14.771 - Problem Set 2

MIT

This problem set is about the NREGA (National Rural Employment Guarantee Act) program in India, which is (mostly likely) the world's largest public jobs program. The NREGA program guarantees to all workers in rural areas up to 100 days per year of government-funded employment, doing manual labour. NREGA labour is paid at the official statutory minimum wage, which is **above** the typical prevailing wage for rural labour in most areas of India. The program was rolled out nationally in three annual waves between 2006 to 2008, with approximately one-third of districts starting the program in each year, although we will consider slightly different dates for the purpose of the exercises.

Exercise 1

In this exercise, we will derive predictions to what should happen to this local labour market when NREGA is introduced.

1. First, let us derive the labour demand of a generic local labour market in this economy. We will assume that the labour market is competitive (i.e., firms are price-takers) and frictionless. The representative firm's production function is given by

$$F(L) = \log(L) \tag{1}$$

   where L is the labour employed at the firm. Assume that wage is given by w and that the firm maximises profits.

   Show that the labour demand curve in this economy is given by $D(w) = w^{-1}$. (tip: even though there is a single firm, we are assuming this firm is price-taker!)

2. We will now compute the labour market equilibrium in this local labour market. Let us take a

reduced-form approach, by simply assuming that labour supply is given by

$$S(w) = \alpha \cdot w \tag{2}$$

What is the equilibrium wage and employment in this labour market?

3. Let us now introduce NREGA to our model. One way to think about it is the following: Assume that the wage offered by NREGA is higher than the market equilibrium price. [1] So whenever a given worker can work for NREGA, they will prefer to work there, and will then supply 0 labour to the private sector. Conversely, on days they cannot work for NREGA (remember, they can only work 100 days per year), they supply labour according to equation 2. We can model this by assuming labour supply curve is now

$$S^N(w) = (1-p) \cdot \alpha \cdot w \tag{3}$$

where p is the proportion of workers that are working at NREGA at any given day (which might be something like $p = 100/365$). Show that the equilibrium wage in the **private sector** will be

$$w^{**} = \left(\frac{1}{(1-p)\alpha} \right)^{1/2}$$

4. Provide a brief explanation of why private wages increase. (max. 4 lines).

5. From the previous question, show how each of the following aggregate quantities is affected by NREGA. Provide a brief explanation for each of the comparative statics (4 lines max per item).
   (a) Private sector employment
   (b) Overall employment

Exercise 2 Suppose you wanted to estimate impact of the program on rural wages using a difference-in-difference design. Let us assume for this exercise that the program had the following rollout: in 2006 part of the municipalities receive NREGA and kept being treated until the end of your data set in 2008. Assume also you only have data from 2005-2008. Please answer the following questions **briefly**.

1. What is the estimating equation you would use to estimate the impact of NREGA by diff-in-diff?

[1] We are cheating a bit by assuming that. This is an equilibrium condition that we should check, but we will ignore this issue in this problem set.

2. (i) What is the **name** of the key assumption behind diff-in-diff models? (ii) Express it formally using potential outcomes notation. (iii) Is this a testable assumption?

3. A suggestive test we can make is that of *pre-trends*. Why can't we do this test with data from 2005-2008?

4. Assume we get data from 2002 to 2008. Write down the estimating equation you would use to test for pre-trends and explain what your estimates would need to look like for you to conclude that there are no pre-trends.

5. In this setting, what are possible threats to the identification assumption behind the diff-in-diff? Provide 2 examples of such threats and explain why it would matter in this setting (max. 3 lines per example).

Exercise 3 In this question we will estimate the impact of NREGA on private sector employment using an event study (staggered rollout) design in an artificial dataset. This dataset contains data for 100 municipalities between 2001 and 2010, where the programme was rolled out between 2003 and 2007. [2] You should use the dataset attached with the problem set, *ps2_data.csv*.

1. In this type of design, it is often more useful to think of the effect of a program h years after it was implemented (sometimes called "event time"), rather than in terms of calendar years. Let E_i be the adoption ("event") year of unit i and $K_{it} = t - E_i$ the number of periods unit i has been treated for. Define the variable, $D_{it}^h = \mathbb{1}_{[K_{it}=h]}$.

   Write the estimating equation, with the maximum number of lags and leads possible given the data you were given. The outcome variable should be employment in the private sector.

   Hint: In event study designs, you cannot estimate every lag and lead (see the notes for Recitation 4 for more detail).

2. Which type of standard error estimator should you use here? (Hint: check Bertrand, Duflo, and Mullainathan 2004 if you are unsure)[3]. Relate this briefly to your findings about clustering in problem set 1.

3. Estimate the equation you proposed on the item above and present an event study graph with

[2] We discussed event study designs in Recitations 4 and 5. Note that there is no never-treated group in this setting, because the programme was eventually rolled out everywhere.

[3] Bertrand, M., Duflo, E. and Mullainathan, S., 2004. How much should we trust differences-in-differences estimates?. The Quarterly journal of economics, 119(1), pp.249-275.

your results. Your graph should have the coefficients estimated for "pre" and "post" trends and confidence intervals for each of them. Figure 1 shows an example of one such graph with **a different** simulated dataset.

Note: You do not need to implement the fancy suggestions from Freyaldenhoven, Hansen, Perez, and Shapiro (2021)[4], although you are welcome to do so if you would like to.

4. (i) What does your graph tell about the presence of pre-trends? (ii) Does it provide supporting evidence that the identification strategy is valid? (iii) Propose and conduct a formal test of absence of pre-trends.

5. (i) What is the impact of NREGA on private sector employment? (max 3 lines). (ii) Discuss how the results contrast with the model's prediction (max 2 lines). (iii) Please give at least **two** reasons why you might be observing this result that are not modeled in Exercise 1 (max. 3 lines per explanation)

6. In Recitation 5, we discussed how event study estimators like the one we considered here can go wrong in the presence of treatment effect heterogeneity. Implement an alternative estimator of your choice that is robust to this issue and briefly discuss the results.

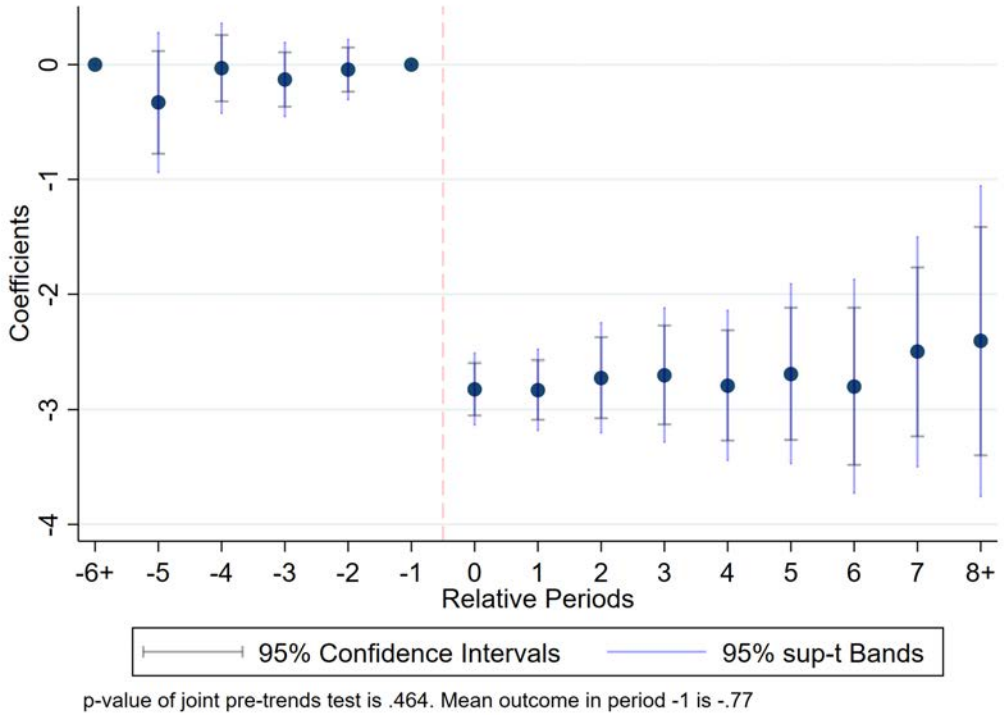
   Note 1: You do not need to implement the alternative estimator from scratch: feel free to use one of the Stata or R packages discussed in recitation.

   Note 2: You do not need to figure out and/or explain the source of any differences, although you are welcome to do so if you would like to. If you do, a good start would be to estimate either the estimator in Sun and Abraham (2021) or the estimator in Wooldridge (2021), both of which explicitly estimate the full set of identified $\tau_{e,h}$ (the coefficient for cohort e, relative time h), at least for $h \geq 0$.[5] Note that, for these estimators, you should drop any time periods $t \geq \max E_i$ because nothing after this point is identified in those models. Do any of the coefficients look much larger or smaller than the others? Then, take a look at the weights placed on the different (e, h) pairs using Sun and Abraham's Stata command, eventstudyweights. If you observed any particularly large or small coefficients, what weights are being placed on them?

[4] Freyaldenhoven, S., Hansen, C., Perez, J.P. and Shapiro, J.M., 2021. Visualization, Identification, and Estimation in the Linear Panel Event-Study Design (No. w29170). National Bureau of Economic Research.

[5] The references are as follows, but both of these estimators–including the estimating equations–are discussed explicitly in the notes to Recitation 5. References: Sun, L. and Abraham, S., 2020. Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. Journal of Econometrics. Wooldridge, J., 2021. Two-Way Fixed Effects, the Two-Way Mundlak Regression, and Difference-in-Differences Estimators. Available at SSRN 3906345.

Figure 1: Example of Event Study Graph **With Another Data Set**



14.771: Development Economics
Fall 2021