Propensity-Score Matching with Instrumental Variables

By HIDEHIKO ICHIMURA AND CHRISTOPHER TABER*

Propensity-score matching is a nonexperimental method for estimating the average effect of social programs (see William Cochran, 1968; Paul Rosenbaum and Donald Rubin, 1983; James Heckman et al., 1998b). The method compares average outcomes of participants and nonparticipants, conditioning on the propensity-score value. The average comparison measures the average impact of a program. This methodology has received much attention recently in econometrics (see Heckman et al., 1996, 1997, 1998a, b; Jinyong Hahn, 1998; Rajeev Dehejia and Sadek Wahba, 1999; Jeffrey Smith and Petra Todd, 2000; Keisuke Hirano et al., 2000).

The underlying identification requirement of the matching methodology is that the program choice is independent of outcomes conditional on a certain set of observables. While intuitively attractive in that the method replicates features of randomized experiments within observational data, the identification requirement excludes a possibility that the program-choice decision could be correlated with the outcomes given the set of observables (see Heckman et al., 1997, 1998b). Unobservables that are correlated both with an outcome and the program choice are not allowed.

There are some efforts to estimate more general models using nonparametric methods (see Whitney Newey and James Powell, 1989; Heckman, 1997; Alberto Abadie, 2000; Serge Darolles et al., 2000; Matali Das, 2000; Jean-Pierre Florens, 2000; Ichimura and Taber, 2000). One such effort is the use of the instrumental-variable methods. Heckman (1997) has shown that the set of assumptions to justify instrumental-variable methods are very restrictive from the perspective of behavioral models of program participation. We show that his conditions justifying instrumental-variable methods actually justify the matching method as a special case.[1] This observation ties the limitations of the matching method in line with those of instrumental-variable methods and also is useful in constructing specification tests for matching methods when valid instrumental variables are available. This is analogous to testing the validity of the identification conditions for ordinary least-squares (OLS) estimators when there are overidentifying instrumental variables.

We then present two different propensity-score methods that are based on instrumental variables. Both methods include standard propensity-score matching as special cases. They help reduce the dimension of the conditioning variables without invoking functional-form assumptions in the same way that the standard propensity-score matching helps reduce the dimension of the conditioning variables. We show how to use these ideas to construct estimators that can be easily implemented.

I. OLS Propensity-Score Matching for the Standard Case

Following the standard program evaluation literature we let D be a binary variable indicating participation in a program. We let Y_1 denote the outcome for an individual if the person participates, and Y_0 the outcome if she does not.[2] The observed outcome variable is thus defined as $Y = Y_0 + D(Y_1 - Y_0)$. The propensity score is defined as $P(X) \equiv \Pr(D = 1|X)$.

The identification condition of matching for the average treatment effect is

$$E(Y_0|D, X) = E(Y_0|X) \tag{1}$$

* Ichimura: Department of Economics, University College London, Gower Street, London WC1E 6BT, United Kingdom, and Institute for Fiscal Studies; Taber: Department of Economics, Northwestern University, 2003 Sheridan Road, Evanston, IL 60208. We thank Josh Angrist, Jim Heckman, Guido Imbens, and Chuck Manski for comments. Ichimura acknowledges NSF SBR-9730688 for support, and Taber acknowledges NSF SBR-9709873.

[1] Heckman and Smith (1998) and Heckman et al. (1999) discuss a framework that includes both the instrumental-variable method and the matching method as special cases.

[2] We assume that all the moments written are finite.

(2) $$E(Y_1|D, X) = E(Y_1|X)$$

(3) $$0 < P(X) < 1$$

or equivalently

(4) $$E(Y|D, X) = E(Y_0|X) + E(Y_1 - Y_0|X)D$$

(5) $$\text{Var}[E(D|D, X)|X] \neq 0.$$

Note that the symmetry between $D = 0$ and $D = 1$ seems to be lost in the formulation, but given that the index 0 and 1 are arbitrary, it is not. Under these conditions it is easy to show that $E[Y|D, P(X)] = E[Y_0|P(X)] + D \times E[Y_1 - Y_0|P(X)]$. Thus OLS of Y on D given $P(X)$ yields $E[Y_1 - Y_0|P(X)]$ as the coefficient on D (see Rosenbaum and Rubin, 1983). Note that the OLS estimator of the coefficient of D is the propensity-score matching estimator of $E[Y_1 - Y_0|P(X)]$. We can estimate the average treatment effect $\Delta = E(Y_1 - Y_0)$ by estimating $E[Y_1 - Y_0|P(X)]$ by OLS and then integrating up over the distribution of $P(X)$. That is $\Delta = E(E[Y_1 - Y_0|P(X)])$.

Note that the matching identification conditions (1) and (2) are equivalent to assuming that D itself is a valid exclusion restriction in the mean of the outcomes Y_0 and Y_1 conditional on observables X.

II. Generalizing Matching Assumptions

Heckman (1997) shows that the average treatment effect can be identified through an instrumental variable, Z under the following conditions:

(4′) $$E(Y|Z, X) = E(Y_0|X) + E(Y_1 - Y_0|X)P(Z, X)$$

(5′) $$\text{Var}[E(D|Z, X)|X] \neq 0$$

where $P(Z, X) \equiv \Pr(D = 1|Z, X)$. Under these conditions, OLS of Y on a constant term and $P(Z, X)$ given X will yield $E(Y_1 - Y_0|X)$.[3]

[3] Heckman (1997) discusses a Wald-type estimator.

Conditioning variable D in the assumptions (4) and (5) is replaced by Z. This formulation makes clear that assumptions (4′) and (5′) generalize assumptions (4) and (5). Note that when Z is D, $P(Z, X) = D$.

Heckman (1997) discusses how these conditions are restrictive from the perspective of behavioral model of a program participation. However, he also points out that they are weaker than those specified by the dummy endogenous variable model, as they allow for some heterogeneity in treatment effects that are not correlated with Z. For example, the condition will be satisfied if we can write

$$Y = F(D, X) + U$$

with $E(U|Z, X) = 0$. This is the case Das (2000) examines.[4] The assumption specified above is more general in that it allows the treatment impact $Y_1 - Y_0$ to depend on unobservables so long as it is uncorrelated with D given Z and X. The generality allows us to embed the matching assumptions (4) and (5) as a special case.

III. Regression-Based Propensity-Score Matching

We consider the situation where Z takes on discrete values (one can always discretize a continuous Z). In this case given X, $P(Z, X)$ takes on discrete values. We discuss two different ways to generalize matching methods. First, we show that when assumptions (4′) and (5′) hold, rather than running OLS of Y on a constant term and $P(Z, X)$ conditional on the entire X vector, one can condition on two arguments:

$$P(X) = E[P(Z, X)|X]$$

$$Q(X) \equiv E[P^2(Z, X)|X].$$

Note that when Z is D, $P(Z, X) = D$, and both $E[P(Z, X)|X]$ and $E[P^2(Z, X)|X]$ reduce to $E(D|X)$, the propensity score. Thus, in the

[4] Newey and Powell (1989), Darolles et al. (2000), and Florens (2000) discuss nonparametric estimation of models with continuous endogenous variables and additive error terms.

special case, the method can be seen as a generalization of the standard propensity-score matching.

To see how this works observe that

$$\frac{\mathrm{Cov}[Y, P(Z, X)|P(X), Q(X)]}{\mathrm{Var}[P(Z, X)|P(X), Q(X)]}$$

$$= (E[Y\, P(Z, X)|P(X), Q(X)]$$

$$- E[Y|P(X), Q(X)]E[P(Z, X)|P(X), Q(X)])$$

$$\times (\mathrm{Var}[P(Z, X)|P(X), Q(X)])^{-1}.$$

Note that

$$E[Y\, P(Z, X)|P(X), Q(X)]$$

$$= E[E(Y_0|X)P(Z, X)|P(X), Q(X)]$$

$$+ E[E(Y_1 - Y_0|X)E(P^2(Z, X)|X)|P(X), Q(X)]$$

$$= E[Y_0|P(X), Q(X)]P(X)$$

$$+ E[Y_1 - Y_0|P(X), Q(X)]Q(X)$$

and analogously

$$E[Y|P(X), Q(X)]$$

$$= E[Y_0|P(X), Q(X)]$$

$$+ E[Y_1 - Y_0|P(X), Q(X)]P(X)$$

and

$$\mathrm{Var}[P(Z, X)|P(X), Q(X)] = Q(X) - P^2(X).$$

Thus,

$$\frac{\mathrm{Cov}[Y, P(Z, X)|P(X), Q(X)]}{\mathrm{Var}[P(Z, X)|P(X), Q(X)]}$$

$$= E[Y_1 - Y_0|P(X), Q(X)].$$

Therefore, in the same way that OLS is a special case of two-stage least squares, the simple matching estimator is a special case of this method when one can use Z and X as instruments for D.

This method can be implemented by estimating the mean of $Y_1 - Y_0$ conditional on $P(X) = E[P(Z, X)|X]$ and $Q(X) = E[P^2(Z, X)|X]$ by the sample analogue of

$$\frac{\mathrm{Cov}[Y, P(Z, X)|P(X), Q(X)]}{\mathrm{Var}[P(Z, X)|P(X), Q(X)]}$$

and then averaging over the sample distribution of $[P(X), Q(X)]$. $\mathrm{Cov}[Y, P(Z, X)|P(X), Q(X)]$ and $\mathrm{Var}[P(Z, X)|P(X), Q(X)]$ can be nonparametrically estimated using estimated P and Q. This estimator is quite similar to the kernel-based matching estimator discussed in Heckman et al. (1997, 1998b).

IV. Difference-Based Propensity-Score Matching

An alternative way to generalize the matching estimator is to view matching as a Wald-type estimator. For any two values $Z = z$ and $Z = z'$, define

$$Q(X) = \Pr\{Z = z|Z = z \text{ or } Z = z', X\}.$$

Then,

$$E(Y_1 - Y_0|Q(X), Z = z \text{ or } Z = z')$$

$$= E\left(\frac{Y}{P(z, X) - P(z', X)}\middle| Z = z, Q(X), Z = z \text{ or } Z = z'\right)$$

$$- E\left(\frac{Y}{P(z, X) - P(z', X)}\middle| Z = z', Q(X), Z = z \text{ or } Z = z'\right).$$

Note that when Z is D, $P(Z, X) = D$, and $Q(X) = P(X)$. Thus, this method reduces to the standard propensity-score matching when one uses D as the instrumental variable.

The key insight of Rosenbaum and Rubin (1983) is that, for any function $g(X)$, $E[g(X)|D, P(X)] = E[g(X)|P(X)]$. We exploit an analogous result for Z. To see how this works, note that, writing $\Delta(X) = E(Y_1 - Y_0|X)$,

$$E\left(\frac{Y}{P(z, X) - P(z', X)}\middle| Z = z, Q(X)\right)$$

$$- E\left(\frac{Y}{P(z, X) - P(z', X)}\middle| Z = z', Q(X)\right)$$

$$= E\left(\frac{E(Y_0|X) + \Delta(X)P(z, X)}{P(z, X) - P(z', X)}\middle| Q(X), Z = z \text{ or } z'\right)$$

$$- E\left(\frac{E(Y_0|X) + \Delta(X)P(z', X)}{P(z, X) - P(z', X)}\middle| Q(X), Z = z \text{ or } z'\right)$$

$$= E[\Delta(X)|Q(X), Z = z \text{ or } z'].$$

We have exploited

$$E[g(X)|Z, Q(X), Z = z \text{ or } z'] = E[g(X)|Q(X), Z = z \text{ or } z']$$

in the first equality above.

If Z takes on two values, then one can estimate the average treatment effect by averaging over $Q(X)$. If Z takes on more than two values, there are different ways to proceed. For example, one could create binary values using Z and simply averaging over $Q(X)$, or one can average over different combinations of z and z'. However, there may be efficiency losses in discretizing Z when it is not binary to begin with.

This method can be implemented in the same way as above using nonparametric regression or using a nearest-neighbor approach. To implement the idea, for each z we need an alternative value z' for comparison. We do so by defining a transformation T from the support of Z to itself conditional on X such that $T(Z) \neq Z$ almost surely. In the special case in which the instrument is binary $T(z) = z'$ and $T(z') = z$. We also generalize Q and P so that $Q(X; z) = \Pr(Z = z|X, z \in \{z, T(z)\})$ and $E(D|X, Z = z) = P_z(X)$.

For each X_i in the sample we match on propensity scores $Q(X_i)$ by finding another observation j in the sample with $Q(x_j) = Q(x_i)$ and $Z_j = T(Z_i)$. That is, we choose

$$j(i) = \operatorname*{argmin}_{\{j = 1, \dots, N\}} \{[Q(X_i; z) - Q(X_j; z)]^2 + [Z_j - T(Z_i)]^2\}.$$

We can estimate Δ using

$$\hat{\Delta}_2 = \frac{1}{N} \sum_{i=1}^{N} \left(\frac{Y_i}{P_{Z_i}(X_i) - P_{T(Z_i)}(X_i)} - \frac{Y_{j(i)}}{P_{Z_i}(X_{j(i)}) - P_{T(Z_i)}(X_{j(i)})}\right).$$

This is a straight generalization of the standard nearest-neighbor matching. If we take $Z = D$, $T(1) = 0$, and $T(0) = 1$, then we obtain the familiar estimator. To see this, notice that when $D_i = 1$, $P_{Z_i} - P_{T(Z_i)} = 1$ and when $D_i = 0$, $P_{Z_i} - P_{T(Z_i)} = -1$, and conditioning on $Q(X_i; Z)$ is equivalent to conditioning on the propensity score.

There are many different ways to form the transformation T. Under some regularity conditions $\hat{\Delta}_2$ should be a consistent estimate of the average treatment effect regardless of the transformation that we use. This can form the basis of a specification test. A finding that the estimated parameter varies considerably with the transformation we choose can be taken as evidence against the model.

V. Conclusion

We have shown that the conditions justifying instrumental-variable estimation of the average treatment effect justify the matching method as a special case. This observation can be used to construct a test for the identifying assumptions that justify the standard matching method.

We also have constructed two propensity-score methods that use instrumental variables. These methods have the same advantages and disadvantages and should be used with the same concern for the support conditions as those of the standard propensity-score methods.

In this paper we have assumed that the parameter of interest is the average treatment effect and have focused on strong identification conditions. There are other models explored in the literature that allow for much more general unobservable effects than the conditions studied in this paper (see Heckman, 1990; Charles Manski, 1990; Guido Imbens and Joshua Angrist, 1994; Ichimura and Taber, 2000). The ideas in this paper can be extended to show that these approaches also have propensity-score matching analogues and thus can be implemented with the reduced dimension when certain conditional probabilities are estimated.

REFERENCES

Abadie, Alberto. "Semiparametric Estimator of Instrumental Variable Models for Causal Effects." Unpublished manuscript, Harvard University, 2000.

Cochran, William. "The Effectiveness of Adjustment by Subclassification in Removing Bias in Observational Studies." *Biometrics*, June 1968, *24*(2), pp. 295–313.

Darolles, Serge; Florens, Jean-Pierre and Renault, Eric. "Nonparametric Instrumental Regression." Unpublished manuscript, Université Toulouse, France, 2000.

Das, Mitali. "Instrumental Variables Estimation of Nonparametric Models with Discrete Endogenous Regressors." Unpublished manuscript, Université Toulouse, France, 2000.

Dehejia, Rajeev and Wahba, Sadek. "Causal Effects in Non-experimental Studies: Reevaluating the Evaluation of Training Programs." *Journal of the American Statistical Association*, December 1999, *94*(448), pp. 1053–62.

Florens, Jean-Pierre. "Inverse Problems and Structural Econometrics: The Example of Instrumental Variables." Unpublished manuscript, Université Toulouse, France, 2000.

Hahn, Jinyong. "On the Role of Propensity Score in Efficient Semiparametric Estimation of Average Treatment Effects." *Econometrica*, March 1998, *66*(2), pp. 315–31.

Heckman, James. "Varieties of Selection Bias." *American Economic Review*, May 1990 (*Papers and Proceedings*), *80*(2), pp. 313–18.

______. "Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations." *Journal of Human Resources*, Summer 1997, *32*(3), pp. 441–62.

Heckman, James; Ichimura, Hidehiko; Smith, Jeffrey and Todd, Petra. "Sources of Selection Bias in Evaluating Social Programs: An Interpretation of Conventional Measures and Evidence on the Effectiveness of Matching as a Program Evaluation Method." *Proceedings of the National Academy of Sciences (USA)*, November 1996, *93*(23), pp. 13416–20.

______. "Characterization of Selection Bias Using Experimental Data." *Econometrica*, September 1998a, *66*(5), pp. 1017–98.

Heckman, James; Ichimura, Hidehiko and Todd, Petra. "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme." *Review of Economic Studies*, October 1997, *64*(4), pp. 605–54.

______. "Matching as an Econometric Estimator." *Review of Economic Studies*, April 1998b, *65*(2), pp. 261–94.

Heckman, James; LaLonde, Robert and Smith, Jeffrey. "The Economics and Econometrics of Active Labor Market Programs," in O. Ashenfelter and D. Card, eds., *Handbook of labor economics*, Vol. 3A. Amsterdam: North-Holland, 1999, pp. 1865–2097.

Heckman, James and Smith, Jeffrey. "Evaluating The Welfare State," in S. Strom, ed., *Econometrics and economic theory in the 20th century: The Ragnar Frisch Centennial*, Econometric Society Monograph Series. Cambridge, U.K.: Cambridge University Press, 1998, pp. 241–318.

Hirano, Keisuke; Imbens, Guido and Ridder, Geert. "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score." National Bureau of Economic Research (Cambridge, MA) Technical Working Paper No. 251, 2000.

Ichimura, Hidehiko and Taber, Christopher. "Direct Estimation of Policy Impacts." National Bureau of Economic Research (Cambridge, MA) Technical Working Paper No. 254, 2000.

Imbens, Guido and Angrist, Joshua. "Identification and Estimation of Local Average Treatment Effects." *Econometrica*, March 1994, *62*(2), pp. 467–75.

Manski, Charles F. "Nonparametric Bounds on Treatment Effects." *American Economic Review*, May 1990 (*Papers and Proceedings*), *80*(2), pp. 319–23.

Newey, Whitney and Powell, James. "Instrumental Variables for Nonparametric Models." Unpublished manuscript, Princeton University, 1989.

Rosenbaum, Paul and Rubin, Donald. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika*, April 1983, *70*(1), pp. 41–55.

Smith, Jeffrey and Todd, Petra. "Is Propensity Score Matching the Answer to LaLonde's Critique of Non-experimental Estimators?" Unpublished manuscript, University of Western Ontario, 2000.