5304 Econometrics Autumn 2023

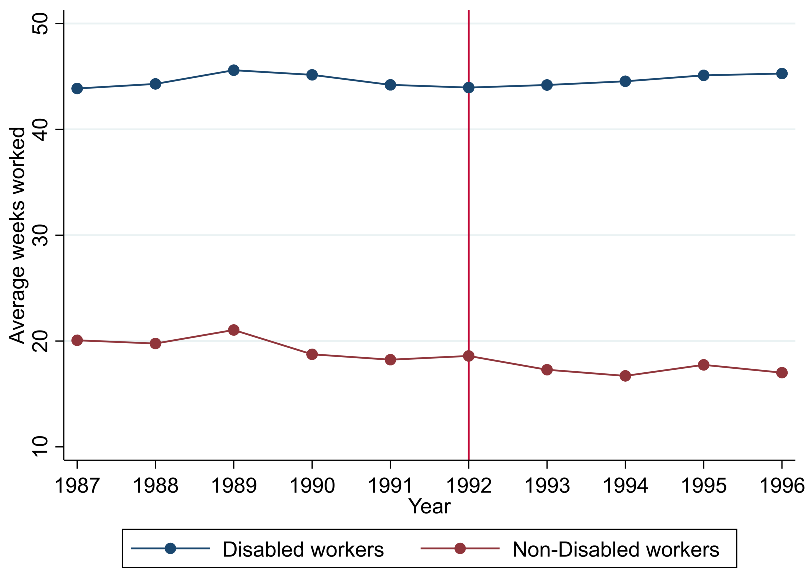
Problem set 5: Group 15[1]

Part 1a: Basic and dynamic differences-in-differences

In this assignment, we are using an excerpt from the data set used in Acemoglu and Angrist (2001). The paper investigates the labor market effects for disabled workers from the Americans with Disabilities Act that came into effect in 1992. The authors use a DiD design where non-disabled workers act as a control group for disabled workers affected by the new law. We focus on the sample of males aged 21–39.

1. First, you need to prepare the data set. Keep only workers with a business income equal to zero (incbus). Also, the variable year refers to the survey year, which asks questions about work related matters in the previous year. Generate a new year variable that captures the working year instead of the survey year. Also generate a dummy for being disabled. Hint: check value labels to figure out how the variables are coded.

2. It is good practice to begin by plotting the data. Show the time trends in weeks worked (wkswork1) for disabled and non-disabled workers and indicate with a vertical line when the ADA came into effect. Describe what you see and assess whether the parallel trends assumption seem to hold.



The parallel trends assumption states that the group of untreated and treated individuals should follow a similar trend in the counterfactual case that none of the individuals receive treatment. This is impossible to test since the treated group receive treatment, but an indication that the parallel trends assumption holds is that the group of treated and untreated follow a common trend before the treatment intervention (i.e. rule out pre-treatment trends). From the plot, the time trends in weeks worked seem to change similarly for the disabled and non-disabled up until the time of intervention (1992). Hence the parallel trends assumption seems to hold.

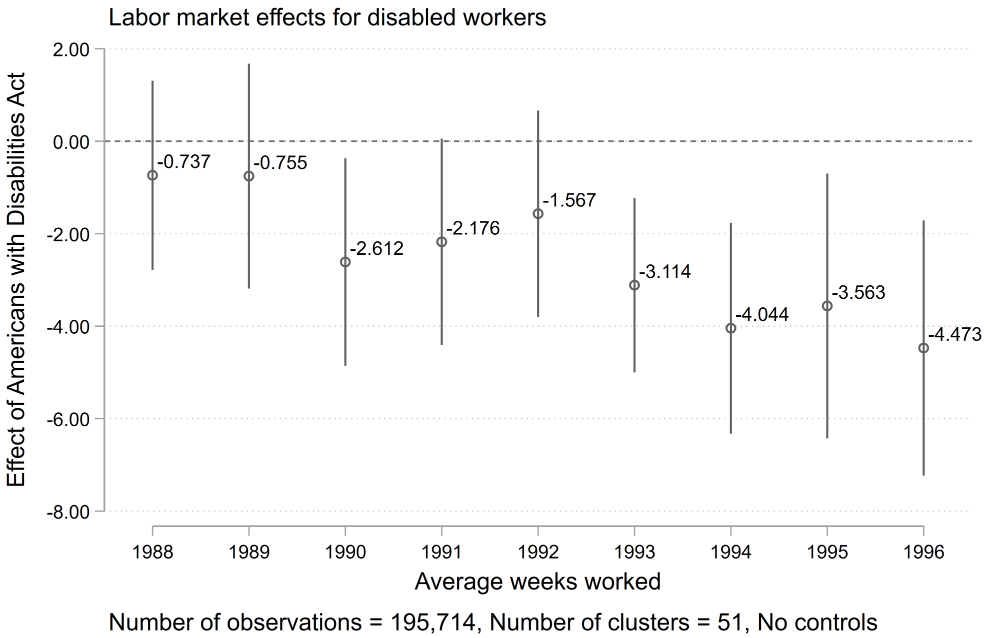
[1] 42624@student.hhs.se, 42613@student.hhs.se, 42632@student.hhs.se, 25164@student.hhs.se

After the treatment intervention, the non-disabled seem to decrease their average weeks worked slightly while we observe a (very slight) increase for the disabled. This difference between the groups could be an effect from the ADA, other be the result of a confluence of other factors. Without considering the unlikely general equilibrium effects of the impact of an increase of disabled workers increasing labour supply that may decrease the labour demand for non-disabled workers, we expect there to be no changes in the average worked weeks by the non-disabled; the ADA only outlaws discriminatory labour practices targeting the disabled. Thus, looking at the entire graph for the non-disabled, we can observe a gradually decreasing trend. We are unsure if this gradually decreasing trend is specific to the non-disabled group, or affects the disabled group as well.

3. Denote weeks worked of individual i at working year t as Y_{it}. Let D_i be a dummy for whether an individual is disabled or not. Let y_t be a dummy for year t (so that y_{1988} is a dummy that takes on the value 1 in year 1998, and 0 otherwise). Now, write out a specification of Y_{it} on being disabled, year dummies, and a set of interactions that allow the coefficients on the year dummies to vary depending on whether the individual is disabled or not.2 Let year 1987 be the omitted category.

$$Y_{it} = \beta_0 + \beta_1 \times \text{Disabled}_i + \sum_{\tau=1988}^{1996} \beta_\tau \times \text{y}_t + \sum_{\tau=1988}^{1996} \beta_\tau^D \times \text{y}_t \times \text{D}_i + \epsilon_{it}$$

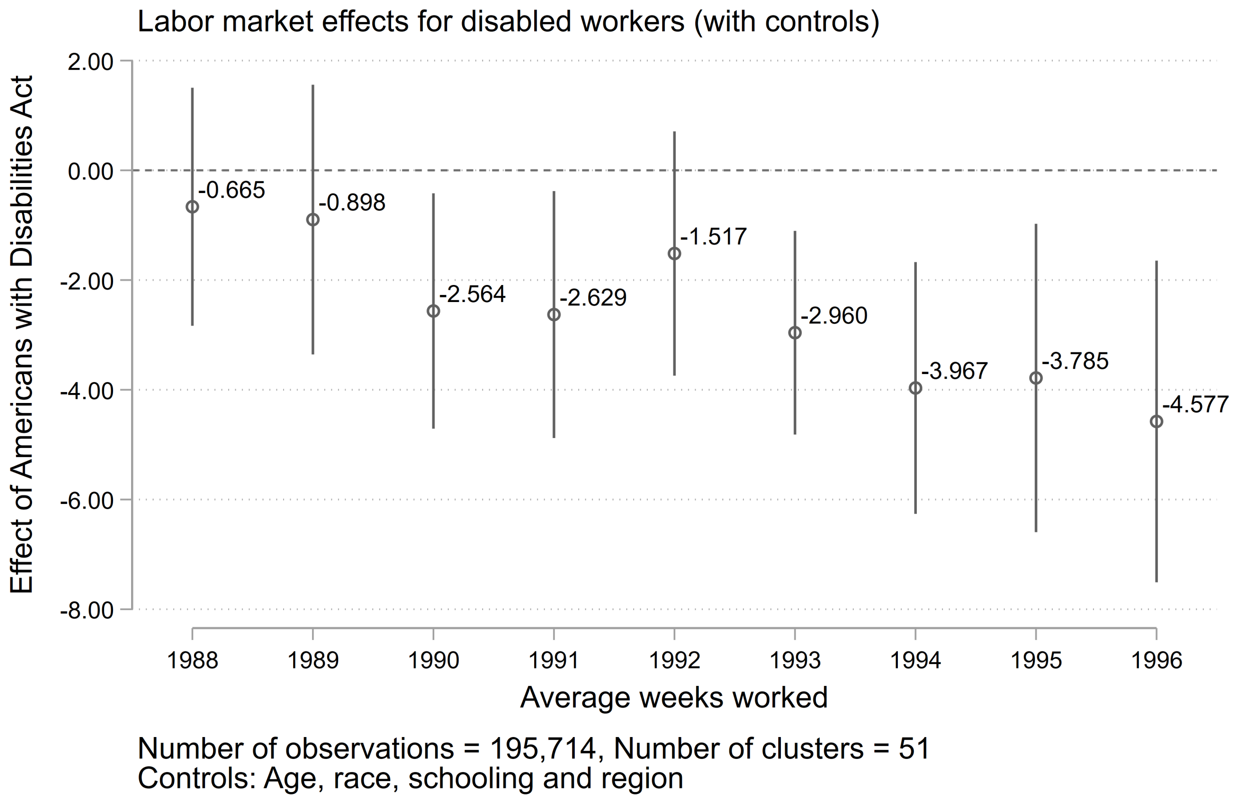
4. Estimate the dynamic DiD model that you wrote down. Use the function coefplot to show the coefficients on the interactions between year and disability in a figure. Does there seem to be an effect of ADA on weeks worked for disabled workers? Does the parallel trends assumption seem to hold?



The coefficients on the interactions between year and disability indicate the effect of being disabled on working hours for each year 1988-1996. From the plot, we observe smaller values of the estimated coefficients in the years 1993-1996 than 1988-1992, so working hours for disabled seems to have decreased after the ADA intervention in 1992 compared to before. However, the coefficients before the intervention are within the 95% confidence interval of the coefficients after the intervention.

In question 2, we said that the differences in time trends in weeks worked between disabled and non-disabled could be an effect from the ADA, while it could also be due to different trends for the different groups stemming from some other reason. This plot does not give us any more information than what we obtained from the plot in question 2, except that it might be easier to interpret numbers than colored lines, so we can't say anything more on whether the parallel trends assumption seems to hold from this.

5. Now add year×age, year×race, year×schooling and year×region dummies. Intuitively, what do these interactions control for? Output another coefficient plot. Do the controls change the results compared to the simple specification?



Intuitively, these interactions control for how the changes in composition of the workforce, i.e. the age, racial, educational and regional characteristics of the workforce, differ between the groups of disabled and non-disabled. By adding these controls, we account for different reasons that could lead to different trends in weeks worked for the two groups and can hence isolate the component of the difference in trends stemming from the ADA. The coefficient plot when controlling for year×age, year×race, year×schooling and year×region is very similar to the coefficient plot of the simple specification without the controls. This is an indication that the parallel trends assumption hold, i.e. in the counterfactual case without the ADA, disabled and non-disabled would have followed the same trend in how weeks worked changed over the time period, since possible differences in age, race, schooling and region between the groups does not seem to have affected differences in changes in weeks worked. This makes us more confident that the downward shift in weeks worked for disabled after the ADA is a causal effect from the ADA.

Part 2: Interpreting published results

In Kleven et al. (2019), the authors investigate the differential impact of having children on labor market outcomes for men and women (i.e. fathers and mothers) respectively. To do so, they run event-study regressions with the event of interest being defined as the year t in which the man/woman had its first child. They estimate the following baseline regression *separately* for both genders g:

$$Y_{ist}^{g} = \pi + \sum_{j \neq -1} \alpha_j^g \cdot \mathbf{1}[j = t] + \sum_{k} \beta_k^g \cdot \mathbf{1}[k = age_{is}] + \sum_{y} \gamma_y^g \cdot \mathbf{1}[y = s] + u_{ist}^g$$

where i refers to an individual, s refers to an actual year and t refers to the event-time, i.e. the number of years *until* or after the individual's first child-birth. Hence, the coefficients of interest are α_g^j for time periods $j \geq 0$, which measures the effect on labor market outcomes in time periods *after* child birth. The reference period is the year just before child birth, i.e. $t = -1$. The intercept is given by π.

1. The coefficients β_k^g are dummies for the age of an individual i at time s. Given that the authors are investigating the impact on labor market outcomes, such as earnings, for men and women of having their first child, why is it important to control for age?

Labor market outcomes are highly dependent on age, e.g. salary usually increases with age. At the same time, age affects the probability of having children. Therefore, if we are not controlling for age, $Cov(u_{ist}^g, t) \neq 0$ since $Cov(u_{ist}, age_{is}) \neq 0$ and $Cov(age_{is}, t) \neq 0$, which is a violation of the endogeneity assumption. We therefore control for age to ensure $Cov(u_{ist}^g, t) = 0$.

Given that the researchers seek to estimate the impact on labour market outcomes (e.g. earnings), age is an important control. It is a confounding variable that is positive correlated with earnings: it takes time for career progression to occur, and thereby substantially increase an individual's income. Moreover, age determines both the incentive and ability for men and women to have children. From a practical standpoint, without the assistance of IV therapy, a woman's fertility is limited beyond a certain age (e.g. menopause). By contrast, men may either desire to start families earlier or later depending on their personal inclinations (and the age of their spouses / partner).

Moreover, the impact of having a child on wages may differ based on age. Since men or women or a higher age will likely be in more senior positions, they are more able to balance both the responsibilities of their career and care giving. Problematically, such perceptions by employers influence the decision of women and men of the timing to have a child. This is because the rational decision to have a child will be based on the impact on the parent's and household's projected lifetime earnings.

Thus, when we understand controls as an artificial attempt at matching, the problem is that we are comparing individuals of different ages with different expected wage growth.

2. Let's assume that childbirth is correlated with the business cycle in such a way that when the economy is doing poorly, more women decide to have their first child. Further, assume that the timing of the first child for men is uncorrelated with the business cycle. Would this be a problem for the specification given above? Explain why or why not.

If we understand the business cycle to be negatively correlated with a woman's incentive to have a child, but not for a man then we would need to account for it in our approach.

Using Figure 1, naively carrying out the event study regression without accounting for the correlation with the business cycle means that female earnings after having a child are negatively biased.

Since the event period starts from the observation choosing to have a child, then (compared to men) women are more likely to have a child in a recession such that their lifetime earnings would be skewed downwards not just due to having a child, but because a recession depresses wage growth for several years. Thus, the specification must include the business cycle as a control if we want to make comparisons between men and women.

On a side note, such a possibility where women timing is correlated but men timing is uncorrelated with the business cycle is unrealistic. What affects their spouse's timing will also affect their timing so it should be correlated with the both of them.

Figure 1: Main results

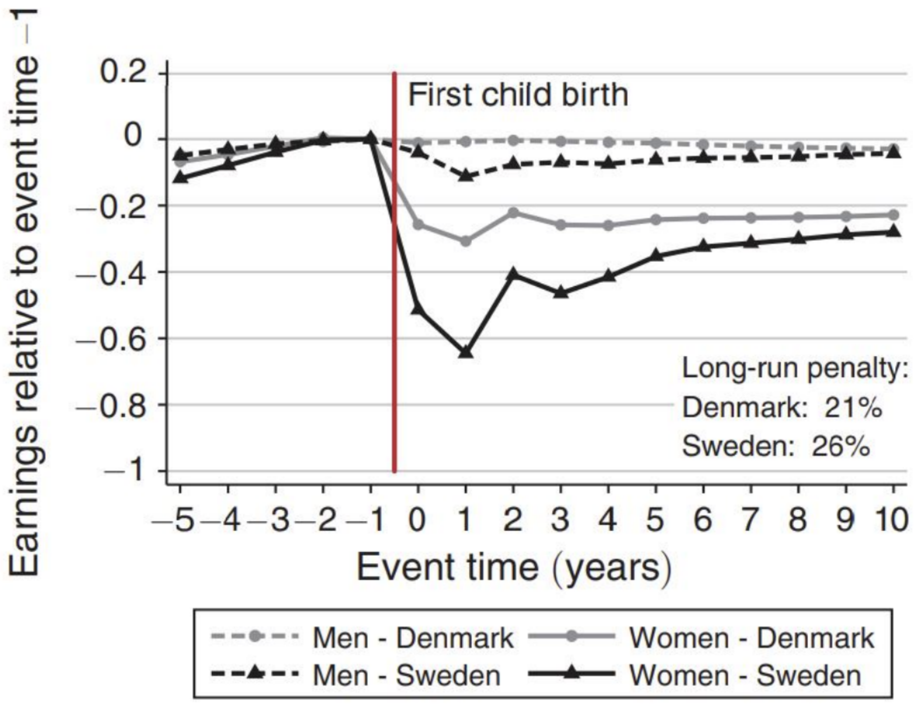


FIGURE 1. CHILD PENALTIES IN EARNINGS IN SCANDINAVIAN COUNTRIES

References

Acemoglu, D. and J. Angrist (2001): "Consequences of Employment Protection? The Case of the Americans with Disabilities Act," *The Journal of political economy*, 109, 915–957.

Kleven, H., C. Landais, J. Posch, A. Steinhauer, and J. Zweimüller (2019): "Child Penalties across Countries: Evidence and Explanations," *AEA papers and proceedings*, 109, 122–126.