

Juanna Schrøter Joensen  
Department of Economics  
University of Chicago  
[jjoensen@uchicago.edu](mailto:jjoensen@uchicago.edu)

February 21, 2022

Monday **March 7**, 2022, by 9:30am

The assignment gives a maximum of 100 points which amounts to a maximum of 6% of the total exam score. All questions within each of the three problems are equally weighted in the evaluation. Remember to hand in (i) a write-up with answers to posed questions and (ii) a clearly commented do file that shows how you have solved each of the problems. Remember to present parameter estimates in tables or figures rather than as raw output. See the Syllabus and the course Canvas page for further instructions regarding how the solutions are to be structured. The key to a good score is to be clear, precise, and concise in your answers.

### Assignment 6

#### Problem 1 (Difference-in-Differences, DID) 72 points

In this problem you will implement part of the analysis from the paper “[Workers’ Compensation and Injury Duration: Evidence from a Natural Experiment](#)” by Meyer, Viscusi & Durbin (1995). This paper examines the effect of workers’ compensation on time out of work. It compares individuals injured *before* and *after* increases in the maximum weekly benefit amount. The increases examined in Kentucky and Michigan raised the benefit amount for “high-earnings” individuals by approximately 50%, while “low-earnings” individuals, who were unaffected by the benefit maximum, did not experience a change in their incentives. You naturally do not need to read the paper, but having a quick look at it may help you better understand the setting in which this policy change was implemented. Analyze that data `INJURY.dta` to answer the following questions. You will get full points if you use the data for Kentucky only for questions (a)-(g), but you can naturally also do these steps for Michigan if that helps you better understand what is going on in question (h). Note that you are encouraged to present all regression results in appropriately formatted tables such that you can easily compare results across specifications.

- (a) Write down the regression you will run to get the **Before-After** (BA) estimator for the *treatment* group. Run this regression. What do you conclude about the effect of increased workers’ compensation on time out of work based on this BA estimator? Clearly state the main identifying assumption that would need to be fulfilled in order for this to be a causal effect. Explain why (or why not) you find this identifying assumption credible in this setting.

- (b) Now also run the regression to calculate the BA estimator for the *control* group. Explain whether this raises doubts about the identifying assumption stated in (a). Explain how you can calculate the **Difference-in-Differences** (DID) estimator from the two regressions you ran in (a) and (b). Note that it may be helpful to use graphical illustrations to sharpen your explanation.
- (c) Write down the regression you will run to get the **Cross-Section** (CS) estimator of the difference in outcomes between the treatment and the control group in the period *after* benefits were raised for “high-earnings” individuals. Run this regression. What do you conclude about the effect of increased workers’ compensation on time out of work based on this CS estimator? Clearly state the main identifying assumption that would need to be fulfilled in order for this to be a causal effect. Explain why (or why not) you find this identifying assumption credible in this setting.
- (d) Now also run the regression to calculate the CS estimator in the period *before* benefits were raised for “high-earnings” individuals. Explain whether this raises doubts about the identifying assumption stated in (c). Explain how you can calculate the DID estimator from the two regressions you ran in (c) and (d). Note that it may be helpful to use graphical illustrations to sharpen your explanation.
- (e) Replicate the analysis in **Example 13.4** in Wooldridge (2016). What do you conclude about the effect of increased workers’ compensation on time out of work based on the DID estimator? Clearly state the main identifying assumption for this DID estimator to be interpreted as a causal effect. Is this the same DID estimator you calculated in (b) and (d)?
- (f) Now compare the BA estimate from (a), the CS estimate from (c), and the DID estimate from (e). Substantiate which one of these three estimators you find most credible in this setting. Remember: which estimator you find most credible should come down to which identifying assumption you find most credible.
- (g) Replicate the analysis in **Problem 13.7** in Wooldridge (2016) and answer each of the questions posed in (i) and (ii).
- (h) Answer each of the posed questions (i), (ii), and (iii) in **Computer Exercise C13.4** in Wooldridge (2016).
- (i) State what data you would collect and which regressions you would run in order to assess the credibility of the DID estimator in (e).

**Problem 2 (Repeated Cross-Section Data) 24 points**

What additional analysis can you do when adding a time dimension to your data? In this problem you will be analyzing how the wage return to education and the gender wage gap have changed over time. Use the pooled cross-section data set `CPS78_85.dta` to answer the following questions. Note that you are encouraged to present all regression results in appropriately formatted tables such that you can easily compare results across specifications.

- (a) Replicate the analysis in **Example 13.2** in Wooldridge (2016). That is, estimate (13.1) and compare your OLS estimates to those presented in (13.2). What was the wage return to education in 1978? How much has the wage return to education changed over this time-period? What was the gender wage gap in 1978? How much has the gender wage gap changed over this time-period?
- (b) Answer each of the posed questions (i)-(vii) in **Computer Exercise C13.2** in Wooldridge (2016).
- (c) (**Bonus question, 0 points** that you will have a deeper understanding of after Lecture 7) Explain what additional analysis you could have done if this were true panel data. Substantiate whether this would have made your analysis more credible.

**Problem 3 (Final Project EXTENSION) 4 points**

This problem helps you take the next step for the final project. Now that you are (hopefully!) well on your way to replicate the main empirical results in the paper. In Problem 3 in Assignment 5 you demonstrated that you are getting familiar with the relevant theory and research question. Now you have to show us that you are ready to start thinking about how to extend this paper. Answer the following questions:

- (a) Do you think the empirical results of this paper are externally valid in the sense that the population model considered would be the same one in a different time-period or if sampling from a different population? Would it be possible for you to extend the data to cover a longer time-period or a different time-period? Would it be possible for you to get data from a different environment; e.g. a different state or a different country?
- (b) Do you think the main empirical specification is reasonable? Could you think about modifying it; e.g. to map more directly to the relevant economic theory?
- (c) Do you think the main variables are measuring the “real world” counterparts of the theoretical constructs they are supposed to measure? Could you propose alternative or better measures?

In Assignment 2, Problem 3, you stated the “ideal” randomized experiment to answer the causal question of interest in the paper you are replicating. Now assume that this “ideal” randomized experiment is not feasible in reality.

- (d) Propose a DID estimation strategy that would allow you to estimate the causal effect of  $X$  on  $Y$ .
- (e) Clearly state the identifying assumption(s) and discuss how you could collect data to assess the credibility of your DID estimate.