

Difference in differences

This problem describes an impact evaluation of the French enterprise zone program which was initiated in 1997 to help unemployed workers find employment. It granted a significant wage-tax exemption (about one third of total labor costs) to firms hiring at least 20% of their labor force from local residents. These enterprise zone program are becoming popular in western countries to help out disadvantaged areas. The problem is inspired from Gobillon, Magnac and Selod, 2012, "Do unemployed workers benefit from enterprise zones? The French experience", *Journal of Public Economics* 96 (2012) 881–892. The program seems to have a small but significant effect on the rate at which unemployed workers find a job. This effect is localized and significant only in the short run.

The dataset consists of variables for 271 municipalities in the Paris region over an extended period of time from 1993 to 2003. These municipalities are not Paris itself and have more than 8,000 inhabitants and less than 100,000. The enterprise zone program was implemented in 13 municipalities only. Those were chosen using a "score" (which remains unknown) made up of variables defined below (*txcho90*, *tx25*, and *pfi_1996*). The employment zones are only sub-areas of the 13 municipalities, on average slightly less than half of their total area. For lack of data at a finer level, we consider that the treatment is given to municipalities. The main outcome is the logarithm of the probability to exit unemployment to a job for residents in each municipality in each half-year in this time span.

The dataset comprises the following variables:

dc : Municipality id

nomcom : Name of the municipality

per : Period (t= 1 to 20)

ez : Enterprise zone (1 = yes)

stock : Number of unemployed

esem : Number of exits to a job from unemployment

asem : Number of exits to a non-employment from unemployment

isem : Number of other exits from unemployment

entry : Number of entries into unemployment

rtc_all60 : Density of jobs within 1 hour in public transport

rvp_all60 : Density of jobs within 1 hour driving

tx_nodip : School drop outs (Frequency)

tx_tech : Technical diplomas (Frequency)

tx_univ : College educated (Frequency)

txcho90 : Unemployment rate in 1990

tx25 : Less than 25 years old in the population (Frequency)

pop90 : Population in 1990

dmin : Distance to the nearest EZ (kms)

pfi_1996 : Average net household income in 1996

Part I: Raw difference-in-differences

1. What are the economic channels through which the policy is acting on the propensity to exit unemployment (= leave the pool of unemployed)? Does it affect the entry rate into unemployment (= propensity to enter the pool of unemployed)?

The policy makes the employment of residents in the employment zone less costly thanks to the wage-tax exemption. It should therefore affect the propensity to exit unemployment in two ways:

- Unemployed people living in the zone should find a job more easily since it is cheaper for companies to hire them than someone else \Rightarrow Propensity to exit unemployment \nearrow
- Some people have given up on finding a job, they are not actively looking for a job and are therefore not counted as unemployed. These are called the *non-employed*. These people may see the propensity to exit unemployment increase after the implementation of the policy, and their hopes of finding a job may rise. They would then start looking for a job, and be counted as unemployed again. \Rightarrow Entry rate in unemployment \nearrow

It is not a priori clear which is the strongest effect, and therefore we can't say whether we expect the pool of unemployed to be bigger or smaller as a result of the reform. What is clear is that if you want to answer the question "Do employment zones help locals finding a job?", you do not want to capture the effect explained in the second bullet point. This is the reason why we will take the propensity to exit unemployment as the outcome variable, as opposed to, for instance the probability to be unemployed.

2. The exact treated area is not observed. We only observe the municipality in which it is located. Explain how the fact that we do not observe the exact treated area affects the treatment effect parameters?

The observational level is the municipality. (And, just to compare with previous TPs, there is no randomization level here since there is no randomization, treatment was given according to a score instead). However, the level at which treatment is given is the employment zone, which is smaller than the municipality. But in this analysis, we don't observe the employment zone (since I just said the observational level is the municipality).

What are the consequences of this difference between treatment and observation levels? Well, it averages the treatment effects over the whole municipality. Given that we expect the effect of treatment to be zero in non-treated areas, the parts of the municipality which aren't treated should stay unchanged after treatment. However, should the treatment have an effect, treated zones within the municipality should see a positive change in the outcome variable (the propensity to exit unemployment). Therefore, if you average the treatment effect at the municipality level, you mix the positive change in outcome observed in treated areas of the municipality with a zero change in untreated ones. As a result, you dampen the treatment effect. So we will probably underestimate the effect of treatment.

3. The treatment was implemented at period 8. Estimate by difference-in-differences the impact of the treatment on the logarithm of the exit probability. Repeat using the

exit probability itself instead of the logarithm. Repeat using weights given by the population (and renormalized so that the number of municipalities is equal to the original number). Comment.

Let's define $R_m = 1$ if the m is one of the selected municipalities for treatment and 0 otherwise, and t^* the period in which the treatment was implemented (per=8). The regressions are following the equation:

$$Y_{mt} = \alpha + \beta \underbrace{R_m * \mathbb{1}\{t \geq t^*\}}_{=d_{mt}} + \gamma R_m + \delta_t + \epsilon_{mt}$$

Here, β is the treatment effect, since it is "activated" for municipalities that were chosen to be treated, at the time of treatment. So $d_{mt} = 1$ if municipality m is treated *at time* t . Therefore, this regression is implicitly assuming that the effect of treatment is the same for treated municipalities on the first year after treatment and the following years. $d_{mt} = 1$ if the region is treated at the moment, no matter when it started being treated. So implicitly, this equation assumes that the effect of treatment is static, and that it is the same every year.

Let's explain what the population-weighted regression is doing. Municipalities all have a different number of inhabitants. The first two regressions (without the weights) forget about the number of inhabitants in each municipality and just consider 1 observation is 1 municipality at time t . The weights are here to make municipalities with more inhabitants have a bigger impact on the estimation. So instead of counting only for one observation, they would count for a bit more than one. But we need the overall number of observations to be unchanged after making this modification, otherwise we artificially increase n ! So to compensate, we make municipalities with little inhabitants count for a bit less than 1, so that overall the sum of the weighted observations is still n . As a result of this weighted regression, if 2 big municipalities see no effect of treatment and 10 small municipalities see a big effect, we will conclude that the treatment effect is small. This wouldn't be true for the non-weighted regressions which would conclude that the treatment effect is big, since they wouldn't take population into account.

Concretely, the weights on each observation m (municipality) are:

$$Population_m \times \frac{Number_Municipalities}{Total_Population} = \frac{Population_m}{Total_Population} \times Number_Municipalities$$

So the weight on m is the percentage of the total population it represents $\times n$, the number of municipalities. The multiplication by n makes sure that

$$\sum_{m=1}^n Weight_m = n$$

Now let's look at the estimations' results. Note that the index of a treated municipality (*ez*) always has a negative and significant coefficient, meaning that treated municipalities have a relatively low unemployment exit probability *before treatment*. This was to be expected, since they were chosen according to a score showing poor labor market outcomes. The treatment effect is never significant and can be positive or negative. Note that the coefficients of time dummies indicate a decrease in the exit

probability from year 7. This is unrelated with treatment since the effect of treatment is captured in the coefficient for d . It reflects an evolution of the general economic situation at the time: labor outcomes seem to have worsened around Paris at the end of the 90s.

NB: What is the main identifying assumption allowing us to draw causal conclusions from these regressions? Parallel trends (also referred to as "equal trend" in the lecture slides)! In this context, it means that Y was following a similar *trend* before treatment in treated and untreated municipalities. This means that, although Y could have a different value across treatment groups, the gradient between Y_t and Y_{t+1} had to be the same.

Part II: Using the propensity score

We are wondering whether our results are driven by the fact that the control group is very different from the treated one. This would indeed make the parallel trend assumption likely to be violated, which would bias our treatment effect. In this part, we try to "make" treatment and control groups more similar, so that we can credibly assume that trends are parallel. we thus estimate the propensity score and get rid of control observations with too small a value. If control observations with a very low PS indeed follow a different trend than the rest, getting rid of them should help us estimate the treatment effect if there is one.

1. Estimate and predict the estimated propensity score using weighted Probit and using covariates *rtc_all60*, *rvp_all60*, *tx_nodip*, *tx_tech*, *tx_univ*, *txcho90*, *tx25*, *pop90* and *dmin*. Compare its distribution function in the treatment and control groups. Comment.

The propensity score is the probability to be treated according to some covariates. Because treatment groups are determined in period 8, only a cross-section in $t=8$ is necessary to estimate the propensity score.

The distributions are very different. The propensity score is much higher for treated municipalities (min, mean, median and quartiles). This is not surprising given that the selection of treatment zones wasn't random. However, one problematic finding is that the common support assumption is violated. Indeed, in the control group, no observation has a propensity score (PS) above 0.72 whereas in the treatment group you have 25% of observations with a PS above that value. As for low values of the PS, although we do have municipalities within the treated group that have a very low PS value, the lowest value is 0.001. The problem is that the median PS in the control sample is 0.001, which means that 50% of control observations are below that value. So 25% of treated observation have no control counterpart in terms of PS, and almost half of control observations are useless since their PS is so low that they can't be compared with treatment observations.

2. Select the final control group using only municipalities whose value of the propensity score is above half of the smallest propensity score in the treatment group.

So here we are getting rid of observations in the control group that have a super low PS value, since they can't match the PS value of any observation in the treatment group. As a result, we hope that the difference in the distribution of the PS across

group will be less terrible, so that the differences in covariates between treated and control groups may be less stark.

Why do we want treatment and control groups to be similar here, given that we control for ez_m ? The treatment effect will be identified from variation in Y_{mt} among treated observations, removing the part of this variation that can be explained by time (δ_t) or treatment group (ez_m). For this treatment effect to be well estimated, we need for δ_t and ez_m to be well estimated as well. The estimation of δ_t is done from time variation on the entire sample (control + treatment groups), and we need that the effect of time on control and treatment groups is the same so that this coefficient is meaningful (this is the parallel trend assumption). If half of the control group is so different that its time trend differs from the rest, the parallel trend assumption is violated and δ_t is meaningless, biasing the treatment estimate. Therefore, by removing part of the control group we hope to avoid such problems on the parallel trend assumption and get an unbiased estimate of the average treatment effect.

Part III: Diff-in diffs using the common support

1. Reestimate by difference-in-differences the impact of the treatment on the logarithm of the exit probability. Compare results to the ones in I.3.

The DiD using the restricted control group still doesn't yield any significant treatment effect. Controlling for the PS doesn't help.

2. Our results so far suggest that there is no effect of treatment. We wonder whether this result could be in fact driven by unobserved heterogeneity at the municipality level: after all, treatment zones were chosen according to a score we don't observe, based on variables that we don't observe either. This leads us to re-think our model so that we can capture some potential unobserved heterogeneity. This is done through adding municipality-specific dummies to the regression.

$$Y_{mt} = \alpha + \beta d_{mt} + \gamma R_m + \delta_t + u_m + \varepsilon_{mt}$$

Why is γ not identified? We consider first differences $\Delta y_{mt} = y_{mt} - y_{mt-1}$ to get rid of the fixed effect. Rewrite the treatment equation and estimate it.

Before answering the question, note that this equation as well is implicitly assuming that β , the effect of treatment is static and the same over years. See discussion in I.3. Now, γ is not identified because the term γR_m is absorbed into the fixed effect u_m . Concretely, u_m is:

$$u_m = u_1 \times \mathbb{1}\{m = 1\} + u_2 \times \mathbb{1}\{m = 2\} + \dots$$

Therefore, all the part of Y_m that is explained by municipality-level specifics is captured by u_m . This means all the information that concerns m but that doesn't change over time is captured in u_m . This is the case of R_m , whose value doesn't change over time (it is either =1 forever or =0 forever since it says whether the municipality is part of those *which will be/have been selected for the program*). Therefore, all the information contained in R_m will be captured in u_m , which means γ isn't identified. This is also true if one includes the propensity score in the equation.

The estimation in first differences is the following (just compute $Y_{m,t} - Y_{m,t-1}$ using the regression specified in the question):

$$\Delta Y_{mt} = \beta \Delta d_{mt} + \Delta \delta_t + \Delta \varepsilon_{mt}$$

$\hat{\beta}$ is now positive and significant. It seems that there was a significant amount of selection on unobservables, since adding municipality fixed effects changes the results so much! It went undetected by the balancing test, which is by definition with selection on unobservables.

Note: There is no need to difference the time dummies to get $\Delta \delta_t$. You can just rename it $\Delta \delta_t \equiv \lambda_t$ to convince yourself of it! You just need to make sure that you don't get confused when interpreting the coefficient: it is no longer the time trend of the evolution of Y but the difference between the impact of t and that of $t - 1$. However, note that you need to get rid of one extra time dummy when making this regression since λ_t exists only from $t=2$ on. Therefore, to avoid collinearity, you get rid of the first time dummy (λ_2 in this case).

3. Repeat the previous question by keeping only observations at the period at which the treatment effect is implemented. Comment.

We mentioned in III.2 that the regression was assuming a static fixed treatment effect across years. This is also true for the first difference regression since the β we recover is the same for both equations.

Now let's look at the first-difference regression by period.

$$\Delta Y_{mt} = \beta \Delta d_{mt} + \Delta \delta_t + \Delta \varepsilon_{mt}$$

I claim that $\Delta d_{mt} = 1$ only for $t=8$. Indeed, $\Delta d_{mt} = d_{mt-1} - d_{mt}$ which is $=0$ always for non-treated municipalities since $d_{mt} = 0$. But for treated municipalities, since $d_{mt} = 1$ for all $t > 7$, $\Delta d_{mt} = 0$ for all t except $t=8$! Therefore, $\hat{\beta}$ is the same as in III.2.

Nonetheless, in III.2, the number of observations is much larger. Therefore, all the other coefficients than β will be more precisely estimated in III.2: when you restrict the sample to $t = 8$, you lose variation for other variables than the treatment.

Note that when taking all periods with the first-difference equation, the estimation of standard errors could be biased because of the presence of autocorrelation ($\Delta \varepsilon_{mt}$ and $\Delta \varepsilon_{mt-1}$ both contain ε_{t-1}). When we take this into account by computing standard errors that are robust to serial correlation (for instance using a jackknife or bootstrap method), we get slightly larger standard errors but the treatment effect is still significant at the 1% level. So overall focusing on period 8 is better for the precision of estimation of the treatment effect since it gets rid of inbuilt autocorrelation (and the number of observations on which β is estimated is the same in III.2 and III.3).

Part IV: Robustness checks

1. Plot the average outcome variable over time for treated and control group separately. What can we say about the equal trends assumption? Formally test the validity of the equal trend assumption.

The first thing to do is to visually compare the evolution of the outcome across treatment groups before treatment. When doing so, we see that trends seem roughly parallel, except for a few differences (sharper drop for treatment between period 2 and 4, drop in period 7 for treatment only...). Reducing the control group to make it more similar to treatment based on the PS doesn't seem to help.

To formally test the equal trend assumption, we add "placebo" variables to our regression. These variables are similar to d , only for another year than $t = 8$. Concretely, we are trying to find an effect of treatment before treatment happened. If groups have parallel trends, this effect should be close to zero. In our regression, we add these placebo variables for years before treatment, and others after treatment. If those before treatment yield significant coefficients, we reject the equal trend assumption. However, coefficients of placebo variables after treatment are allowed to be significant, since they are a consequence of treatment.

Based on those results, we can't formally reject the equal trends assumption.

2. Repeat the first-difference regression by looking now at observations at longer gaps in periods i.e for instance the outcome:

$$\Delta_h y_{it} = y_{it} - y_{it-h}$$

for $h = 2$ (one year) or $h = 6$ (3 years)

Note that you need to get rid of the time dummy for $t = 3$, which is the first that exists in the second-difference setting. This is to avoid collinearity, see discussion in III.3.

In both cases, the treatment effect is insignificant. This is unexpected, and can come from several causes:

- The effect of treatment occurs only on the first year: somehow companies realize after the first hires that they prefer forgoing tax cuts and hiring people from outside the employment zone. Therefore, in the case of $h = 2$, when comparing $t = 9$ and $t = 7$ we find no effect since at $t = 9$ there is no effect of treatment.
- There is an announcement effect of treatment: in $t = 7$, companies heard that this treatment would be implemented next period. Therefore, they stopped hiring locals and waited for treatment to happen so they could benefit from the tax credits. As a result, when comparing $t = 8$ and $t = 7$ we find an effect that we associate to treatment, but it isn't actually the treatment effect, it is the effect of the announcement. Therefore, for $h = 2$, when comparing $t = 8$ and $t = 6$ we find no effect since in $t = 6$ this announcement doesn't happen (companies don't yet know about treatment).

A similar logic applies to $h = 6$.

Another explanation could simply be that DiD isn't an appropriate estimation method here because the parallel trend assumption is violated! To check this, we would need to plot outcomes separately for treated and untreated municipalities, and see whether the time trends seem to be the same (whether the curves seem parallel).

3. Change the control groups by excluding municipalities which are too close ($d_{min} < 5$ km, < 10 km) to an enterprise zone. Which assumption do we want to evaluate?

We want to evaluate the impact of spillovers. The line of thinking is: control municipalities which are close to treated ones could be negatively affected by treatment. How? Well you could imagine that companies which are in the treated municipality would usually hire workers from the neighboring municipalities. However, following the reform, they have an advantage in hiring from the treated area, therefore, when looking at 2 job applicants that are identical except one lives in the treated area and the other one doesn't, they will choose the one in the treated group.

Therefore, we imagine that including these municipalities in the control group could overestimate the treatment effect. However, we see that once we get rid of them the treatment effect basically doesn't change. We conclude that such spillovers weren't much of an issue after all.