A Note on Adapting Propensity Score Matching and Selection Models to Choice Based Samples*

James J. Heckman
University of Chicago and American Bar Foundation

Petra E. Todd
University of Pennsylvania and NBER

First version: August, 1995; This version: August, 2008

1   Introduction

The probability of selection into a treatment, or the propensity score, plays a central role in classical selection models and in matching methods (see, e.g, Heckman (1980), Rosenbaum and Rubin (1983), Hirano, Imbens and Ridder (2003)). Heckman and Robb (1986) discuss how the propensity score is used differently in matching and selection models. They also show that, given the propensity score, both matching and selection models are robust to choice based sampling, which occurs when treatment group members are over or underrepresented relative to their frequency in the population. Choice-based

sampling designs are frequently chosen in evaluation studies to reduce the costs of data collection and to get a larger number of observations on treated individuals. Given a consistent estimate of the propensity score, matching and classical selection methods are robust to choice-based sampling, because both are "control function" methods and are defined conditional on treatment and comparison group status.

This note extends Heckman and Robb's (1986) analysis to consider the case where population weights are unknown so that the propensity score cannot be consistently estimated. In evaluation settings, the population weights are often unknown or cannot easily estimated.[1] For example, for the NSW training program studied in Lalonde (1986), Dehejia and Wahba (1998, 1999) and in Smith and Todd (2005), the population consists of all persons eligible for the program, which was targeted at drug-addicts, ex-convicts, and welfare recipients. Few datasets would have the information necessary to determine whether a person is eligible for the program, so it would be difficult to estimate the population weights needed to consistently estimate propensity scores.

In this note, we establish that matching and selection procedures can still be applied when the propensity score is estimated on unweighted choice based samples. The idea is simple. To implement both matching and classical selection models, only a monotonic transformation of the propensity score is required. In choice based samples, the odds ratio of the propensity score estimated using misspecified weights is monotonically related to the

---

[1]The methods of Manski and Lerman (1977) and Manski (1986) for adjusting for choice-based sampling in estimating the discrete choice probabilities cannot be applied when the weights are unknown and cannot be estimated from the data.

2

odds ratio of the true propensity scores. Thus, selection and matching procedures can identify population treatment effects using misspecified estimates of propensity scores fit on choice based samples.

2   Discussion of the Proposition

Let $D = 1$ if a person is a treatment group member; $D = 0$ if a member of the comparison group. $X = x$ is a realization of X. In the population generated from random sampling, the joint density is

$$g(d, x) = [\Pr(D = 1 \mid x)]^d [Pr(D = 0 \mid x)]^{1-d} g(x) \qquad \text{for}$$
$$D = d, \qquad d \in \{0, 1\},$$

where g is the density of the data. By Bayes's theorem, we have, letting $Pr(D = 1) = P$,

(1a) $g(x \mid D = 1)P = g(x)Pr(D = 1 \mid x)$

and

(1b) $g(x \mid D = 0)(1 - P) = g(x)Pr(D = 0 \mid x).$

Take the ratio of (1a) to (1b)

(2) $\dfrac{g(x \mid D = 1)}{g(x \mid D = 0)} \left(\dfrac{P}{1 - P} \right) = \dfrac{Pr(D = 1 \mid x)}{Pr(D = 0 \mid x)}.$

Assume $0 < Pr(D = 1 \mid x) < 1$. From knowledge of the densities of the data in the two samples, $g(x \mid D = 1)$ and $g(x \mid D = 0)$, one can form a scalar multiple of the ratio of the propensity score without knowing P. The odds ratio is a monotonic function of the propensity score that does not require knowledge of the true sample weights. In a choice-based sample, both the numerator and denominator of the first term in (2) can be consistently

3

estimated. This monotonic function can replace $P(x)$ in implementing both matching and nonparametric selection models.

However, estimating $g(x \mid D = d)$ is demanding of the data when X is of high dimension. Instead of estimating these densities, we can substitute for the left hand side of (2) the odds ratio of the estimated conditional probabilities obtained using the choice-based sample with the wrong weights. (*i.e.* for example, ignoring the fact that the data are a choice based sample). The odds ratio of the estimated probabilities is a scalar multiple of the true odds ratio. It can therefore be used instead of $Pr(D = 1 \mid X)$ to match or construct nonparametric control functions in selection bias models.

In the choice-based sample, let $\tilde{P}r(D = 1 \mid x)$ be the conditional probability that $D = 1$ and P^* be the unconditional probability of sampling $D = 1$, where $P^* \neq P$, the true population proportion. The joint density of the data from the sampled population is

$$[g(x \mid D = 1)P^*]^d [g(x \mid D = 0)(1 - P^*)]^{1-d}.$$

Using (1a) and (1b) to solve for $g(x \mid D = 1)$ and $g(x \mid D = 0)$ one may write the data density as

$$\left[\frac{Pr(D = 1 \mid x)g(x)}{P} P^* \right]^d \left[\frac{Pr(D = 0 \mid x)g(x)}{(1 - P)} (1 - P^*) \right]^{1-d}$$

so

(3a) $\qquad \tilde{P}r(D = 1 \mid x) = \dfrac{Pr(D = 1 \mid x)g(x)\dfrac{P^*}{P}}{g(x \mid D = 1)P^* + g(x \mid D = 0)(1 - P^*)}$

and

(3b) $\qquad \tilde{P}r(D = 0 \mid x) = \dfrac{Pr(D = 0 \mid x)g(x)\dfrac{1 - P^*}{1 - P}}{g(x \mid D = 1)P^* + g(x \mid D = 0)(1 - P^*)}.$

4

Under random sampling, the right-hand sides of (3a) and (3b) are the limits to which the choice-based probabilities converge. Taking the ratio of (3a) to (3b), assuming the latter is not zero, one obtains

$$(4) \quad \frac{\tilde{Pr}(D = 1 \mid x)}{\tilde{Pr}(D = 0 \mid x)} = \frac{Pr(D = 1 \mid x)}{Pr(D = 0 \mid x)} \left(\frac{P^*}{1 - P^*} \right) \left(\frac{1 - P}{P} \right).$$

Thus, one can estimate the ratio of the propensity score up to scale (the scale is the product of the two terms on the right-hand side of (4)). Instead of matching or estimating semiparametric selection models using $\Pr(D = 1 \mid x)$ (as in, for example, Ahn and Powell (1993), Heckman, Ichimura, Smith and Todd (1998), and Powell (1999)), we can instead use the odds ratio of the estimate $\tilde{Pr}(D = 1 \mid x)$, which is monotonically related to the true $Pr(D = 1 \mid x)$. In the case of a logit $P(x)$, $P(x) = \exp(x\beta)/(1 + \exp(x\beta))$, the log of this ratio becomes

$$\ln \frac{\tilde{Pr}(D = 1 \mid x)}{\tilde{Pr}(D = 0 \mid x)} = x\tilde{\beta}$$

where the slope coefficients are the true values and the intercept $\tilde{\beta}_0 = \beta_0 + \ell n(P^*/(1 - P^*)) + \ell n((1 - P)/P))$, where β_0 is the true value.[2]

In implementing nearest neighbor matching estimators, matching on the log odds ratio gives identical estimates to matching on the (unknown) $\Pr(D = 1 \mid x)$, because the odds ratio preserves the ranking of the neighbors. In application of either matching or classical selection bias correction methods, one must account for the usual problems of using estimated log odds ratios instead of true values.[3]

---

[2] See Manski and McFadden (1981), p. 26.

[3] For discussion related to using estimated propensity scores, see Hahn (1998), Heckman, Ichimura, Smith and Todd (1998), Ichimura, Smith and Todd (1998), Imbens, Hirano and Ridder (2003).

References

[1] Ahn, H. and J. Powell (1993), "Semiparametric Estimation of Censored Selection Models With Nonparametric Selection Mechanism," *Journal of Econometrics*, 58, 3-29.

[2] Dehejia, Rajeev and Sadek Wahba (1998): "Propensity Score Matching Methods for Nonexperimental Causal Studies," NBER Working Paper No. 6829.

[3] Dehejia, Rajeev and Sadek Wahba (1999): "Causal Effects in Noexperimental Studies: Reevaluating the Evaluation of Training Programs," *Journal of the American Statistical Association*, 94(448), 1053-1062.

[4] Hahn, J. (1998), "On the Role of the Propensity Score in Efficient Estimation of Average Treatment Effects," *Econometrica*, 66(2), 315-331.

[5] Heckman, J., (1980), "Addendum To Sample Selection Bias As A Specification Error," in E. Stromsdorfer and G. Farkas, eds., *Evaluation Studies Review Annual*, 5, Beverly Hills, Sage Publications.

[6] Heckman, J. and R. Robb (1986), "Alternative Methods For Solving The Problem of Selection Bias in Evaluating The Impact of Treatments on Outcomes," in H. Wainer, eds., *Drawing Inferences From Self Selected Samples*. Berlin: Springer Verlag.

[7] Heckman, J., H. Ichimura, J. Smith and P. Todd (1998), "Characterizing Selection Bias Using Experimental Data," *Econometrica*, 66(5), 1017-1098.

[8] Heckman, J., H. Ichimura and P. Todd (1998), "Matching As An Econometric Evaluation Estimator," *Review of Economic Studies,* 65, 261-294.

[9] Heckman, James, Robert Lalonde and Jeffrey Smith (1999): "The Economics and Econometrics of Active Labor Market Programs" in Orley Ashenfelter and David Card, eds., *Handbook of Labor Economics Volume 3A* (Amsterdam: North-Holland), 1865-2097.

[10] Hirano, K., Imbens, G., and Ridder, G. (2003). "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score," *Econometrica*, Vol. 71, No. 4, pp. 1161-1191.

[11] LaLonde, R. (1986), "Evaluating the Econometric Evaluations of Training Programs with Experimental Data," *American Economic Review,* 76, 604-620.

[12] Manski, C. and S. Lerman (1977): The Estimation of Choice Probabilities from Choice-Based Samples, Econometrica*,* 45, 1977-1988.

[13] Manski, C. F. and McFadden, D. (1981): "Statistical Analysis of Discrete Probability Models," in *Structural Analysis of Discrete Data with Econometric Applications*, MIT Press, p. 2-49.

[14] Manski, C. F. (1986): Semiparametric Analysis of Binary Response From Response-Based Samples, Journal of Econometrics, 31, 31-40.

[15] Powell, J., (1999), "Semiparametric Estimation of Bivariate Latent Variable Models," in C. Hsiao and J. Powell, eds., *Essays in Honor of Takeshi Amemiya.*

[16] Rosenbaum, P. and D. Rubin (1982), "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, 70, 41-55.

[17] Smith, Jeffrey and Petra Todd (2005): "Does Matching Overcome Lalonde's Critique of Nonexperimental Estimators?" *Journal of Econometrics*, 125 (1-2), p. 305-353.