FSOE) of $\psi$ in the paper. Note that $\hat{\psi}_{FSO}$ is an unbiased estimator of $\psi$ and

$$Var_\theta\left(\hat{\psi}_{FSO}\right) = \frac{\sigma^2(\theta)}{\sum_{i=1}^n m_i} \to 0$$

as $n \to \infty$. Thus, $\hat{\psi}_{FSO}$ is an MSE consistent estimator of $\psi$ as $n \to \infty$, which implies that $\hat{\psi}_{FSO}$ is also a weakly consistent estimator of $\psi$ as $n \to \infty$. Consequently, $\hat{\psi}_{FSO}$ could be used for good estimation of $\psi$. ∎

**Example 2.** First of all, consider a special case as follows: Assume that $Y$ $(\equiv (Y_1^T, \ldots, Y_n^T)^T)$ is the response of experiment $\mathcal{E}$ for some $n \in \mathcal{N}$ such that

(1) $Y_1, \ldots, Y_n$ are independent random vectors and

(2) for each $i \in \{1, \ldots, n\}$, $Y_i$ $(\equiv (Y_{i0}, Y_{i1}, \ldots, Y_{ip})^T)$ is distributed as the $(p+1)$-variate normal distribution with mean vector $\lambda_i 1_{(p+1)\times 1} + (0, \psi_1, \ldots, \psi_p)^T$ and covariance matrix $\lambda_0 I_{p+1}$ for some $p \in \mathcal{N}$,

where $\psi$ is the parameter of interest in $\mathcal{R}^p$ $(\equiv \Psi)$, $(\lambda_0, \lambda_1, \ldots, \lambda_n)^T$ $(\equiv \tilde{\lambda})$ is a nuisance parameter in $(0,\infty)\times\mathcal{R}^n$, $(\psi, \tilde{\lambda})$ is the parameter with parameter space $\Psi \times (0,\infty) \times \mathcal{R}^n$, and $I_{p+1}$ denotes the identity matrix of order $p+1$.

Note that the MLE $\hat{\psi}_{ML}$ of $\psi$ is given by

$$\hat{\psi}_{ML} = \left(\bar{Y}_{\cdot 1} - \bar{Y}_{\cdot 0}, \ldots, \bar{Y}_{\cdot p} - \bar{Y}_{\cdot 0}\right)^T, \tag{22}$$

where $\bar{Y}_{\cdot j} \equiv \sum_{i=1}^n Y_{ij}/n$ for each $j \in \{0, 1, \ldots, p\}$. Then $\hat{\psi}_{ML}$ is distributed as the $p$-variate normal distribution with mean vector $\psi$ and covariance matrix $\lambda_0(I_p + 1_{p\times 1}1_{p\times 1}^T)/n \to 0_{p\times p}$ as $n \to \infty$, where $I_p$ denotes the identity matrix of order $p$. Thus, $\hat{\psi}_{ML}$ is an MSE consistent estimator of $\psi$ as $n \to \infty$, i.e., $E_\theta((\hat{\psi}_{ML} - \psi)^T(\hat{\psi}_{ML} - \psi)) \to 0$ as $n \to \infty$. It follows from the CLT that

$$\sqrt{\frac{n}{\lambda_0}}\left(\hat{\psi}_{ML} - \psi\right) \xrightarrow{d} N\left(0_{p\times 1}, I_p + 1_{p\times 1}1_{p\times 1}^T\right)$$

as $n \to \infty$, where $N(0_{p\times 1}, I_p + 1_{p\times 1}1_{p\times 1}^T)$ denotes the $p$-variate normal distribution with mean vector $0_{p\times 1}$ and covariance matrix $I_p + 1_{p\times 1}1_{p\times 1}^T$. Moreover, it follows from the SLLN that $\hat{\psi}_{ML}$ is a strongly consistent estimator of $\psi$ as $n \to \infty$, which implies that it is also a weakly consistent estimator of $\psi$ as $n \to \infty$. Consequently, $\hat{\psi}_{ML}$ could be used for good estimation of $\psi$.

Next, consider a more general case as follows: Assume that $Y(\equiv(Y_1^T, \ldots, Y_n^T)^T)$ is the response of experiment $\mathcal{E}$ for some $n \in \mathcal{N}$ such that

(i) *for each* $i \in \{1, \ldots, n\}$, $Y_i$ *(*$\equiv (Y_{i0}, Y_{i1}, \ldots, Y_{ip})^T$*) has mean vector* $\lambda_i 1_{(p+1)\times 1} + (0, \psi_1, \ldots, \psi_p)^T$ *and covariance matrix* $\sigma^2(\theta) I_{p+1}$ *for some* $p \in \mathcal{N}$;

(ii) *for each* $i \neq i'$ *and* $i, i' \in \{1, \ldots, n\}$, $Cov_\theta(Y_i, Y_{i'}) = 0_{(p+1)\times(p+1)}$,

where $(\psi_1, \ldots, \psi_p)$ $(\equiv \psi)$ is the parameter of interest in $\mathcal{R}^p$ $(\equiv \Psi)$, $(\lambda_0, \lambda_1, \ldots, \lambda_n)$ $(\equiv \lambda)$ is a nuisance parameter in $\Lambda_0 \times \mathcal{R}^n$ $(\equiv \Lambda)$ for some known or unknown non-empty set $\Lambda_0$, $(\psi, \lambda)$ $(\equiv \theta)$ is the parameter with parameter space $\Psi \times \Lambda$ $(\equiv \Theta)$, and $\sigma^2 : \Theta \mapsto (0, \infty)$ is a known or unknown function. For example, $\Lambda_0 = (0, \infty)$ and $\sigma^2 : \Theta \mapsto (0, \infty)$ is a known function such that $\sigma^2(\theta) = \lambda_0$ for all $\theta \in \Theta$ for the special case of Example 2.

For each $i \in \{1, \ldots, n\}$, set $h_i(Y) \equiv (Y_{i1} - Y_{i0}, \ldots, Y_{ip} - Y_{i0})^T$. Assume that one of conditions (i)-(v) given in Example 1 holds. For each $\theta \in \Theta$ and $i, i' \in \{1, \ldots, n\}$, assume that

$$E_\theta(h_i(Y)|\mathcal{F}_i) = E_\theta(h_i(Y)) \quad P_\theta\text{-a.s.} \tag{23}$$

and

$$\begin{aligned} &E_\theta\left([h_i(Y) - E_\theta(h_i(Y))]\,[h_{i'}(Y) - E_\theta(h_{i'}(Y))]^T \,\middle|\, \mathcal{F}_i \vee \mathcal{F}_{i'}\right) \\ &\quad = E_\theta\left([h_i(Y) - E_\theta(h_i(Y))]\,[h_{i'}(Y) - E_\theta(h_{i'}(Y))]^T\right) \quad P_\theta\text{-a.s.} \end{aligned} \tag{24}$$

For example, equations (23) and (24) hold under any one of Conditions 1-5 for the special case of Example 2.

For each $i \in \{1, \ldots, n\}$, let $\mu_i : \Psi \times \mathcal{Y} \mapsto \mathcal{R}^p$ and $V_i : \Psi \times \mathcal{Y} \mapsto \mathcal{M}_{p\times p}(\mathcal{R})$ be known functions such that $\mu_i(\psi, y) = \psi$ and $V_i(\psi, y) = I_p + 1_{p\times 1}1_{p\times 1}^T$ for all $\psi \in \Psi$ and $y \in \mathcal{Y}$. Then $\dot{\mu}_i(\psi, Y) = I_p$ and $V_i^+(\psi, Y) = I_p - 1_{p\times 1}1_{p\times 1}^T/(p+1)$ for all $\psi \in \Psi$ and $i \in \{1, \ldots, n\}$. By equations (23) and (24), it is easy to verify that equations (3)-(5) hold.

Let $C : \Psi \mapsto \mathcal{M}_{p\times p}(\mathcal{R})$ be a known differentiable function such that $C(\psi)$ is non-singular for each $\psi \in \Psi$. Let $G_C^*$ be the estimating function for $\psi$ given in Lemma 6. Then

$$\begin{aligned} G_C^*(\psi, Y) = n\, C(\psi) \left(I_p - \frac{1}{p+1}\, 1_{p\times 1}1_{p\times 1}^T\right) \\ \left(\bar{Y}_{\cdot 1} - \bar{Y}_{\cdot 0} - \psi_1, \ldots, \bar{Y}_{\cdot p} - \bar{Y}_{\cdot 0} - \psi_p\right)^T \end{aligned}$$

for all $\psi \in \Psi$, where $\bar{Y}_{\cdot j} \equiv \sum_{i=1}^n Y_{ij}/n$ for each $j \in \{0, 1, \ldots, p\}$. By Corollary 2, $G_C^*$ is fixed-sample optimal within $\mathcal{G}_1$. Moreover, by Theorem 2, the following are equivalent:

(i) *$G^*$ is a fixed-sample optimal estimating function for $\psi$ within $\mathcal{G}_1$;*

(ii) *$G^*$ is in $\mathcal{G}_1$ and for each $\theta \in \Theta$,*

$$\begin{aligned} &G^{*(s)}(\theta, Y) \\ &= n\,\sigma^{-2}(\theta)\left(I_p - \tfrac{1}{p+1}\, 1_{p\times 1}1_{p\times 1}^T\right)\left(\bar{Y}_{\cdot 1} - \bar{Y}_{\cdot 0} - \psi_1, \ldots, \bar{Y}_{\cdot p} - \bar{Y}_{\cdot 0} - \psi_p\right)^T \\ &\quad P_\theta\text{-a.s.}; \end{aligned}$$

(iii) *$G^*$ is in $\mathcal{G}_1$ and for each $\theta \in \Theta$,*

$$I_{G^*}(\theta) = n\,\sigma^{-2}(\theta)\left(I_p - \tfrac{1}{p+1}\, 1_{p\times 1}1_{p\times 1}^T\right).$$

Set

$$\hat{\psi}_{FSO} \equiv \left(\bar{Y}_{\cdot 1} - \bar{Y}_{\cdot 0}, \ldots, \bar{Y}_{\cdot p} - \bar{Y}_{\cdot 0}\right)^T .$$

Then $\hat{\psi}_{FSO} = \hat{\psi}_{ML}$, where $\hat{\psi}_{ML}$ is given in equation (22). Since $G^*(\psi, Y)|_{\psi=\hat{\psi}_{FSO}}$ = $0_{p\times 1}$ $P_\theta$-a.s. for each fixed-sample optimal estimating function $G^*$ for $\psi$ within $\mathcal{G}_1$, $\hat{\psi}_{FSO}$ is the FSOE of $\psi$. Note that $\hat{\psi}_{FSO}$ is an unbiased estimator of $\psi$ and

$$Cov_\theta\left(\hat{\psi}_{FSO}\right) = \frac{\sigma^2(\theta)}{n}\left(I_p + 1_{p\times 1}1_{p\times 1}^T\right) \to 0_{p\times p}$$

as $n \to \infty$. Thus, $\hat{\psi}_{FSO}$ is an MSE consistent estimator of $\psi$ as $n \to \infty$, which implies that $\hat{\psi}_{FSO}$ is also a weakly consistent estimator of $\psi$ as $n \to \infty$. Consequently, $\hat{\psi}_{FSO}$ could be used for good estimation of $\psi$. ■

## Acknowledgements

The authors would like to thank the referees for many valuable comments and suggestions on revising the paper.

## References

1. O. E. Barndorff-Nielsen and D. R. Cox, *Inference and Asymptotics*, Chapman & Hall, London, 1994.
2. D. R. Cox and D. V. Hinkley, *Theoretical Statistics*, Chapman & Hall, London, 1974.
3. T. M. Durairajan, Optimal estimating function for non-orthogonal model, *J. Statist. Plann. Inference* **33** (1992), 381-384.
4. V. P. Godambe and C. C. Heyde, Quasi-likelihood and optimal estimation, *Int. Statist. Rev.* **55** (1987), 231-244.
5. C. C. Heyde, Fixed sample and asymptotic optimality for classes of estimating functions, *Contemp. Math.* **80** (1988), 241-247.
6. C. C. Heyde, *Quasi-Likelihood And Its Application*: *A General Approach to Optimal Parameter Estimation*, Springer-Verlag, New York, 1997.
7. P. McCullagh and J. A. Nelder, *Generalized Linear Models*, 2nd ed., Chapman & Hall, London, 1989.
8. J. Neyman and E. L. Scott, Consistent estimates based on partially consistent observations. *Econometrica* **16** (1948), 1-32.
9. L. Pace and A. Salvan, *Principles of Statistical Inference: from a Neo-Fisherian Perspective*, World Scientific, Singapore, 1997.
10. B. L. S. Prakasa Rao, *Semimartingales and their Statistical Inference*, Chapman & Hall/CRC, Boca Raton, 1999.

11. C. R. Rao, *Linear Statistical Inference and Its Applications*, 2nd ed., John Wiley & Sons, New York, 1973.

12. R. W. M. Wedderburn, Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method, *Biometrika* **61** (1974), 439-447.

Chih-Rung Chen
Institute of Statistics,
National Chiao Tung University,
Hsinchu 30050,
Taiwan
E-mail: cchen@stat.nctu.edu.tw

Lih-Chung Wang
Department of Applied Mathematics,
National Dong Hwa University,
Hualien 974,
Taiwan
E-mail: lcwang@mail.ndhu.edu.tw