REPARAMETRIZATION OF THE LEAST FAVORABLE SUBMODEL IN SEMI-PARAMETRIC MULTI-SAMPLE MODELS

YUICHI HIROSE* **and ALAN LEE**

School of Mathematics, Statistics and Computer Science, Victoria University of Wellington, New Zealand, and Department of Statistics, University of Auckland, New Zealand

October 20, 2008

The method of estimation in Scott and wild (1986,1997,2001) uses an reparametrisation of the profile likelihood and Lee and Hirose (2008) demonstrated the estimator is fully efficient. As an extension of these works, we investigate conditions under which the efficient score function and the efficient information matrix can be expressed in terms of the parameters in the reparametrised model.

Key words: Multi-sample; Semi-parametric model; Profile likelihood; Case-control study; Efficiency; M-estimator; Maximum likelihood estimator; Efficient score; Efficient information bound.

E-mail: Yuichi.Hirose@mcs.vuw.ac.nz
Telephone: 64 - 4 - 4635341 extn 5275
Fax: 64 - 4 - 4635045

1 Introduction

In a series papers, Scott and wild (1986,1997,2001) and Wild (1991) have developed a methodology which can be applied to a variety of response-selective sampling method. The efficiency of these methods has been demonstrated in special cases by several authors. For example, Breslow, Robins and Wellner (2000) consider case-control sampling, assuming that the data are generated by Bernoulli sampling, where either a case or control is selected by a randomisation device with known selection probabilities, and the covariates of the resulting case or control are measured. In the case of two-phase outcome-dependent sampling, Breslow, McNeney and Wellner (2003) apply the missing value theory of Robins, Rotnitzky and Zhao (1994) and Robins, Hsieh and Newey (1995). Here, individuals in the population are selected at random and their status (e.g. case or control) is determined. Then with a probability depending on their status, the covariates are measured or not. The unobserved covariates are treated as missing data. Lee and Hirose (2008) used an adaptation of the the profile likelihood method due to Newey (1994) to derive a semi-parametric efficiency bound, and then show that this bound coincides with the asymptotic variance of the Scott-Wild estimator, hence demonstrating the efficiency of the estimator.

In Lee and Hirose (2008), they demonstrated that, in the case of the Scott-Wild estimator, it is possible to reparametrise the least favorable submodel so that the efficient score function and the efficient information matrix can be expressed in terms of the parameters in the reparametrised model. The aim of this paper is to investigate conditions under which a reparametrization of the least favorable submodel yields this situation.

We consider an S-vector of semi-parametric models $(\mathcal{P}_1, \ldots, \mathcal{P}_S)$ where, for each $s = 1, \ldots, S$,

$$\mathcal{P}_s = \{p_s(x; \beta, \eta) : \ \beta \in \Theta_\beta \subset \mathbb{R}^m, \ \eta \in \Theta_\eta\}$$

is a probability model on the sample space $\mathcal{X}_s$ with the parameter of interest β, an m-dimensional parameter, and the nuisance parameter η, which may be an infinite-dimensional parameter. Let (β_0, η_0) be the true value of (β, η). We assume Θ_β is a compact set containing an open neighborhood of β_0 in $\mathbb{R}^m$, and Θ_η is a convex set containing η_0 in a Banach space $\mathcal{B}$. We refer the S-vector of semi-parametric models $(\mathcal{P}_1, \ldots, \mathcal{P}_S)$ as the multi-sample model.

Under the model, we observe S independent samples $X_{s1}, \ldots, X_{sn_s}$, $s = 1, \ldots, S$, where $X_{s1}, \ldots, X_{sn_s}$ are independently and identically distributed (i.i.d.) according to the model $\mathcal{P}_s$. Let $n = \sum_{s=1}^{S} n_s$. We assume the sample size proportions $(\frac{n_1}{n}, \ldots, \frac{n_S}{n})$ converge to weight probabilities $(w_1, \ldots, w_S)$:

$$(\frac{n_1}{n}, \ldots, \frac{n_s}{n}) \to (w_1, \ldots, w_s) \tag{1}$$

where $w_i > 0$ and $\sum_{i=1}^{s} w_i = 1$.

The log-likelihood for the multi-sample data is

$$\ell_n(\beta, \eta) = \sum_{s=1}^{S} \sum_{i=1}^{n_s} \log \, p_s(X_{si}; \beta, \eta).$$

The *log-likelihood function* for a one observation is

$$\ell(s, x, \beta, \eta) = \log\, p_s(x; \beta, \eta). \tag{2}$$

The expectation with respect to the density $p_s(x; \beta, \eta)$ is denoted by $E_{s,\beta,\eta}$.

The efficient score function We assume that there is a differentiable function $\hat{\eta}(\beta)$ such that

$$\hat{\eta}(\beta_0) = \eta_0 \tag{3}$$

and

$$\dot{\ell}^*_\beta(s, x, \beta_0) = \frac{\partial}{\partial \beta}\bigg|_{\beta=\beta_0} \ell(s, x, \beta, \hat{\eta}(\beta)) \tag{4}$$

is the efficient score function. We call the model

$$p_s(x; \beta, \hat{\eta}(\beta)),\ \beta \in \Theta_\beta\ s = 1, \ldots, S,$$

the least favorable submodel for the multi-sample model $(\mathcal{P}_1, \ldots, \mathcal{P}_S)$.

2 Main result

Definition of reparametrised model: Suppose the density for the least favorable submodel is of the form

$$p_s(x; \beta, \hat{\eta}(\beta)) = p'_s(x; \beta, q(\beta)), \quad \text{for } \beta \in \Theta_\beta, \quad s = 1, \ldots, S, \tag{5}$$

where the function $p'_s(x; \beta, q)$ is twice continuously differentiable with respect to (β, q) and q is a finite dimensional parameter. Further, suppose

$$\sum_{s=1}^{S} w_s \int p'_s(x; \beta, q)dx = 1, \quad \text{for all } (\beta, q) \in \Theta_\beta \times D_q \tag{6}$$

where Θ_β and D_q are neighborhoods of β_0 and $q(\beta_0)$, respectively. Then the model

$$p'_s(x; \beta, q) :\ \beta \in \Theta_\beta,\ q \in D_q, \quad s = 1, \ldots, S,$$

is called a *reparametrised model* for the least favorable submodel. The score functions for β and q in the reparametrised model are denoted by $\dot{\ell}_\beta(s, x; \beta, q) = \frac{\partial}{\partial \beta} \log p'_s(x; \beta, q)$ and $\dot{\ell}_q(s, x; \beta, q) = \frac{\partial}{\partial q} \log p'_s(x; \beta, q)$, respectively.

Remark 2.1: In general, we may not have the condition

$$\int p'_s(x; \beta, q)dx = 1, \quad \text{for all } (\beta, q) \in \Theta_\beta \times D_q,\ s = 1, \ldots, S.$$

Therefore, there is no guarantee that each $p'_s(x; \beta, q) :\ \beta \in \Theta_\beta,\ q \in D_q$ is a probability model.

Remark 2.2: Note that, since $\hat{\eta}(\beta_0) = \eta_0$, we have $p_s(x; \beta_0, \eta_0) = p'_s(x; \beta_0, q(\beta_0))$, $s = 1, \ldots, S$. Therefore for the reparametrised model, the notation $E_{s,0}$, $s = 1, \ldots, S$ is used for the expectations at the true value $(\beta_0, q(\beta_0))$.

Centering: For a measurable function $f(s,x;\beta,q)$, define the *centering* of $f(s,x;\beta,q)$ by

$$f^c(s,x;\beta,q) = f(s,x;\beta,q) - E_{s,0}f(s,x;\beta_0,q(\beta_0)).$$

The function $f^c(s,x;\beta,q)$ is called the *centered* $f(s,x;\beta,q)$.

THEOREM 1. *[Efficiency in a reparametrised model] We assume that the least favorable submodel and the corresponding reparametrised model are as in Equations (3), (4), (5) and (6). Further, assume that*

$$\left.\frac{\partial}{\partial q}\right|_{q=q(\beta)} \sum_{s=1}^{S} w_s E_{s,0}[\log p_s'(x;\beta,q)] = 0 \ \textit{for}\ \beta \in \Theta_\beta. \tag{7}$$

Then, the efficient score function and the efficient information matrix in the original multi-sample model $(\mathcal{P}_1,\ldots,\mathcal{P}_s)$ are given by

$$\dot\ell^*_\beta(s,x;\beta_0,\eta_0) = \dot\ell^c_\beta - \sum_{s=1}^{S} w_s E_{s,0}(\dot\ell^c_\beta \dot\ell^{cT}_q)\left(\sum_{s=1}^{S} w_s E_{s,0}(\dot\ell^c_q \dot\ell^{cT}_q)\right)^{-1} \dot\ell^c_q, \tag{8}$$

and

$$\begin{aligned} I^*_\beta(\beta_0,\eta_0) &= \sum_{s=1}^{S} w_s E_{s,0}(\dot\ell^c_\beta \dot\ell^{cT}_\beta) \\ &\quad - \sum_{s=1}^{S} w_s E_{s,0}(\dot\ell^c_\beta \dot\ell^{cT}_q)\left(\sum_{s=1}^{S} w_s E_{s,0}(\dot\ell^c_q \dot\ell^{cT}_q)\right)^{-1} \sum_{s=1}^{S} w_s E_{s,0}(\dot\ell^c_q \dot\ell^{cT}_\beta) \end{aligned} \tag{9}$$

where $\dot\ell^c_\beta(s,x;\beta,q)$ and $\dot\ell^c_q(s,x;\beta,q)$ are the centered score functions for β and q in the reparametrised model, respectively.

PROOF. By Equation (5), the efficient score function is given by

$$\begin{aligned} \dot\ell^*_\beta(s,x;\beta_0,q(\beta_0)) &= \left.\frac{\partial}{\partial\beta}\right|_{\beta=\beta_0} \log p_s'(x;\beta,q(\beta)) \\ &= \dot\ell_\beta(s,x;\beta_0,q(\beta_0)) + \dot q(\beta_0)^T \dot\ell_q(s,x;\beta_0,q(\beta_0)) \end{aligned} \tag{10}$$

Since $E_{s,\beta_0\eta_0}\dot\ell^*_\beta(s,x;\beta_0,q(\beta_0)) = 0$, for $s = 1,\ldots,S$, we have

$$E_{s,\beta_0\eta_0}\dot\ell_\beta(s,x;\beta_0,q(\beta_0)) + \dot q(\beta_0)^T E_{s,\beta_0\eta_0}\dot\ell_q(s,x;\beta_0,q(\beta_0)) = 0, \quad s = 1,\ldots,S. \tag{11}$$

Therefore, Equations (10) and (11) imply

$$\dot\ell^*_\beta(s,x;\beta_0,q(\beta_0)) = \dot\ell^c_\beta(s,x;\beta_0,q(\beta_0)) + \dot q(\beta_0)^T \dot\ell^c_q(s,x;\beta_0,q(\beta_0)). \tag{12}$$

By differentiating Equation (6) with respect to q, for all $(\beta,q) \in \Theta_\beta \times D_q$, we have

$$\sum_{s=1}^{S} w_s \int \dot\ell_q(s,x;\beta,q) p_s'(x;\beta,q)dx = 0.$$

In particular, for all $\beta \in \Theta_\beta$,

$$\sum_{s=1}^{S} w_s \int \dot{\ell}_q(s,x;\beta,q(\beta))p'_s(x;\beta,q(\beta))dx = 0.$$

By differentiating with respect to β at β_0,

$$\begin{aligned} & \sum_{s=1}^{S} w_s \int \left(\frac{\partial}{\partial \beta}\bigg|_{\beta=\beta_0} \dot{\ell}_q(s,x;\beta,q(\beta)) \right) p'_s(x;\beta_0,q(\beta_0))dx \\ = & -\sum_{s=1}^{S} w_s \int \dot{\ell}_q(s,x;\beta_0,q(\beta_0)) \left(\frac{\partial}{\partial \beta}\bigg|_{\beta=\beta_0} p'_s(x;\beta,q(\beta)) \right) dx. \end{aligned}$$

Since $p'_s(x;\beta_0,q(\beta_0)) = p_s(x;\beta_0,\hat{\eta}(\beta_0,F_0)) = p_s(x;\beta_0,\eta_0)$ and by the first equality in Equation (10), this equation is equivalent to

$$\sum_{s=1}^{S} w_s E_{s,0}\left[\frac{\partial}{\partial \beta}\bigg|_{\beta=\beta_0} \dot{\ell}_q(s,x;\beta,q(\beta)) \right] = -\sum_{s=1}^{S} w_s E_{s,0}[\dot{\ell}_q \dot{\ell}^{*T}_\beta(s,x;\beta_0,q(\beta_0))]. \tag{13}$$

By differentiating Equation (7) with respect to β at β_0, we get

$$\begin{aligned} 0 &= \frac{\partial}{\partial \beta}\bigg|_{\beta=\beta_0} \frac{\partial}{\partial q}\bigg|_{q=q(\beta)} \sum_{s=1}^{S} w_s E_{s,0}[\log p'_s(x;\beta,q)] \\ &= \sum_{s=1}^{S} w_s E_{s,0}\left[\frac{\partial}{\partial \beta}\bigg|_{\beta=\beta_0} \dot{\ell}_q(i,x,\beta,q(\beta)) \right] \\ &= -\sum_{s=1}^{S} w_s E_{s,0}[\dot{\ell}_q \dot{\ell}^{*T}_\beta(s,x,\beta_0,q(\beta_0))] \quad \text{(by Equation (13))} \\ &= -\sum_{s=1}^{S} w_s E_{s,0}[\dot{\ell}^c_q \dot{\ell}^{*T}_\beta(s,x,\beta_0,q(\beta_0))] \quad \text{(since } E_{s,0}\dot{\ell}^*_\beta(s,x,\beta_0,q(\beta_0)) = 0,\ s = 1,\ldots,S). \end{aligned}$$

Therefore, the centered score function $\dot{\ell}^c_q(s,x,\beta_0,q(\beta_0))$ and the efficient score function $\dot{\ell}^*_\beta(s,x,\beta_0,q(\beta_0))$ are uncorrelated. Since $\dot{\ell}^*_\beta = \dot{\ell}^c_\beta + \dot{q}(\beta_0)^T\dot{\ell}^c_q$ (cf. Equation (12)), by the projection theorem (cf. Appendix A), we have

$$\dot{q}(\beta_0)^T\dot{\ell}^c_q = -\sum_{s=1}^{S} w_s E_{s,0}(\dot{\ell}^c_\beta \dot{\ell}^{cT}_q) \left(\sum_{s=1}^{S} w_s E_{s,0}(\dot{\ell}^c_q \dot{\ell}^{cT}_q) \right)^{-1} \dot{\ell}^c_q.$$

The rest of the claims follow by substituting this expression into Equation (12). □

Remark 2.3: Under the usual regularity conditions, the solution $(\hat{\beta}_n, \hat{q}_n)$ to the system of the score equations,

$$\begin{cases} \sum_{s=1}^{S}\sum_{i=1}^{n_i} \dot{\ell}_\beta(s,X_{si};\hat{\beta}_n,\hat{q}_n) = 0 \\ \sum_{s=1}^{S}\sum_{i=1}^{n_i} \dot{\ell}_q(s,X_{si};\hat{\beta}_n,\hat{q}_n) = 0, \end{cases}$$

is asymptotically distributed as

$$\begin{pmatrix} \sqrt{n}(\hat{\beta}_n - \beta_0) \\ \sqrt{n}(\hat{q}_n - q_0) \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, [I(\beta_0,q_0)]^{-1} \right)$$

where

$$I(\beta_0, q_0) = \begin{pmatrix} \sum_{s=1}^{S} w_s E_{s,0}(\dot{\ell}_\beta^c \dot{\ell}_\beta^{cT}) & \sum_{s=1}^{S} w_s E_{s,0}(\dot{\ell}_\beta^c \dot{\ell}_q^{cT}) \\ \sum_{s=1}^{S} w_s E_{s,0}(\dot{\ell}_q^c \dot{\ell}_\beta^{cT}) & \sum_{s=1}^{S} w_s E_{s,0}(\dot{\ell}_q^c \dot{\ell}_q^{cT}) \end{pmatrix}.$$

Then the asymptotic variance of $\sqrt{n}(\hat{\beta}_n - \beta_0)$ is given by $[I_\beta^*(\beta_0, \eta_0)]^{-1}$ where $I_\beta^*(\beta_0, \eta_0)$ is the efficient information for β given by Equation (9) (cf. Bickel, Klaassen, Ritov and Wellner (1993), page 28). In this case, the estimator $\hat{\beta}_n$ is efficient.

3 Example: Stratified sampling

We assume that the underlying data generating process on the sample space $\mathcal{Y} \times \mathcal{X}$ is a model

$$\mathcal{Q} = \{p(y, x; \theta, G) = f(y|x; \theta)g(x) : \ \theta \in \Theta, \ G \in \mathcal{G}\}$$

where $f(y|x; \theta)$ is a conditional density of Y given X which depends on a finite dimensional parameter θ, $G(x)$ is an unspecified distribution function of X which is an infinite-dimensional nuisance parameter ($g(x)$ is the density of $G(x)$). We assume the set Θ is a compact set containing a neighborhood of the true value θ_0 and $\mathcal{G}$ is the set of all distribution functions of x. Unless stated otherwise Y may be a discrete or continuous variable.

For a partition of the sample space $\mathcal{Y} \times \mathcal{X} = \cup_{s=1}^{S} \mathcal{S}_s$, let

$$Q_s(\theta, G) = \int f(y|x; \theta) \, 1_{(y,x) \in \mathcal{S}_s} \, dy \, dG(x)$$

be the probability of (Y, X) belonging to stratum $\mathcal{S}_s$. Also, let

$$Q_{s|X}(x; \theta) = \int f(y|x; \theta) \, 1_{(y,x) \in \mathcal{S}_s} \, dy.$$

In standard stratified sampling, for each $s = 1, \ldots, S$, a random sample of size n_i is taken from the conditional distribution

$$p_s(y, x; \theta, G) = \frac{f(y|x; \theta) g(x) 1_{(y,x) \in \mathcal{S}_s}}{Q_s(\theta, G)}$$

of (X, Y) given stratum $\mathcal{S}_s$.

3.1 Finding the least favorable submodel

The aim of this section is to find the form of the density function for the least favorable submodel in stratified sampling.

THEOREM 2. *[The least favorable submodel] For $\theta \in \Theta$, let*

$$\hat{g}(\theta) = \hat{g}(x, \theta, \hat{Q}(\theta)) = \frac{f_0^*(x)}{\sum_{s=1}^{S} w_s \frac{Q_{s|X}(x;\theta)}{\hat{Q}_s(\theta)}}, \tag{14}$$

where

$$f_0^*(x) = \sum_{s=1}^{S} w_s \frac{Q_{s|X}(x;\theta_0)g_0(x)}{Q_s(\theta_0, g_0)}, \tag{15}$$

and

$$\hat{Q}_s(\theta) = \int Q_{s|X}(x;\theta)\hat{g}(x,\theta,\hat{Q}(\theta))dx, \quad s = 1,\ldots,S. \tag{16}$$

Then the efficient score function is given by

$$\dot{\ell}^*_\beta(s,y,x;\theta_0) = \frac{\partial}{\partial\beta}\bigg|_{\theta=\theta_0} \log p_s(y,x;\theta,\hat{g}(\theta)). \tag{17}$$

The proof is given in Appendix C.

Remark 3.1 Note that the equations (Equation 14 and Equation 16) are consistent at θ_0: Equation 16 at $\theta = \theta_0$ has a solution when $\hat{Q}_s(\theta_0) = Q_s(\theta_0, g_0)$. But, since $\hat{g}(\theta_0) = g_0$, $\hat{Q}_s(\theta_0) = \int Q_{s|X}(x;\theta_0)g_0(x)dx = Q_s(\theta_0, g_0)$.

3.2 Efficiency of reparametrised model

In this section, we study the efficiency of the Scott & Wild estimator in the context of stratified sampling using a reparametrised form of the least favorable submodel.

By Theorem 2, the least favorable submodel is given by

$$p_s(y,x;\theta,\hat{g}(\theta)) = \frac{f(y|x;\theta)1_{(y,s)\in\mathcal{S}_s}\hat{g}(x,\theta,\hat{Q}(\theta))}{\hat{Q}_s(\theta)}.$$

By replacing $\hat{Q}(\theta) = (\hat{Q}_1(\theta),\ldots,\hat{Q}_{S-1}(\theta),\hat{Q}_S(\theta))$ with $q = (q_1,\ldots,q_{S-1},1)$, we consider a reparametised model of the form

$$p'_s(y,x;\theta,q) = \frac{f(y|x;\theta)1_{(y,s)\in\mathcal{S}_s}\hat{g}(x,\theta,q)}{q_s}, \tag{18}$$

where

$$\hat{g}(x,\theta,q) = \frac{f_0^*(x)}{\sum_{s=1}^{S} w_s \frac{Q_{s|X}(x;\theta)}{q_s}} \tag{19}$$

with $f_0^*(x)$ is given by Equation (15).

The true value of (θ, q) is

$$(\theta_0, q_0) = \left(\theta_0, \left(\frac{Q_1(\theta_0,g_0)}{Q_S(\theta_0,g_0)},\ldots,\frac{Q_{S-1}(\theta_0,g_0)}{Q_S(\theta_0,g_0)},1\right)\right).$$

Let D_q be some neighborhood of q_0.

We will demonstrate that the conditions in Theorem 1 are satisfied, so that we can apply the theorem to identify the efficient score function and the efficient information matrix in the example.

First, we will show that

$$\sum_{s=1}^{S} w_s \int p'_s(y,x;\theta,q)dydx = 1, \quad \text{for all } (\theta,q) \in \Theta_0 \times D_q.$$

For any (θ,q), since $Q_{s|X}(x;\theta) = \int f(y|x;\theta)1_{(y,s)\in\mathcal{S}_s}dy$,

$$\begin{aligned}
\sum_{s=1}^{S} w_s \int p_s(y,x;\theta,q)dydx &= \sum_{s=1}^{S} w_s \int \frac{f(y|x;\theta)1_{(y,s)\in\mathcal{S}_s}\hat{g}(x,\theta,q)}{q_s}dydx \\
&= \sum_{s=1}^{S} w_s \int \frac{Q_{s|X}(x;\theta)\hat{g}(x,\theta,q)}{q_s}dx \\
&= \int \sum_{s=1}^{S} w_s \frac{Q_{s|X}(x;\theta)}{q_s}\hat{g}(x,\theta,q)dx \\
&= \int f_0^*(x)dx \quad \text{(by Equation 19)} \\
&= 1.
\end{aligned}$$

Second, we will show that for all $\theta \in \Theta_0$,

$$\left.\frac{\partial}{\partial q}\right|_{q=\hat{Q}(\theta)} \sum_{s=1}^{S} w_s E_{s,0} \log p_s(y,x;\theta,q) = 0. \tag{20}$$

For $j = 1,\ldots,S-1$, the derivative is

$$\begin{aligned}
&\frac{\partial}{\partial q_j}\sum_{s=1}^{S} w_s E_{s,0} \log p_s(y,x;\theta,q) \\
= &-\frac{\partial}{\partial q_j}\sum_{s=1}^{S} w_s E_{s,0}\left(\log \sum_{s'=1}^{S} w_{s'}\frac{Q_{s'|X}(x;\theta)}{q_{s'}} + \log q_s\right) \\
= &\sum_{s=1}^{S} w_s E_{s,0}\left(\frac{w_j \frac{Q_{j|X}(x;\theta)}{q_j^2}}{\sum_{s'=1}^{S} w_{s'}\frac{Q_{s'|X}(x;\theta)}{q_{s'}}}\right) - \frac{w_j}{q_j} \\
= &\sum_{s=1}^{S} w_s \int \frac{w_j \frac{Q_{j|X}(x;\theta)}{q_j^2}}{\sum_{s'=1}^{S} w_{s'}\frac{Q_{s'|X}(x;\theta)}{q_{s'}}}\frac{Q_{s|X}(x;\theta_0)g_0(x)}{Q_s(\theta_0,g_0)}dx - \frac{w_j}{q_j} \\
= &\int \frac{w_j \frac{Q_{j|X}(x;\theta)}{q_j^2} f_0^*(x)}{\sum_{s'=1}^{S} w_{s'}\frac{Q_{s'|X}(x;\theta)}{q_{s'}}}dx - \frac{w_j}{q_j} \quad \text{(by Equation (15))} \\
= &\frac{w_j}{q_j^2}\left(\int Q_{j|X}(x;\theta)\hat{g}(x,\theta,q)dx - q_j\right).
\end{aligned}$$

Therefore by Equation (16), at $q = (q_1,\ldots,q_{S-1},1) = \left(\frac{\hat{Q}_1(\theta)}{\hat{Q}_S(\theta)},\ldots,\frac{\hat{Q}_{S-1}(\theta)}{\hat{Q}_S(\theta)},1\right)$, we have Equation (20).

By Theorem 1, the efficient score function and the efficient information matrix in the example are calculated by Equation (8) and (9), respectively, where the score functions $\dot{\ell}_\theta$ and $\dot{\ell}_q$ are given

by

$$\dot{\ell}_\theta(s,y,x;\theta,q) = \frac{\frac{\partial}{\partial\theta}f(y|x;\theta)}{f(y|x;\theta)} - \frac{\sum_{s'=1}^{S} w_{s'} \frac{\frac{\partial}{\partial\theta}Q_{s'|X}(x;\theta)}{q_{s'}}}{\sum_{s'=1}^{S} w_{s'} \frac{Q_{s'|X}(x;\theta)}{q_{s'}}}$$

and

$$\dot{\ell}_{q_j}(s,y,x;\theta,q) = \frac{w_j}{q_j^2}\left(\frac{Q_{j|X}(x;\theta)}{\sum_{s'=1}^{S} w_{s'}\frac{Q_{s'|X}(x;\theta)}{q_{s'}}} - q_j\right), \quad j = 1,\ldots,S-1.$$

3.3 Identifiability of the parameter in stratified sampling

For a reference, see Breslow, Robins & Wellner (2000).

3.3.1 Non-identifiability in the semiparametric model

In stratified sampling, let $\theta = (\alpha,\beta)$ where $\alpha \in \mathbf{R}$ and $\beta \in \mathbf{R}^J$, and assume the logistic regression model

$$f(y|x;\alpha,\beta) = \frac{\exp(y(\alpha + x^T\beta))}{1+\exp(\alpha+x^T\beta)}. \tag{21}$$

The sample space is partitioned as $\mathcal{Y}\times\mathcal{X} = (\{0\}\times\mathcal{X}) \cup (\{1\}\times\mathcal{X})$ and let

$$Q_s(\alpha,\beta,g) = \int f(y=s|x;\alpha,\beta)g(x)dx, \quad s=0,1.$$

Then the parameters $(\theta,g) = (\alpha,\beta,g)$ in the multi-sample model for stratified sampling are not identifiable: Let

$$g^*(x) = \frac{g(x)[1+\exp(x^T\beta)]}{1+\exp(\alpha+x^T\beta)}.$$

Then

$$\begin{aligned} p_s(x;\alpha,\beta,g) &= \frac{f(y=s|x;\alpha,\beta)g(x)}{Q_s(\alpha,\beta,g)} \\ &= \frac{f(y=s|x;0,\beta)g^*(x)}{Q_s(0,\beta,g^*)} \\ &= p_s(x;0,\beta,g^*), \end{aligned}$$

but $(\alpha,\beta,g) \neq (0,\beta,g^*)$ for any $\alpha \neq 0$, β and g.

3.3.2 Non-identifiability in the reparametrised model

A reparametrised model for the logistic regression model (Equation (21)) is

$$\begin{aligned} p(s,x;\alpha,\beta,\rho_1) &= w_s p_s(x;\alpha,\beta,\rho_1) \\ &= \frac{\rho_s f(y=s|x;\alpha,\beta)}{\sum_{s'=0}^{1}\rho_{s'}f(y=s'|x;\alpha,\beta)} f_0^*(x) \\ &= \frac{\rho_s\exp(s(\alpha+x^T\beta))}{1+\rho_1\exp(\alpha+x^T\beta)} f_0^*(x) \end{aligned}$$

where $f_0^*(x) = \sum_{s=0}^{1} \rho_{0,s} f(y = s|x; \alpha_0, \beta_0) g_0(x)$ and we took a parameterization so that $\rho_0 = 1$.

The score functions for α, β, and ρ_1 are

$$\dot{\ell}_\alpha(s, x; \alpha, \beta, \rho_1) = s - \frac{\rho_1 \exp(\alpha + x^T\beta)}{1 + \rho_1 \exp(\alpha + x^T\beta)},$$

$$\dot{\ell}_\beta(s, x; \alpha, \beta, \rho_1) = x\left(s - \frac{\rho_1 \exp(\alpha + x^T\beta)}{1 + \rho_1 \exp(\alpha + x^T\beta)}\right),$$

and

$$\dot{\ell}_{\rho_1}(s, x; \alpha, \beta, \rho_1) = \frac{1}{\rho_1}\left(s - \frac{\rho_1 \exp(\alpha + x^T\beta)}{1 + \rho_1 \exp(\alpha + x^T\beta)}\right).$$

Therefore the score function $\dot{\ell}_{(\alpha,\beta,\rho_1)}$ for (α, β, ρ_1) is

$$\dot{\ell}_{(\alpha,\beta,\rho_1)}(s, x; \alpha, \beta, \rho_1) = \begin{pmatrix} 1 \\ x \\ \frac{1}{\rho_1} \end{pmatrix}\left(s - \frac{\rho_1 \exp(\alpha + x^T\beta)}{1 + \rho_1 \exp(\alpha + x^T\beta)}\right).$$

The components of the score function $\{\dot{\ell}_\alpha, \dot{\ell}_\beta, \dot{\ell}_{\rho_1}\}$ are not linearly independent. This implies:

(1) The parameterization $(\alpha, \beta, \rho_1) \to p(s, x; \alpha, \beta, \rho_1)$ is not one-to-one (i.e. the parameterization is not identifiable).

(2) The information matrix $I(\alpha_0, \beta_0, \rho_{1,0}) = P_{\alpha_0,\beta_0,g_0}(\dot{\ell}_{(\alpha,\beta,\rho_1)}\dot{\ell}^T_{(\alpha,\beta,\rho_1)})$ is not invertible.

(3) The nuisance tangent space (the tangent space for ρ_1) in the reparametrised model coincides with the tangent space for α.

However, the parameter β is identifiable and can be estimated in this model. (See Breslow, Robins & Wellner (2000).)

4 Conclusion

Theorem 1 gives conditions under which the efficient score function and the efficient information matrix can be expressed in terms of the parameters in the reparametrised model, namely, Equation 8 and Equation 9, respectively. This result extend the result of Lee and Hirose (2008) in more general situation.

Appendix A: The projection theorem

Let $\mathcal{H}$ be the Hilbert space of m-dimensional measurable functions with zero mean and finite variance:

$$\mathcal{H} = \left\{ f(s, x) : \ E_{s,\beta_0,\eta_0} f = 0, s = 1, \ldots, S, \ \sum_{s=1}^{S} w_s E_{s,\beta_0,\eta_0} f^2 < \infty \right\}.$$

The covariance of $f, g \in \mathcal{H}$ is defined by $\text{Cov}(f,g) = \sum_{s=1}^{S} w_s E_{s,\beta_0,\eta_0}(fg^T)$. We say f and g are uncorrelated (orthogonal) if $\text{Cov}(f,g) = 0$. For a set of functions $\mathcal{G}$ in $\mathcal{H}$, $\mathcal{G}^\perp$ is the set of all functions $f \in \mathcal{H}$ with $\text{Cov}(f,g) = 0$. The projection $\pi(f|\mathcal{G})$ of $f \in \mathcal{H}$ onto a closed subspace $\mathcal{G}$ is characterized by

$$\pi(f|\mathcal{G}) \in \mathcal{G} \quad \text{and} \quad f - \pi(f|\mathcal{G}) \in \mathcal{G}^\perp.$$

THEOREM A. *[The projection theorem] Suppose $g(s,x)$ is a l-dimensional vector of measurable functions such that*
(1) for $s = 1, \ldots, S$, $E_{s,\theta_0}(g) = 0$;
(2) $\sum_{s=1}^{S} w_s E_{s,\theta_0}(g^T g) < \infty$;
(3) $\left[\sum_{s=1}^{S} w_s E_{s,\theta_0}(gg^T)\right]^{-1}$ exists.

Let $\mathcal{G} = \{Ag : A \in \mathbf{R}^{m\times l}\}$ be the closed subspace of $\mathcal{H}$ generated by g. Then, for each $f \in \mathcal{H}$, the projection of f onto the closed subspace $\mathcal{G}$ is given by

$$\pi(f|\mathcal{G}) = \left[\sum_{s=1}^{S} w_s E_{s,\theta_0}(fg^T)\right]\left[\sum_{s=1}^{S} w_s E_{s,\theta_0}(gg^T)\right]^{-1} g.$$

PROOF. Note that $\pi(f|\mathcal{G}) \in \mathcal{G}$ implies $\pi(f|\mathcal{G}) = A_0 g$ for some $A_0 \in \mathbf{R}^{m\times l}$. Then, by the properties of the projection, we have

$$A_0 g \in \mathcal{G} \quad \text{and} \quad f - A_0 g \in \mathcal{G}^\perp.$$

It follows that, since $g \in \mathcal{G}$,

$$\sum_{s=1}^{S} w_s E_{s,0}[(f - A_0 g)g^T] = 0.$$

This implies

$$\sum_{s=1}^{S} w_s E_{s,0}(fg^T) - A_0 \sum_{s=1}^{S} w_s E_{s,0}(gg^T) = 0.$$

By the assumption that $[\sum_{s=1}^{S} w_s E_{s,0}(gg^T)]^{-1}$ exists, we have

$$A_0 = \left[\sum_{s=1}^{S} w_s E_{s,0}(fg^T)\right]\left[\sum_{s=1}^{S} w_s E_{s,0}(gg^T)\right]^{-1}.$$

Therefore,

$$\pi(f|\mathcal{G}) = A_0 g = \left[\sum_{s=1}^{S} w_s E_{s,0}(fg^T)\right]\left[\sum_{s=1}^{S} w_s E_{s,0}(gg^T)\right]^{-1} g.$$

□

Appendix B: Theorem to identify the least favorable submodel

To verify Condition ($R0$), the following theorem may be useful.

THEOREM 2. *A path $\eta(t)$ is a continuously differentiable map in a neighborhood of 0 such that $\eta(0) = \eta_0$. Define $\alpha(t) = \eta(t) - \eta_0$. If $\hat{\eta}(\beta)$ is a differentiable function such that*

$$\hat{\eta}(\beta_0) = \eta_0 \tag{22}$$

and, for each $\beta \in \Theta_\beta$, and for each path $\eta(t)$,

$$\left.\frac{\partial}{\partial t}\right|_{t=0} E_{\beta_0,\eta_0}\left[\log p(x;\beta,\hat{\eta}(\beta)+\alpha(t))\right] = 0, \tag{23}$$

then the function $\dot{\ell}^*_\beta(x,\beta_0) = \frac{\partial}{\partial\beta}|_{\beta=\beta_0} \log p(x;\beta,\hat{\eta}(\beta))$ is the efficient score function.

PROOF. See Hirose(2008) for the proof. □

Appendix C: Proof of Theorem 2

For each $s = 1,\ldots,S$, let F_{s0} be the cdf for the true distribution for the model

$$\{p_s(y,x;\theta,G) : \theta \in \Theta, G \in \mathcal{G}\}.$$

The expected likelihood in the model is

$$\sum_{s=1}^{S} w_s \int \log p_s(y,x;\theta,G) dF_{s0}(y,x).$$

In Step 1, we find a function $\hat{g}(\theta)$ by using the method of Scott and Wild (1997, 2001). In Step 2, we show that $\sum_{s=1}^{S} w_s \int \log p(y,x;\theta,\hat{g}(\theta)) dF_{s0}$ satisfies Conditions (22) and (23) in THEOREM 2 in Appendix B so that the claim follows form this theorem.

Step 1: First, we find a function $\hat{g}(x,\theta)$ under the assumption that the support of the distribution of X is finite: i.e. $\text{supp}(X) = \{v_1,\ldots,v_K\}$. Let $(g_1,\ldots,g_K) = (g(v_1),\ldots,g(v_K))$, then $\log g(x)$ and $Q_s(\theta,g)$ can be expressed as $\log g(x) = \sum_{k=1}^{K} 1_{x=v_k} \log g_k$ and $Q_s(\theta,g) = \int Q_{s|X}(x;\theta) g(x) dx = \sum_{k=1}^{K} Q_{s|X}(v_k;\theta) g_k$.

To find the maximizer $(g_1,\ldots,g_K)$ of

$$\sum_{s=1}^{S} w_s \int \log p(y,x;\theta,\hat{g}(\theta)) dF_{s0} \;=\; \sum_{s=1}^{S} w_s \left\{ \int \left(\log f(y|x;\theta) + \log g(X)\right) dF_{s0} - \log Q_s(\theta,g) \right\}$$

at θ, differentiate this expression with respect to g_k and set the derivative equal to zero,

$$\frac{\partial}{\partial g_k} \sum_{s=1}^{S} w_s \int \log p(y,x;\theta,\hat{g}(\theta)) dF_{s0} = \sum_{s=1}^{S} w_s \left\{ \frac{\int 1_{X=v_k} dF_{s0}}{g_k} - \frac{Q_{s|X}(v_k;\theta)}{Q_s(\theta,g)} \right\} = 0.$$

The solution g_k to the equation is

$$\hat{g}(v_k, \theta) = g_k \quad = \quad \frac{\sum_{s=1}^{S} w_s \int 1_{X=v_k} dF_{s0}}{\sum_{s=1}^{S} w_s \frac{Q_{s|X}(v_k;\theta)}{Q_s(\theta,g)}}.$$

Since $f_0^*(v_k) = \sum_{s=1}^{S} w_s \int 1_{X=v_k} dF_{s0}$ by Equation (15), this expression is of the form in Equation (14).

Step 2: Condition (22) is verified in REMARK 3.3. Now, we verify Condition (23). Let $g(x,t)$ be a path in the space of density functions with $g(x,0) = g_0(x)$. Define $\alpha(t) = \alpha(x,t) = g(x,t) - g_0(x)$ and write $\alpha'(x,0) = \frac{\partial}{\partial t}|_{t=0}\, \alpha(x,t)$. Then

$$\begin{aligned}
& \frac{\partial}{\partial t}\Big|_{t=0} \sum_{s=1}^{S} w_s \int \log p_s(y,x;\theta,\hat{g}(\theta)+\alpha(t)) dF_{s0} \\
= \; & \frac{\partial}{\partial t}\Big|_{t=0} \sum_{s=1}^{S} w_s \left\{ \int \log(\hat{g}(x,\theta)+\alpha(t)) dF_{s,0} - \log Q_s(\theta, \hat{g}(\theta)+\alpha(t)) \right\} \\
= \; & \frac{\partial}{\partial t}\Big|_{t=0} \left\{ \int \log(\hat{g}(x,\theta)+\alpha(t)) f_0^*(x) dx - \sum_{s=1}^{S} w_s \log Q_s(\theta, \hat{g}(\theta)+\alpha(t)) \right\} \\
= \; & \int \frac{\alpha'(x,0)}{\hat{g}(x,\theta)} f_0^*(x) dx - \sum_{s=1}^{S} w_s \frac{\int Q_{s|X}(x;\theta)\alpha'(x,0) dx}{\hat{Q}_s(\theta)} = 0
\end{aligned}$$

by Equations (14) and (15). By Theorem in Appendix B, the claim follows. □

References

Begun, J. M., Hall, W. J., Huang, W. M. and Wellner, J. A. (1983). Information and asymptotic efficiency in parametric–nonparametric models. *Ann. Statist.* **11** 432–452.

Bickel, P.J., Klaassen, C.A.J., Ritov, Y. and Wellner, J.A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models.* Johns Hopkins Univ. Press, Baltimore.

BRESLOW, N.E. AND CHATTERJEE, N. (1999). Design and analysis of two-phase studies with binary outcome applied to Wilms tumor prognosis. *Appl. Statist.* **48** 457–468.

BRESLOW, N.E. AND HOLUBKOV, R. (1997). Maximum likelihood estimation of logistic regression parameters under two-phase, outcome-dependent sampling. *J. Roy. Statist. Soc. ser. B* **59** 447–461.

BRESLOW, N.E., MCNENEY, B. AND WELLNER, J.A. (2003). Large sample theory for semiparametric regression models with two-phase, outcome dependent sampling. *Ann. Statist.* **31** 1110–1139.

BRESLOW, N.E., ROBINS, J.M. AND WELLNER, J.A. (2000). On the semi-parametric efficiency of logistic regression under case-control sampling. *Bernoulli* **6** 447–455.

Dudley, R.M. (1989). *Real analysis and probability*. Pacific Grove, California.

Godambe, V.P. (1991). Orthogonality of estimating functions and nuisance parameters. *Biometrika* **78** 143–151.

Hirose, Y. (2008). Efficiency of profile likelihood in semi-parametric models, Submitted to Annals of the Institute of Statistical Mathematics.

Lawless, J.L., Kalbfleish, J.D. and Wild, C.J. (1999). Estimation for response-selective and missing data problems in regression. *J. Roy. Statist. Soc. Ser. B* **61** 413–438.

Lee, A.J. (2004). Semi-parametric efficiency bounds for regression models under choice-based sampling. Unpublished manuscript, Univ. Auckland.

Lee, A.J. and Hirose, Y. (2008). Semi-parametric efficiency bounds for regression models under case-control sampling: the profile likelihood approach, to appear in Annals of the Institute of Statistical Mathematics.

McLeish, D. L. and Small, C. G. (1992). A projected likelihood function for semiparametric models. *Biometrika* **79** 93–102.

Murphy, S.A. and van der Vaart, A.W. (1999). Observed information in semi-parametric models. *Bernoulli* **5** 381–412.

Murphy, S.A. and van der Vaart, A.W. (2000). On profile likelihood (with discussion). *J. Amer. Statist. Assoc.* **95** 449–485.

Nan, B., Emond, M. and Wellner, J.A. (2004). Information bounds for cox regression models with missing data. *Ann. Statist.* **32** 723–753.

Newey, W.K. (1990). Semi-parametric efficiency bounds. *J. Appl. Econ* **5** 99–135.

Newey, W.K. (1994). The asymptotic variance of semi-parametric estimators. *Econometrica* **62** 1349–1382.

Prentice, R.L. and Pyke, R. (1979). Logistic disease incidence models and case-control studies. *Biometrika* **66** 403–411.

Robins, J.M., Hsieh, F. and Newey, W.K. (1995). Semiparametric efficient estimation of a conditional density with missing or mismeasured covariates. *J. Roy. Statist. Soc. Ser. B* **57** 409–424.

Robins, J.M., Rotnitzky, A. and Zhao, L.P. (1994). Estimation of regression coefficients when some regressors are not always observed. *J. Amer. Statist. Assoc.* **89** 846–866.

Seber, G.A.F. and Lee, A.J. (2003). *Linear Regression Analysis, Second Edition*. Wiley, New York.

Scott, A.J. and Wild, C.J. (1997). Fitting regression models to case-control data by maximum likelihood. *Biometrika* **84** 57–71.

Scott, A.J. and Wild, C.J. (2001). Maximum likelihood for generalised case-control studies. *J. Stat. Plann. Inference* **96** 3–27.

Tsiatis, A.B. (2003). *Semiparametric Theory and Missing Data Problems*. Course Notes. Unpublished manuscript, North Carolina State Univ.

van de Geer, S.A. (2000). *Empirical Processes in M-Estimation*. Cambridge Univ. Press, Cambridge.

van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes*. Springer, New York.

van der Vaart, A.W. (1998). *Asymptotic Statistics*. Cambridge Univ. Press, Cambridge.