IZA DP No. 10055

Estimating More Precise Treatment Effects in Natural and Actual Experiments

Harriet Duleep
Xingfei Liu

July 2016

Estimating More Precise Treatment Effects in Natural and Actual Experiments

Harriet Duleep
College of William and Mary and IZA

Xingfei Liu
University of Alberta and IZA

Discussion Paper No. 10055
July 2016

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 10055
July 2016

ABSTRACT

Estimating More Precise Treatment Effects in Natural and Actual Experiments

The importance of using natural experiments and experimental data in economic research has long been recognized. Yet, it is only in recent years that these approaches have become an integral part of the economist's analytical toolbox, thanks to the efforts of Meyer, Card, Peters, Krueger, Gruber, and others. This use has shed new light on a variety of public policy issues and has already caused a major challenge to some tightly held beliefs in economics, most vividly illustrated by the finding of a positive effect of a minimum wage increase on the employment of low-wage workers. Although currently in vogue in economic research, the analysis of experimental data and natural experiments could be substantially strengthened. This paper discusses how analysts could increase the precision with which they measure treatment effects. An underlying theme is how best to measure the effect of a treatment on a variable, as opposed to explaining a level or change in a variable.

JEL Classification: C1, J1

Keywords: precision of treatment effects, differences in averages, average of differences, experimental approach, natural experiment, policy evaluation

Corresponding author:

Xingfei Liu
Department of Economics
University of Alberta
8-14 HM Tory Building
Edmonton Alberta T6G 2H4
Canada
E-mail: xingfei@ualberta.ca

Estimating More Precise Treatment Effects in Natural and Actual Experiments

Harriet Duleep and Xingfei Liu

The importance of using natural and actual experiments in economic research has long been recognized (Campbell and Cook, Simon, 1966, Orcutt, 1970). Yet, it is only in recent years that they have become an integral part of the economist's analytical toolbox, thanks to the efforts of Meyer, Card, Peters, Krueger, Gruber, and others. Their increased use promises new understanding and has already challenged one of economics' most tightly held beliefs: an increase in the minimum wage should have a negative effect on the employment of low-wage workers.

Although currently in vogue in economic research, the analysis of actual and natural experiments could be substantially strengthened. This paper presents an overlooked fact that would increase the precision of the results from their analysis. Precision can be increased by measuring the mean of the individual differences rather than the difference in means between treatment and control group observations, a fact with foundations in statistics so deep that it is difficult to cite its origins. Nevertheless, despite its long history, it is overlooked in almost all analyses of natural and actual experiments by economists. We seek to highlight this issue within the context of the predominant method of analysis of natural and actual experiments by economists. The leitmotif of this paper is how best to measure the effect of a treatment on a variable, as opposed to how best to explain a level or change in a variable.

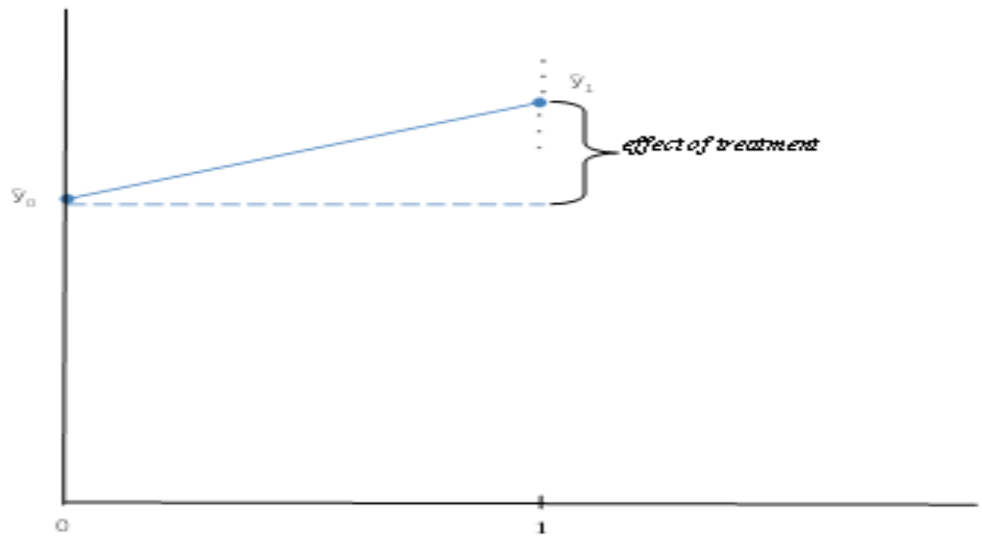
I. The Analysis of Data from Natural and Actual Experiments: The Difference in Averages versus the Average of the Differences

In studies by economists, the analysis of natural and actual experiments is generally put into a regression format. Meyer (1995) provides a comprehensive presentation and review of the analysis of experimental data using a regression format. For a before-treatment/after-treatment analysis, we may estimate the following regression:

$$y_{it} = \alpha + \beta d_t + \varepsilon_{it}$$

Where i refers to the individual and t refers to the time period (t=0 for the initial period, t=1 for the post-treatment period). The dummy variable, d, equals 1 if the observation is after the treatment, and equals 0 if the observation is for the initial period, before the treatment. Estimating this equation, we get $\hat{\alpha} = \bar{y}_0$, or the average value of the before-treatment observations. As shown in Figure 1, the estimated effect of the treatment is $\hat{\beta} = \bar{y}_1 - \bar{y}_0$, or the average value of the after-treatment observations minus the average value of the before-treatment observations.

Figure 1



Similarly, Meyer shows that the effect of the treatment from an over-time natural or actual experiment that includes a control group can be estimated as

$$y_{it}^{j} = \alpha + \beta_1 d_t + \beta_2 d_j + \beta_3 d_t^{j} + \varepsilon_{it}^{j}$$

where i refers to the individual; t refers to the time period, with 0 for the initial period and 1 for the post-treatment period; and j refers to the group membership of the individual, with j=1 if individual i is in the group that receives the treatment in time period 1, and j=0 if the individual i is in the control group.

The dummy variables are defined as follows:

d_t = 1, if the time period is the post-treatment period (e.g. if t=1)
d_j = 1 if the group is the treatment group (e.g. if j=1)
d_t^j = 1 if the group is the treatment group and the time period is the post-treatment period (e.g. if t=1 and if j=1)

Then,

$\hat{\alpha} = \bar{y}_{00}$, the average value of the control observations in time period 0;

$\hat{\alpha} + \hat{\beta}_1 = \bar{y}_{01}$, the average value of the control observations in time period 1;

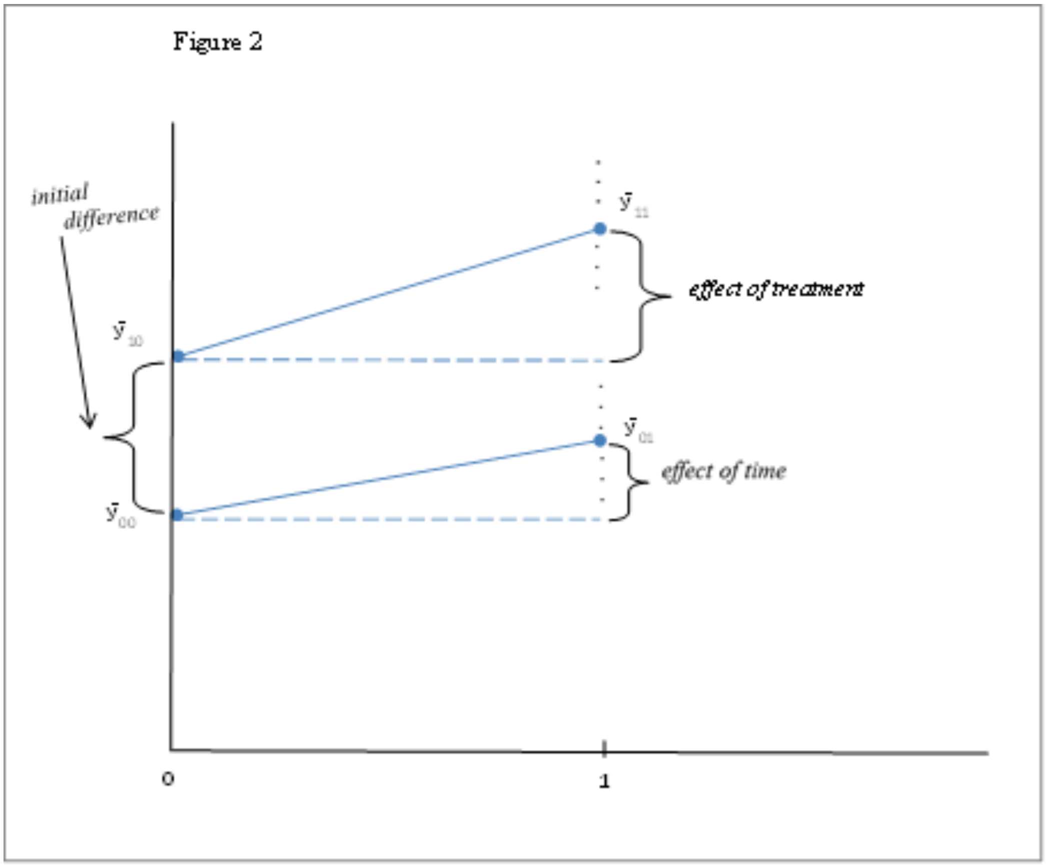
$\hat{\alpha} + \hat{\beta}_2 = \bar{y}_{10}$, the average value of the treatment observations in time period 0; and

$\hat{\alpha} + \hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_3 = \bar{y}_{11}$, the average value of the treatment observations in time period 1.

These points are displayed in Figure 2.

As shown in Figure 2, $\hat{\beta}_2$, or ($\bar{y}_{10}$ - $\bar{y}_{00}$), is the difference between the average values of the treatment and control group in the initial period. This difference is assumed to be time invariant. $\hat{\beta}_1$ or ($\bar{y}_{01}$ - $\bar{y}_{00}$), is assumed to be that part of the change with time that is common to both the control and treatment groups. And, the estimated effect of the treatment is:

$$[(\hat{\alpha} + \hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_3) - (\hat{\alpha} + \hat{\beta}_2)] - [(\hat{\alpha} + \hat{\beta}_1) - \hat{\alpha}] = \hat{\beta}_3 \text{ or } (\bar{y}_{11} - \bar{y}_{10}) - (\bar{y}_{01} - \bar{y}_{00})$$



Given the goal of explaining the level of a variable (a pursuit that lends itself to adopting a regression format) rather than measuring the effect of a treatment, there has been a tendency in the analysis of natural and actual experiments by economists to measure the *difference in the averages* between treatment and control group outcomes, measured by the coefficients in the estimated regressions shown above, rather than the *average of the individual differences*. Thus, in analyzing the outcome of interest in a before and after natural experiment, where y_{1i} is the after-treatment value of the ith observation and y_{0i} is the before-treatment value of the ith observation, the measure of the effect of the treatment has typically been the difference in the averages, $1/N\Sigma(y_{1i}) - 1/N\Sigma(y_{0i})$, rather than the average of the differences, $1/N\ \Sigma\ (y_{1i} - y_{0i})$.

Of course, the difference in the averages, $1/N\Sigma(y_{1i}) - 1/N\Sigma(y_{0i})$, equals the average of the individual differences, $1/N\Sigma(y_{1i} - y_{0i})$. However, the variance of the average of the individual differences is smaller than the variance of the difference in the averages. Hence, the precision with which we measure the effect of the treatment will be greater if we measure the average of the individual differences. Intuitively, the variance of the average of the differences is less than the difference in the averages because separately averaging the before-treatment values and the after-treatment values before taking the difference throws away the within-group variance. A more formal proof that the variance of the average of the individual differences is less than the variance of the difference in the averages follows.

The variance of the difference in the averages equals

$\text{var}(\bar{y}_1 - \bar{y}_0) = \text{var}\,\bar{y}_1 + \text{var}\,\bar{y}_0 - 2\text{cov}\,\bar{y}_1\bar{y}_0$

The variance of the average of the individual differences equals

$\text{var}\,1/N\Sigma(y_{1i} - y_{0i})$ or $\text{var}(\overline{y_{1i} - y_{0i}}) = \text{var}(y_1 - y_0)/N$ (since $\text{var}\,\bar{y} = (\text{var}\,y)/N$).

$\text{var}(y_1 - y_0)/N = (\text{var}y_1 + \text{var}y_0 - 2\text{cov}y_1y_0)/N = (\text{var}\,y_1)/N + (\text{var}\,y_0)/N - (2\text{cov}\,y_1y_0)/N$

$= \text{var}\,\bar{y}_1 + \text{var}\,\bar{y}_0 - (2\text{cov}y_1y_0)/N$

Since $(2\text{cov}y_1y_0)/N$, the term subtracted off $\text{var}\,\bar{y}_1 + \text{var}\,\bar{y}_0$ in the formula for the $\text{var}\,1/N\Sigma(y_{1i} - y_{0i})$, is greater than $2\text{cov}\,\bar{y}_1\bar{y}_0$, the term subtracted off $\text{var}(\bar{y}_1 - \bar{y}_0)$, it follows that $\text{var}\,1/N\Sigma(y_1 - y_0) < \text{var}(\bar{y}_1 - \bar{y}_0)$.

<u>Proof that $(2\text{cov}y_1y_0)/N > 2\text{cov}\,\bar{y}_1\bar{y}_0$ or $(\text{cov}y_1y_0)/N > \text{cov}\,\bar{y}_1\bar{y}_0$:</u>

$\text{Cov}y_1y_0 = E[y_1 - Ey_1][y_0 - Ey_0] = E[y_1y_0 - y_0Ey_1 - y_1Ey_0 + Ey_1Ey_0]$

$= E(y_1y_0) - 2Ey_1Ey_0 + Ey_1Ey_0 = E(y_1y_0) - Ey_1Ey_0$ Thus, $(\text{cov}y_1y_0)/N = [E(y_1y_0) - Ey_1Ey_0]/N$.

Cov $\bar{y}_1\bar{y}_0$ = E[$\bar{y}_1$ - E$\bar{y}_1$][$\bar{y}_0$ - E$\bar{y}_0$] = E($\bar{y}_1\bar{y}_0$) - 2E$\bar{y}_1$E$\bar{y}_0$ + E$\bar{y}_1$E$\bar{y}_0$

= E($\bar{y}_1\bar{y}_0$) - E$\bar{y}_1$E$\bar{y}_0$ = E($\bar{y}_1\bar{y}_0$) - Ey_1Ey_0 (since E$\bar{y}$ = Ey)

But E($\bar{y}_1\bar{y}_0$) = $E(1/N\Sigma y_{1i}\cdot 1/N\Sigma y_{0i}) = 1/N^2\, E\Sigma y_{1i}\Sigma y_{0i} = 1/N^2\, E\Sigma\Sigma y_{1i}y_{0i}$

$= 1/N^2\Sigma\Sigma E(y_{1i}y_{0i}) = 1/N^2(N\cdot E(y_1y_0)) = 1/N\ E(y_1y_0)$.

So, cov $\bar{y}_1\bar{y}_0$ = E($\bar{y}_1\bar{y}_0$) - $Ey_1Ey_0 = 1/N\ E(y_1y_0) - Ey_1Ey_0$

$= [E(y_1y_0) - N\cdot Ey_1Ey_0]/N$ versus $(\mathrm{cov}\, y_1y_0)/N = [E(y_1y_0) - Ey_1Ey_0]/N$.

Thus, $(\mathrm{cov}\, y_1y_0)/N >$ cov $\bar{y}_1\bar{y}_0$.

The preceding discussion suggests that in a before-treatment/after-treatment design with no control group, rather than adopting the regression format, $y_{it} = \alpha + \beta d_t + \varepsilon_{it}$, and measuring the difference between the average of the after-treatment values and the average of the before-treatment values, a superior approach that will yield more precise estimates of the treatment effect is the more straightforward one of measuring $1/N\Sigma(y_{1i} - y_{0i})$.

For exactly the same reason, in a before-treatment/after-treatment analysis with a control group, rather than estimating $y_{it}^j = \alpha + \beta_1 d_t + \beta_2 d_j + \beta_3 d_t^j + \varepsilon_{it}^j$ the preferred approach is to measure $1/N\Sigma[(y_{11i} - y_{10i}) - (y_{01i} - y_{00i})]$ where the first subscript refers to the group (treatment versus control), the second subscript refers to the time period, and the third subscript denotes each individual treatment-control observation pair. This presumes that the treatment and control observations are matched. However, even when the control and treatment observations are not matched, precision of the estimated treatment effect can be improved by estimating

$1/N\Sigma(y_{11i} - y_{10i}) - 1/N\Sigma(y_{01j} - y_{00j})$, rather than $(\bar{y}_{11} - \bar{y}_{10}) - (\bar{y}_{01} - \bar{y}_{00})$.

Note, that if the regression were set up as y_1-$y_0 = \alpha + \beta T$ where y_1-y_0 is an observation pair from either the control or treatment group and T = 1 if the observation pair belongs to the treatment group and 0 if it belongs to the control group, then $\hat{\beta}$ will be measuring the average of the individual differences of the treatment and, separately, the average of the individual differences of the control observation pairs, or $1/N\Sigma(y_{11i} - y_{10i}) - 1/N\Sigma(y_{01j} - y_{00j})$.

The advantage of using the regression format and estimating $y_{it}^j = \alpha + \beta_1 d_t + \beta_2 d_j + \beta_3 d_t^j + \varepsilon_{it}^j$ is that it conveniently decomposes the effects on the outcome variable of the initial difference between the control and treatment groups (β_2), their assumed common time trend effect on the outcome (β_1), and the effect of the treatment on the outcome (β_3). Yet, if our goal is to measure as accurately as possible the effect of a treatment, rather than to explain the level of a variable, then our preferred approach should always be to measure the average of the individual differences, rather than the difference in the averages.

II. Isolating the Treatment's Effect versus Explaining the Dependent Variable

Precision can be increased by analyzing the average of the individual differences between control and treatment observations, as opposed to the difference in the means of control and treatment groups. To examine how commonly economists use the average of the differences versus the difference in the averages, we conducted a literature review of nearly 60 studies of actual and natural experiments. About 40 percent of our sample consisted of analyses of the Negative Income Tax experiments, a third were analyses of experiments other than the Negative Income Tax experiments, and another third were analyses of natural experiments. We found only one study that used the average of the individual differences between control and treatment observations, as opposed to the difference in the means of control and treatment groups. Table 1 gives a sampling of these studies that estimate the effect of a treatment within a regression format that includes several explanatory variables.

So why, in the analysis of actual and natural experiments, have economists by and large eschewed analyzing the average of the individual differences between control and treatment observations in favor of the difference in the means of control and treatment groups, despite the greater precision of the former? One reason is that, despite being an everyday statistician's issue, it simply may not have occurred to social scientists given that the difference in averages equals the average of differences.

A second reason, and perhaps the dominant one, is the generally embraced goal of economists to explain levels of variables or changes in levels of variables—the appropriate domain of the regression format—rather than accurately measuring the effect of a treatment on an outcome. If, for instance, our goal is to explain individual earnings, then we will want to estimate the effect on earnings of variables such as schooling, region, and age. Explaining the level of a variable, or a change in a variable, requires estimating the effects of all relevant variables on the dependent variable. This, however, is a very different and more ambitious goal than trying to estimate, as precisely as possible, the effect of a treatment on a variable.

Finally, people are drawn to the regression format because in non-experimental data--e.g. natural experiments--they need to control for variables that may be correlated with who gets the treatment, which in true experiments is taken care of by random assignment. Using a regression is an easy way to do this. However, it removes the possibility of getting the average of the differences.

If our goal is to measure as accurately as possible the effect of the treatment, rather than to explain the level of a variable, then the control for relevant variables should be done in the least restrictive way without imposing assumptions. For instance, without assuming, as is often done in regression analysis, that the control variables have an additive, linear effect on the outcome variable. Furthermore, a technique should be used that permits measuring the average of the differences as opposed to the difference in the averages. These objectives can be achieved by matching observations in the control and treatment groups.

III. Using a Non-Regression Strategy to Increase Precision by Using Matched Data

Underlying several analyses of natural experiments is the idea that if the control and treatment groups are similar in respects other than the imposition of the treatment, then we will be more likely to detect the effect of the treatment; its effect will be more likely to surface above the other noise.

In this vein, the effect of a treatment is generally measured by comparing the before-treatment/after-treatment experience of the treatment group with the experience of other, non-treatment, units over the same period of time. To pick the control group from another time period would inject into the comparison an additional source of variation in the dependent variable and make it more difficult to detect the effect of the treatment. For instance, in a natural experiment analysis of the effect of state-imposed price changes on liquor consumption, Simon (1966) compares the change in liquor consumption in state i, which experienced a liquor tax change, with the changes in liquor consumption occurring over the same time period in states that did not experience a liquor tax change.

Analysts have also sought to have a comparison group that was similar to the treatment group in characteristics other than a shared time period. For instance, Card and Krueger (1994) sought to shed light on the effect of a minimum wage change on low-wage employment by comparing the over time employment of fast-food restaurant workers in a state in which a minimum wage increase was legislated, during this time period, to a state with no minimum wage change: the employment in fast-food restaurants before and after a minimum wage increase in New Jersey is compared with changes in the employment of fast-food restaurants over the same time period of neighboring Pennsylvania, which had not instituted a minimum wage increase. Presumably, the economies and populations of neighboring states share more in common than more geographically dispersed states. Similarly, in an analysis of the effect of the sudden influx (with the Mariel boatlift) of Cuban immigrants into Miami on the unemployment and wages of the low-skilled in Miami, Card (1990) compared Miami's employment and wage experience preceding and following the Mariel boatlift with that of another Florida city that was similar to Miami in a number of respects.

As these examples show, the idea of trying to eliminate sources of variation other than the treatment by matching has been incorporated to a varying extent in analyses of natural experiments. Yet, when the analyst has observations on individual units, this fundamental idea can be taken much further.

The ideal experimental design to estimate the effect of a treatment would entail the following specifications. From the relevant population, a sample of pairs of individuals would be selected who were matched in terms of an assortment of characteristics. From this sample of matched individuals, one person (or other unit) in each pair would be randomly assigned to the treatment, the other would act as a control. Matching individuals increases the precision of the estimate of the effect of the treatment. Randomly assigning members of matched pairs to treatment and control groups insures that the treatment is uncorrelated with other variables that affect the outcome.

The advantage of matching before-treatment observations with after-treatment observations (and control observations with treatment observations) can be easily seen from the formula for the variance of a difference. Let us first consider simply a before and after analysis where y_0 is the before treatment outcome and y_1 is the after treatment outcome.

$$\text{Var } (y_1 - y_0) = \text{Var } y_1 + \text{Var } y_0 - 2\text{cov}y_1y_0$$

Evidently, the greater the covariance between y_1 and y_0, the smaller the variance of the difference. Independently drawn samples will have the largest variance. Thus, matching should be incorporated wherever possible in the collection of information for the analysis of a natural experiment. If possible the same individual units should be followed before and after the treatment, and control observations should be matched with treatment observations. The matched data should then be analyzed by taking the average of the individual differences; not to do so would fail to take advantage of the matching.

However, even if the analyst of a natural, as opposed to real, experiment has two random samples at their disposal, one before the treatment and one after, the precision of the estimated effect of the treatment can be increased by matching the before/after observations on the basis of their characteristics. Thus, in Peters' before-after study of the effect of a legislative change on divorce, the preciseness of her estimated effect of the treatment—the change in legislation—could have been improved if she divided her before and after samples according to characteristics such as age, education, and year of marriage. The difference in the probability of divorce for these pairs of before-after observation groups could be measured. Her estimated effect of the divorce law would then be the average of the differences for these pairs of before-after observations.

Similarly, in analyzing control-treatment groups, observations can be paired across the groups according to similar characteristics. Ideally, matching should be incorporated in the initial data collection for analyzing a natural experiment. Thus, in the Card/Krueger analysis of the effect on low-wage employment of the increase in the minimum wage in New Jersey versus Pennsylvania, pairs of New Jersey-Pennsylvania fast-food restaurants could have been chosen that were similar in several characteristics, such as the income level of the neighborhoods they served, how long they had been in business, the number of employees in the initial period, etc. Such pairs of restaurants could have been drawn from the New Jersey-Pennsylvania sample of restaurants, rather than drawing the New Jersey and Pennsylvania samples separately. The analysts could then have computed the difference in the before-after employment for each restaurant pair and taken the average of the differences.

There is, however, no reason why matching treatment and control observations cannot be incorporated into the analysis of a natural experiment even if it was not part of the original data collection effort. In his study of the effect of military service on earnings (in which the random assignment of service numbers was used to distinguish the treatment group from the control group), Angrist (1990) notes a lack of statistical significance in his results. However, the preciseness of his estimated treatment effect might be increased if the members of the control group (those with non-draft selective service numbers) were matched according to several characteristics (such as initial earnings) to members of the treatment group (those with draft selective service numbers). The differences in the before-after outcomes between the matched individuals would then be measured and averaged across all treatment-control observation pairs. Doing so would likely increase the statistical significance of Angrist's results.

Matching would not be done if we were trying to explain the level of a variable, or changes in a variable. To match in this case would be to throw out relevant information. Thus, matching individuals on the basis of their education in the Peters' study eliminates the possibility of estimating the effect of education on the probability of divorce. However, if our goal is to estimate the effect of a treatment—the effect of a certain legislative provision on the probability of divorce—rather than explaining the level of divorce, then matching before and after observations, and treatment and control observations, will increase the precision with which we can estimate that effect. If one wants to know how the treatment effect varies by the level of characteristics, then one can still adopt a non-regression framework by subdividing the data by the relevant characteristics.

IV. Summary

The analysis of experimental data, whether from natural or actual experiments, offers considerable promise for shedding light on a number of policy issues. Yet, in economic studies, natural and actual experiments could be better exploited in terms of the precision of the estimated effects from their analysis.

Precision can be increased by analyzing the average of the individual differences between control and treatment observations, as opposed to the difference in the means of control and treatment groups. The advantage of measuring the average of the differences, rather than the difference in averages, may have gone unnoticed by most economists because of the generally embraced goal by economists of trying to explain levels of variables or changes in levels of variables—the appropriate domain of the regression format—rather than accurately measuring the effect of a treatment on an outcome.

Matching before and after observations also increases the precision of the estimate. Ideally the same units would be followed over time. But in absence of this, before and after observations could be matched on characteristics. Forming matched pairs between the control and treatment groups also increases the potential precision with which an effect can be measured.

In what was arguably the most important social science experiment done in the U.S., it appears that none of the analyses of the Negative Income Tax experiments used the average of the individual differences. Reanalyzing NIT experimental data using a non-regression perspective could uncover important truths that have as yet remain uncovered.

References

Achen, Christopher H., *The Statistical Analysis of Quasi-Experiments*, Berkeley and Los Angeles, California: University of California Press, 1986.

Angrist, Joshua D., "Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence form Social Security Administrative Records, American Economic Review, vol. 80, 1990, pp. 313-336.

Card, David, "The Impact of the Mariel Boatlift on the Miami Labor Market," *Industrial and Labor Relations Review*, vol. 43, January 1990, pp. 245-57.

Card, David and Alan Krueger, " Minimum Wages and Employment: A Case Study of the Fast Food Industry in New Jersey and Pennsylvania, American Economic Review, vol. 84, pp. 772-793.

Hunt, Jennifer. 1992. The impact of the 1962 repatriates from algeria on the French labor market. *Industrial and Labor Relations Review* 45(3): 556-572.

Lyon, Herbert L. and Julian L. Simon, "Price Elasticity of the Demand for Cigarettes in the United States," *American Journal of Agricultural Economics*, vol. 50, November 1968, pp. 888-895.

Meyer, Bruce D. "Natural and Quasi-Experiments in Economics," *Journal of Business and Economic Statistics*, April 1995, pp. 151-161.

Orcutt, Guy H., "Data, Research, and Government," *American Economic Review, Proceedings*, May 1970.

Peters, H. Elizabeth. 1986. "Marriage and Divorce: Informational Constraints and Private Contracting." American Economic Review, 76(3): 437–54

Simon, Julian L., "The Price Elasticity of Liquor in the U.S. and a Simple Method of Determination," *Econometrica*, vol. 34, January 1966, pp. 193-205.

Table 1: Various Studies Using Regression Technique to Analyse Treatment Effects

I. THE ANALYSIS OF EXPERIMENTAL DATA EXCEPT FOR NIT

Studies	Regression Models
Effects of beneficiary participation on learning outcome of children. Indian Household level data were used. Banerjee, A, Banerji, B, Duflo, E, Glennerster, R. and Khermani, S. (2010) "Pitfalls of Participatory Programs: Evidence from a Randomized Evaluation in Education in India" *American Economic Journal: Economic Policy,* 2 (1), 1-30.	$$y_{ijk}=\alpha+\beta_{1k}T_1+\beta_{2k}T_2+\beta_{3k}T_3+X\gamma_k+\varepsilon_{ijk}$$ y_{ijk}: children's reading ability. T_1 : Whether individual resides in a village where mobilization only intervention occurred. T_2 : mobilization and information intervention. T_3 : mobilization, information and "Real India" camps intervention. X : baseline values for all the outcomes in the family.
Peer effects on academic performance from randomly assigned roommates using U.S. student level data. Sacerdote, B. (2001) "Peer Effects with Random Assignment: Results for Dartmouth Roommates" *Quarterly Journal of Economics,* 116 (2), 681-704.	$$(1): GPA_i=\sigma+\alpha^*(ACA_i+\mu_i)+\beta^*(ACA_j+\mu_j)+\gamma^*GPA_j+\varepsilon_i$$ $$(2): GPA_j=\sigma+\alpha^*(ACA_j+\mu_j)+\beta^*(ACA_i+\mu_i)+\gamma^*GPA_i+\varepsilon_j$$ GPA: grade point average. ACA: single academic index to measure ability.
Effects on corruption from randomized field experiment of monitoring practices on road projects. Indonesian village level data. Olken, B. (2007) "Monitoring Corruption: Evidence from a Field Experiment in Indonesia" *Journal of Political Economy,* 115 (2), 200-249.	$$Y_{ijk}=\alpha_1+\alpha_2 Audit_{jk}+\alpha_3 Invitations_{ijk}+\alpha_4 InvitationsandComments_{ijk}+\varepsilon_{ijk}$$ y: differences in project expenditures between what is reported and what actually occurred. Audit: project being externally audited. Invitation: participation in accountability meetings. InvitationandComments: providing an anonymous comment form to villages.
Effects of disclosing information about corruption practices on electoral accountability. Brazilian municipal level data were used. Ferraz, C. and Finan, F. (2008) "Exposing Corrupt Politicians: The Effects of Brazilian Publicly Released Audits on Electoral Outcomes" *Quarterly Journal of Economics,* 123 (2), 703-745.	$$E_{ms}=\alpha+\beta A_{ms}+X_{ms}\gamma+v_s+\varepsilon_{ms}$$ E_{ms} : electoral performance of incumbent mayor in municipality m and state s. A_{ms} : whether municipality was audited or not. X_{ms} : municipality and mayor characteristics. v : state fixed effects.

II. THE ANALYSIS OF NATURAL EXPERIMENTS

Studies	Regression Models
Effects of terrorist conflict on economic performance of Basque Country in Spain. Firm level and region level data in Spain were used. Abadie, A. and Gardeazabal, J. (2003) "The Economic Costs of Conflict: A Case Study of the Basque Country" *American Economic Review,* 93 (1), 113-132.	$R_t^j = \alpha^j + \beta_1^j R_t^m + \beta_2^j SMB_t + \beta_3^j HML_t + \gamma_1^j Goodnews_t + \gamma_2^j Baddnews_t + AR_t^j$ R: excess return on a buy-and-hold portfolio j: Basque or non-Basque. t: date. SMB: difference between returns of portfolios composed by small and big size stocks. HML: difference between the returns of portfolios composed by high and low book-to-market stocks. Goodnews: periods when "cease fire" was valid. Badnews: periods when "cease fire" was invalid.
Effects of compulsory schooling on education and earning. U.S. individual level data. Angrist, J. and Krueger, A. (1991) "Does Compulsory School Attendance Affect Schooling and Earnings?" *Quarterly Journal of Economics*, 106 (4), 979-1014.	$(1): E_i = X_i\pi + \sum_c Y_{ic}\sigma_c + \sum_c \sum_j Y_{ic}Q_{ij}\theta_{jc} + \varepsilon_i$ $(2): \ln W_i = X_i\beta + \sum_c Y_{ic}\xi_c + \rho E_i + \mu_j$ E: education. X: vector of covariates. Q_{ij} : whether individual was born in quarter j (j=1,2,3). Y_{ic} : whether individual was born in year c (c=1,2,3…,10). W: weekly wage.
Effects of Colombia's PACES program (secondary school vouchers) on longer-run educational outcomes of those treated. Colombian administrative individual level data. Angrist, J. Bettinger, E. and Kremer, M. (2006) "Long-Term Educational Consequences of Secondary School Vouchers: Evidence from Administrative Records in Colombia" American Economic Review, 96 (3), 847-862.	$(1): y_i = X_i^{'}\beta + \alpha D_i + \eta_i$ $(2): Y_i(\tau) \equiv 1[T_i y_i \geq \tau]y_i + 1[T_i y_i < \tau]\tau$ y: Latent scores. D: whether student won the lottery or not. X: student characteristics, age, sex. τ : positive threshold level for Tobit $y_i(\tau)$.
Effects of Head Start funding program on health and schooling. U.S. county and individual level data were used. Ludwig, J. and Miller, D. (2007) "Does Head Start Improve Children's Life Chances? Evidence from a Regression Discontinuity Design" Quarterly Journal of Economics, 122 (1), 159-208.	$(1): G_c = 1(P_c \geq P_{300}),$ $(2): Y_c = m(P_c) + G_c\alpha + v_c$ G_c : whether grant-writing assistance (for Head Start funding) is provided or not. P_c : poverty rate for ranking c, c=1 means poorest county in the U.S. $P_{300} = 59.1984$ Y_c : average outcome for county c, i.e. children mortality rate. $m(P_c)$: some smooth function around the cutoff point .

III. THE ANALYSIS OF NIT EXPERIMENTAL DATA

Studies	Regression Models
Effects of negative income tax plans relative to Aid to Families with Dependent Children program on the rate of marital dissolutions. U.S. individual (couples) level data. Cain, G. and Wissoker, D. (1990) “A Reanalysis of Marital Stability in the Seattle-Denver Income Maintenance Experiment” *American Journal of Sociology,* 95 (5), 1235-1269.	$\ln r_t = E^{'}\alpha + X^{'}\beta + \gamma t$ r: log of the instantaneous rate of marital dissolution. E: vector of experimental treatment variables: eligible for NIT, or NIT for 3 yrs, or AFDC, or training, education, job counseling treatment. X: personal and family control variables. t: time.
Effects of Gary Income Maintenance Experiment on the school enrolment and labor supply decisions of black teenagers. U.S. individual level data. McDonald, J. and Stphenson Jr, S. (1979) “The Effects of Income Maintenance on the School-Enrollment and Labor-Supply Decisions of Teenagers” *Journal of Human Resources,* 14 (4), 488-495.	$Logit(y_i) = X^{'}\alpha + E_i^{'}\alpha_1 + \beta_1 NIT_i + \beta_2 Breakeven_i + \beta_3 NIT_i * Breakeven_i + \varepsilon_i$ y: represents logistic outcomes including school enrolment, labor force participation, working etc. X: incorporates family background and personal characteristics as well as ability measure. E: income measures, including family income net of teenager income, AFDC receipts, food stamp etc. NIT: participating in Gary experiment (treatment dummy). Breakeven: dummy for income level reaching break-even point.
Effects of Gary Negative Income Tax Experiment on household labor supply, i.e. husbands; female hh heads; wives. U.S. household and individual level data. Moffitt, R. (1979) “The Labor Supply Response in the Gary Experiment” *Journal of Human Resources,* 14 (4), 477-487.	$H = \gamma + \sigma^{'} W(1-r) + \sigma^{''}(-Wt) + \eta^{'} N + \eta^{''} B_0$ H: labor supply measure, employment rate, working hours etc. W: gross wage rate. r: average non-NIT tax rate on earnings. t: average NIT tax rate on earnings. B: NIT benefit per month at zero hours. N: non-NIT nonwage income.
Effects of Denver Income Maintenance Experiment on labor supply decisions of 20-year eligible families. U.S. individual/household level data. Robbins, P. (1984) “The Labor Supply Response of Twenty-Year Families in the Denver Income Maintenance Experiment” *Review of Economics and Statistics,* 66 (3), 491-495.	$H_t = \alpha_0 + \alpha_1 H_p + \alpha_2 E + \alpha_3 X_p + \alpha_4 T + e_t$ H_t: Labor supply in period t. H_p: labor supply in the pre-experimental period. **E:** a set of normal income dummy variables used to assign families to experimental treatment. X_p : exogenous pre-experimental variables. **T:** dummy for experimental treatment.

Note: The listed papers are only selected examples from the literature of NIT and Treatment effects. We do not intend to survey the entire literature so that the list is far less than exhaustive.