7.5-7.6 Robust PS method

Jae-Kwang Kim

Basic Setup

- Observe $(\mathbf{x}_i, \delta_i, \delta_i y_i)$ for $i = 1, \cdots, n$.
- We are interested in estimating $\theta = E(Y)$ from the observed data.
- In particular, we are interested in constructing ω_i such that

$$\hat{\theta}_{\mathrm{PS}} = \frac{1}{n} \sum_{i=1}^{n} \delta_i \omega_i y_i \tag{1}$$

  is a consistent estimator of θ.
- A popular approach is to use

$$\omega_i = \{\pi(\mathbf{x}_i; \hat{\phi})\}^{-1} \tag{2}$$

  where $\hat{\phi}$ is a consistent estimator of ϕ_0 in the response probability (RP) model $P(\delta = 1 \mid \mathbf{x}) = \pi(\mathbf{x}; \phi_0)$.
- It is based on the assumption that the RP model is correctly specified.

Doubly robustness property

- Two models
  - Response Probability (RP) model: model about δ

$$P(\delta = 1 \mid \mathbf{x}, y) = \pi(\mathbf{x}; \phi_0)$$

  - Outcome Regression (OR) model: model about y

$$E(Y \mid \mathbf{x}) = m(\mathbf{x}_i; \boldsymbol{\beta}_0)$$

- Doubly robust (DR) estimation aims to achieve (asymptotic) unbiasedness under either RP model or OR model.
- For estimation of $\theta = E(Y)$, a doubly robust estimator is

$$\hat{\theta}_{\mathrm{DR}} = \frac{1}{n} \sum_{i=1}^{n} \left\{ \hat{y}_i + \frac{\delta_i}{\hat{\pi}_i} (y_i - \hat{y}_i) \right\}$$

where $\hat{y}_i = m(\mathbf{x}_i; \hat{\boldsymbol{\beta}})$ and $\hat{\pi}_i = \pi(\mathbf{x}_i; \hat{\phi})$.

Doubly robustness property

- Writing $\hat{\theta}_n = n^{-1} \sum_{i=1}^{n} y_i$, we have

$$\hat{\theta}_{\mathrm{DR}} - \hat{\theta}_n = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{\delta_i}{\hat{\pi}_i} - 1 \right) (y_i - \hat{y}_i) \, . \tag{3}$$

- Taking an expectation of the above, we note that the first term has approximate zero expectation if the RP model is true. The second term has approximate zero expectation if the OR model is true. Thus, $\hat{\theta}_{\mathrm{DR}}$ is approximately unbiased when either RP model or OR model is true.
- When both models are true, then the choice of $\hat{\boldsymbol{\beta}}$ and $\hat{\phi}$ does not make any difference in the asymptotic sense. Robins et al. (1994) called the property the local efficiency of the DR estimator.

Covariate balancing

- Suppose that the OR model is a linear regression model:

$$y_i = \mathbf{x}_i'\boldsymbol{\beta} + e_i$$

where $E(e_i \mid \mathbf{x}_i) = 0$.

- Since

$$\hat{\theta}_{\mathrm{PS}} - \hat{\theta}_n = n^{-1}\sum_{i=1}^{n}(\delta_i\omega_i - 1)\,\mathbf{x}_i'\boldsymbol{\beta} + n^{-1}\sum_{i=1}^{n}(\delta_i\omega_i - 1)\,e_i,$$

the PS estimator in (1) is unbiased for θ if

$$\frac{1}{n}\sum_{i=1}^{n}(\delta_i\omega_i - 1)\,\mathbf{x}_i = 0. \tag{4}$$

- Property in (4) is often called the covariate-balancing property (Imai and Ratkovic, 2014). In survey sampling, (4) is often called the calibration equation (Deville and Särndal, 1992).

Motivation

- The covariate-balancing constraint (4) provides an unbiased estimator under the OR model regardless of the RP model.
- While the PS weights can be motivated by the RP model, it may be a good idea to impose (4) as a constraint for finding the PS weights to incorporate the OR model.
- To uniquely determine ω_i, we can consider minimizing $Q(\omega)$, an objective function for optimization, subject to (4).
- Hainmueller (2012) proposed using

$$Q(\omega) = \sum_{i=1}^{n} \delta_i \omega_i \log\left(\omega_i / \omega_i^{(0)}\right)$$

where $\omega_i^{(0)}$ are the baseline weights such as $\omega_i^{(0)} = 1/\hat{\pi}_i$. He called this method as the entropy balancing method.

Recall: Information projection

- Let Π be a (non-empty) closed, convex set of distributions
- The information projection of Q onto Π is $P^* \in \Pi$ such that

$$D_{\mathrm{KL}}(P^* \parallel Q) = \min_{P \in \Pi} D_{\mathrm{KL}}(P \parallel Q),$$

where

$$D_{\mathrm{KL}}(P \parallel Q) = \int P(x) \log \left\{ \frac{P(x)}{Q(x)} \right\} d\mu(x).$$

- One important family of distributions is a linear family:

$$\mathcal{L}(\alpha) = \left\{ P; \int T_i(x) P(x) d\mu(x) = \alpha_i, i = 1, \cdots, k \right\} \subset \Pi.$$

Information projection (Csiszár and Shields, 2004)

- Under some conditions, P^* satisfying

$$D_{\mathrm{KL}}(P^* \parallel Q) = \min_{P \in \mathcal{L}(\alpha)} D_{\mathrm{KL}}(P \parallel Q).$$

  exists and is unique.
- Moreover, P^*, the information projection of Q onto $\mathcal{L}(\alpha)$ is of the form

$$P^*(x) = Q(x) \frac{\exp\left\{\sum_{i=1}^{K} \theta_i T_i(x)\right\}}{E_Q\left[\exp\left\{\sum_{i=1}^{K} \theta_i T_i(x)\right\}\right]}. \tag{5}$$

- Thus, the exponential family of distributions can be derived as the information projection onto the space $\mathcal{L}$ using $Q(\cdot)$ as the baseline distribution.
- Note that there is an one-to-one correspondence between $\theta_1, \ldots, \theta_k$ (canonical parameter) and $\alpha_1, \ldots, \alpha_k$ (natural parameter).

PS model using information projection

- Recall that
$$\frac{1}{\pi(\mathbf{x})} = 1 + c \cdot \frac{f_0(\mathbf{x})}{f_1(\mathbf{x})},$$
where $c = p^{-1} - 1$ and $p = P(\delta = 1)$.
- For a fixed $f_1(\mathbf{x})$, the PS function is completely determined by $f_0(\mathbf{x})$. That is, there is an one-to-one relationship between $f_0(\mathbf{x})$ and $\pi(\mathbf{x})$.
- If $\pi(\mathbf{x})$ is correct, then the covariate balancing property holds automatically in the population level. That is, $\omega(\mathbf{x}) = \{\pi(\mathbf{x})\}^{-1}$ satisfies
$$E\left[\{\delta\omega(\mathbf{x}) - 1\}\,\mathbf{x}\right] = 0. \tag{6}$$
Thus, there is no need to add this constraint to the PS model if it is really correct.

- In practice, we do not know the true response probability and $\pi^{(0)}(\mathbf{x})$ is an working PS model.
- In this case, we can express

$$\frac{1}{\pi^{(0)}(\mathbf{x})} = 1 + c \cdot \frac{f_0^{(0)}(\mathbf{x})}{f_1(\mathbf{x})}, \tag{7}$$

where $f_0^{(0)}(\mathbf{x})$ is the baseline density for $f_0(\mathbf{x})$ derived from $\pi^{(0)}(\mathbf{x})$.
- Note that $\omega^{(0)}(\mathbf{x}) = \{\pi^{(0)}(\mathbf{x})\}^{-1}$ does not necessarily satisfy (6).
- Constraint (6) can be understood as a constraint for f_0:

$$p \int \mathbf{x} \cdot f_1(\mathbf{x}) d\mu(\mathbf{x}) + (1-p) \int \mathbf{x} \cdot f_0(\mathbf{x}) d\mu(\mathbf{x}) = E(\mathbf{x}). \tag{8}$$

- Thus, how to find $\omega^*(\mathbf{x})$ minimizing a distance between $\omega(\mathbf{x})$ and $\omega^{(0)}(\mathbf{x})$ under constraint (6) can be expressed as finding $f_0^*(\mathbf{x})$ minimizing a distance between f_0 and $f_0^{(0)}$ under constraint (8).

- Because f_0 and $f_0^{(0)}$ are densities, a natural choice for distance function is the Kullback-Leibler divergence measure:

$$D_{\mathrm{KL}}\left(f_0 \parallel f_0^{(0)}\right) = \int f_0(\mathbf{x}) \log\left(\frac{f_0(\mathbf{x})}{f_0^{(0)}(\mathbf{x})}\right) d\mu(\mathbf{x}). \tag{9}$$

- Thus, for a given $f_0^{(0)}$, we wish to find the minimizer of (9) subject to (8). The solution is given by

$$f_0^*(\mathbf{x}) = f_0^{(0)}(\mathbf{x}) \cdot \frac{\exp\left(\boldsymbol{\lambda}_1' \mathbf{x}\right)}{\int \exp\left(\boldsymbol{\lambda}_1' \mathbf{x}\right) f_0^{(1)}(\mathbf{x}) d\mu(\mathbf{x})}, \tag{10}$$

where $\boldsymbol{\lambda}_1$ is the Lagrange multiplier for constraint (8).

- The solution in (10) is obtained by the information projection technique.

- Using (10) and (7), we can obtain

$$\begin{aligned} \{\pi^*(\mathbf{x})\}^{-1} &= 1 + c \cdot f_0^*(\mathbf{x})/f_1(\mathbf{x}) \\ &= 1 + c \cdot f_0^{(0)}(\mathbf{x}) \cdot \exp\left(\lambda_0 + \lambda_1'\mathbf{x}\right)/f_1(\mathbf{x}) \\ &= 1 + \left\{1/\pi^{(0)}(\mathbf{x}) - 1\right\} \exp\left(\lambda_0 + \lambda_1'\mathbf{x}\right), \end{aligned}$$

  where $\exp(\lambda_0) = \{\int \exp\left(\lambda_1'\mathbf{x}\right) f_0^{(1)}(\mathbf{x}) d\mu(\mathbf{x})\}^{-1}$.

- Thus, we can express

$$\pi^*(\mathbf{x}) = \frac{\pi^{(0)}(\mathbf{x})}{\pi^{(0)}(\mathbf{x}) + \{1 - \pi^{(0)}(\mathbf{x})\} \exp\left(\lambda_0 + \lambda_1'\mathbf{x}\right)}, \tag{11}$$

  which is called the augmented propensity score model (Kim and Riddles, 2012).

- Basically, the information projection approach gives an expression for the final PS model satisfying the constraint (6).

Parameter estimation: Two-step procedure

- Step 1: For the working response model
$$P(\delta = 1 \mid \mathbf{x}) = \pi_0(\mathbf{x}; \phi), \tag{12}$$
we can estimate ϕ by maximizing
$$\ell(\phi) = \sum_{i=1}^{n} [\delta_i \log\{\pi_0(x_i; \phi) + (1 - \delta_i) \log\{1 - \pi_0(x_i; \phi)\}]$$
with respect to ϕ.
- Step 2: From the augmented PS model in (11), we obtain
$$\omega^*(\mathbf{x}_i; \hat{\boldsymbol{\lambda}}) = 1 + \left(\hat{\pi}_i^{-1} - 1\right) \exp\left(\hat{\lambda}_0 + \hat{\boldsymbol{\lambda}}_1' \mathbf{x}_i\right), \tag{13}$$
where $\hat{\pi}_i = \pi(x_i; \hat{\phi})$ and $\hat{\lambda}_0$ and $\hat{\boldsymbol{\lambda}}_1$ are computed from the calibration equation
$$\sum_{i=1}^{n} \delta_i \omega_i^*(\boldsymbol{\lambda}) \left(1, \mathbf{x}_i'\right) = \sum_{i=1}^{n} \left(1, \mathbf{x}_i'\right). \tag{14}$$

Remark

- The final PS estimator under the augmented PS model (11) is

$$\hat{\theta}_{\mathrm{APS}} = \frac{1}{n}\sum_{i=1}^{n} \delta_i \widehat{\omega}_i^* y_i, \tag{15}$$

  where $\widehat{\omega}_i^* = \omega^*(\mathbf{x}_i; \hat{\boldsymbol{\lambda}})$.
- If the working PS model (12) is correct, then $\hat{\boldsymbol{\lambda}}$ converges in probability to zero and we will have

$$\hat{\theta}_{\mathrm{APS}} \cong n^{-1}\sum_{i=1}^{n} \delta_i \hat{\pi}_i^{-1} y_i$$

- Also, the calibration constraints in (14) guarantees that the final PS estimator is unbiased under the linear regression outcome model.
- Thus, the proposed PS estimator $\hat{\theta}_{\mathrm{APS}}$ is doubly robust.

Computational Details

- The calibration equation in (14) can be expressed as

$$\sum_{i=1}^{n} \delta_i \left(\hat{\pi}_i^{-1} - 1\right) \exp\left(\hat{\lambda}_0 + \hat{\boldsymbol{\lambda}}_1' \mathbf{x}_i\right) \begin{pmatrix} 1 \\ \mathbf{x}_i \end{pmatrix} = \sum_{i=1}^{n} (1 - \delta_i) \begin{pmatrix} 1 \\ \mathbf{x}_i \end{pmatrix}$$

- The first equation is

$$\sum_{i=1}^{n} \delta_i \left(\hat{\pi}_i^{-1} - 1\right) \exp(\lambda_0 + \mathbf{x}_i' \boldsymbol{\lambda}_1) = n_0, \tag{16}$$

  which computes $\hat{\lambda}_0$ in terms of $\hat{\boldsymbol{\lambda}}_1$, where $n_0 = \sum_{i=1}^{n} (1 - \delta_i)$.
- By (16), the second estimating equation can be written as

$$\sum_{i=1}^{n} \delta_i w_i(\boldsymbol{\lambda}_1) \mathbf{x}_i = n_0^{-1} \sum_{i=1}^{n} (1 - \delta_i) \mathbf{x}_i, \tag{17}$$

  where

$$w_i(\boldsymbol{\lambda}_1) = \frac{\left(\hat{\pi}_i^{-1} - 1\right) \exp(\mathbf{x}_i' \boldsymbol{\lambda}_1)}{\sum_{i=1}^{n} \left(\hat{\pi}_i^{-1} - 1\right) \delta_i \exp(\mathbf{x}_i' \boldsymbol{\lambda}_1)}.$$

- Note that (17) can be written as

$$\bar{\mathbf{x}}_1(\boldsymbol{\lambda}_1) = \bar{\mathbf{x}}_0,$$

where $\bar{\mathbf{x}}_1(\boldsymbol{\lambda}_1) = \sum_{i=1}^n \delta_i w_i(\boldsymbol{\lambda}_1)\mathbf{x}_i$ and $\bar{\mathbf{x}}_0 = n_0^{-1}\sum_{i=1}^n (1-\delta_i)\mathbf{x}_i$.

- The solution may not exist if $\bar{\mathbf{x}}_0$ is outside the convex hull of $\mathbf{x}_i$ in the responding units ($\delta_i = 1$).
- Newton's method for solving equation (17):

$$\hat{\boldsymbol{\lambda}}_1^{(t+1)} = \hat{\boldsymbol{\lambda}}_1^{(t)} + \left\{\hat{\Sigma}_{xx}\left(\hat{\boldsymbol{\lambda}}_1^{(t)}\right)\right\}^{-1}\left\{\bar{\mathbf{x}}_0 - \bar{\mathbf{x}}_1(\hat{\boldsymbol{\lambda}}_1^{(t)})\right\}, \tag{18}$$

where

$$\hat{\Sigma}_{xx}(\boldsymbol{\lambda}_1) = \sum_{i=1}^n \delta_i w_i(\boldsymbol{\lambda}_1)\{\mathbf{x}_i - \bar{\mathbf{x}}_1(\boldsymbol{\lambda}_1)\}^{\otimes 2}.$$

- Use $\hat{\boldsymbol{\lambda}}_1^{(0)} = 0$ as an initial value in (18).

Remark

- We can show that the augmented PS estimator can be expressed as an imputation estimator. That is, we can show that

$$\frac{1}{n}\sum_{i=1}^{n}\delta_i\widehat{\omega}_i^* y_i = \frac{1}{n}\sum_{i=1}^{n}\left\{\delta_i y_i + (1-\delta_i)\widehat{y}_i\right\}, \tag{19}$$

where $\widehat{y}_i = \mathbf{x}_i'\hat{\boldsymbol{\beta}}^*$ (with intercept is included in $\mathbf{x}_i$) and

$$\hat{\boldsymbol{\beta}}^* = \left\{\sum_{i=1}^{n}\delta_i\left(\hat{\pi}_i^{-1}-1\right)\exp(\hat{\lambda}'\mathbf{x}_i)\mathbf{x}_i\mathbf{x}_i'\right\}^{-1}\sum_{i=1}^{n}\delta_i\left(\hat{\pi}_i^{-1}-1\right)\exp(\hat{\lambda}'\mathbf{x}_i)\mathbf{x}_i y_i. \tag{20}$$

- Equation (19) means that the final PS weights $\widehat{\omega}_i^*$ do not directly use the outcome regression model for imputation, but it implements regression imputation indirectly.

Justification

- Since $\hat{\boldsymbol{\lambda}}$ satisfies (14), we can express $\hat{\theta}_{\rm APS}$ in (15) as

$$\begin{aligned}\hat{\theta}_{\rm APS} &= \frac{1}{n}\sum_{i=1}^{n}\delta_i\widehat{\omega}_i^* y_i + \frac{1}{n}\sum_{i=1}^{n}\left(1-\delta_i\widehat{\omega}_i^*\right)\mathbf{x}_i'\boldsymbol{\beta} \\ &= \frac{1}{n}\sum_{i=1}^{n}\left\{\delta_i y_i + (1-\delta_i)\mathbf{x}_i'\boldsymbol{\beta}\right\} \\ &\quad + \frac{1}{n}\sum_{i=1}^{n}\delta_i\left(\widehat{\omega}_i^*-1\right)\left(y_i-\mathbf{x}_i'\boldsymbol{\beta}\right)\end{aligned}$$

  for any $\boldsymbol{\beta}$.
- Thus, we choose $\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}$ where $\hat{\boldsymbol{\beta}}$ satisfies

$$\hat{U}(\boldsymbol{\beta} \mid \hat{\boldsymbol{\lambda}}) \equiv \frac{1}{n}\sum_{i=1}^{n}\delta_i\left(\widehat{\omega}_i^*-1\right)\left(y_i-\mathbf{x}_i'\boldsymbol{\beta}\right)=0, \tag{21}$$

  then we have an imputation representation of $\hat{\theta}_{\rm APS}$.

Remark

- Equation (21) is a key condition for deriving the imputation representation from the PS estimator, but it does not give a unique solution for $\hat{\boldsymbol{\beta}}$.
- To uniquely determine $\hat{\boldsymbol{\beta}}$, we can impose that

$$\frac{\partial}{\partial \boldsymbol{\lambda}} \hat{U}(\hat{\boldsymbol{\beta}} \mid \hat{\boldsymbol{\lambda}}) = 0 \tag{22}$$

  which leads to $\hat{\boldsymbol{\beta}}^*$ in (20).
- By condition (22), the uncertainty of $\hat{\boldsymbol{\lambda}}$ can be completely transferred to $\hat{\boldsymbol{\beta}}$. In this case, there is an one-to-one relationship between $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\lambda}}$.

Asymptotic Results

- We are now interested in investigating the asymptotic properties of $\hat{\theta}_{\mathrm{APS}}$ in (15).
- Note that we can express

$$\begin{aligned}
\hat{\theta}_{\mathrm{APS}}(\hat{\phi}, \hat{\boldsymbol{\lambda}}) &= n^{-1}\sum_{i=1}^{n} \delta_i \widehat{\omega}_i^*(\hat{\phi}, \hat{\boldsymbol{\lambda}}) y_i \\
&= n^{-1}\sum_{i=1}^{n} \delta_i \omega_i^*(\hat{\phi}, \hat{\boldsymbol{\lambda}}) y_i + \frac{1}{n}\sum_{i=1}^{n}\left\{1 - \delta_i \omega_i^*(\hat{\phi}, \hat{\boldsymbol{\lambda}})\right\} \mathbf{x}_i' \boldsymbol{\kappa}_1 \\
&:= \hat{\theta}_1(\hat{\phi}, \hat{\boldsymbol{\lambda}}; \boldsymbol{\kappa}_1)
\end{aligned}$$

  for any $\boldsymbol{\kappa}_1$.
- Let $\boldsymbol{\lambda}^*$ be the probability limit of $\hat{\boldsymbol{\lambda}}$.
- Taylor expansion wrt $\boldsymbol{\lambda} = \boldsymbol{\lambda}^*$ can be obtained by finding $\boldsymbol{\kappa}_1^*$ from

$$E\left\{\frac{\partial}{\partial \boldsymbol{\lambda}} \hat{\theta}_1(\hat{\phi}, \boldsymbol{\lambda}^*; \boldsymbol{\kappa}_1)\right\} = 0. \tag{23}$$

- Condition (23) can be written as

$$
\begin{aligned}
& E\left\{n^{-1} \sum_{i=1}^{n} \delta_i \left(\hat{\pi}_i^{-1} - 1\right) \exp\left(\mathbf{x}_i' \boldsymbol{\lambda}^*\right) \mathbf{x}_i y_i\right\} \\
= \quad & E\left\{n^{-1} \sum_{i=1}^{n} \delta_i \left(\hat{\pi}_i^{-1} - 1\right) \exp\left(\mathbf{x}_i' \boldsymbol{\lambda}^*\right) \mathbf{x}_i \mathbf{x}_i' \boldsymbol{\kappa}_1^*\right\}
\end{aligned}
$$

  which implies that $\hat{\boldsymbol{\kappa}}_1 = \hat{\boldsymbol{\beta}}^*$ in (20).
- Now, it remains to apply Taylor expansion on $\hat{\theta}_1(\hat{\boldsymbol{\phi}}, \boldsymbol{\lambda}^*, \boldsymbol{\kappa}_1^*)$ with respect to $\boldsymbol{\phi} = \boldsymbol{\phi}^*$, where $\boldsymbol{\phi}^*$ is the probability limit of $\hat{\boldsymbol{\phi}}$.
- Let $\hat{\boldsymbol{\phi}}$ be the solution to

$$
\hat{U}(\boldsymbol{\phi}) \equiv \frac{1}{n} \sum_{i=1}^{n} \left(\frac{\delta_i}{\pi(\mathbf{x}_i; \boldsymbol{\phi})} - 1\right) \mathbf{b}_i = \mathbf{0}.
$$

- Write

$$
\begin{aligned}
\hat{\theta}_1(\hat{\phi}, \boldsymbol{\lambda}^*; \boldsymbol{\kappa}_1^*) &= \hat{\theta}_1(\hat{\phi}, \boldsymbol{\lambda}^*; \boldsymbol{\kappa}_1^*) - \{\hat{U}(\hat{\phi})\}' \boldsymbol{\kappa}_2 \\
&:= \hat{\theta}_2(\hat{\phi}; \boldsymbol{\kappa}_2)
\end{aligned}
$$

- Taylor expansion wrt $\phi = \phi^*$ can be obtained by finding $\boldsymbol{\kappa}_2^*$ from

$$
E\left\{ \frac{\partial}{\partial \phi} \hat{\theta}_2(\phi^*; \boldsymbol{\kappa}_2) \right\} = 0. \tag{24}
$$

- Condition (24) can be written as

$$
\begin{aligned}
& E\left\{ \sum_{i=1}^n \delta_i \pi_i^{-2} \pi_i (1-\pi_i) \exp(\mathbf{x}_i' \boldsymbol{\lambda}^*) \mathbf{h}_i \left(y_i - \mathbf{x}_i' \boldsymbol{\kappa}_1^* \right) \right\} \\
= \; & E\left\{ \sum_{i=1}^n \delta_i \pi_i^{-2} \pi_i (1-\pi_i) \mathbf{h}_i \mathbf{b}_i' \boldsymbol{\kappa}_2^* \right\}
\end{aligned}
$$

where $\mathbf{h}_i = \partial \text{logit}\{\pi(\mathbf{x}_i; \phi)\} / \partial \phi$.

Variance estimation

- Recall that

$$d(x_i, y_i, \delta_i) = \mathbf{x}_i' \boldsymbol{\kappa}_1^* + \mathbf{b}_i' \boldsymbol{\kappa}_2^* + \delta_i \omega_i^*(\boldsymbol{\phi}^*, \boldsymbol{\lambda}^*) \left(y_i - \mathbf{x}_i' \boldsymbol{\kappa}_1^* - \mathbf{b}_i' \boldsymbol{\kappa}_2^* \right)$$

  is the influence function for $\hat{\theta}_{\mathrm{APS}}$.

- Thus, a consistent variance estimator is

$$\hat{V}(\hat{\theta}_{\mathrm{APS}}) \cong \hat{V}(\bar{d}_n) = \frac{1}{n} S_{\hat{d}}^2$$

  where

$$S_{\hat{d}}^2 = \frac{1}{n} \sum_{i=1}^{n} \left(\hat{d}_i - \bar{\hat{d}}_n \right)^2$$

  and $\hat{d}_i = \mathbf{x}_i' \hat{\boldsymbol{\kappa}}_1^* + \mathbf{b}_i' \hat{\boldsymbol{\kappa}}_2^* + \delta_i \hat{\omega}_i^* \left(y_i - \mathbf{x}_i' \hat{\boldsymbol{\kappa}}_1^* - \mathbf{b}_i' \hat{\boldsymbol{\kappa}}_2^* \right)$.

REFERENCES

Csiszár, Imre and P. C. Shields (2004), *Information theory and Statistics: A tutorial*, Now Publishers Inc.

Deville, J. C. and C. E. Särndal (1992), 'Calibration estimators in survey sampling', *Journal of the American Statistical Association* **87**, 376–382.

Hainmueller, J. (2012), 'Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies', *Political Analysis* **20**, 25–46.

Imai, K. and M. Ratkovic (2014), 'Covariate balancing propensity score', *Journal of the Royal Statistical Society: Series B* **76**, 243–263.

Kim, J. K. and M. K. Riddles (2012), 'Some theory for propensity-score-adjustment estimators in survey sampling', *Survey Methodology* **38**, 157–165.

Robins, J. M., A. Rotnitzky and L. P. Zhao (1994), 'Estimation of regression coefficients when some regressors are not always observed', *Journal of the American Statistical Association* **89**, 846–866.