The International Journal of Biostatistics

Volume 4, Issue 1

2008

Article 10

Comment: Improved Local Efficiency and Double Robustness

Zhiqiang Tan, *Rutgers University*

Recommended Citation:

Tan, Zhiqiang (2008) "Comment: Improved Local Efficiency and Double Robustness," *The International Journal of Biostatistics*: Vol. 4: Iss. 1, Article 10.

DOI: 10.2202/1557-4679.1109

Comment: Improved Local Efficiency and Double Robustness

Zhiqiang Tan

Abstract

For missing data and causal inference problems, Rubin and van der Laan (2008) proposed estimators to achieve so-called improved local efficiency. We show that their estimators agree with existing estimators in the case of linear models, point out that one particular version of their estimators is also doubly robust, and suggest an extension for where the propensity score is estimated.

KEYWORDS: causal inference, double robustness, local efficiency

Author Notes: Department of Statistics, Rutgers University. Address: 110 Frelinghuysen Road, Piscataway, NJ 08854. e-mail: ztan@stat.rutgers.edu.

1 Introduction

For a population, let X be a vector of (pretreatment) covariates, T be the treatment status, Y be the observed outcome given by $(1 - T)Y_0 + TY_1$, where (Y_0, Y_1) are potential outcomes. The observed data consist of independent and identically distributed copies $O_i = (X_i, T_i, Y_i)$, $i = 1, \dots, n$. Assume that T and (Y_0, Y_1) are conditionally independent given X . The objective is to estimate

$$\mu_1 = E(Y_1), \quad \mu_0 = E(Y_0),$$

and their difference, $\mu_1 - \mu_0$, which gives the average causal effect.

For estimating μ_1 , the following results are established in the literature (e.g. van der Laan and Robins 2003):

(R1) The efficient influence curve is given by $D(O|\pi, m_1) - \mu_1$, where

$$D(O|\pi, m_1) = \frac{T}{\pi(X)}Y - \left(\frac{T}{\pi(X)} - 1\right)m_1(X),$$

depending on the true regression function $m_1(X) = E(Y|T = 1, X)$ and the true propensity score $\pi(X) = P(T = 1|X)$.

(R2) $D(O|\pi, m_1^*)$ remains to have mean μ_1 given the true propensity score $\pi(X)$ and a possibly misspecified regression function $m_1^*(X)$.

(R3) $D(O|\pi^*, m_1)$ remains to have mean μ_1 given the true regression function $m_1(X)$ and a possibly misspecified propensity score $\pi^*(X)$.

The method of Rubin and van der Laan (RV) makes use of the first two results. Suppose that our knowledge about $\pi(X)$ is given by $\pi^*(X)$. Until Section 4, assume that $\pi^*(X)$ is correct and equals $\pi(X)$ as in controlled experiments. For an outcome regression model

$$m_1(X) = m_1(X; \alpha_1)$$

with α_1 a vector of unknown parameters, RV's proposal is to find

$$\hat{\alpha}_1 = \operatorname{argmin}_{\alpha_1} \tilde{E} \left[\left\{ \frac{T}{\pi^*(X)}Y - \left(\frac{T}{\pi^*(X)} - 1 \right) m_1(X; \alpha_1) \right\}^2 \right]$$

and estimate μ_1 by

$$\hat{\mu}_1 = \tilde{E} \left[\frac{T}{\pi^*(X)}Y - \left(\frac{T}{\pi^*(X)} - 1 \right) m_1(X; \hat{\alpha}_1) \right]$$

where $\tilde{E}(\cdot)$ denotes sample average. The estimator $\hat{\mu}_1$ enjoys the following properties:

- (P1) $\hat{\mu}_1$ is locally efficient, i.e., its influence function equals $D(O|\pi, m_1) - \mu_1$ if the model $m_1(X; \alpha_1)$ is correctly specified, and
- (P2) $\hat{\mu}_1$ asymptotically achieves the minimum variance among the class of estimators indexed by α_1

$$\bar{\mu}_{1, \alpha_1} = \tilde{E} \left[\frac{T}{\pi^*(X)} Y - \left(\frac{T}{\pi^*(X)} - 1 \right) m_1(X; \alpha_1) \right],$$

whether or not the model $m_1(X; \alpha_1)$ is correctly specified.

The two properties together can be called “improved local efficiency.”

In contrast, a common locally efficient estimator of μ_1 is given by (e.g. van der Laan and Robins 2003)

$$\check{\mu}_1 = \tilde{E} \left[\frac{T}{\pi^*(X)} Y - \left(\frac{T}{\pi^*(X)} - 1 \right) m_1(X; \check{\alpha}_1) \right],$$

where $\check{\alpha}_1$ is obtained by fitting the model $m_1(X; \alpha_1)$ to $\{(X_i, Y_i) : T_i = 0\}$ by least squares or, if Y is binary, by maximum likelihood. The estimators $\hat{\mu}_1$ and $\check{\mu}_1$ are equally efficient if the model $m_1(X; \alpha_1)$ is correctly specified, but $\hat{\mu}_1$, due to property P2, gains efficiency over $\check{\mu}_1$ if the model $m_1(X; \alpha_1)$ is misspecified.

Theorem 1 of RV can be simplified to the identity [assuming that $\pi^*(X) = \pi(X)$]

$$\begin{aligned} E[D^2(O|\pi^*, m_1(\cdot; \alpha_1))] &= E \left[\left\{ \frac{T}{\pi^*(X)} Y - \left(\frac{T}{\pi^*(X)} - 1 \right) m_1(X; \alpha_1) \right\}^2 \right] \\ &= E \left[\frac{T}{\pi^*(X)} Y^2 \right] + E \left[\frac{T}{\pi^*(X)} \left(\frac{T}{\pi^*(X)} - 1 \right) \{Y - m_1(X; \alpha_1)\}^2 \right], \end{aligned}$$

where the first term is free of α_1 and the second term is a weighted mean squared error. By minimizing the second term, RV proposed an alternative estimator to $\hat{\alpha}_1$:

$$\tilde{\alpha}_1 = \operatorname{argmin}_{\alpha_1} \tilde{E} \left[\frac{T}{\pi^*(X)} \left(\frac{T}{\pi^*(X)} - 1 \right) \{Y - m_1(X; \alpha_1)\}^2 \right].$$

The resulting estimator of μ_1 is

$$\tilde{\mu}_1 = \tilde{E} \left[\frac{T}{\pi^*(X)} Y - \left(\frac{T}{\pi^*(X)} - 1 \right) m_1(X; \tilde{\alpha}_1) \right].$$

The estimator $\tilde{\mu}_1$ has properties P1 and P2 described above as $\hat{\mu}_1$ does.

The following is a summary of the main points of our discussion.

- (1) If the model $m_1(X; \alpha_1)$ is linear in α_1 , then $\hat{\mu}_1$ agrees with the regression estimator of Robins, Rotnitzky, and Zhao (1995) and Tsiatis (2006) and $\tilde{\mu}_1$ agrees with the regression estimator of Tan (2006).
- (2) For linear outcome regression models, the estimator of the vector (μ_0, μ_1) of Tan (2006) achieves improved local efficiency. Therefore, the resulting estimator of $\mu_1 - \mu_0$ achieves improved local efficiency.
- (3) $\tilde{\mu}_1$ is doubly robust, but $\hat{\mu}_1$ is not. An estimator is doubly robust if it remains consistent if either a model for $\pi(X)$ is correct or a model for $m_1(X)$ is correct.
- (4) We propose a computationally convenient extension of $\hat{\mu}_1$ and $\tilde{\mu}_1$ in the case where the propensity score $\pi(X)$ need to be estimated.
- (5) Even in the case of binary outcomes, it is appropriate to use a linear outcome regression model based on the fitted value $m_1(X; \hat{\alpha}_1)$ or a linearly extended outcome regression model.

2 Linear outcome regression model

Suppose that the outcome regression model $m_1(X; \alpha_1)$ is linear in α_1 , i.e.,

$$m_1(X; \alpha_1) = \alpha_1^\top g(X)$$

where $g(X)$ is a vector of known functions of X , typically including the constant.

The first-order condition in the minimization leading to $\hat{\alpha}_1$ implies that

$$\tilde{E}\left[\left\{\frac{T}{\pi^*(X)}Y - \left(\frac{T}{\pi^*(X)} - 1\right)\hat{\alpha}_1^\top g(X)\right\}\left(\frac{T}{\pi^*(X)} - 1\right)g(X)\right] = 0,$$

and $\hat{\alpha}_1 = \tilde{E}^{-1}(\xi^* \xi^{*\top}) \tilde{E}(\xi^* \eta^*)$ is the vector of estimated regression coefficients of $\eta^* = \pi^{*-1}(X)TY$ against $\xi^* = (\pi^{*-1}(X)T - 1)g(X)$. Therefore,

$$\hat{\mu}_1 = \tilde{E}(\eta^*) - \hat{\alpha}_1^\top \tilde{E}(\xi^*)$$

agrees with the regression estimator of Robins, Rotnitzky, and Zhao (1995) and Tsiatis (2006). Similarly, the first-order condition in the minimization leading to $\tilde{\alpha}_1$ implies that

$$\tilde{E}\left[\left\{\frac{T}{\pi^*(X)}Y - \frac{T}{\pi^*(X)}\tilde{\alpha}_1^\top g(X)\right\}\left(\frac{T}{\pi^*(X)} - 1\right)g(X)\right] = 0,$$

and $\tilde{\alpha}_1 = \tilde{E}^{-1}(\xi^* \zeta^{*\top}) \tilde{E}(\xi^* \eta^*)$ with $\zeta = \pi^{*-1}(X) Tg(X)$. Therefore,

$$\tilde{\mu}_1 = \tilde{E}(\eta) - \tilde{\alpha}_1^\top \tilde{E}(\xi^*)$$

agrees with the regression estimator of Tan (2006). The estimators $\hat{\mu}_1$ and $\tilde{\mu}_1$ are algebraically similar except that different regression coefficients, $\hat{\alpha}_1$ and $\tilde{\alpha}_1$, are used in front of ξ^* . Under the assumption that $\pi^*(X)$ is correct, $\hat{\alpha}_1$ and $\tilde{\alpha}_1$ converge to the same limit in probability, and $\hat{\mu}_1$ and $\tilde{\mu}_1$ have the same influence function.

It is interesting that $\hat{\mu}_1$ and $\tilde{\mu}_1$ are considered in different ways by RV and Tan (2006). The estimators $\hat{\mu}_1$ and $\tilde{\mu}_1$ are derived by RV from two equivalent representations of $E[D^2(O|\pi, m_1(\cdot; \alpha_1))]$, and are in fact asymptotically equivalent in the setting of RV where $\pi^*(X)$ is correct. In contrast, an important point of Tan (2006) is that $\hat{\alpha}_1$ is the classical least-squares coefficient in the regression of η^* on ξ^* , whereas $\tilde{\alpha}_1$ is specially constructed by exploiting the structure of the regressors ξ^* , so that $\tilde{\mu}_1$ is doubly robust but $\hat{\mu}_1$ is not. See Section 4 for further discussion of double robustness.

3 Estimation in two treatment arms

For causal inference, estimation of both μ_1 and μ_0 are of interest. Consider an outcome regression model

$$m_0(X) = m_0(X; \alpha_0)$$

where $m_0(X) = E(Y|T = 0, X)$, and α_0 is a vector of unknown parameters. Estimators of μ_0 , $\hat{\mu}_0$ and $\tilde{\mu}_0$, can be defined by substituting $\hat{\alpha}_0$ and $\tilde{\alpha}_0$ for α_0 in

$$\bar{\mu}_{0,\alpha_0} = \tilde{E} \left[\frac{1-T}{1-\pi^*(X)} Y - \left(\frac{1-T}{1-\pi^*(X)} - 1 \right) m_0(X; \alpha_0) \right]$$

where $\hat{\alpha}_0$ and $\tilde{\alpha}_0$ are obtained similarly as $\hat{\alpha}_1$ and $\tilde{\alpha}_1$:

$$\begin{aligned} \hat{\alpha}_0 &= \operatorname{argmin}_{\alpha_0} \tilde{E} \left[\left\{ \frac{1-T}{1-\pi^*(X)} Y - \left(\frac{1-T}{1-\pi^*(X)} - 1 \right) m_0(X; \alpha_0) \right\}^2 \right], \\ \tilde{\alpha}_0 &= \operatorname{argmin}_{\alpha_0} \tilde{E} \left[\frac{1-T}{1-\pi^*(X)} \left(\frac{1-T}{1-\pi^*(X)} - 1 \right) \{Y - m_0(X; \alpha_0)\}^2 \right]. \end{aligned}$$

By symmetry, $\hat{\mu}_0$ and $\tilde{\mu}_0$ have properties corresponding to (i) and (ii).

The estimators $(\hat{\mu}_0, \hat{\mu}_1)$ and $(\tilde{\mu}_0, \tilde{\mu}_1)$ of the vector (μ_0, μ_1) are locally efficient. However, as pointed out by RV, $(\hat{\mu}_0, \hat{\mu}_1)$ and $(\tilde{\mu}_0, \tilde{\mu}_1)$ do not necessarily achieve minimum variance in the order on positive definite matrices among the

class of estimators $(\bar{\mu}_{1,\alpha_1}, \bar{\mu}_{0,\alpha_0})$. In fact, there may exist no minimum variance in this class. As a result, $\hat{\mu}_1 - \hat{\mu}_0$ and $\tilde{\mu}_1 - \tilde{\mu}_0$ are not guaranteed to achieve the minimum variance among the class of estimators $\bar{\mu}_{1,\alpha_1} - \bar{\mu}_{0,\alpha_0}$. The minimum variance of $\bar{\mu}_{1,\alpha_1} - \bar{\mu}_{0,\alpha_0}$ is achieved asymptotically by the estimator $\widehat{\mu_1 - \mu_0}$, obtained by substituting

$$\begin{aligned} (\widehat{\alpha_0}, \widehat{\alpha_1}) = \operatorname{argmin}_{(\alpha_0, \alpha_1)} \tilde{E} \left[\left\{ \frac{T}{\pi^*(X)} Y - \left(\frac{T}{\pi^*(X)} - 1 \right) m_1(X; \alpha_1) \right. \right. \\ \left. \left. - \frac{1-T}{1-\pi^*(X)} Y + \left(\frac{1-T}{1-\pi^*(X)} - 1 \right) m_0(X; \alpha_0) \right\}^2 \right] \end{aligned}$$

for (α_0, α_1) in $\bar{\mu}_{1,\alpha_1} - \bar{\mu}_{0,\alpha_0}$ as in RV (Section 4.1). That is, $\widehat{\mu_1 - \mu_0} \neq \hat{\mu}_1 - \hat{\mu}_0$.

For linear outcome regression models, $m_1(X; \alpha_1) = \alpha_1^\top g(X)$ and $m_0(X; \alpha_0) = \alpha_0^\top g(X)$, Tan (2006) considered an extended class of estimators and obtained an estimator of (μ_0, μ_1) that achieves improved local efficiency among that class. Specifically, let $m_1(X; \alpha_{10}, \alpha_{11}) = -\pi^*(X)/[1 - \pi^*(X)]\alpha_{10}^\top g(X) + \alpha_{11}^\top g(X)$ and $m_0(X; \alpha_{00}, \alpha_{01}) = \alpha_{00}^\top g(X) - [1 - \pi^*(X)]/\pi^*(X)\alpha_{01}^\top g(X)$ such that

$$\begin{aligned} \left(\frac{T}{\pi^*(X)} - 1 \right) m_1(X; \alpha_{10}, \alpha_{11}) &= \left(\frac{1-T}{1-\pi^*(X)} - 1 \right) \alpha_{10}^\top g(X) + \\ &\quad \left(\frac{T}{\pi^*(X)} - 1 \right) \alpha_{11}^\top g(X), \\ \left(\frac{1-T}{1-\pi^*(X)} - 1 \right) m_0(X; \alpha_{00}, \alpha_{01}) &= \left(\frac{1-T}{1-\pi^*(X)} - 1 \right) \alpha_{00}^\top g(X) + \\ &\quad \left(\frac{T}{\pi^*(X)} - 1 \right) \alpha_{01}^\top g(X) \end{aligned}$$

are linear combinations of a common set of variables $(\pi^{*-1}(X)T - 1)g(X)$ and $((1 - \pi^*(X))^{-1}(1 - T) - 1)g(X)$. With $m_1(X; \alpha_{10}, \alpha_{11})$ and $m_0(X; \alpha_{00}, \alpha_{01})$ in place of $m_1(X; \alpha_1)$ and $m_0(X; \alpha_0)$, define $(\bar{\mu}_{0,\alpha_{00},\alpha_{01}}, \bar{\mu}_{1,\alpha_{01},\alpha_{11}})$ as $(\bar{\mu}_{1,\alpha_1}, \bar{\mu}_{0,\alpha_0})$, and define $(\hat{\mu}_0^\dagger, \hat{\mu}_1^\dagger)$ and $(\tilde{\mu}_0^\dagger, \tilde{\mu}_1^\dagger)$ as $(\hat{\mu}_0, \hat{\mu}_1)$ and $(\tilde{\mu}_0, \tilde{\mu}_1)$. Then the class of estimators $(\bar{\mu}_{0,\alpha_{00},\alpha_{01}}, \bar{\mu}_{1,\alpha_{01},\alpha_{11}})$ contains the class of estimators $(\bar{\mu}_{1,\alpha_1}, \bar{\mu}_{0,\alpha_0})$. The estimators $(\hat{\mu}_0^\dagger, \hat{\mu}_1^\dagger)$ and $(\tilde{\mu}_0^\dagger, \tilde{\mu}_1^\dagger)$ asymptotically achieve the minimum variance of $(\bar{\mu}_{0,\alpha_{00},\alpha_{01}}, \bar{\mu}_{1,\alpha_{01},\alpha_{11}})$.

The extended models $m_1(X; \alpha_{10}, \alpha_{11})$ and $m_0(X; \alpha_{00}, \alpha_{01})$ become equivalent to $m_1(X; \alpha_1)$ and $m_0(X; \alpha_0)$, with $-\pi^*/(1 - \pi^*)\alpha_{10} + \alpha_{11} = \alpha_1$ and $\alpha_{00} - (1 - \pi^*)/\pi^*\alpha_{01} = \alpha_0$, if $\pi^*(X)$ is a constant, free of X . In this special case, the estimators $(\hat{\mu}_0, \hat{\mu}_1)$ and $(\tilde{\mu}_0, \tilde{\mu}_1)$ in fact asymptotically achieve the minimum variance among the class of estimators $(\bar{\mu}_{1,\alpha_1}, \bar{\mu}_{0,\alpha_0})$.

If the model $m_1(X; \alpha_1)$ is not linear in α_1 and $m_0(X; \alpha_0)$ not linear in α_0 , this sort of augmentation does not seem to yield improved local efficiency jointly for (μ_0, μ_1) . It remains an interesting open problem whether and how this property can be achieved for nonlinear outcome regression models.

4 Double robustness

We relax the assumption that our knowledge $\pi^*(X)$ about $\pi(X)$ is correct, and suppose that $\pi^*(X)$ may differ from $\pi(X)$. Then $\hat{\mu}_1$ and $\tilde{\mu}_1$ are generally inconsistent, and moreover converge to different limits. The limits of $\hat{\alpha}_1$ and $\tilde{\alpha}_1$ are no longer the same, because Theorem 1 of RV does not hold if $\pi^*(X) \neq \pi(X)$.

Nevertheless, $\tilde{\mu}_1$ remains consistent when the model $m_1(X; \alpha_1)$ is correctly specified, that is, $\tilde{\mu}_1$ is doubly robust. Under standard regularity conditions, the limit of $\tilde{\alpha}_1$ equals

$$\alpha_1^* = \operatorname{argmin}_{\alpha_1} E \left[\frac{T}{\pi^*(X)} \left(\frac{T}{\pi^*(X)} - 1 \right) \{Y - m_1(X; \alpha_1)\}^2 \right].$$

If the model $m_1(X; \alpha_1)$ is correctly specified, then α_1^* equals the true value of α_1 [i.e., $m_1(X; \alpha_1^*) = m_1(X)$] and, by result R3, $\tilde{\mu}_1$ converges to μ_1 asymptotically. The double robustness of $\tilde{\mu}_1$ relies on the fact that $\tilde{\alpha}_1$ is defined as a minimizer to a weighted mean squared error.

In contrast, $\hat{\mu}_1$ is generally not doubly robust. The limit of $\hat{\alpha}_1$ equals

$$\operatorname{argmin}_{\alpha_1} E \left[\left\{ \frac{T}{\pi^*(X)} Y - \left(\frac{T}{\pi^*(X)} - 1 \right) m_1(X; \alpha_1) \right\}^2 \right],$$

which does not necessarily give the true value of α_1 even when the model $m_1(X; \alpha_1)$ is correctly specified.

Tan (2006) showed that $\tilde{\mu}_1$ is doubly robust, but $\hat{\mu}_1$ is not, in the case of linear outcome regression models. The present discussion generalizes that result to nonlinear or generalized linear outcome regression models.

5 Estimated propensity score

Suppose that $\pi(X)$ need to be estimated. Consider a propensity score model

$$\pi(X) = \pi(X; \gamma),$$

where γ is a vector of unknown parameters. Let $\hat{\gamma}$ be the maximum likelihood estimator and $\hat{\pi}(X) = \pi(X; \hat{\gamma})$. If the model $\pi(x; \gamma)$ is correct, then $\bar{\mu}_{1, \alpha_1}$ using $\hat{\pi}(X)$

$$\tilde{E}\left[\frac{T}{\hat{\pi}(X)}Y - \left(\frac{T}{\hat{\pi}(X)} - 1\right)m_1(X; \alpha_1)\right]$$

is at least efficient as $\bar{\mu}_{1, \alpha_1}$ using the true $\pi(X)$. It is well known (e.g. van der Laan and Robins 2003) that the influence function is

$$D(O|\pi, m_1(\cdot; \alpha_1)) - \mu_1 - E(Ds^\top)V^{-1}s,$$

where $s = (T - \pi(X))/[\pi(X)(1 - \pi(X))](\partial\pi/\partial\gamma)$ and $V = E(ss^\top)$.

An extension of the RV method is to search over α_1 to minimize

$$\tilde{E}[\{\hat{D} - \tilde{E}(\hat{D}\hat{s}^\top)\hat{V}^{-1}\hat{s}\}^2],$$

where $\hat{D} = D(O|\hat{\pi}, m_1(\cdot; \alpha_1))$, $\hat{s} = s(\cdot; \hat{\gamma})$, and $\hat{V} = V(\cdot; \hat{\gamma})$. This exercise seems computationally demanding, because α_1 is involved in $\hat{D}$ and $\tilde{E}(\hat{D}\hat{s}^\top)$.

We propose a computationally more convenient approach for efficiency maximization. Specifically, let $m_1(X; \alpha_1, \beta_1, \beta_2) = \beta_1 m_1(X; \alpha_1) + \beta_2 (\partial\pi/\partial\gamma)|_{\gamma=\hat{\gamma}}/[1 - \hat{\pi}(X)]$ so that

$$\left(\frac{T}{\hat{\pi}(X)} - 1\right)m_1(X; \alpha_1, \beta_1, \beta_2) = \left(\frac{T}{\hat{\pi}(X)} - 1\right)\beta_1 m_1(X; \alpha_1) + \beta_2^\top \hat{s}$$

where (β_1, β_2) are unknown parameters like α_1 . With $\hat{\pi}(X)$ in place of $\pi^*(X)$ and $m_1(X; \alpha_1, \beta_1, \beta_2)$ in place of $m_1(X; \alpha_1)$, define $\bar{\mu}_{1, \alpha_1, \beta_1}$ as $\bar{\mu}_{1, \alpha_1}$, define $(\hat{\alpha}_1, \hat{\beta}_1, \hat{\beta}_2)$ as $\hat{\alpha}_1$ and $\hat{\mu}_1$ as $\hat{\mu}_1$, and define $(\tilde{\alpha}_1, \tilde{\beta}_1, \tilde{\beta}_2)$ as $\tilde{\alpha}_1$ and $\tilde{\mu}_1$ as $\tilde{\mu}_1$. Note that $\tilde{E}(\hat{s}) = 0$, and hence

$$\bar{\mu}_{1, \alpha_1, \beta_1} = \tilde{E}\left[\frac{T}{\hat{\pi}(X)}Y - \left(\frac{T}{\hat{\pi}(X)} - 1\right)\beta_1 m_1(X; \alpha_1)\right].$$

If the model $\pi(X; \gamma)$ is correct, then $\hat{\mu}_1$ and $\tilde{\mu}_1$ achieve the minimum asymptotic variance among the class of estimators $\bar{\mu}_{1, \alpha_1, \beta_1}$ for the following reasons. First, the minimum asymptotic variance of $\bar{\mu}_{1, \alpha_1, \beta_1}$ over (α_1, β_1) is given by

$$\min_{\alpha_1} \min_{\beta_1, \beta_2} n^{-1} E\left[\left\{\frac{T}{\pi(X)}Y - \left(\frac{T}{\pi(X)} - 1\right)\beta_1 m_1(X; \alpha_1) - \beta_2^\top s\right\}^2\right],$$

because for each fixed α_1 , the minimum asymptotic variance of $\bar{\mu}_{1, \alpha_1, \beta_1}$ is given by the inner minimum in the above display (Tan 2006). Second, under standard regularity conditions, $(\hat{\alpha}_1, \hat{\beta}_1, \hat{\beta}_2)$ and $(\tilde{\alpha}_1, \tilde{\beta}_1, \tilde{\beta}_2)$ converge to the population minimizer, and substitution of $(\hat{\alpha}_1, \hat{\beta}_1, \hat{\beta}_2)$ or $(\tilde{\alpha}_1, \tilde{\beta}_1, \tilde{\beta}_2)$ into $\bar{\mu}_{1, \alpha_1, \beta_1}$ does not affect the asymptotic variance of the resulting estimator.

The discussion in Sections 2–4 is applicable in the present situation. The estimator $\tilde{\mu}_1$ is doubly robust, but $\hat{\mu}_1$ is not. For linear outcome regression models, β_1 becomes redundant and $\bar{\mu}_{1,\alpha_1,\beta_1}$ reduces to $\bar{\mu}_{1,\alpha_1}$ using $\hat{\pi}(X)$. In this case, $\hat{\mu}_1$ agrees with the regression estimator of Robins, Rotnitzky, and Zhao (1995) and Tsiatis (2006), and $\tilde{\mu}_1$ agrees with the regression estimator of Tan (2006).

6 Use of outcome regression models

Given an outcome regression model with parameters α_1 , the proposal of RV is to use an estimator of α maximizing the efficiency among a class of estimators of μ_1 , not maximizing the efficiency of itself, assuming that a propensity score model is correctly specified. This proposal is very insightful and relevant for many applications.

Our discussion indicates that improved local efficiency and double robustness can be achieved simultaneously. As a result, we understand that outcome regression models are working models, primarily used for variance and bias reduction (Tan 2006). From this perspective, linear outcome regression models can be used even in the case of binary outcomes. For example, Tan (2006) effectively used the linear outcome regression model $m_1(X) = \beta_0 + \beta_1 m_1(X; \check{\alpha}_1)$ with the unknown parameters (β_0, β_1) and the regressors 1 and $m_1(X; \check{\alpha}_1)$, and considered the class of estimators

$$\tilde{E}\left[\frac{T}{\pi^*(X)}Y - \left(\frac{T}{\pi^*(X)} - 1\right)\{\beta_0 + \beta_1 m_1(X; \check{\alpha}_1)\}\right].$$

The efficient estimator among this class improves upon $\check{\mu}_1$ and the unadjusted estimators in the raw and ratio forms

$$\tilde{E}\left[\frac{TY}{\pi^*(X)}\right] \quad \text{and} \quad \tilde{E}\left[\frac{TY}{\pi^*(X)}\right] / \tilde{E}\left[\frac{T}{\pi^*(X)}\right],$$

but does not necessarily dominate $\hat{\mu}_1$ and $\tilde{\mu}_1$ or vice versa. The indefiniteness can be resolved by using the extended outcome regression model $m_1(X) = \beta_0 + \beta_1 m_1(X; \alpha_1)$ with the unknown parameters $(\alpha_1, \beta_0, \beta_1)$ and considering the class of estimators

$$\tilde{E}\left[\frac{T}{\pi^*(X)}Y - \left(\frac{T}{\pi^*(X)} - 1\right)\{\beta_0 + \beta_1 m_1(X; \alpha_1)\}\right].$$

The efficient estimator among this class gains efficiency over $\check{\mu}_1$, $\hat{\mu}_1$, and $\tilde{\mu}_1$. Furthermore, this construction can be adopted smoothly in the proposed approach in the case of estimated propensity score in Section 5.

References

- Robins, J.M., Rotnitzky, A., and Zhao, L.P. (1995) “Analysis of semiparametric regression models for repeated outcomes in the presence of missing data,” *Journal of the American Statistical Association*, 90, 106–121.
- Tan, Z. (2006) “A distributional approach for causal inference using propensity scores,” *Journal of the American Statistical Association*, 101, 1619–1637.