Université Paris-Saclay
Master of Economics – 1st year
Advanced Econometrics – Tutorial #7 *

April 04th 2022

Context of session and objectives

This session is an introduction to the difference in differences (DiD) estimation strategy based on the paper: "*Minimum Wages and Employment: A Case Study of the FastFood Industry in New Jersey and Pennsylvania*", by David Card and Alan B. Krueger, American Economic Review (1994). In this exercise, we are interested in estimating the effect of an increase in minimum wage on the demand for low skilled labor. The dataset for this exercise, the codebook presenting the dataset, as well as the research article of Card and Krueger(1994), are available on eCampus.

1 Presentation of the research question and the data

Before moving on to the practice, we address some preliminary questions on the research question and on the research design to identify a causal effect of a variation in minimum wage on the demand for labor.

1. What does economic theory predict on the effect on an increase in minimum wage on employment, in perfectly competitive markets?

> **Solution:** In a competitive labor market, increases in the minimum wage move up a downward-sloping demand curve. Higher minimums therefore reduce employment, perhaps hurting the very workers minimum-wage policies were designed to help. This is the prediction from economic theory, which is unambiguous: a rise in the minimum wage leads perfectly competitive employers to cut employment.
>
> Empirical results are rather unconclusive on this question, finding mixed results.

2. We want to estimate the effect empirically. We observe that minimum wage regulations vary across States in the US (some States have higher minimum wage regulations than others). We have a database with wage and employment for firms in different States.
   (a) Write the regression of employment on minimum wage.

*Contact: thibault.richard(at)ens-paris-saclay.fr

Solution: We here look at all state minimum wages in the United States, and its link with employment at the firm-level. We have firm-level data in the different US states, for one year. Denote i the firm index and s the US state where firm i is located. We have the following linear model for the employment of firm i in state s, denoted Y_{is}:

$$Y_{is} = \beta w_s + X_{is}'\delta + \varepsilon_{is}$$

where w_{is} is the minimum wage in state s and X_{is} can include several controls at the firm level.

(b) Do you think that this simple regression might give us the causal effect of the minimum wage on employment? Explain.

Solution: β here intends to capture the effect of minimum wage on employment using cross-sectional variations of employment and minimum wage across states. We may capture state effects and not the pure effect of states, if some unobserved variables are correlated with the the minimum wage and with employment.

In an improved perspective, we could have a multi-year panal dataset, and estimate another equation as such

$$Y_{is} = \lambda_s + \lambda_t + \gamma w_s + X_{is}'\delta + \varepsilon_{is}$$

We can here add state fixed effect, because we have several years of observations allowing us to derive a state fixed-effect. There are still sources of omitted variables, such as time-varying variables measured at the state level.

In order to assess the impact of minimum wage laws on employment, David Card and Alan Krueger exploited the exogenous policy change in minimum wage that occurred in April 1992 in New Jersey (NJ): the hourly minimum wage was raised from 4.25 to 5.05 dollars in NJ but not in nearby States. Card and Krueger collected data at fastfood stores in NJ, before (in February 1992) and after the policy change (in November 1992) and also collected data at fast food stores in a nearby State, Pennsylvania (PA), where the minimum wage was not changed.

3. Consider the policy change:

(a) What is the interest of looking at a policy change in order to estimate the effect of minimum wage on employment?

Solution: An exogeneous shock which thus should be distinguished of any other macroeconomic shocks. However, there might still be some identification issues if some shocks jointly occur.

(b) What is the interest to look at fast food stores?

Solution: These restaurants (Burger King, Wendys, and so on) are expected to be big minimum-wage employers. Thus, we expect to observe an impact of the change in regulation on employment. Also, these restaurants comply with minimum wage regulations and would be expected to raise wages in response to a rise in the minimum wage. Also, job requirements and products are relatively homogeneous, which make it easier to obtain reliable measures of employment and wages.

(c) What is the interest of having stores in different US states? Compare the characteristics of the stores by states and by time period.

Solution: Collecting data in the two states to have a subsample of stores affected by the regulation, and a subsample of store unaffected by the regulation. We can thus have treated observations and a comparison group, but the characteristics of the stores in PA needs to be very similar to the ones in NJ to be a good comparison group.

See STATA do-file for replication code.

Across variables, we observe that stores across states are mostly similar, with some variations, in particular in terms of employment. We notice that starting wage were similar in February but they change in November (stores in NJ seems to have complied with the regulation).

(d) We need to create unique variables measuring full-time equivalent employment and starting wage for the two time periods. Generate a variable denominated `fte` which measures the full-time equivalent employment in the store (including managers and the number of part-time employees count for half of the full-time employees), and another variable denominated `starting_wage`. *Note: be careful not to forget that the same variables observed in February 1992 and in November 1992 are stored in different variables.*

Solution: See STATA do-file for replication code.

(e) How can you check in the data that the policy has indeed been implemented in NJ stores and not in PA stores?

Solution: Before answering this question, let's explore the dataset, its structure, and discover the relevant variables. Notice that observations by time period are stacked one above each others, and that variables associated to the first and second interviews are stored in different variables, that is, `empft` measures the number of full-time employees in February in the store, and `empft2` measures the number of full-time employees in November.

Now, we present three alternative ways to observe that the implementation of the policy is in NJ and not in PA stores.

- Investigate the distribution of the starting wage between state and between time period. To do so, we generate four variables, for the starting wage for each pair of state and time period. We generate two histograms, for the two states, where we distinguish the distribution of the starting wage in February and and in November. We observe some differences in Pennsylvania betwee both time periods, with some more observations in the lower end of the distribution, but overall, these are not very clear patterns. While for New Jersey, we clearly observe that the vast majority of stores follow the new state law in November 1992 relative to February 1992. They set their starting wage at the new minimum wage (5.05$) in November 1992, while before there was more heterogeneity in starting wage, with the majority of stores set starting wages below 5$.
- We can test for statistical differences in means between time periods for both subsamples of states. For Pennsylvania, we observe that there are no differences in means between the starting wages in February and in November 1992. However, for New Jersey, there is a significant increase in the mean of the starting wages, underlining the implementation of the policy in New Jersey and not in Pennsylvania.
- We can count the number of observations offering a starting wage above 5.05$ by states and by time period, to compare their evolutions. And this approach leads to similar conclusions.

4. We now investigate the variation in full time employment in New Jersey stores. Compute the difference in average full-time employment employment in NJ stores before and after the policy change. Do you think this calculation provides an unbiased estimate of the effect the increase in minimum wage on employment? Why?

Solution:

$$\mathbb{E}\left[Y_{ist} \mid s = NJ, t = Nov\right] - \mathbb{E}\left[Y_{ist} \mid s = NJ, t = Feb\right]$$

Computing this for New Jersey, we get the following results:

	NJ
FTE employment before	20.44
	(0.51)
FTE employment after	21.03
	(0.52)
Change in mean FTE	0.59
	(0.54)

Standard errors of the mean are reported in parenthesis.

From these differences, we fail to reject the absence of a difference in full-time employment in NJ between the first and the second time period. In other words, it seems that FTE has not clearly changed in NJ before and after the change in minimum wage. There is an increase in FTE of 0.59 in average in NJ across period but this change is not significant. In other words, NJ store grew slightly after the increase in the minimum wage.

However, this difference captures the change in employment both between time period due to any macro-economic event or seasonal employment effects, ***and*** the policy change. FTE can change over time for reasons unrelated to the policy change. This makes impossible to distinguish the effect of the change in minimum wage on other time-varying factors. Thus, this does not provide an unbiased estimate of the effect the increase in minimum wage on employment.

An alternative solution would be to compare the means for NJ and PA in the second time period only. But using this approach, there might be systematic unmeasured differences in the two states that have nothing to do with the change in minimum wage. Thus attributing the difference in averages to the change in minimum wage might be misleading.

2 The Difference-in-Difference estimation strategy

Card and Krueger used their data set to compute differences-in-differences (DD) estimates of the effects of the New Jersey minimum wage increase. That is, they compared the change in employment in New Jersey to the change in employment in Pennsylvania around the time New Jersey raised its minimum wage.

5. Card and Krueger apply the following strategy: they compute the difference in average full-time employment before and after the policy change in NJ and Pennsylvania and calculate the difference in these average differences.

   (a) Compute this "difference-in-differences". How do you interpret this number? Under which hypothesis this strategy can estimate the causal effect of the minimum wage on employment?

Solution: See STATA do-file for replication code.
We obtain the following results:

	PA	NJ	NJ – PA
FTE employment before	23.33	20.44	-2.89
	(1.35)	(0.51)	(1.44)
FTE employment after	21.16	21.03	- 0.14
	(0.94)	(0.52)	(1.08)
Change in mean FTE	-2.16	0.59	2.76
	(1.25)	(0.54)	(1.36)

The relative gain (the "difference in differences" of the changes in employment) is 2.76 FTE employees (or 13 percent), with a t statistic of 2.03. To be the causal effect of the minimum wage on employment, there should be no other sources of variation in employment growth, such as differences across chains. Also, Pennsylvania should be a good comparison group, meaning with similar trends in employment before the change in minimum wage.

(b) Can you compute the "counterfactual" average FTE employment in NJ stores after April 1992 if the policy had not been implemented?

Solution: To compute the counterfactual, we can estimate a simple linear model of the FTE as a function of the state, of the time period, and of the interaction variable. And then we compute the counterfactual value of FTE as being the mean value of FTE in November, in New Jersey, without the interaction variable which measures the effect of the regulation.

See STATA do-file for replication code.

Now by comparing the distribution of the observed FTE and the counterfactual FTE, we see that, according to this prediction, there should have been a drop in employment in NJ, which follows the trend of the drop in Pennsylvania, with FTE in average of 18.27 in November 1992. Which is clear lower than the observed mean in November in New Jersey of 21.03 unit of FTE.

6. Lets perform the same analysis in a regression:

(a) Write the corresponding regression model that would give you the same DiD estimator of the effect of the policy change, and perform the regression. Show how the results change when you include additional controls for: location within State, chain ownership, type of chain (i.e. KFC, Wendys, Roys or BK).

Solution: Let NJ_s be a dummy for restaurants in New Jersey, and d_t be a time-dummythat swiches on for observations obtained in November (that is, after the minimum wage change). Then:

$$y_{ist} = \alpha + \gamma NJ_s + \lambda_t + \beta\left(NJ_s \times d_t\right) + \varepsilon_{ist}$$

$NJ_s \times d_t$ is the variable which identifies treated observations. This simple model includes two main effects for state and year and an interaction term that marks observation from New Jersey in November. γ captures possible differences between the treatment and control states prior to the policy change. λ captures the aggregate factors that would cause changes in FTE even in the absence of a policy change. The conditional mean function $\mathbb{E}\left[y_{ist} \mid s, t\right]$ takes on four possible values and there are four parameters. β is the parameter associated to the treatment effect:

$$\beta = \{\mathbb{E}[Y_{ist} \mid s = NJ, t = Nov] - \mathbb{E}[Y_{ist} \mid s = NJ, t = Feb]\}$$
$$- \{\mathbb{E}[Y_{ist} \mid s = PA, t = Nov] - \mathbb{E}[Y_{ist} \mid s = PA, t = Feb]\}$$

Using this approach, we obtain the following results:

	(1)
	fte
1 if NJ; 0 if Pa	-2.892
	(1.194)
t	-2.166
	(1.516)
state_post	2.754
	(1.688)
Constant	23.33
	(1.072)
Observations	794
R^2	0.007

Standard errors in parentheses

We thus recover the 2.75 in front of the $d \times NJ$ variable (coefficient for `state_post`), which measure the average increase of 2.75 of FTE in NJ relatively to in PA after the change in minimum wage. The standard error of the mean has changed compared than to the one obtained in the previous table, as we here account for state means, and time mean values, which thus change the variance of the unexplained component.

However, to obtain a consistent estimation of the effect of the change in minimum wage on employment, we need to account for other factors that may affect employments. We can control for other variables, such that the location within State, chain ownership, type of chain (i.e. KFC, Wendys, Roys or bk). We thus have the following regression:

$$y_{ist} = \alpha + \gamma NJ_s + \lambda_t + \beta (NJ_s \times d_t) + X_{is}^{'} \delta + \varepsilon_{ist}$$

where X_{is} are variables varying at the individual level of at the state level. By adding controls, we are thus comparing stores with similar types of chains and ownerships after the change in regulation across states. We thus obtain the following outcome:

	(1) fte	(2) fte	(3) fte
1 if NJ; 0 if Pa	-2.892	-2.377	-0.908
	(1.194)	(1.079)	(1.272)
t	-2.166	-2.224	-2.212
	(1.516)	(1.368)	(1.349)
state_post	2.754	2.845	2.815
	(1.688)	(1.523)	(1.502)
chain 2=kfc		-10.45	-10.06
		(0.849)	(0.845)
chain 3=roy		-1.625	-1.693
		(0.860)	(0.859)
chain 4=wendy		-1.064	-1.065
		(0.929)	(0.921)
1 if company owned		-1.169	-0.716
		(0.716)	(0.719)
1 if in southern NJ			-3.702
			(0.780)
1 if in central NJ			0.00788
			(0.897)
1 if in PA, ne suburbs of Phil			0.924
			(1.385)
Constant	23.33	25.95	25.32
	(1.072)	(1.038)	(1.211)
Observations	794	794	794
R^2	0.007	0.196	0.221

Standard errors in parentheses

Thus, the estimation is rather robust to the introduction of controls. This was quite predictable, since we observed similar characteristics of store between states.

(b) Compare the results with a standard OLS regression of employment on wages.

Solution: We here estimate the number of full-time employees on the starting wage by stores.

See STATA do file for replication.

We obtain the following results:

	(1) fte	(2) fte
starting_wage	1.633	2.068
	(0.942)	(0.849)
chain 2=kfc2		-10.83
		(0.844)
chain 3=roy		-1.756
		(0.858)
chain 4=wendy		-1.442
		(0.944)
1 if company owned		-1.182
		(0.715)
Constant	13.40	14.55
	(4.539)	(4.060)
Observations	760	760
R^2	0.004	0.212

Standard errors in parentheses

Where we include control variables in the second column. This estimations tells that a one unit increase in the starting wage, leads to an increase in full-time employment by 1.6 (or 2 depending on the specification), and this effect is significant. However, this strategy does not identify the proper effect of the policy change on employment, since we do not account for differences in starting wage accross states. If we introduce state dummies and time dummies, we still observe a positive effect of the starting wage on full-time employment.

7. Discuss the validity of the DiD strategy implemented. What checks can be done to test the validity of the strategy?

Solution: How convincing is this evidence against the standard labor-demand story? The key identifying assumption here is that employment trends would be the same in both states in the absence of treatment. Treatment induces a deviation from this common trend. Although the treatment and control states differ, this difference is captured by the state fixed effect, which plays the same role as the unobserved

individual effect in a fixed effect model. The common trends assumption can be investigated using data on multiple periods. Pennsylvania may not provide a good measure of counterfactual employment rates in New Jersey in the absence of a policy change, if these trends are different. Thus, using multiple periods and not just one period might help in investigating the robustness of the strategy. In an update of their work, in 2020, obtained data for restaurants in NJ and PA for several years. These data show the slight decline in employment from February to November in PA, and little change in NJ over the same period. However, the data reveal year-to-year variations which are more important for other years, and with variations that are quite opposite in the two states. This can give evidence that PA may not provide a very good conterfactual group for employment rates in NJ in the absence of a policy change.

We also might want to control for sources of omitted state-specific trends, for instance adult employment. However, this needs to be *good control*, that is, variables which do not respond to the minimum wage change.

We might also question the definition of Full-time employment use (partial employee counting hald full-time employee). We may also investigate the results on the structure of labor (hours worked, the fraction of full-time workers), and on other working benefits or bonuses (free-meal for employess for instance of the wage rises).