The Economics and Econometrics of Active Labor Market Programs: Generalized Differences Estimators

by James J. Heckman, Robert J. LaLonde, and Jeffrey Smith

James J. Heckman

THE UNIVERSITY OF CHICAGO

Econ 312, Spring 2023

Identification Assumptions for Cross-Section Estimators

The Method of Matching

- To operationalize the method of matching, assume two samples: "t" for treatment and "c" for comparison group.

- Unless otherwise noted, observations are statistically independent. Simple matching methods are based on the following idea: For each person i in the treatment group, we find some group of "comparable" persons.

- The same individual may be in both groups if that person is treated at one time and untreated at another.

- We denote outcomes in the treatment group by Y_i^t and we match these to the outcomes of a subsample of persons in the comparison group to estimate a treatment effect.

- In principle, we can use a different subsample as a comparison group for each person.

- In practice, we can construct matches on the basis of a neighborhood $X(X_i)$where X_i is a vector of characteristics for person i.

- Neighbors to treated person i are persons in the comparison sample whose characteristics are in neighborhood $C(X_i)$.

- Suppose that there are N_C persons in the comparison sample and N_t in the treatment sample.

- Thus, the persons in the comparison sample who are neighbors to i, are persons j for whom $X_j \in C(X_i)$, i.e., the set of persons $A_i = \{j | X_j \in C(X_i)\}$.

- Let $W(i,j)$ be the weight placed on observation j in forming a comparison with observation i and further assume that the weights sum to one,

$$\sum_{j=1}^{N_c} W(i,j) = 1,$$

and that $0 \leq W(i,j) \leq 1$. Then we form a weighted comparison group mean for person i, given by

$$\bar{Y}_i^c = \sum_{j=1}^{N_c} W(i,j) Y_j^c, \qquad (7.8)$$

and the estimated treatment effect for person i is $Y_i - \bar{Y}_i^c$.

- Heckman et al. (1997a) survey a variety of alternative matching schemes proposed m the literature.

- Here we briefly introduce two widely used methods.

- The nearest-neighbor matching estimator defines Ai such that only one j is selected so that it is closest to Xi in some metric:

$$A_i = \{j \mid \underset{j \in \{1,\ldots,N_c\}}{\text{Min}} \| X_i - X_j \| \},$$

where $\| \ \|$ is a metric measuring distance in the X characteristics space.

- The Mahalanobis metric is one widely used metric for implementing the nearest-neighbor matching estimator.

- The metric used to define neighborhoods for i is

$$\| \ \| = (X_i - X_j)' \sum_c{}^{-1} (X_i - X_j),$$

Where $\sum_C$ is the covariance matrix in the comparison sample.

- The weighting scheme for the nearest neighbor matching estimator is

$$W(i,j) = \begin{cases} 1 & \text{if } j \in A_i, \\ 0 & \text{otherwise.} \end{cases}$$

- A version of nearest-neighbor matching, called "caliper" matching (Cochran and Rubin, 1973), makes matches to person i only if

$$\| X_i - X_j \| < \varepsilon,$$

where s is a pre-specified tolerance.

- Otherwise, person i is bypassed and no match is made to him or her.

- Kernel matching uses the entire comparison sample, so that A i = { 1 ..... N,:}, and sets

$$W(i,j) = \frac{K(X_j - X_i)}{\sum_{j=1}^{N_c} K(X_j - X_i)}, \qquad (7.9)$$

where K is a kernel.

- In practice, kernels are typically a standard distribution function such as that for the normal.
- Kernel matching is a smooth method that reuses and weights the comparison group sample observations differently for each person i in the treatment group with a different Xi.
- Kernel matching can be defined pointwise at each sample point X~ or for broader intervals.
- The impact of treatment on the treated is estimated by forming the mean difference across the I

$$m = \frac{1}{N_t} \sum_{i=1}^{N_t} (Y_i^t - \hat{Y}_i^c) = \frac{1}{N_t} \sum_{i=1}^{N_t} (Y_i^t - \sum_{j=1}^{N_c} W(i,j) Y_j^c). \quad (7.10)$$

The instrumental variable estimator as a matching estimator

The Instrumental Variable Estimator as A Matching-Comparison Group Estimator

- Heckman (1998c) shows how most evaluation estimators, including IV estimators, can be interpreted as matching estimators using the weighting framework of Eqs. (7.8) and (7.10).

- To see the basic idea, consider the simple random coefficient model

$$Y = \beta(X) + \alpha D + U.$$

- We define β and α as functions of X where $E(U|X,D) \neq 0$. Assume a valid instrumental Z that satisfies conditions (7.17a)-(7.17c). Then

$$\mathrm{E}(Y \mid X, Z) = \beta(X) + \mathrm{E}(\alpha \mid X, D = 1)\mathrm{E}(D \mid X, Z) + \mathrm{E}(U \mid X, Z).$$

- Now we can express the outcome equation as follows:

$$Y = \beta(X) + \mathrm{E}(\alpha \mid X, D = 1)\mathrm{E}(D \mid X, Z) + U$$

$$+[\alpha - \mathrm{E}(\alpha \mid X, D = 1)][\mathrm{E}(D \mid X, Z) + W] + \mathrm{E}(\alpha \mid X, D = 1)W,$$

- where $D = E(D|X,Z) + W$ and where, under our assumptions, the error terms have mean zero conditional on X and Z.

- If we have a valid instrument, then $E(U|X, Z) = E(U|X)$ and $E(\alpha|X, Z, D = 1) = E(\alpha|X, D = 1)$.

- To identify $E(\alpha|X, D = 1)$ we may form pairwise comparisons between person i and anyone else, provided that the matched partner for i, say i', has the same X but a different $Z = Z'$, where

$$\mathrm{E}(D \mid X, Z) \neq \mathrm{E}(D \mid X, Z').$$

- If this condition is satisfied, we may match a suitable i / to form the pairwise estimate of the gains as follows:

$$\frac{Y_i - Y_{i'}}{\mathrm{E}(D_i \mid X, Z_i) - \mathrm{E}(D_{i'} \mid X, Z_{i'})}.$$

Therefore,

$$\mathrm{E}\left[\frac{Y_i - Y_{i'}}{\mathrm{E}(D_i \mid X, Z_i) - \mathrm{E}(D_{i'} \mid X, Z_{i'})}\right] = \mathrm{E}(\alpha \mid X, D = 1).$$

- Accordingly, we can write our estimate of $E(\alpha|X, D = 1)$ as a weighted average of contrasts:

$$\hat{\alpha} = \sum_{i,i'} \left[\frac{(Y_i - Y_{i'})}{\mathrm{E}(D_i \mid X, Z_i) - \mathrm{E}(D_{i'} \mid X, Z_{i'})} \right] W(i, i') \tag{7.20}$$

for i, i' such that $E(D_i|X, Z_i) \neq E(D_i'|X, Z_i')$, and where the weights are given by

$$W(i, i') = \frac{(\mathrm{E}(D_i \mid X, Z_i) - \mathrm{E}(D_{i'} \mid X, Z_{i'}))^2}{\sum_{i,i'} (\mathrm{E}(D_i \mid X, Z_i) - \mathrm{E}(D_i \mid X, Z_{i'}))^2}.$$

- Formally, we set

$$\frac{Y_i - Y_{i'}}{\mathrm{E}(D_i \mid X, Z_i) - \mathrm{E}(D_{i'} \mid X, Z_{i'})} = 0$$

for i, i', where $E(D_i|X, Z_i) = E(D_{i'}|X, Z_{i'},)$ and we get the same result summed over all i, i' since for these cases $W(i, i') = 0$.

- Eq. (7.20) reveals that propensity score matching with Z as the propensity score estimates $E(\alpha|X, D = 1)$ by taking a weighted average of all i, i' contrasts for values of (X, Z) with distinct probability values.

- Instrumental variable estimation is just a weighted average of contrasts of conditional means constructed in terms of propensity scores.

- Observe that this method only requires (7.17b) and not that E(U I X, Z) = 0.

- Thus, like matching and randomized trials, the IV method does not eliminate conventional econometric exogeneity bias - it just balances the bias.

Panel data estimators as matching estimators

- The simple before-after estimator can be written as a matching estimator using the weighting scheme introduced in Section 7.4.1.

- To begin, accept assumption (4.A.I) as valid.

- For person i at time $t > k$ (k is the program participation period in the notation of Section 4) who has participated in the program, the match is with himself/herself in period $t' < k$.

- Assume a stationary environment.

- Letting the match partner be the same individual at time $t' < k$, we match $Y_{0,i,t'}, t' < k$ to obtain the following:

$$Y_{1,i,t} - W(i,t')Y_{0,i,t'}, \qquad \text{for } t' < k,$$

where the weight $W(i,t') = 1$. More generally if we have access to more than one preprogram observation per person, one can weight the various terms by functions of the variances determined using the optimal weighting schemes in minimum distance estimation (see Heckman, 1998c, for details.)

- Thus, the comparison group for person i at time t is a weighted average of the available observations for that person over the pre-program observation period:

$$Y_{0,i,t}^{c} = \sum_{j=0}^{k-1} W(i,j) Y_{0,i,j}, \qquad \text{for } j < k, \tag{7.31}$$

where

$$\sum_{j=0}^{k-1} W(i,j) = 1.$$

- Each post-program period can be matched in this way with the pre-program observations.

- The weights can be chosen to minimize the variance in the sum of the contrasts. (Heckman, 1998c).

- Assuming that the same treatment effect characterizes all post-program periods, and summing over all post-program observations, we can estimate the treatment on the treated parameter by the sample analog of

$$\sum_{t=k+1}^{T} (Y_{1,i,t} - Y_{0,i,t}^{c})\varphi(i,t),$$

where

$$\sum_{t=k+1}^{T} \varphi(i,t) = 1$$

and $\varphi(i,t)$ are weights chosen to minimize the variance of this expression.

If the treatment effects are different for each post-program period, there is no point in summing across post-program periods.

- There is no necessary reason why the weights should be the same on tile components.

- Thus, we may write

$$\sum_{t=k+1}^{T} (\alpha(i,t) Y_{1,i,t} - \beta(i,t) Y_{i,t}^{c}),$$

provided that

$$\sum_{t=k+1}^{T} \alpha(i,t) = 1 \qquad \text{and} \qquad \sum_{t=k+1}^{T} \alpha(i,t) = \sum_{t=k+1}^{T} \beta(i,t),$$

for all i.

- These conditions enable us to difference out common components and retain identification of $E(\alpha | X, D = 1)$.

- If there are trends operating on participants, it is necessary to eliminate them to estimate the parameter of interest.

- If the trends are common across participants, we are led to using the differences-in-differences method as long as assumption (4.A.2) is valid.

- In this setting, it is necessary to use a group of persons who do not receive treatment.

- Accordingly, we can think of creating a comparison person i' for treatment person i:

$$Y_{0,i',t} - \sum_{j=1}^{k-1} W(i',j) Y_{0,i',j}, \qquad \text{for } t > k > j,$$

where

$$\sum_{j=1}^{k-1} W(i',j) = 1 \qquad \text{and} \qquad W(i,j) = W(i',j),$$

for all i, i' and j.

- This transforms the comparison group to be conformable with the treatment group.

- We thus create a pairing $i \rightarrow i'$, such that persons i and i' have the same weights, i is in the treatment group and i' is in the comparison group, and we can form the difference-in-differences estimator for person i paired with person i' as follows:

$$\left[Y_{1,i,t} - \sum_{j=1}^{k-1} W(i,j) Y_{0,i,j}\right] - \left[Y_{0,i',t} - \sum_{j=0}^{k-1} W(i',j) Y_{0,i',j}\right] \quad (7.32)$$

and $W(i,j) = W(i',j)$ for any (i,i') and all j and where

$$\sum_{i} W(i,j) = 1 \qquad \text{and} \qquad \sum_{i'} W(i',j) = 1.$$

- This procedure eliminates common trends and weights the comparison group and treatment group symmetrically.

- Different weights are required for models with different serial correlation properties (Heckman, 1998c).

- More generally, we can form other pairings in the comparison group and compare i to an entire collection of non-treated persons who are operated on by a common trend.

- For example, we can form an alternative difference-in-differences estimator as follows:

$$\left[Y_{1,i,t} - \sum_{j=0}^{k-1} W(i,j) Y_{0,i,j} \right] - \frac{1}{N_c} \sum_{i'=1}^{N_c} \left[Y_{0,i',t} - \sum_{j=0}^{k-1} W(i',j) Y_{0,i',j} \right] \varphi(i'), \tag{7.33}$$

where N_C, is the number of persons in the comparison sample, q)(i ~) is a weight and where

$$\frac{1}{N_c} \sum_{i'=1}^{N_c} \varphi(i') = 1 \qquad \text{and} \qquad \frac{1}{N_c} \sum_{i'=1}^{N_c} W(i',j) \varphi(i') = W(i,j).$$

- Difference (7.33) eliminates age- or period-specific common trends or year effects.

- We can form variance weighted versions of (7.33) to pool information across i to estimate $E(Y_1 - Y_0|X, D = 1)$ efficiently if the effect is constant (see Heckman, 1998c).

- The same scheme can be used to estimate models with person-specific, time-varying variables.

- Time-invariant variables are eliminated by subtraction. Consider the before-after estimator.

- Let $A_{it}(Y_{it})$ be an "adjustment" to Y_{it}, where

$$A_{it}(Y_{i,t}) = Y_{i,t} - g(X_{i,t}).$$

- Then the comparison group for person i based on his preprogram adjusted outcomes can be written as

$$A_{it}^{c}(Y_{i,t}) = \sum_{j=0}^{k-1} W(i,j) A_{jt}(Y_{0,i,j})$$

and the before-after estimator can now be written in terms of adjusted outcomes as follows:

$$A_{it}(Y_{1,i,t}) - A_{i,t}^{c}(Y_{i,t}).$$

- We can make a similar modification to the difference-in-differences scheme:

$$\left[A_{it}(Y_{1,i,t}) - \sum_{j=0}^{k-1} W(i,j) A_{jt}(Y_{0,j,t})\right] - \left[A_{i',t}(Y_{1,i',t}) - \sum_{j=0}^{k-1} W(i',j) A_{i',j}(Y_{0,i',t})\right],$$

where $W(i,j) = W(i',j)$ for all i, i', and

$$\sum_{j=0}^{k-1} W(i',j) = 1 \qquad \text{and} \qquad \sum_{j=0}^{k-1} W(i,j) = 1.$$

- This modification eliminates non-invariant components.
- This enables us to generalize the simple before-after estimator to a case where person-specific and period-specific shocks operate on agents.
- This produces a large class of longitudinal estimators as special cases of the weighting scheme introduced in our discussion and is the basis for a unified treatment of a variety of evaluation estimators.
- Heckman (1998a) presents a comprehensive analysis and many examples of weights for different traditional econometric estimators.