Web-based Supplementary Materials for "Semiparametric estimation of the covariate-specific ROC curve in presence of ignorable verification bias"

by Danping Liu and Xiao-Hua Zhou

1 Technical details of the proof

Regularity conditions for Theorem 1.

C1. There exist some θ_1, α_1 and α_2 satisfying that $EU_{12,i} = 0$, $EB_{1i} = 0$ and $EB_{2i} = 0$.

C2. $EU_{12,i}$ is second-order differentiable with respect to θ_1, α_1 and α_2, and all the derivatives are uniformly bounded in the neighborhood of θ_1.

C3. The information matrices I, K_1 and K_2 are positively definite.

Proof of Theorem 1.

By the Taylor expansion, we have

$$\begin{aligned} 0 &= \frac{1}{\sqrt{n}}\sum_i S_{12,i}(\hat{\theta}_1, \hat{\alpha}_1, \hat{\alpha}_2) \\ &= \frac{1}{\sqrt{n}}\sum_i S_{12,i}(\theta_1, \alpha_1, \alpha_2) + \left(\frac{1}{n}\sum_i \frac{\partial}{\partial\theta_1^T} S_{12,i}(\theta_1, \alpha_1, \alpha_2)\right)\sqrt{n}\left(\hat{\theta}_1 - \theta_1\right) \\ &\quad + \left(\frac{1}{n}\sum_i \frac{\partial}{\partial\alpha_1^T} S_{12,i}(\theta_1, \alpha_1, \alpha_2)\right)\sqrt{n}\left(\hat{\alpha}_1 - \alpha_1\right) \\ &\quad + \left(\frac{1}{n}\sum_i \frac{\partial}{\partial\alpha_2^T} S_{12,i}(\theta_1, \alpha_1, \alpha_2)\right)\sqrt{n}\left(\hat{\alpha}_2 - \alpha_2\right) + o_p(1). \end{aligned} \quad (1)$$

By the law of large numbers, $-\frac{1}{n}\sum_i \frac{\partial}{\partial\theta_1^T} S_{12,i}(\theta_1, \alpha_1, \alpha_2) \stackrel{a.s}{\to} I$, $-\frac{1}{n}\sum_i \frac{\partial}{\partial\alpha_1^T} S_{12,i}(\theta_1, \alpha_1, \alpha_2) \stackrel{a.s}{\to} J_1$, and $-\frac{1}{n}\sum_i \frac{\partial}{\partial\alpha_2^T} S_{12,i}(\theta_1, \alpha_1, \alpha_2) \stackrel{a.s}{\to} J_2$. Then $\sqrt{n}\left(\hat{\alpha}_1 - \alpha_1\right)$ and $\sqrt{n}\left(\hat{\alpha}_2 - \alpha_2\right)$ can be written in the form of their influence functions:

$$\sqrt{n}\,(\hat{\alpha}_1 - \alpha_1) = \frac{1}{\sqrt{n}} \sum_i K_1^{-1} B_{1i} + o_p(1), \quad (2)$$

$$\sqrt{n}\,(\hat{\alpha}_2 - \alpha_2) = \frac{1}{\sqrt{n}} \sum_i K_2^{-1} B_{2i} + o_p(1) \quad (3)$$

Plugging into (1) the influence functions (2) and (3), as well as the almost sure convergence components, we have

$$\begin{aligned} 0 &= \frac{1}{\sqrt{n}} \sum_i U_{12,i}(\theta_1, \alpha_1, \alpha_2) - I\sqrt{n}\left(\hat{\theta}_1 - \theta_1\right) \\ &\quad -\frac{1}{\sqrt{n}} \sum_i \left[J_1 K_1^{-1} B_{1i} + J_2 K_2^{-1} B_{2i}\right] + o_p(1) \end{aligned}$$

Hence the influence functions for $\sqrt{n}\left(\hat{\theta}_1 - \theta_1\right)$ is

$$\sqrt{n}\left(\hat{\theta}_1 - \theta_1\right) = \frac{1}{\sqrt{n}} \sum_i I^{-1} Q_i + o_p(1),$$

which completes the proof.

Proof of Lemma 1.

For the influence function of $\hat{G}_0^{-1}(1-t)$, note that

$$\begin{aligned}
0 &= \frac{1}{\sqrt{n}}\sum_i S_{5i}(\hat{s},\hat{\theta}_1,\hat{\alpha}_1,\hat{\alpha}_2) \\
&= \frac{1}{\sqrt{n}}\sum_i S_{5i}(s,\theta_1,\alpha_1,\alpha_2) + \left(\frac{1}{n}\sum_i \frac{\partial}{\partial s}S_{5i}(s,\theta_1,\alpha_1,\alpha_2)\right)\sqrt{n}\,(\hat{s}-s) \\
&\quad + \left(\frac{1}{n}\sum_i \frac{\partial}{\partial \theta_1^T}S_{5i}(\theta_1,\alpha_1,\alpha_2)\right)\sqrt{n}\left(\hat{\theta}_1-\theta_1\right) \\
&\quad + \left(\frac{1}{n}\sum_i \frac{\partial}{\partial \alpha_1^T}S_{5i}(\theta_1,\alpha_1,\alpha_2)\right)\sqrt{n}\,(\hat{\alpha}_1-\alpha_1) \\
&\quad + \left(\frac{1}{n}\sum_i \frac{\partial}{\partial \alpha_2^T}S_{5i}(\theta_1,\alpha_1,\alpha_2)\right)\sqrt{n}\,(\hat{\alpha}_2-\alpha_2) + o_p(1) \\
&= \frac{1}{\sqrt{n}}\sum_i S_{5i}(s,\theta_1,\alpha_1,\alpha_2) + \left(\frac{\partial}{\partial s}ES_{5i}(s,\theta_1,\alpha_1,\alpha_2)\right)\sqrt{n}\,(\hat{s}-s) \\
&\quad + \left(\frac{\partial}{\partial \theta_1^T}ES_{5i}(\theta_1,\alpha_1,\alpha_2)\right)\sum_i I^{-1}Q_i + \left(\frac{\partial}{\partial \alpha_1^T}ES_{5i}(\theta_1,\alpha_1,\alpha_2)\right)\sum_i K_1^{-1}B_{1i} \\
&\quad + \left(\frac{\partial}{\partial \alpha_2^T}ES_{5i}(\theta_1,\alpha_1,\alpha_2)\right)\sum_i K_2^{-1}B_{2i} + o_p(1).
\end{aligned}$$

The second equality comes from the Taylor expansion; the third equality comes from the law of large numbers and the influence function of $\hat{\theta}_1$, $\hat{\alpha}_1$ and $\hat{\alpha}_2$. Some rearrangements give us the final expression of A_{1i}:

$$\begin{aligned}
A_{1i}(t) &= \left(-\frac{\partial}{\partial s}ES_{5i}\right)^{-1}\left[S_{5i} + \left(\frac{\partial}{\partial \theta_1^T}ES_{5i}\right)I^{-1}Q_i + \left(\frac{\partial}{\partial \alpha_1^T}ES_{5i}\right)K_1^{-1}B_{1i}\right. \\
&\quad \left. + \left(\frac{\partial}{\partial \alpha_2^T}ES_{5i}\right)K_2^{-1}B_{2i}\right].
\end{aligned}$$

The influence function for A_{2i} is obtained using exactly the same techniques.

Proof of Theorem 2.

Denote $u = \frac{\sigma(x,0;\gamma)}{\sigma(x,1;\gamma)}G_0^{-1}(1-t) + \frac{\mu(x,0;\beta)-\mu(x,1;\beta)}{\sigma(x,1;\gamma)}$, which is also referred to as the "placement value" in some literatures. This is interpreted as the standardized test result for cases to the control distribution. Replacing β, γ and G_0^{-1} with the estimated version yields the estimated placement value $\hat{u}$. Note that $\sqrt{n}\left(\widehat{ROC}_x(t) - ROC_x(t)\right) = \sqrt{n}\left(G_1(u) - G_1(\hat{u})\right) - \sqrt{n}\left(\hat{G}_1(\hat{u}) - G_1(\hat{u})\right)$. The first part can be further written as the influence function of $\hat{\beta}$, $\hat{\gamma}$ and $\hat{G}_0^{-1}$:

$$\sqrt{n}\left(G_1(u) - G_1(\hat{u})\right) = G_1'(u)\sqrt{n}\left(u - \hat{u}\right) + o_p(1)$$

$$\begin{aligned} &= G_1'(u)\left[\frac{\partial u}{\partial \theta_1^T}\sqrt{n}\left(\theta_1 - \hat{\theta}_1\right) + \frac{\partial u}{\partial s}\sqrt{n}\left(s - \hat{s}\right)\right] + o_p(1) \\ &= -\frac{1}{\sqrt{n}}\sum_i G_1'(u)\left[\frac{\partial u}{\partial \theta_1^T}I^{-1}Q_i + \frac{\partial u}{\partial s}A_{1i}(t)\right] + o_p(1). \end{aligned}$$

By the uniform law of large numbers, the second term can be expresses as the influence function of $\hat{G}_1$:

$$\sqrt{n}\left(\hat{G}_1(\hat{u}) - G_1(\hat{u})\right) = \sqrt{n}\left(\hat{G}_1(u) - G_1(u)\right) + o_p(1)$$

$$= \frac{1}{\sqrt{n}}\sum_i A_{2i}(u) + o_p(1).$$

Hence we have

$$\sqrt{n}\left(\widehat{ROC}_x(t) - ROC_x(t)\right) = -\frac{1}{\sqrt{n}}\sum_i\left\{G_1'(u)\left[\frac{\partial u}{\partial \theta_1^T}I^{-1}Q_i + \frac{\partial u}{\partial s}A_{1i}(t)\right] + A_{2i}(u)\right\} + o_p(1),$$

and the asymptotic variance is given by

$$\Omega_2 = \text{var}\left\{G_1'(u)\left[\frac{\partial u}{\partial \theta_1^T}I^{-1}Q_i + \frac{\partial u}{\partial s}A_{1i}(t)\right] + A_{2i}(u)\right\}.$$

Proof of Corollary 1.

(1) When $\hat{\rho}_i$ is estimated with $\sqrt{n}$ consistency, it suffices to prove that the partial derivative of S_{ki} with respect to α_2 has zero expectation. Note that

$$\begin{aligned} \frac{\partial}{\partial \alpha_2^T}S_{ki} &= \frac{\partial}{\partial \alpha_2^T}\left(\frac{V_i}{\pi_i}U_{ki} + \left(1 - \frac{V_i}{\pi_i}\right)E_{D_i|T_i,X_i}U_{ki}\right) \\ &= \left[-\frac{V_i}{\pi_i^2}\left(U_{ki} - E_{D_i|T_i,X_i}U_{ki}\right)\right]\frac{\partial \pi_i}{\partial \alpha_2^T}. \end{aligned}$$

Conditional on V_i, $U_{ki} - E_{D_i|T_i,X_i} U_{ki}$ has zero expectation if the disease model is correctly specified.

(2) When $\hat{\pi}_i$ is estimated with $\sqrt{n}$ consistency, it suffices to prove that the partial derivative of S_{ki} with respect to α_1 has zero expectation. Similarly we can write out the partial derivative as

$$\begin{aligned} \frac{\partial}{\partial \alpha_1^T} S_{ki} &= \frac{\partial}{\partial \alpha_1^T} \left(\frac{V_i}{\pi_i} U_{ki} + \left(1 - \frac{V_i}{\pi_i}\right) E_{D_i|T_i,X_i} U_{ki} \right) \\ &= \left(1 - \frac{V_i}{\pi_i}\right) \frac{\partial}{\partial \alpha_1^T} E_{D_i|T_i,X_i} U_{ki}. \end{aligned}$$

As the verification model is correctly specified, $1 - \frac{V_i}{\pi_i}$ has zero expectation, and hence $\frac{\partial}{\partial \alpha_1^T} S_{ki}$ has zero expectation.

2 Additional simulation results

We conduct further simulations to study the efficiency of the three proposed estimators. The proposed estimators are also compared with the counterparts with the true selection and disease probability π and ρ. We consider a total of eight estimators:

(1) DR_1: the DR estimator with both estimated $\hat{\pi}$ and $\hat{\rho}$.

(2) DR_2: the DR estimator with estimated $\hat{\rho}$ and known π.

(3) DR_3: the DR estimator with estimated $\hat{\pi}$ and known ρ.

(4) DR_4: the DR estimator with both known π and ρ.

(5) IPW_1: the IPW estimator with estimated $\hat{\pi}$.

(6) IPW_2: the IPW estimator with known π.

(7) IB_1: the IB estimator with estimated $\hat{\rho}$.

(8) IB_2: the IB estimator with known ρ.

The sample size is taken to be $n = 1000$, and the parameters are the same as in simulation one. The results are summarized in Table 1. The DR_1 estimator serves as the benchmark to compute the relative efficiency. Obviously DR_2, DR_3 and DR_4 estimators are close to DR_1 estimator and the relative efficiency is close to 1. This confirms the statement in Corollary 1: when both selection and disease models are correctly specified, estimating the two probabilities π and ρ does not introduce extra variability in the ROC curve estimation. We note that IPW estimator is only about 26-75% efficient as compared to the DR_1 estimator, no matter whether π is estimated or known. The IB_1 estimator is about 5-28% more efficient than DR_1 estimator. This moderate improvement of efficiency is subject to the risk of mis-specification of the disease model, and hence is less robust than the DR_1 estimator. It is not surprising that IB_2 estimator is much more efficient than the DR_1 estimator, because knowing the disease probability indeed provides much information.

Table 1: Comparison of different DR, IPW and IB estimators.

	$RMSE\ (Rel.eff)$							
	DR_1	DR_2	DR_3	DR_4	IPW_1	IPW_2	IB_1	IB_2
β_0	0.091 (1.00)	0.090 (1.01)	0.090 (1.00)	0.090 (1.01)	0.132 (0.47)	0.146 (0.39)	0.086 (1.10)	0.070 (1.67)
β_1	0.152 (1.00)	0.151 (1.01)	0.151 (1.01)	0.150 (1.02)	0.176 (0.75)	0.179 (0.72)	0.138 (1.21)	0.074 (4.19)
β_2	0.130 (1.00)	0.129 (1.00)	0.130 (0.99)	0.130 (1.00)	0.206 (0.40)	0.204 (0.40)	0.124 (1.09)	0.104 (1.55)
β_3	0.112 (1.00)	0.112 (1.00)	0.113 (0.99)	0.112 (1.00)	0.191 (0.34)	0.191 (0.35)	0.105 (1.14)	0.089 (1.57)
β_4	0.187 (1.00)	0.187 (1.00)	0.189 (0.99)	0.188 (0.99)	0.242 (0.60)	0.239 (0.62)	0.172 (1.18)	0.116 (2.62)
β_5	0.164 (1.00)	0.163 (1.01)	0.164 (1.00)	0.162 (1.02)	0.215 (0.58)	0.214 (0.59)	0.145 (1.28)	0.096 (2.93)
γ_0	0.036 (1.00)	0.036 (1.00)	0.036 (1.01)	0.036 (1.01)	0.070 (0.26)	0.071 (0.26)	0.035 (1.05)	0.030 (1.43)
γ_1	0.064 (1.00)	0.064 (1.00)	0.063 (1.03)	0.063 (1.03)	0.085 (0.57)	0.085 (0.57)	0.059 (1.18)	0.040 (2.52)
$ROC_{(1,0.5)}(0.1)$	0.082 (1.00)	0.082 (1.00)	0.082 (1.00)	0.082 (1.01)	0.115 (0.51)	0.118 (0.48)	0.075 (1.20)	0.062 (1.77)
$ROC_{(1,0.5)}(0.2)$	0.055 (1.00)	0.055 (1.00)	0.055 (0.99)	0.055 (1.00)	0.087 (0.40)	0.089 (0.38)	0.050 (1.21)	0.041 (1.81)