Some Mechanics on the Method of Matching

by James J. Heckman, Robert J. LaLonde and Jeffrey Smith
From Section 7 of *Handbook of Labor Economics*, 1999

James J. Heckman

THE UNIVERSITY OF CHICAGO

Econ 312, Spring 2021

$$(Y_0, Y_1) \perp\!\!\!\perp D|X \qquad \text{(A-1)}$$

- "$\perp\!\!\!\perp$" denote independence.

$$F(Y_0|D=1,X) = F(Y_0|D=0,X) = F(Y_0|X)$$

and

$$F(Y_1|D=1,X) = F(Y_1|D=0,X) = F(Y_1|X)$$

$$0 < Pr(D = 1|X) < 1 \qquad \text{(A-1)}$$

- Assumptions A-1 and A-2 imply that

$$E(Y_0|D = 1, X) = E(Y_0|D = 0, X).$$

- In addition,

$$E(Y_1|D = 1, X) = E(Y_1|D = 0, X).$$
$$E(U_0|D = 1, X) = E(U_0|D = 0, X) = E(U_0|X).$$

- Assumptions A-1 and A-2 imply that

$$E(Y_0|D = 1, X) = E(Y_0|D = 0, X).$$

- In addition,

$$E(Y_1|D = 1, X) = E(Y_1|D = 0, X).$$
$$E(U_0|D = 1, X) = E(U_0|D = 0, X) = E(U_0|X).$$

- If A-1 and A-2 are true, it is possible to construct both the "treatment on the treated" parameter

$$E(Y_1 - Y_0 | X, D = 1)$$

and the effect of "nontreatment on the nontreated" parameter using the same data.

$$E(Y_0 - Y_1 | X, D = 0).$$

- In fact, TOT=TUT, MTE is flat.

$$Y = Y_0 + D(Y_1 - Y_0)$$

$$E(Y|X) = E(Y_0|X) + E(Y_1 - Y_0|X, D = 1)D$$

by matching

$$E(Y_1 - Y_0|X, D = 1) = E(Y_1 - Y_0|X)$$

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

- Observe no exclusion restriction needed.

- Under exogeneity for X and $E(U_0) = 0$

$$\begin{aligned} E(U_0|X, D = 1) &= E(U_0|X, D = 0) \\ &= E(U_0|X) \\ &= 0. \end{aligned}$$

- Also under exogeneity and $E(U_1) = 0$

$$\begin{aligned} E(U_1|X, D = 1) &= E(U_1|X, D = 1) \\ &= E(U_1|X) \\ &= 0. \end{aligned}$$

$$E(Y_1 - Y_0|X, D = 1) = E(Y_1 - Y_0|X).$$

- **But exogeneity not required**

How to Construct Matches

- Matches constructed on the basis of a neighborhood C_i around X_i.
- $C(X_i)$ defines the neighborhood.
- Let X_i be a vector of characteristics for person i.
- Thus, the persons in sample C who are neighbors to i are persons j, for whom $X_j \in C(X_i)$ i.e., it is the set of persons A_i for whom.

$$A_i = \{j | X_j \in C_i\}$$

- Let $W(i,j)$ be a weight.

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

$$\bar{Y}_i^C = \sum_{j=1}^{N_{ij}} W(i,j) Y_j^C$$

- Estimated treatment effect for person i is $Y_i - \bar{Y}_i^C$.

- **Nearest-neighbor matching estimator** defines A_i

$$A_i = \left\{ j \middle| Min \parallel X_i - X_j \parallel \right\}$$
$$j \in \{1, \dots, N_c\}$$

where "‖ ‖" is a metric.

- The weighting scheme for the nearest-neighbor estimator is

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

- "Caliper" matching adds a "closeness" requirement:

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

- The overall mean difference is the treatment effect:

$$m = \frac{1}{N_t}\sum_{i=1}^{N_t}(Y_i^t - \bar{Y}_i^C)$$
$$= \frac{1}{N_t}\sum_{i=1}^{N_t}(Y_i^t - \sum_{j=1}^{N_C} W(i,j)Y_j^C)$$

- **Kernel matching** $A_i = \{1, \dots, N_C\}$

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

- K is a kernel

- Mahalanobis Metric:

$$\|\| = (X_i - X_j)' \sum_{C}^{-1} (X_i - X_j).$$

Regression-adjusted matching

- Heckman, Ichimura and Todd (1997, 1998)

$$A(Y_i) = Y_i - X_i\beta$$

- Rosenbaum and Rubin (1983),

$$\begin{aligned} (Y_1, Y_0) \quad &\perp\!\!\!\perp \quad D|P(X), \qquad \text{for } X \in \mathcal{x}C \\ B\big(P(X)\big) \ &= \ E(Y_0|P(X), D = 1) - E(Y_0|P(X), D = 0) \\ &= 0 \end{aligned}$$

- Can construct counterfactual

$$E(Y_0|P(X), D = 1)$$

- Matching is sometimes used to estimate $E(Y_1 - Y_0|X, D = 1)$ at points of $X = x$.

- An averaged version

$$E(Y_1 - Y_0|D = 1) = \frac{\int_{S(X)} E(Y_1 - Y_0|D = 1, X) dF(X|D = 1)}{\int_{S(X)} dF(X|D = 1)}.$$

- *S(X)* is common support of *X* for *D=1* and *D=0* samples

Instrumental Variable Estimator as Matching-Comparison Group Estimator

$$Y = \beta(X) + \alpha(X)D + U$$
$$E(Y|X,Z) = \beta(X) + E(\alpha(X)|X,D=1)E(D|X,Z) + E(U|X,Z)$$
$$Y = \beta + E(\alpha(X)|X,D=1)E(D|X,Z) + (U + \alpha W)$$

where $D = E(D|Z) + W$

$$E(\alpha(X)|X,Z,D=1) = E(\alpha(X)|X,XD=1)$$
$$E(D|X,Z) \neq E(D|X,Z').$$

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

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

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

$$W(i, i') = \frac{\left(E(D_i|Z_i) - E(D_{i'}|Z_{i'}) \right)^2}{\sum_{i,i'} \left(E(D_i|Z_i) - E(D_i|Z_{i'}) \right)^2}$$

$\frac{Y_i - Y_{i'}}{E(D_i|X, Z_i) - E(D_{i'}|X, Z_{i'})} = 0$, if denominator zero.

- The IV method does not eliminate conventional econometric exogeneity bias – it just balances the bias.

- By the same token, OLS is a matching estimator (see notes on Theil weights).

Panel Data Estimators as Matching Estimators
(Difference in Differences Estimators)

- Consider an intervention in period k.
- For person i at time $t > k$ (k is the program participation period).
- Assume a stationary environment.

Fixed Effect

- We match $Y_{0,i,t'}, t' < k$.

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

where $W(i,t') = 1$.

General Form

$$Y^c_{0,i,t} = \sum_{j=0}^{k-1} W(i,j) Y_{0,i,j}, \qquad j < k$$

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

- $t = k+1, \ldots, T$, the summed comparison group-controls are

$$\sum_{t=k+1}^{T} \left[\left(Y_{1,i,t} - Y^c_{i,t}\right)\right] \phi(i,t), \qquad \sum_{t=k+1}^{T} \phi(i,t) = 1$$

More Generally

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

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

and

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

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

More Generally

where

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

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

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

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

- This eliminates common trends.

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

- N^c = # in control group.

$$\frac{1}{N}\sum_{i'=1}^{N_c}\varphi(i')$$

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

- This eliminates age-or-period-specific common trends or year effects. We can form variance weighted versions.

- The same scheme can be used to estimate models with person-specific, time varying variables. Let $A_{it}(Y_{it})$ be an "adjustment" to Y_{it}.

- An example is

$$A_{it}(Y_{it}) = Y_{it} - X_{it}\beta$$

or for more general models we may write

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

- Then the comparison group for person i can be written as
- $A_{it}^{c}(Y_i, t) = \sum_{j=0}^{k-1} W(i,j) A_{jt}(Y_{0,i,j})$

$$A_{it}\left(Y_{1,it}\right) - A_{i,t}^{c}(Y_{it}) = \text{ estimator.}$$

Similar Modification to Differences in Differences

$$\left[A_{it}(Y_1, i, t) - \sum_{j=0}^{k-1} W(i,j) A_{jt}\left(Y_{0,j,t}\right)\right] - \left[A_{i',t}\left(Y_{0,i',t}\right) - \sum_{j=0}^{k-1} W(i',j) A_{i',j}\left(Y_{0,i',j}\right)\right].$$