Web-based Supplementary Materials for Propensity Score Matching and Subclassification in Observational Studies with Multi-level Treatments

Shu Yang¹,*, Guido W. Imbens², Zhanglin Cui and Douglas E. Faries³, and Zbigniew Kadziola⁴

¹Department of Biostatistics, Harvard T. H. Chan School of Public Health, Boston, Massachusetts 02115, U.S.A.
²Graduate School of Business, Stanford University, and NBER, Stanford, California 94305, U.S.A.
³Real World Analytics, Eli Lilly and Company, Indianapolis, Indiana 46285, U.S.A.
⁴Real World Analytics, Eli Lilly and Company, Vienna, Austria.

*email: shuyang@hsph.harvard.edu

This paper has been submitted for consideration for publication in Biometrics
1. Mathematical Details for Inference

Here we focus on large sample approximations to the distribution of the matching estimators, extending the results in Abadie & Imbens (2006, 2012) to the multi-level treatment case. We will present results for matching estimators for average effects \( \tau(w, w') \), where we match either on covariates, on the true generalized propensity score, or on the estimated generalized propensity score. In the latter case we assume the true generalized propensity score has a parametric, multinomial logit, form.

Define the conditional means and variances given covariates and given the generalized propensity scores:

\[
\mu(w, x) = \mathbb{E}[Y_i(w) \mid W_i = w, X_i = x],
\]

\[
\bar{\mu}(w, p) = \mathbb{E}[Y_i(w) \mid W_i = w, p(w|X_i) = p],
\]

\[
\sigma^2(w, x) = \mathbb{V}(Y_i(w) \mid W_i = w, X_i = x),
\]

\[
\bar{\sigma}^2(w, p) = \mathbb{V}(Y_i(w) \mid W_i = w, p(w|X_i) = p).
\]

We make the following assumptions.

**Assumption 1:** We have a random sample of size \( N \) from a large population.

**Assumption 2:** \( p(w \mid x) > c \) for some \( c > 0 \) for all \( w, x \).

**Assumption 3:** \( X \) has a continuous distribution with compact support \( \mathbb{X} \) with a continuous density function. \( \mu(w, x) \) is Lipschitz-continuous in \( x \). For some \( \delta > 0 \), \( \mathbb{E}[|Y_i|^{2+\delta} \mid W_i = w, X_i = x] \) is uniformly bounded.

**Assumption 4:** \( p(w \mid X_i) \) has a continuous distribution with compact support \([\underline{p}, \bar{p}]\) with a continuous density function. \( \bar{\mu}(w, p) \) is Lipschitz-continuous in \( p \). For some \( \delta > 0 \), \( \mathbb{E}[|Y_i|^{2+\delta} \mid W_i = w, p(w \mid X_i) = p] \) is uniformly bounded.

First we consider the matching estimator where we match on the covariates.
Theorem 1: Suppose Assumptions A1-A3 hold and the assignment mechanism is weakly unconfounded. If $X_i$ is a scalar, then

$$N^{1/2}(\hat{\tau}(w, w') - \tau(w, w')) \to N(0, \sigma_{cov}^2(w, w')),$$

$$\sigma_{cov}^2(w, w') = \mathbb{E}[(\mu(w', X_i) - \mu(w, X_i) - \tau(w, w'))^2]$$

$$+ \mathbb{E}[\sigma^2(w, X_i) \times (3/(2p(w | X_i)) - p(w | X_i)/2)]$$

$$+ \mathbb{E}[\sigma^2(w', X_i) \times (3/(2p(w' | X_i)) - p(w' | X_i)/2)].$$

Next we consider the case where we match on the true generalized propensity score.

Theorem 2: Suppose Assumptions A1-A3 hold and the assignment mechanism is weakly unconfounded. If $X_i$ is a scalar, then

$$N^{1/2}(\hat{\tau}_{gps}(w, w') - \tau(w, w')) \to N(0, \sigma_{gps}^2(w, w')),$$

$$\sigma_{gps}^2(w, w') = \mathbb{E}[(\bar{\mu}(w', p(w' | X_i)) - \bar{\mu}(w, p(w | X_i)) - \tau(w, w'))^2]$$

$$+ \mathbb{E}[\bar{\sigma}^2(w, p(w | X_i)) \times (3/(2p(w | X_i)) - p(w | X_i)/2)]$$

$$+ \mathbb{E}[\bar{\sigma}^2(w', p(w' | X_i)) \times (3/(2p(w' | X_i)) - p(w' | X_i)/2)].$$

Finally, we consider the case where we match on the estimated generalized propensity score. We use a parametric model for the generalized propensity score. Suppose our parametric model for the generalized propensity score is

$$p(w | x, \theta) = p(w | x^T \theta_1, \ldots, x^T \theta_{T-1}) = \exp(x^T \theta_w) \times \left\{1 + \sum_{w'=1}^{T-1} \exp(x^T \theta_{w'})\right\}^{-1}.$$ 

where $\theta = (\theta_1, \ldots, \theta_{T-1})$, and let $p'(w | X_i; \theta)$ denote the derivatives:

$$p'(w | a_1, \ldots, a_{T-1}) = \frac{\partial p}{\partial a_w}(w | a_1, \ldots, a_{T-1}),$$

$$p'(T | a_1, \ldots, a_{T-1}) = -\sum_{w=1}^{T-1} p'(w | a_1, \ldots, a_{T-1}).$$

We estimate $\theta$ by maximum likelihood. Let $I_\theta$ be the information matrix, and $\hat{p}(w | X_i)$ is the estimated generalized propensity score. Define the estimator that matches
on the estimated generalized propensity score as

\[ \hat{\tau}_{egps}(w, w') = N^{-1} \sum_{i=1}^{N} (Y_{m(w', \hat{w}(w'|X_i))} - Y_{m(w, \hat{w}(w'|X_i))}) \]

Also define

\[ c(w, w') = \mathbb{E} \left[ \text{Cov}(X_i, \mu(w', X_i) \mid p(w' \mid X_i; \theta))p'(w' \mid X_i; \theta)/p(w' \mid X_i; \theta) \right] \]

\[- \mathbb{E} \left[ \text{Cov}(X_i, \mu(w, X_i) \mid p(w \mid X_i; \theta))p'(w \mid X_i; \theta)/p(w \mid X_i; \theta) \right] \]

**Theorem 3:** Suppose Assumptions A1-A4 hold and the assignment mechanism is weakly unconfounded. Then,

\[ N^{1/2} (\hat{\tau}_{egps}(w, w') - \tau(w, w')) \rightarrow \mathcal{N} \left( 0, \sigma^2_{egps}(w, w') \right), \]

where \( \sigma^2_{egps}(w, w') = \sigma^2_{gps}(w, w') - c(w, w')^T I^{-1}_\theta c(w, w'). \)

2. Extended Simulation Study

In the extended simulation, we compare the performance of the estimators under the combinations of (w/o) trimming and (correct/incorrect) generalized propensity score model.

The simulation design is the same as the second design of the simulation study presented in the manuscript, except for the true propensity score design. Specifically, we consider the true propensity score generated from a multinomial regression model

\[(D_i(1), D_i(2), D_i(3), D_i(4), D_i(5), D_i(6)) \sim \text{Multinom}(p(1 \mid X_i), p(2 \mid X_i), p(3 \mid X_i), p(4 \mid X_i), p(5 \mid X_i), p(6 \mid X_i)), \]

with two propensity score designs: \( p(w \mid X_i) \propto \exp(f_w(X_i)), \) with (i) \( f_1(X_i) = 0, f_2(X_i) = 0.4(X_{1i} + X_{2i} + 2X_{3i} + X_{4i} + X_{5i} + X_{6i}), f_3(X_i) = 0.6(X_{1i} + X_{2i} + X_{3i} + X_{4i} + X_{5i} - 5X_{6i}), f_4(X_i) = 0.8(X_{1i} + X_{2i} + X_{3i} + X_{4i} + X_{5i} + 5X_{6i}), f_5(X_i) = 1.0(X_{1i} + X_{2i} + X_{3i} - 2X_{4i} + X_{5i} + X_{6i}), f_6(X_i) = 1.2(X_{1i} + X_{2i} + X_{3i} - 2X_{4i} - X_{5i} + X_{6i}); \) and (ii) \( f_1(X_i) = 0, f_2(X_i) = 0.5(X_{1i} + X_{2i} + X_{3i} + X_{4i} + X_{5i} + X_{6i}), f_3(X_i) = 0.25(X_{1i} + X_{2i} + X_{3i} + X_{4i} + X_{5i} - 2X_{6i})^2, f_4(X_i) = 0.3(X_{1i} + X_{2i} + X_{3i} + X_{4i} + X_{5i} + X_{6i})^2, f_5(X_i) = 0.15(X_{1i} + X_{2i} + X_{3i} - 2X_{4i} + X_{5i} + X_{6i})^2, f_6(X_i) = 0.2(X_{1i} + X_{2i} + X_{3i} - 2X_{4i} - X_{5i} + X_{6i})^2. \) In both scenarios, the propensity score
is estimated by a multinomial regression model with linear predictors which only include
the main effect. So the propensity score model is correctly specified for design (i), but it
is misspecified for design (ii). We consider the seven estimators applied to the full sampled
data and also the trimmed data.

In Figure 1, 2, 3 and 4, we present the results including bias, root MSE, and coverage of
95% confidence interval for the fifteen average treatment effects under the combination of
(w/o) trimming and (correct/incorrect) generalized propensity score model.

Comparing Figure 1 and 2 (or Figure 3 and 4), when the propensity score model is
incorrect, the performance for all methods based on the propensity score deteriorates. In
particular, weighting shows huge bias and variance and poor coverage for all parameters.
GPSM is inferior to COV in terms of bias and variance; however, it presents better coverage
for ten parameters out of fifteen parameters. This suggests that when covariates are high
dimensional, the inference for COV is not satisfactory.

Comparing Figure 1 and 3, after trimming, bias and variance are greatly reduced and
coverage is improved for all parameters for GPSM, GPSS and weighting, which suggests
that trimming can improve the performance of GPS based methods.

[Figure 1 about here.]

[Figure 2 about here.]

[Figure 3 about here.]

[Figure 4 about here.]

References

ABADIE, A. & IMBENS, G. (2006). Large sample properties of matching estimators for
average treatment effects. *Econometrica* **74**, 235–67.
ABADIE, A. & IMBENS, G. (2012). A martingale representation for matching estimators.

*J. Am. Statist. Assoc.* **107**, 833–43.

Received June 2015. Revised XXXX 2015. Accepted XXXX 2015.
Figure 1: Simulation Results, Design II - Correct Propensity Score Model & No Trimming.
Figure 2: Simulation Results, Design II - Incorrect Propensity Score Model & No Trimming.
Figure 3: Simulation Results, Design II - Correct Propensity Score Model & WithTrimming.
Figure 4: Simulation Results, Design II - Incorrect Propensity Score Model & With Trimming.