Biostatistics (2010), **0**, 0, *pp.* 1–12
doi:10.1093/biostatistics/

Supplemental materials for semiparametric inference for a two-stage outcome-auxiliary-dependent sampling design with continuous outcome

HAIBO ZHOU*
Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7420, USA
zhou@bios.unc.edu

YUANSHAN WU
Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7420, USA, and School of Mathematics and Statistics, Wuhan University, Wuhan, Hubei 430072, China
yswu@bios.unc.edu

YANYAN LIU
School of Mathematics and Statistics, Wuhan University, Wuhan, Hubei 430072, China
liuyy@whu.edu.cn

JIANWEN CAI
Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7420, USA
cai@bios.unc.edu

*To whom correspondence should be addressed.

1. FROM LIKELIHOOD TO ESTIMATED LOG-LIKELIHOOD

In this section, we intend to derive estimated log-likelihood from the likelihood function. It can be shown that the full likelihood based on all the observations under two-stage OADS design is proportional to

$$L_F(\beta) = \left[\prod_{k=0}^{K}\prod_{i\in\tilde{V}_k} f(Y_i|Z_i,X_i;\beta)g(X_i|Z_i,W_i)\right]\left[\prod_{k=1}^{K}\prod_{i\in\bar{V}_k}\int_{\mathcal{X}} f(Y_i|Z_i,x;\beta)dG(x|Z_i,W_i)\right]. \quad \text{(A.1)}$$

Let S denote the informative components of (Z,W) in the sense that $G(X|Z,W) = G(X|S)$ almost surely. Without loss of generality, assume S is continuous variable with dimension d. Note that

$$G(x|s) = \sum_{k=1}^{K}\pi_k(s)G_k(x|s),$$

where $\pi_k(s) = \Pr((Y,W)\in\Delta_k|s)$ and $G_k(x|s) = G(x|s,(Y,W)\in\Delta_k)$. Then we estimate the $\pi_k(s)$ and $G_k(x|s)$ respectively by

$$\widehat{\pi}_k(s) = \frac{\sum_{i=1}^{N} I((Y_i,W_i)\in\Delta_k)\phi_{h_N}(S_i-s)}{\sum_{i=1}^{N}\phi_{h_N}(S_i-s)}$$

and

$$\widehat{G}_k(x|s) = \frac{\sum_{i\in V_k} I(X_i \leqslant x)\phi_{h_N}(S_i-s)}{\sum_{i\in V_k}\phi_{h_N}(S_i-s)},$$

where $I(\cdot)$ is an indicator function and $\phi_{h_N}(\cdot) = \phi(\frac{\cdot}{h_N})$ is a d-dimensional kernel function with the bandwidth h_N. For simplicity, we suppress the subscript of h_N hereafter. Hence, $G(x|s)$ can be subsequently estimated by

$$\widehat{G}(x|s) = \sum_{k=1}^{K}\widehat{\pi}_k(s)\widehat{G}_k(x|s),$$

which is a consistent estimator as shown below.

Hence, we obtain an estimated likelihood function by substituting $G(x|s)$ in (A.1) with $\widehat{G}(x|s)$ and denote it by

$$\widehat{L}_F(\beta) = \left[\prod_{k=0}^{K}\prod_{i\in\tilde{V}_k} f(Y_i|Z_i, X_i;\beta)\widehat{g}(X_i|S_i)\right]\left[\prod_{k=1}^{K}\prod_{i\in\bar{V}_k}\int_{\mathcal{X}} f(Y_i|Z_i, x;\beta)d\widehat{G}(x|S_i)\right]. \tag{A.2}$$

Then the estimated log-likelihood function is

$$\begin{aligned}
&\widehat{l}_F(\beta) \equiv \log \widehat{L}_F(\beta)\\
=&\sum_{k=0}^{K}\sum_{i\in\tilde{V}_k}\log f(Y_i|Z_i, X_i;\beta) + \sum_{k=0}^{K}\sum_{i\in\tilde{V}_k}\log \widehat{g}(X_i|S_i) + \sum_{k=1}^{K}\sum_{j\in\bar{V}_k}\log \widehat{f}(Y_j|Z_j, W_j;\beta)\\
=&\sum_{k=1}^{K}\sum_{i\in V_k}\log f(Y_i|Z_i, X_i;\beta) + \sum_{k=1}^{K}\sum_{i\in V_k}\log \widehat{g}(X_i|S_i) + \sum_{k=1}^{K}\sum_{j\in\bar{V}_k}\log \widehat{f}(Y_j|Z_j, W_j;\beta)\\
=&\sum_{k=1}^{K}\sum_{i\in V_k}\log f(Y_i|Z_i, X_i;\beta) + \sum_{k=1}^{K}\sum_{j\in\bar{V}_k}\log \widehat{f}(Y_j|Z_j, W_j;\beta) + C,
\end{aligned}$$

where

$$\begin{aligned}
\widehat{f}(Y_j|Z_j, W_j;\beta) &= \int_{\mathcal{X}} f(Y_j|Z_j, x;\beta)d\widehat{G}(x|S_j)\\
&= \sum_{r=1}^{K}\widehat{\pi}_r(S_j)\frac{\sum_{l\in V_r} f(Y_j|Z_j, X_l;\beta)\phi_h(S_l - S_j)}{\sum_{l\in V_r}\phi_h(S_l - S_j)},
\end{aligned}$$

and $C = \sum_{k=1}^{K}\sum_{i\in V_k}\log \widehat{g}(X_i|S_i)$, which is not dependent on β.

2. Regularity conditions

The following conditions are imposed to investigate the asymptotic properties of the estimator $\widehat{\beta}$.

C1. $f(y|z, x;\beta)$ has the 2nd-order continuous derivatives with respect to β, for every $\beta \in \mathcal{B}$, where $\mathcal{B}$ is the parametric space, a compact subset in Euclidean space $\mathcal{R}^q$, containing β^0 as its interior point.

C2. When N goes to ∞, $|V|/N \to \rho_V > 0$ and $n_k/|V| \to \rho_k \geqslant 0$ for $k = 0, \ldots, K$. Let $\gamma_k = \Pr\{(Y, W) \in \Delta_k\}$.

C3. $\phi(\cdot)$ is a αth-order bounded and symmetric kernel function with bounded support and $\int \phi^2 < \infty$. $Nh^{2\alpha} \to 0$ and $Nh^{4d} \to \infty$ as N converges to ∞.

C4. The following expected value matrices are all finite and positive definite at β^0:

$$E\left[\frac{\partial^2 \log(f(Y|Z,X;\beta^0))}{\partial\beta\partial\beta^T}\right], \quad E_k\left[\frac{\partial^2 \log(f(Y|Z,X;\beta^0))}{\partial\beta\partial\beta^T}\right], \quad E_k\left[\frac{\partial^2 \log(f(Y|Z,W;\beta^0))}{\partial\beta\partial\beta^T}\right].$$

3. A USEFUL LEMMA

LEMMA 1. *Let $\xi(\underline{y},\underline{z},\underline{w},\underline{x};\beta)$ be a continuous function of $\beta \in \mathcal{B}$ for every $(\underline{y},\underline{z},\underline{w},\underline{x})$, satisfying that:*

(i). $|\xi(\underline{y},\underline{z},\underline{w},\underline{x};\beta)|$ is bounded, uniformly in β, by some function of $(\underline{y},\underline{z},\underline{w},\underline{x})$, denoted by $\widetilde{\xi}(\underline{y},\underline{z},\underline{w},\underline{x})$;

(ii). For $k = 1,\ldots,K$, $\left|\int_{\mathcal{X}} \widetilde{\xi}(\underline{y},\underline{z},\underline{w},x)G(dx|\underline{s},(y,w)\in\Delta_k)\right| < \infty$, almost surely, given $(Y = y, W = w) \in \Delta_k$ and $S = \underline{s}$.

Then

$$\sup_{\beta\in\mathcal{B}}\left|\frac{\sum_{i\in V_k}\xi(\underline{y},\underline{z},\underline{w},X_i;\beta)\phi_h(S_i-\underline{s})}{\sum_{i\in V_k}\phi_h(S_i-\underline{s})} - \int_{\mathcal{X}}\xi(\underline{y},\underline{z},\underline{w},x;\beta)G(dx|\underline{s},(y,w)\in\Delta_k)\right| = O_p(\eta_N),$$

where $\eta_N = (Nh^{2\alpha} + (Nh^{2d})^{-1})^{1/2}$.

Proof. Denote

$$\mu_k(\underline{\mathcal{O}};\beta) = \frac{1}{(n_k+n_{0k})h^d}\sum_{i\in V_k}\xi(\underline{y},\underline{z},\underline{w},X_i;\beta)\phi_h(S_i-\underline{s})$$

and

$$\nu_k(\underline{\mathcal{O}}) = \frac{1}{(n_k+n_{0k})h^d}\sum_{i\in V_k}\phi_h(S_i-\underline{s}),$$

where $\underline{\mathcal{O}}$ denotes $(\underline{y},\underline{z},\underline{w},\underline{s})$ or its some suitable components. Under conditions (i) and (ii), noting that given $(Y,W)\in\Delta_k$, $(X_i,S_i;i\in V_k)$ are i.i.d., and then using the uniform strong law of large numbers and Taylor expansion, we can show that

$$\mu_k(\underline{\mathcal{O}};\beta) \to \int_{\mathcal{X}} \xi(\underline{\mathcal{O}},x;\beta)q(dx,\underline{s}|(y,w)\in\Delta_k),$$

almost surely, uniformly for all $\beta \in \mathcal{B}$, where

$$q(x,s|(y,w)\in\Delta_k) = \frac{d^2\mathrm{Pr}(X\leqslant x, S\leqslant s|(y,w)\in\Delta_k)}{dxds}$$

is the joint density function of (X,S) given $(Y=y, W=w)\in\Delta_k$.

Specially taking $\xi \equiv 1$, we have

$$\nu_k(\underline{\mathcal{O}}) \to \int_{\mathcal{X}} q(dx,\underline{s}|(y,w)\in\Delta_k),$$

almost surely. Hence,

$$\sup_{\beta\in\mathcal{B}}\left|\frac{\mu_k(\underline{\mathcal{O}};\beta)}{\nu_k(\underline{\mathcal{O}})} - \int_{\mathcal{X}}\xi(\underline{\mathcal{O}},x;\beta)G(dx|\underline{s},(y,w)\in\Delta_k)\right| \to 0, \ \ a.s.$$

Using the Lemma 1 in Wang and Wang (2001) and standard kernel estimation theory, one can further derive that

$$\sup_{\beta\in\mathcal{B}}\left|\frac{\mu_k(\underline{\mathcal{O}};\beta)}{\nu_k(\underline{\mathcal{O}})} - \int_{\mathcal{X}}\xi(\underline{\mathcal{O}},x;\beta)G(dx|\underline{s},(y,w)\in\Delta_k)\right| = O_p(\eta_N).$$

Thus we complete the proof.

Furthermore, it is straightforward to conclude that

$$\sup_{\beta\in\mathcal{B}}\left|\sum_{k=1}^{K}\widehat{\pi}_k(\underline{s})\frac{\mu_k(\underline{\mathcal{O}};\beta)}{\nu_k(\underline{\mathcal{O}})} - \int_{\mathcal{X}}\xi(\underline{\mathcal{O}},x;\beta)G(dx|\underline{s})\right| = O_p(\eta_N).$$

4. Useful conclusions

We introduce the following conclusions that are frequently used in the proving process and their derivations are based on Lemma 1.

(i). For $j \in \bar{V}_k$ and S_j fixed, $\widehat{G}(x|S_j) = G(x|S_j) + O_p(\eta_N)$.

(ii). Let $\frac{\partial^a}{\partial \beta^a} f(Y_j|Z_j, W_j; \beta)$ be the a-th derivative of $f(Y_j|Z_j, W_j; \beta)$ with respect to β, then for $j \in \bar{V}_k$

$$\frac{\partial^a}{\partial \beta^a} \widehat{f}(Y_j|Z_j, W_j; \beta) = \frac{\partial^a}{\partial \beta^a} f(Y_j|Z_j, W_j; \beta) + O_p(\eta_N).$$

(iii). For $j \in \bar{V}_k$,

$$\frac{f(Y_j|Z_j, W_j; \beta)}{\widehat{f}(Y_j|Z_j, W_j; \beta)} = 1 + O_p(\eta_N),$$

and

$$\frac{1}{\sqrt{N}} \sum_{k=1}^{K} \sum_{j \in \bar{V}_k} \left\{ \frac{\frac{\partial}{\partial \beta} \widehat{f}(Y_j|Z_j, W_j; \beta)}{f(Y_j|Z_j, W_j; \beta)} - \frac{\frac{\partial}{\partial \beta} f(Y_j|Z_j, W_j; \beta)}{[f(Y_j|Z_j, W_j; \beta)]^2} \widehat{f}(Y_j|Z_j, W_j; \beta) \right\} \times \left\{ \frac{f(Y_j|Z_j, W_j; \beta)}{\widehat{f}(Y_j|Z_j, W_j; \beta)} - 1 \right\}$$
$$= O_p(\eta_N).$$

Note that the first and second results are obvious from Lemma 1. The third result is followed from the Lemmas 3.7 and 3.8 in Weaver (2001).

5. PROOF OF THEOREM 1

Consistency

By selecting suitable function ξ in Lemma 1, it can be shown that

$$\frac{1}{N} \left[\frac{\partial \widehat{U}_F(\beta)}{\partial \beta^T} - \frac{\partial U_F(\beta)}{\partial \beta^T} \right] \to_p 0,$$

uniformly over $\beta \in \mathcal{B}$, as $N \to \infty$, where $U_F(\beta) = \frac{\partial \log L_F(\beta)}{\partial \beta}$.

On the other hand, it follows from the convergence of $-\frac{1}{N} \frac{\partial U_F(\beta)}{\partial \beta^T}$ to $I(\beta)$ in probability, uniformly for $\beta \in \mathcal{B}$, that

$$-\frac{1}{N} \frac{\partial \widehat{U}_F(\beta)}{\partial \beta^T} \to_p I(\beta), \qquad \text{(A.3)}$$

uniformly for $\beta \in \mathcal{B}$, as $N \to \infty$.

It follows from Lemma 1 that

$$\frac{1}{N}\widehat{U}_F(\beta) - \frac{1}{N}U_F(\beta) \to_p 0,$$

uniformly over $\beta \in \mathcal{B}$, as $N \to \infty$. Furthermore, we can show that by strong law of large numbers

$$\begin{aligned}\frac{1}{N}U_F(\beta) = \rho_0\rho_V E\left[\frac{\partial \log(f(Y|Z,X;\beta))}{\partial\beta}\right] + \sum_{k=1}^{K}\rho_k\rho_V E_k\left[\frac{\partial \log(f(Y|Z,X;\beta))}{\partial\beta}\right] \\ + \sum_{k=1}^{K}[\gamma_k^0(1-\rho_0\rho_V) - \rho_k\rho_V]E_k\left[\frac{\partial \log(f(Y|Z,W;\beta))}{\partial\beta}\right] + o_p(1),\end{aligned}$$

and then $\frac{1}{N}U_F(\beta^0) \to_p 0$ follows directly. Hence,

$$\frac{1}{N}\widehat{U}_F(\beta^0) \to_p 0. \tag{A.4}$$

Therefore, combining conditions C1, C4, and (A.3) and (A.4), the consistency of $\widehat{\beta}$ follows from Foutz (1977) and Weaver and Zhou (2005).

Asymptotic Normality

Firstly, we want to evaluate the difference, induced by kernel smother, between the estimated score function $\widehat{U}_F(\beta)$ and score function $U_F(\beta)$. Using arguments in Pepe and Fleming (1991), we have

$$
\begin{aligned}
&\frac{1}{\sqrt{N}}\widehat{U}_F(\beta)\\
&= \frac{1}{\sqrt{N}}\sum_{k=1}^{K}\sum_{j\in\bar{V}_k}\left\{\frac{\frac{\partial}{\partial\beta}\widehat{f}(Y_j|Z_j,W_j;\beta)}{\widehat{f}(Y_j|Z_j,W_j;\beta)} - \frac{\frac{\partial}{\partial\beta}f(Y_j|Z_j,W_j;\beta)}{f(Y_j|Z_j,W_j;\beta)}\right\} + \frac{1}{\sqrt{N}}U_F(\beta)\\
&= \frac{1}{\sqrt{N}}\sum_{k=1}^{K}\sum_{j\in\bar{V}_k}\left\{\frac{\frac{\partial}{\partial\beta}\widehat{f}(Y_j|Z_j,W_j;\beta)}{f(Y_j|Z_j,W_j;\beta)} - \frac{\frac{\partial}{\partial\beta}f(Y_j|Z_j,W_j;\beta)}{[f(Y_j|Z_j,W_j;\beta)]^2}\widehat{f}(Y_j|Z_j,W_j;\beta)\right\}\\
&\quad\times\frac{f(Y_j|Z_j,W_j;\beta)}{\widehat{f}(Y_j|Z_j,W_j;\beta)} + \frac{1}{\sqrt{N}}U_F(\beta)\\
&= \frac{1}{\sqrt{N}}\sum_{k=1}^{K}\sum_{j\in\bar{V}_k}\left\{\frac{\frac{\partial}{\partial\beta}\widehat{f}(Y_j|Z_j,W_j;\beta)}{f(Y_j|Z_j,W_j;\beta)} - \frac{\frac{\partial}{\partial\beta}f(Y_j|Z_j,W_j;\beta)}{[f(Y_j|Z_j,W_j;\beta)]^2}\widehat{f}(Y_j|Z_j,W_j;\beta)\right\}\\
&\quad+\frac{1}{\sqrt{N}}U_F(\beta) + O_p(\eta_N)\\
&\equiv \frac{1}{\sqrt{N}}D_F(\beta) + \frac{1}{\sqrt{N}}U_F(\beta) + O_p(\eta_N)
\end{aligned}
$$

Secondly, we will establish the weak convergence of $\frac{1}{\sqrt{N}}D_F(\beta)$. We rewrite

$$
\begin{aligned}
&\frac{1}{\sqrt{N}}D_F(\beta)\\
&= \frac{1}{\sqrt{N}}\sum_{k=1}^{K}\sum_{j\in\bar{V}_k}\sum_{r=1}^{K}\widehat{\pi}_r(S_j)\frac{\sum_{i\in V_r} M_{X_i,S_i}(Y_j,Z_j,W_j;\beta)\phi_h(S_i-S_j)}{\sum_{i\in V_r}\phi_h(S_i-S_j)}\\
&= \frac{1}{\sqrt{N}}\sum_{r=1}^{K}\sum_{i\in V_r}\sum_{k=1}^{K}\sum_{j\in\bar{V}_k}\frac{N_r(S_j)}{n_{V_r}(S_j)}\frac{n_{\bar{V}_k}(S_j)}{N(S_j)}\frac{M_{X_i,S_i}(Y_j,Z_j,W_j;\beta)\phi_h(S_i-S_j)}{n_{\bar{V}_k}(S_j)}\\
&= \frac{1}{\sqrt{N}}\sum_{r=1}^{K}\frac{\gamma_r^0}{\rho_r\rho_V+\gamma_r^0\rho_0\rho_V}\sum_{i\in V_r}\sum_{k=1}^{K}[\gamma_k^0(1-\rho_0\rho_V)-\rho_k\rho_V]\pi_k(S_i)E_k(M_{X_i,S_i}(Y,Z,W;\beta)|S_i)\\
&\quad+o_p(1).
\end{aligned}
$$

Using Liapounov's central limit theorem and the Cramér-Wold theorem as Weaver and Zhou (2005), we can show that

$$
\frac{1}{\sqrt{N}}D_F(\beta^0) \to_d \mathcal{N}(0, \sum_{k=1}^{K}\frac{(\gamma_k^0)^2}{\rho_k\rho_V+\gamma_k^0\rho_0\rho_V}\Sigma_k(\beta^0)). \qquad \text{(A.5)}
$$

Thirdly,

$$\frac{1}{\sqrt{N}}U_F(\beta) = \frac{1}{\sqrt{N}}\sum_{k=0}^{K}\sum_{i\in\widetilde{V}_k}\frac{\frac{\partial}{\partial\beta}f(Y_i|Z_i,X_i;\beta)}{f(Y_i|Z_i,X_i;\beta)} + \frac{1}{\sqrt{N}}\sum_{k=1}^{K}\sum_{j\in\bar{V}_k}\frac{\frac{\partial}{\partial\beta}f(Y_j|Z_j,W_j;\beta)}{f(Y_j|Z_j,W_j;\beta)}, \tag{A.6}$$

and from here, it is easy to show that $\frac{1}{\sqrt{N}}U_F(\beta^0)$ converges weakly to a normal distribution with mean zero and variance $I(\beta^0)$. On the other hand, since $\frac{1}{\sqrt{N}}D_F(\beta^0)$ can be regarded as a function of $\{X_i, S_i; i \in V\}$ for large N, it is asymptotically independent of the second term at β^0 in (A.6), which are contributions from the nonvalidation data to the true score function. It can be also shown that $\frac{1}{\sqrt{N}}D_F(\beta^0)$ and the first term of (A.6) at β^0 are asymptotically uncorrelated and, since they are each asymptotically normal, independent. Hence, $\frac{1}{\sqrt{N}}D_F(\beta^0)$ and $\frac{1}{\sqrt{N}}U_F(\beta^0)$ are asymptotically independent, and then combining (A.5), we have

$$\frac{1}{\sqrt{N}}\widehat{U}_F(\beta^0) \rightarrow_d \mathcal{N}(0, I(\beta^0) + \sum_{k=1}^{K}\frac{(\gamma_k^0)^2}{\rho_k\rho_V + \gamma_k^0\rho_0\rho_V}\Sigma_k(\beta^0)). \tag{A.7}$$

Finally, using the first-order Taylor series expansion of the estimated score function around the true parameter β_0, we have

$$\sqrt{N}(\widehat{\beta} - \beta^0) = \left[-\frac{1}{N}\frac{\partial\widehat{U}_F(\beta^*)}{\partial\beta^T}\right]^{-1}\left[\frac{1}{\sqrt{N}}\widehat{U}_F(\beta^0)\right], \tag{A.8}$$

where β^* is on the line segment between $\widehat{\beta}$ and β^0. Using conditions C1 and C4, (A.3), and consistency of $\widehat{\beta}$, it is obvious to conclude that as $N \rightarrow \infty$

$$\left[-\frac{1}{N}\frac{\partial\widehat{U}_F(\beta^*)}{\partial\beta^T}\right]^{-1} \rightarrow_p I^{-1}(\beta^0). \tag{A.9}$$

Combining (A.7), (A.8), and (A.9), we have

$$\sqrt{N}(\widehat{\beta} - \beta^0) \rightarrow_d \mathcal{N}\left(0, \Sigma(\beta^0)\right),$$

which is the desired result.

Additionally, with respect to the proof for Theorem 2, since it is obvious to show the consistency of $-\frac{1}{N}\frac{\partial \widehat{U}_F(\widehat{\beta})}{\partial \beta^T}$ for $I(\beta^0)$ from (A.9), it remains to show that $\widehat{\Sigma}_k(\widehat{\beta})$ is a consistent estimator for $\Sigma_k(\beta^0)$ for every k, which can be proved by using (A.5) and Lemma 1.

6. EFFICIENCY COMPARISONS ALONG THE INFORMATIVE STRENGTH

In this section, we want to investigate the effect of informative strength of W for X on the proposed estimates. Table A.1 listed below summarizes the similarity and difference among these estimators with special comments on each estimator. More specifically, the efficiency difference for methods β_{Y_1}, β_{Y_2}, β_{P_1}, and β_{P_2} should be attributed to the study design instead of estimating procedure. However, β_{P_2} and β_W are different estimating procedures under the same two-stage OADS design.

Table A. 1. Summary for different methods compared in simulation study

Method	Design 1st/2nd	Data structure 1st/2nd	Stage of data used in inference	Comment
β_E	SRS	$\{Y, X, Z\}/-$	1st	Least square estimate
β_W	SRS/OADS	$\{Y, Z, W\}/\{X\|(Y, W) \in \Delta_k\}$	2nd only	Inverse probabilty weight
β_{Y_1}	SRS/ODS	$\{Y\}/\{(X, Z)\|Y \in A_j\}$	1st and 2nd	Weaver and Zhou (2005)
β_{Y_2}	SRS/ODS	$\{Y, Z\}/\{X\|Y \in A_j\}$	1st and 2nd	Modified from β_{Y_1}
β_{P_1}	SRS/ODS	$\{Y, Z, W\}/\{X\|Y \in A_j\}$	1st and 2nd	Proposed method reduced from β_{P_2}
β_{P_2}	SRS/OADS	$\{Y, Z, W\}/\{X\|(Y, W) \in \Delta_k\}$	1st and 2nd	Proposed method

Figure A.1 demonstrates the effect of the strength of W, represented by σ, on the efficiency of estimator $\widehat{\beta}_1$, under the methods considered. It displays the relative efficiency of $\widehat{\beta}_{P_1 1}$, $\widehat{\beta}_{P_2 1}$, $\widehat{\beta}_{Y_1 1}$, $\widehat{\beta}_{Y_2 1}$, $\widehat{\beta}_{W1}$, and $\widehat{\beta}_{R1}$ to $\widehat{\beta}_{E1}$, under varying σ from 0 to 1.5 with $allocation(120, 60)$ and cutpoints $(\frac{1}{3}, \frac{2}{3})$. The other parametric settings remain to be as the same as Table 1. Note that among these estimators only $\widehat{\beta}_{P_1 1}$ and $\widehat{\beta}_{P_2 1}$ depend on σ. Clearly, the efficiency loss of both $\widehat{\beta}_{P_1 1}$ and $\widehat{\beta}_{P_2 1}$ increases when σ increases, that is, when W is less informative for X. However, the asymptotic relative efficiency $ARE(\widehat{\beta}_{P_2 1}|\widehat{\beta}_{E1})$, is

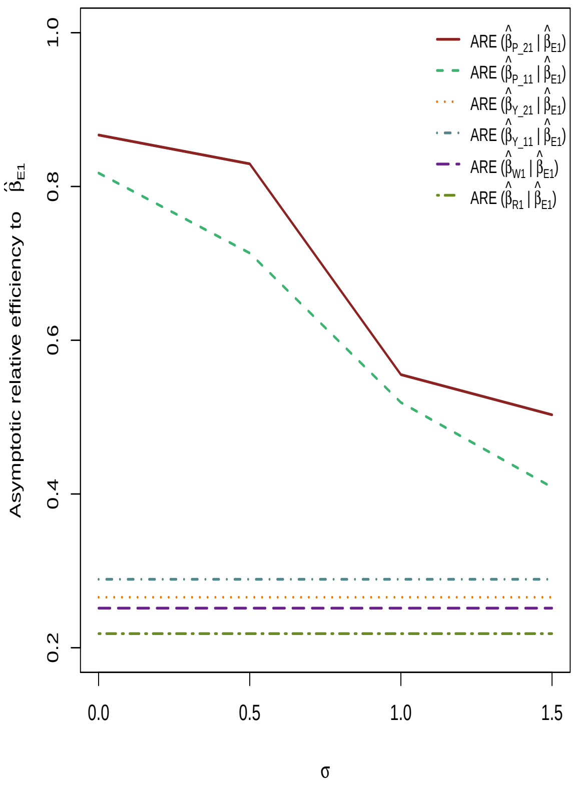

Fig. A.1. Efficiency comparisons of estimator $\widehat{\beta}_1$ along with the informative strength of auxiliary W for covariate X. Y-axis denotes the asymptotic relative efficiency of $\widehat{\beta}_{P_2 1}$, $\widehat{\beta}_{P_1 1}$, $\widehat{\beta}_{Y_2 1}$, $\widehat{\beta}_{Y_1 1}$, $\widehat{\beta}_{W1}$, and $\widehat{\beta}_{R1}$ to $\widehat{\beta}_{E1}$. $ARE(\widehat{\beta}_{P_2 1}|\widehat{\beta}_{E1})$ is defined as the ratio of $var(\widehat{\beta}_{E1})$ over $var(\widehat{\beta}_{P_2 1})$. X-axis denotes the informative strength σ. The larger σ represents weaker information strength of W for X.

always higher than that of the other estimators, which indicates that the proposed two-stage OADS design utilizes W better than the other designs, and that incorporating the auxiliary information into statistical inference can substantially improve the efficiency, especially when W is more informative about X.

REFERENCES

FOUTZ, R. V. (1977). On the unique consistent solution to the likelihood equations. *Journal of the American Statis-*

tical Association **72,** 147–148.

PEPE, M. S. AND FLEMING, T. R. (1991). A nonparametric method for dealing with mismeasured covariate data. *Journal of the American Statistical Association* **86,** 108–113.

WANG, S. AND WANG, C. Y. (2001). A note on kernel assisted estimators in missing covariate regression. *Statistics and Probability Letters* **55,** 439–449.

WEAVER, M. A. (2001). Semiparametric methods for continuous outcome regression models with covariate data from an outcome-dependent subsample, Ph.D. Thesis, University of North Carolina at Chapel Hill.

WEAVER, M. A. AND ZHOU, H. (2005). An estimated likelihood method for continuous outcome regression models with outcome-dependent sampling. *Journal of the American Statistical Association* **100,** 459–469.