Economics 152
Section Notes

Program Evaluation and the Difference in Difference Estimator

1 Program Evaluation

1.1 Notation

We wish to evaluate the impact of a program or **treatment** on an outcome Y over a population of individuals. Suppose that there are two groups indexed by treatment status $T = 0, 1$ where 0 indicates individuals who do not receive treatment, i.e. the **control group**, and 1 indicates individuals who do receive treatment, i.e. the **treatment group**. Assume that we observe individuals in two time periods, $t = 0, 1$ where 0 indicates a time period before the treatment group receives treatment, i.e. **pre-treatment**, and 1 indicates a time period after the treatment group receives treatment, i.e. **post-treatment**. Every observation is indexed by the letter $i = 1, ..., N$; individuals will typically have two observations each, one pre-treatment and one post-treatment. For the sake of notation let $\bar{Y}_0^T$ and $\bar{Y}_1^T$ be the sample averages of the outcome for the treatment group before and after treatment, respectively, and let $\bar{Y}_0^C$ and $\bar{Y}_1^C$ be the corresponding sample averages of the outcome for the control group. Subscripts correspond to time period and superscripts to the treatment status.

1.2 Modeling the Outcome

The outcome Y_i is modeled by the following equation

$$Y_i = \alpha + \beta T_i + \gamma t_i + \delta (T_i \cdot t_i) + \varepsilon_i \qquad \text{(Outcome)}$$

where the coefficients given by the greek letters $\alpha, \beta, \gamma, \delta$, are all unknown parameters and ε_i is a random, unobserved "error" term which contains all determinants of Y_i which our model omits. By inspecting the equation you should be able to see that the coefficients have the following interpretation

$\alpha =$ constant term
$\beta =$ treatment group specific effect (to account for average permanent differences between treatment and control)
$\gamma =$ time trend common to control and treatment groups
$\delta =$ true effect of treatment

The purpose of the program evaluation is to find a "good" estimate of δ, $\hat{\delta}$, given the data that we have available.

Example 1 *Card and Krueger (1994, AER) in "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania" try to evaluate the effect of the minimum wage (the treatment) on employment (the outcome). On April 1, 1992, New Jersey's minimum wage rose from $4.25 to $5.05 per hour. To evaluate the impact of the law, the authors surveyed 410 fast-food restaurants in New Jersey (the treatment group) and eastern Pennsylvania (the control group) before and after the rise. Y_i is the employment of a fast food restaurant, T_i is an indicator of whether or not a restaurant is in New Jersey, and t_i is an indicator of whether the observation is from before or after the minimum wage hike.*

1.3 Assumptions for an Unbiased Estimator

A reasonable criterion for a good estimator is that it be **unbiased** which means that "on average" the estimate will be correct, or mathematically that the expected value of the estimator

$$E\left[\hat{\delta}\right] = \delta$$

The assumptions we need for the difference in difference estimator to be correct are given by the following

1. The model in equation (Outcome) is correctly specified. For example, the additive structure imposed is correct.

2. The error term is on average zero: $E[\varepsilon_i] = 0$. Not a hard assumption with the constant term α put in.

3. The error term is uncorrelated with the other variables in the equation, including the constant:

$$\begin{aligned} cov(\varepsilon_i, T_i) &= 0 \\ cov(\varepsilon_i, t_i) &= 0 \\ cov(\varepsilon_i, T_i \cdot t_i) &= 0 \end{aligned}$$

the last of these assumptions, also known as the **parallel-trend** assumption, is the most critical.

Under these assumptions we can use equation (Outcome) to determine that expected values of the average outcomes are given by

$$\begin{aligned} E\left[Y_0^T\right] &= \alpha + \beta \\ E\left[Y_1^T\right] &= \alpha + \beta + \gamma + \delta \\ E\left[Y_0^C\right] &= \alpha \\ E\left[Y_1^C\right] &= \alpha + \gamma \end{aligned}$$

These equations will prove helpful below.

2 The Difference in Difference Estimator

Before explaining the difference in difference estimator it is best to review the two simple difference estimators and understand what can go wrong with these. Understanding what is wrong about as an estimator is as important as understanding what is right about it.

2.1 Simple Pre versus Post Estimator

Consider first an estimator based on comparing the average difference in outcome Y_i before and after treatment *in the treatment group alone.*[1]

$$\hat{\delta}_1 = \bar{Y}_1^T - \bar{Y}_0^T \tag{D1}$$

Taking the expectation of this estimator we get

$$\begin{aligned} E\left[\hat{\delta}_1\right] &= E\left[\bar{Y}_1^T\right] - E\left[\bar{Y}_0^T\right] \\ &= [\alpha + \beta + \gamma + \delta] - [\alpha + \beta] \\ &= \gamma + \delta \end{aligned}$$

which means that this estimator will be biased so long as $\gamma \neq 0$, i.e. if a time-trend exists in the outcome Y_i then we will confound the time trend as being part of the treatment effect.

[1] This would be the estimate one would get from an OLS estimate on a regression equation of the form

$$Y_i = \alpha_1 + \delta_1 T_i + \varepsilon_i$$

on the sample from the treatment group only.

2.2 Simple Treatment versus Control Estimator

Next consider the estimator based on comparing the average difference in outcome Y_i post-treatment, between the treatment and control groups, *ignoring pre-treatment outcomes*.[2]

$$\hat{\delta}_2 = \bar{Y}_1^T - \bar{Y}_1^C \tag{D2}$$

Taking the expectation of this estimator

$$\begin{aligned} E\left[\hat{\delta}_1\right] &= E\left[\bar{Y}_1^T\right] - E\left[\bar{Y}_1^C\right] \\ &= [\alpha + \beta + \gamma + \delta] - [\alpha + \gamma] \\ &= \beta + \delta \end{aligned}$$

and so this estimator is biased so long as $\beta \neq 0$, i.e. there exist permanent average differences in outcome Y_i between the treatment groups. The true treatment effect will be confounded by permanent differences in treatment and control groups that existed prior to any treatment. Note that in a randomized experiments, where subjects are randomly selected into treatment and control groups, β should be zero as both groups should be nearly identical: in this case this estimator may perform well in a controlled experimental setting typically unavailable in most program evaluation problems seen in economics.

2.3 The Difference in Difference Estimator

The **difference in difference** (or "double difference") estimator is defined as the difference in average outcome in the treatment group before and after treatment *minus* the difference in average outcome in the control group before and after treatment[3]: it is literally a "difference of differences."

$$\hat{\delta}_{DD} = \bar{Y}_1^T - \bar{Y}_0^T - \left(\bar{Y}_1^C - \bar{Y}_0^C\right) \tag{DD}$$

Taking the expectation of this estimator we will see that it is unbiased

$$\begin{aligned} \hat{\delta}_{DD} &= E\left[\bar{Y}_1^T\right] - E\left[\bar{Y}_0^T\right] - \left(E\left[\bar{Y}_1^C\right] - E\left[\bar{Y}_0^C\right]\right) \\ &= \alpha + \beta + \gamma + \delta - (\alpha + \beta) - (\alpha + \gamma - \gamma) \\ &= (\gamma + \delta) - \gamma \\ &= \delta \end{aligned}$$

This estimator can be seen as taking the difference between two pre-versus-post estimators seen above in (D1), subtracting the control group's estimator, which captures the time trend γ, from the treatment group's estimator to get δ. We can also rearrange terms in equation (DD) to get $\hat{\delta}_{DD} = \bar{Y}_1^T - \bar{Y}_1^C - \left(\bar{Y}_0^T - \bar{Y}_0^C\right)$ in which can be interpreted as taking the difference of two estimators of the simple treatment versus control type seen in equation (D2). The difference estimator for the pre-period is used to estimate the permanent difference β, which is then subtracted away from the post-period estimator to get δ.

Another interpretation of the difference in difference estimator is that is a simple difference estimator between the actual $\bar{Y}_1^T$ and the $\bar{Y}_1^T$ that would occur in the post treatment period to the treatment group had there been no treatment $\bar{Y}_{cf}^T = \bar{Y}_0^T + \left(\bar{Y}_1^C - \bar{Y}_0^C\right)$, where the subscript "$cf$" refers to the term "counterfactual," so that $\hat{\delta}_{DD} = \bar{Y}_1^T - \bar{Y}_{cf}^T$. This observation $\bar{Y}_{cf}^T$, which has expectation $E\left[\bar{Y}_{cf}^T\right] = \alpha + \beta + \gamma$, does not exist: it is literally "contrary to fact" since there actually was a treatment in fact. However if our

[2] This would be the estimate one would get from an OLS estimate on a regression equation of the form

$$Y_i = \alpha_2 + \delta_2 t_i + \varepsilon_i$$

on the post-treatment samples only.

[3] This would be the estimate one would get from an OLS estimate of a regression equation of the form given by (Outcome) on the entire sample. If we have each observation before and after we could also estimate δ with the equation

$$\Delta Y_i = \gamma + \delta T_i + u_i$$

where $\Delta Y_i = Y_{i1} - Y_{i0}$ is the post outcome minus the pre outcome for observaton i.

assumption are correct we can construct legitimate estimate of $\bar{Y}_{cf}^T$, taking the pre treatment average $\bar{Y}_0^T$ and adding the our estimate β using the pre versus post difference for the control group.

It is common to find difference in difference estimators presented in a table of the following form.

	Pre	**Post**	**Post-Pre Difference**
Treatment	$\bar{Y}_0^T$	$\bar{Y}_1^T$	$\bar{Y}_1^T - \bar{Y}_0^T$
Control	$\bar{Y}_0^C$	$\bar{Y}_1^C$	$\bar{Y}_1^C - \bar{Y}_0^C$
T-C Difference	$\bar{Y}_0^T - \bar{Y}_0^C$	$\bar{Y}_1^T - \bar{Y}_1^C$	$\bar{Y}_1^T - \bar{Y}_1^C - \left(\bar{Y}_0^T - \bar{Y}_0^C\right)$

Notice that the first row ends with the estimate $\hat{\delta}_1$, the second column ends with estimate $\hat{\delta}_2$, and the lower right hand corner entry gives the estimate $\hat{\delta}_{DD}$.

Example 2 *According to the model, by Card and Krueger (1994) comparisons of employment growth at stores in New Jersey and Pennsylvania (where the minimum wage was constant), provide simple estimates of the effect of the higher minimum wage. Some of the results from Table 3 are shown below with the average employment in the fast-food restaurants, with standard errors in parentheses*

	Before Increase	***After Increase***	***Difference***
New Jersey	20.44	21.03	0.59
(Treatment)	(0.51)	(0.52)	(0.54)
Pennsylvania	23.33	21.17	−2.16
(Control)	(1.35)	(0.94)	(1.25)
Difference	−2.89	−0.14	2.76
	(1.44)	(1.07)	(1.36)

The difference in difference estimator shows a small increase in employment in New Jersey where the minimum wage increased. This came as quite a shock to most economists who thought employment would fall. Notice that we can see that prior to the increase in the minimum wage Pennsylvania had higher employment than New Jersey and that it was bound to fall to a lower level. This may be a failure in the parallel trend assumption. However the small, albeit insignificant increase in employment in New Jersey makes it hard to accept the hypothesis that employment actually decreased in New Jersey over this time. Although still somewhat controversial, this study helped change the common presupposition that a small change in the minimum wage from a low level was bound to cause a significant decrease in employment.

2.4 Problems with Difference in Difference Estimators

If any of the assumptions listed above do not hold then we have no guarantee that the estimator $\hat{\delta}_{DD}$ is unbiased. Unfortunately, it is often difficult and sometimes impossible to check the assumptions in the model as they are made about unobservable quantities. Keep in mind that small deviations from the assumptions may not matter much as the biases they introduce may be rather small, biases are a matter of degree. It is also possible, however, that the biases may be so huge that the estimates we get may be completely wrong, even of the opposite sign of the true treatment effect.

One of the most common problems with difference in difference estimates is the failure of the parallel trend assumption. Suppose that $cov\left(\varepsilon_i, T_i \cdot t_i\right) = E\left(\varepsilon_i \left(T_i \cdot t_i\right)\right) = \phi$ so that Y follows a different trend for the treatment and control group. The control group has a time trend of $\gamma^C = \gamma$, while the treatment group has a trend of $\gamma^T = \gamma + \phi$. In this case the difference in difference estimator will be biased as

$$E\left[\hat{\delta}_{DD}\right] = \left(\gamma^T + \delta\right) - \gamma^C = \gamma + \phi + \delta - \gamma = \delta + \phi \neq \delta$$

The failure of the parallel trend assumption may in fact be a relatively common problem in many program evaluation studies, causing many difference in difference estimators to be biased.

One way to help avoid these problems is to get more data on other time periods before and after treatment to see if there are any other pre-existing differences in trends. It may also be possible to find other control groups which will can provide additional underlying trends. There is a huge literature on this subject, although a good place to start is Meyer (1995, *Journal of Business and Economic Statistics*).