SUPPLEMENTARY MATERIALS

Understanding differences between what alternate propensity score methods estimate

Anirban Basu; Aig Unuigbe; Cristina Masseria

SUPPLEMENTARY EXHIBIT 1

SUPPLEMENTARY EXHIBIT 1

***Stratifying by quintiles of PS*:**

$$Y = \eta' + \varepsilon \text{ and } \eta' = \alpha_0 + \alpha_1 \cdot D + \sum_{j=1}^{4} \left(\alpha_{2j} \cdot I_{Q_j} + \alpha_{3j} \cdot D \cdot I_{Q_j} \right) \quad (A1)$$

where I_{Q_j} is the indicator for the j^{th} quintile. An estimate of ATE is obtained via estimating $(\alpha_1 * 5 + \sum_{j=1}^{4} \alpha_{3j})/5$. An estimate of ATT can also be obtained in this context by estimating $(\alpha_1 * E(D) + \sum_{j=1}^{4} \alpha_{3j})/E(D)$.

***Inverse Weighting with PS*:**

Following Hirano et al,[1] we can estimate the average treatment effect as

$$\hat{\Delta} = \left(\sum_{i=1}^{N} \frac{D_i}{\hat{e}(X_i)} \right)^{-1} \cdot \sum_{i=1}^{N} \frac{D_i \cdot Y_i}{\hat{e}(X_i)} - \left(\sum_{i=1}^{N} \frac{1 - D_i}{1 - \hat{e}(X_i)} \right)^{-1} \cdot \sum_{i=1}^{N} \frac{(1 - D_i) \cdot Y_i}{1 - \hat{e}(X_i)} \quad (A2)$$

Hirano et al[1] claim that this approach would also give an efficient estimate of average treatment effect under all data generating processes, although they provide empirical evidence of this claim only in the context of linear models. Note that this estimator is similar to the Horvitz-Thompson estimator.[2] Equation (A2) shows the normalized version of the reweighted estimator.[1,3]

For ATE estimation, the following weights are used[4]:

$$W_{ATE} = 1/\hat{e}(X_i) \; if \; D_i = 1$$

$$= 1/(1 - \hat{e}(X_i)) \; if \; D_i = 0 \quad (A3)$$

Similarly, using these same weights, a simple weighted comparison of outcomes between the treated and untreated groups with a specific level of X will produce an estimate of CATE.

Alternatively, in order to estimate ATT and TUT, alternative weights based on estimated odds of treatment selection are required. For example,

$$W_{TT} = 1 \; if \; D_i = 1$$

$$= \hat{e}(X_i)/(1 - \hat{e}(X_i)) \; if \; D_i = 0 \quad (A4)$$

while,

$$W_{TUT} = (1 - \hat{e}(X_i))/\hat{e}(X_i)\ if\ D_i = 0$$

$$= 1\ if\ D_i = 0 \qquad \text{(A5)}$$

***Matching with PS*:**

<u>*Nearest-neighbour (with and without a caliper) and radius matching estimator*</u>

The basic idea of matching here based on propensity scores involves[5]:

$$\delta(\hat{e}(X_i), \hat{e}(X_j) < \varepsilon \Rightarrow \delta'(\text{prob}(X_i|\hat{e}(X_i)),\text{prob}(X_j|\hat{e}(X_j))) < \varepsilon', \qquad \text{(A6)}$$

where δ and δ' are distance metrics in the mathematical sense.

Under these assumptions, the ATE estimator is given as:

$\sum_{i=1}^{N}\left(\hat{Y}_{1i} - \hat{Y}_{0i}\right)$, where $\hat{Y}_{ki} = I(D_i = k) \cdot Y_i + I(D_i = 1 - k) \cdot (\sum_{j \in P_{1-k}} I(i, j, \varepsilon) \cdot Y_j /$
$\sum_{j \in P_{1-k}} I(i, j, \varepsilon))$ (A7)

Here,K=0,1 $P_{1\text{-}k}$ represent the set of individuals for whom *D = 1-k, and* I(*i, j,ε*) is an indicator taking the value of 1 if

$|\hat{e}(X_i) - \hat{e}\left(X_j\right)| = min_j|\hat{e}(X_i) - \hat{e}\left(X_j\right)|$ {Nearest Neighbor Matching without Caliper)

$|\hat{e}(X_i) - \hat{e}\left(X_j\right)| = min_j|\hat{e}(X_i) - \hat{e}\left(X_j\right)|$ & $|\hat{e}(X_i) - \hat{e}\left(X_j\right) < d$ (Nearest Neighbor Matching with Caliper)

$|\hat{e}(X_i) - \hat{e}\left(X_j\right) \leq \varepsilon$ (Nearest Neighbor with Specified Radius)

<u>*Kernel-based or Local-linear regression-based matching estimators*</u>

Either the kernel-based or local-linear regression-based matching estimator follows the general expression to calculate ATE:

$\sum_{i=1}^{N}\left(\hat{Y}_{1i} - \hat{Y}_{0i}\right)$, where $\hat{Y}_{ki} = I(D_i = k) \cdot Y_i + I(D_i = 1 - k) \cdot (\sum_{j \in P_{1-k}} W(i, j; \varepsilon) \cdot Y_j, K = 0,1$
(A8)

Here, $P_{1\text{-}k}$ represents the set of individuals for whom *D = 1-k,* and *W*(*i, j*) represents the weights that a specific kernel or the local linear regression computes based on bandwidth □. Typically, a bandwidth of 0.06 for the kernel-based matching estimator and a central band of N*0.25 for the local-linear regression-based matching estimator is used. The bandwidth controls the amount by which the data are smoothed. Large values of bandwidth will lead to large amounts of smoothing, resulting in low variance but high bias. Small values of bandwidth will lead to less smoothing, resulting in high variance but low bias.

Doubly Robust Estimators

The average treatment effect in this case is estimated as

$$\hat{\Delta}_{DR}=n^{-1}\sum_{i=1}^{n}\left[\frac{Z_iY_i}{e(X_i\widehat{\beta})}-\frac{\{Z_i-e(X_i\widehat{\beta})\}}{e(X_i\widehat{\beta})}\,m_1(X_i,\hat{\alpha}_1)\right]-\,n^{-1}\sum_{i=1}^{n}\left[\frac{(1-Z_i)Y_i}{1-\,e(X_i\widehat{\beta})}+\frac{\{Z_i-e(X_i\widehat{\beta})\}}{e(X_i\widehat{\beta})}\,m_0(X_i,\hat{\alpha}_0)\right] \quad \text{(A9)}$$

where $m_z(X_i,\hat{\alpha}_0)$ is the postulated model for the true regression model.

REFERENCES

1. Rosenbaum PR. *Observational Studies.* 2nd ed. Springer-Verlag; 2002.

2. Horvitz DG, Thompson DJ. A generalization of sampling without replacement from a finite universe. *J Am Stat Assoc*. 1952;47(260):663-85. doi:10.2307/2280784

3. Busso M, DiNardo J, McCrary J. New evidence on the finite sample properties of propensity score reweighting and matching estimators. *Rev Econ Stat*. 2014;96(5):885-97.

4. Dehejia RH, Wahba S. Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. *J Am Stat Assoc*. 1999;94(448):1053-62. doi:10.2307/2669919

5. Heckman JJ, Vytlacil E. Policy-relevant treatment effects. *Am Econ Rev*. 2001;91(2):107-11.