Paper Replication Report

1. Results replication of Lalonde (1986)

The results replicated are the estimates of effects of the National Supported Work Demonstration (NSW) employment program, specifically on male participants. Both experimental and non-experimental data is available, and the author used non-experimental method for the estimations, which we are trying to replicate. Also, multiple econometric models are adopted, imposing various assumptions on the dataset.

The data available to us covers the pre-earnings and post-earnings for the tracking period, which is from 1975 to 1978, along with the individual characteristics of the participants and non-participants. The experimental data is from the Manpower Demonstration Research Corporation (MDRC) program, thus containing treated group (participants) and control group (non-participants). The sample sizes of the two groups are 297 and 425, respectively. The non-experimental data is drawn from the Panel Study of Income Dynamics (PSID) and is divided into additional 2 subsets according to some of the samples' characteristics. The sample sizes are respectively 2490, 253 and 128. The individuals covered by PSID here did not participate in the employment program, hence their data functions as 'control group' but in the context of non-experimental analysis, and together with the actual control group from the experiment, they are all called 'comparison group'. As a result, we could estimate the treatment effect using 4 comparison groups, which are the control group of experiment, the full PSID sample and its two sub-samples.

The methodology is composed of both experimental and non-experimental methods. Since the latter has access to bigger size and more abundant categories of data, the non-experimental analysis occupies a larger proportion. Various computations and econometric models are adopted to estimate the treatment effect, which are listed as follows:

(1) Average earnings growth of non-participants

The author first displays the earnings growth of non-participants, that is, to simply subtract post-earnings by pre-earnings and calculate the means. Since there are 4 comparison groups, we generate 4 means of the earnings growth.

(2) Difference in earnings between participants and non-participants

The author then made simple subtractions to compare pre-earnings and post-earnings between participants and non-participants. The differences can be computed by the regressions: $Y_{is} = \beta_0 + \delta D_i + \varepsilon_{is}$ for pre-earnings and $Y_{it} = \beta_0 + \delta D_i + \varepsilon_{it}$ for post earnings. Then, similar comparisons but controlled for variables that represent characteristics are made. The following regressions could generate the estimated differences: $Y_{is} = \beta_0 + \delta D_i + \beta_1 age_i + \beta_2 age2_i + \beta_3 educ_i + \beta_4 black_i + \beta_5 hispanic_i + \beta_6 nodegree_i + \varepsilon_{is}$ for pre-earnings and $Y_{it} = \beta_0 + \delta D_i + \beta_1 age_i + \beta_2 age2_i + \beta_3 educ_i + \beta_4 black_i + \beta_5 hispanic_i + \beta_6 nodegree_i + \varepsilon_{it}$ for post-earnings. Similarly, the parameters are defined as in this paper with β_0 being the constant and the coefficient δ as estimates for the mean differences.

(3) Difference in Difference in earnings growth (participants less non-participants)

Further, earnings growth is compared between participants and non-participants. The author computed both direct subtractions and estimates of treatment effects controlling age. The results could be obtained by $Y_{it} - Y_{is} = \beta_0 + \delta D_i + \varepsilon_i$ and $Y_{it} - Y_{is} = \beta_0 + \delta D_i + \beta_1 age_i + \beta_2 age2_i + \varepsilon_i$,respectively. Again, the parameters are defined as in this paper with β_0 being the constant and the coefficient δ as estimates for the mean differences. Also, we could control the pre-earnings, with or without the characteristic variables. That is, we estimate the effect by $Y_{it} - Y_{is} = \beta_0 + Y_{is} + \delta D_i + \varepsilon_i$ and $Y_{it} - Y_{is} = \beta_0 + Y_{is} + \delta D_i + \beta_1 age_i + \beta_2 age2_i + \beta_3 educ_i + \beta_4 black_i + \beta_5 hispanic_i + \beta_6 nodegree_i + \varepsilon_i$,respectively. Again, the parameters are defined as in this paper with β_0 being the constant and the coefficient δ as estimates for the mean differences.

(4) Controlled for all observable variables and pre-earnings

Finally, post-earnings are compared between participants and non-participants controlling for all observable variables and pre-earnings. The results could be obtained by $Y_{it} = \beta_0 + Y_{is} + \delta D_i + \beta_1 age_i + \beta_2 age2_i + \beta_3 educ_i + \beta_4 black_i + \beta_5 hispanic_i + \beta_6 nodegree_i + \beta_6 married_i + \varepsilon_i$. Again, the parameters are defined as in this paper with β_0 being the constant and the coefficient δ as estimates for the mean differences.

The results of the above regressions are shown in Table 1.

Table 1

Name of comparison group	Comparison Group Earnings Growth 1975-78 (1)	NSW Treatment Earnings Less Comparison Group Earnings				Difference in Differences: Difference in Earnings Growth 1975-1978 Treatments Less Comparisons		Difference in Differences: Difference in Earnings Growth 1975-1978 Treatments Less Comparisons		Controlling for All Observed Variables and Pre-Training Earnings (10)
		Pre-Training Year, 1975		Post-Training Year, 1978						
		Unadjusted (2)	Adjusted (3)	Unadjusted (4)	Adjusted (5)	Without Age (6)	With Age (7)	Unadjusted (4)	Adjusted (5)	
Controls	2063 (6694)	39 (383)	-21 (378)	886 (472)	789 (472)	847 (560)	857 (558)	879 (467)	802 (468)	801 (468)
PSID-1	2490 (10837)	-15997 (795)	-7624 (851)	-15578 (913)	-8067 (990)	420 (651)	-749 (692)	-2380 (680)	-2119 (746)	-1348 (804)
PSID-2	2427 (8965)	-4503 (608)	-3669 (757)	-4020 (781)	-3482 (935)	484 (738)	-650 (850)	-1363 (729)	-1694 (878)	-951 (930)
PSID-3	2669 (8436)	455 (539)	455 (704)	697 (760)	-509 (967)	242 (884)	-1325 (1078)	629 (757)	-552 (967)	-79 (1012)

Note: All formats and units are consistent with those in Table 5 of Lalonde (1986).

2. Reflections on the non-parametric findings

(1) Choosing estimates among the specifications

The estimates-generating process we replicate above encompasses various specifications, which indicates the sensitivity of non-experimental estimates to different specifications. Lalonde (1986) points that this sensitivity itself is not an alarm. The reason is that more observables involved in the regression do not guarantee consistency. i.e., The unobservable could be correlated with the covariate in any case of the specification. However, Lalonde (1986) still justifies the selection for better estimates among these specifications, according to some clear sources of endogeneity. Precisely, columns 8-10 are believed to reveal better estimates compared to columns 4-7. Nevertheless, if we test the endogeneity (for *PSID1* dataset to illustrate) of columns 8 and 9 by the correlation between respective covariates and variables that are not included in the regressions, we could tell that endogeneity still exits (Table 2). For example, the treatment and marriage variables are highly correlated.

Table 2

PSID1: correlations between covariates and excluded variables				
Column 8			Column 9	
	treatment	pre-earnings		married
treatment	1.0000		treatment	-0.5708
pre-earnings	-0.3562	1.0000	pre-earnings	0.2738
age	-0.2980	0.2517	age	0.4924
age square	-0.2704	0.2255	age square	0.4727
education	-0.1774	0.3547	education	0.0091
black	0.3677	-0.3126	black	-0.3137
Hispanic	0.0983	-0.0662	Hispanic	-0.0041
no degree	0.2751	-0.3079	No degree	-0.1428
married	-0.5305	0.3029		

Note: The horizontal variables in 'Column 8' are the excluded variables and the vertical are covariates, while it is the reverse for 'Column 9'.

This test is only for the variables we could observe. There are other variables that these models might fail to capture. Some examples are listed in Ichimura and Todd (1997): Training histories, employment histories, hours worked etc. Given that it is impossible to observe all the variables, our knowledge about the consistency of treatment effect estimates in these regressions is limited. Hence, the best way Lalonde (1986) could propose is to compare the non-experimental estimates with the experimental ones. To draw a conclusion, what may contaminate the non-experimental findings is our inability to adequately control for differences between trainees and the comparison group, given that only non-experimental data is accessible. This could lead to our focus being cast more on the comparison group selection.

(2) Comparison group selection, biases and alternative methods

When we switch our attention to non-experimental data selection, there are also limitations occurring during this process. Ichimura and Todd (1997) introduces some of the drawbacks. First, the support of observables in the eligible project targets is different from that from a non-experimental group, which is the major limitation.

Second, the variables of a comparison group are not measured in the same way as the treatment group, while the experiment imposes on the control group the same questionnaire as that for the treatment group. Third, disparity in labor markets of a comparison group and the treatment group could lead to inefficiency. These factors can all contaminate the non-experimental findings of Lalonde (1986). Specifically, Todd (1997) summarizes whether these conditions are addressed in Lalonde (1986). It is concluded that both the 2nd and 3rd limitations are present in Lalonde (1986).

The biases from different supports are defined as follows in Ichimura and Todd (1997).
Nonoverlapping support bias:

$$B_1 = \int_{S_1 \setminus S_{10}} E(Y_0|X, D=1) f(X|D=1)dX - \int_{S_0 \setminus S_{10}} E(Y_0|X, D=0) f(X|D=0)dX$$

Bias due to different distribution of X :

$$B_2 = \int_{S_{10}} E(Y_0|X, D=0)[\, f(X|D=1) - f(X|D=0)]dX.$$

Bias remaining after controlling for observables and supports:

$$B_2 = \int_{S_{10}} [E(Y_0|X, D=1) - E(Y_0|X, D=0)]\, f(X|D=0)dX.$$

Hence, an alternative method for treatment effect estimation is to compute the sample analog of $E(\widehat{\alpha}) = \int_{S_0} E(Y_1|X, D=1) f(X|D=1)dX - \int_{S_0} E(Y_0|X, D=0) f(X|D=0)dX$ and B_1 to B_3, subtracting the first analog by the sum of $\widehat{B_1}$ to $\widehat{B_3}$. However, given the multidimensionality and dummy variables, it is very difficult to use this estimation.

Alternatively, Ichimura and Todd (1997) proposes another matching estimator. This estimator involves a weighting process that matches comparison group observations to treatment participants. It allows us to assign positive weights only to comparison group observations that are more similar to a participant and allocate zero weight to those who diverge away from the participant. The 'similarity' here could be measured against a norm threshold that is pinned by the researcher. This estimator resolves the difficulty to identify the difference of supports in last estimation method by directly eliminating the comparison observations that are more likely to have a support disparity.

To conclude, non-experimental methods have various limitations in terms of efficiency compared with the experimental method. The major limitation arises from different supports of samples, leading to biases while provoking alternative methods aimed at minimizing the biases.

Appendix (Sata Codes)

https://github.com/ZiqingY/TB2-Research-Methods/blob/60fd14653b761cb5ce60b48e9ee956bc0b793f7d/codes.do