Applied Microeconometrics **Fall 2022**
Manuel Bagues

Problem set 3
(suggested answers)

A. Impact of minimum wages on teen employment

Prior to starting this exercise, please make sure you have installed the csdid Stata package by running the following command: ssc install csdid

If you are using R, install the did package with packages.install("did").

The exercises are conducted using `mpdta.dta`. The dataset consists of (logs of) teen employment in counties in the United States in 2003-2007. The dataset contains the following variables:

- year: Year of observation, 2003-2007.
- countyreal: County identifier.
- lpop: Log of the county population in 1000s.
- lemp: Log of county teen employment.
- first_treat: Year when the state where the county is located first raised it's minimum wage. 0 for states where minimum wage equals the federal minimum wage for the entire period.
- treat: Indicator for whether the county is treated.

1. Estimate the effect of increasing the minimum wage on teen employment using a standard two-way fixed-effects (TWFE) specification. Under which assumptions would this estimator provide a consistent estimate of the Average Effect of the Treatment on the Treated (ATT)? Discuss whether in this particular context (i.e. effect of minimum wage) we should expect these assumptions to hold.

   Note: please cluster errors at the county level, we will discuss later on whether this is the relevant level of clustering.

   Answer: The TWFE estimate is equal to -0.037, with st. error=0.013.
   Key to the validity of difference-in-differences is the parallel trends assumption, that is, the assumption that there are no time-varying differences between the treatment and control groups (i.e. the control group provides a

good counterfactual for how employment would have evolved in the treatment group in the absence of the treatment). Implicitly, this requires that there are no asymmetric time-variant shocks, no anticipation effects and that the SUTVA is satisfied. Moreover, since there more than two periods and/or more than two groups, the two-way fixed-effects specification also requires that there is no heterogeneity in the treatment across groups or across time. In this particular context, this assumption is unlikely to hold. There might be dynamic effects: the impact of minimum wages is likely to be stronger in the longer term due to the existence of short-term rigidities. Similarly, it is also possible that the impact of minimum wages is stronger in states where salaries are relatively lower (e.g. a 10$ minimum wage is likely to be more binding in Mississippi than in San Francisco).

2. Let us now apply the method proposed by Callaway and Sant'Anna (2021), using `mpdta.dta` and the `csdid` package.

   (a) Controlling for county population, estimate the group-time average treatment effects, using never-treated units as the comparison group. Plot the estimates for 2004 and 2006 group. Estimate the average treatment effect on the treated (ATT).

   Answer: We estimate an ATT of ≈ -0.42 (st. error=0.12). This suggests that raising minimum wage had a negative effect on employment.

   (b) Controlling for county population, estimate the group-time average treatment effects, using not-yet-treated units as the comparison group. Plot the estimates for 2004 and 2006 group. Estimate the average treatment effect on the treated (ATT).

   Answer:

   The ATT is very similar, ≈ -0.41 (st. error=0.11), confirming that raising minimum wage had a negative effect on employment.

   (c) Perform a test for whether all pre-treatment group-time average treatment effects were statistically equal to zero. Interpret the result.

   Answer: Results from calculating the χ^2-statistic do not reject the hypothesis that all pre-treatment effects were equal to zero. This lends support to the parallel trends assumption, and thus lends support for the validity of the estimation strategy.

   (d) Let us now discuss clustering. What would you say is the level at which standard errors should be clustered here? (i.e. county or state) Discuss

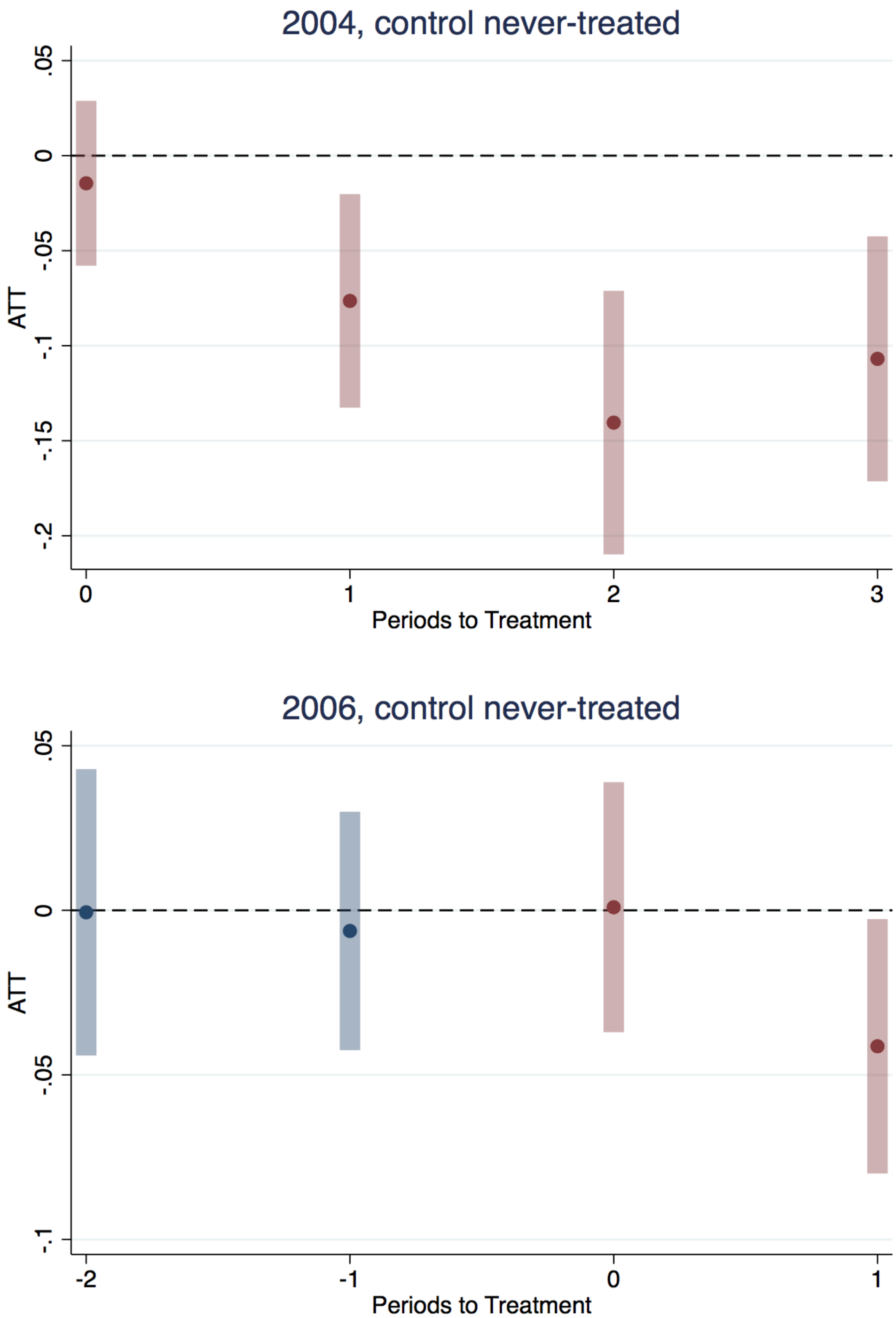

Figure 1: Group-time effects, never-treated as the comparison group.

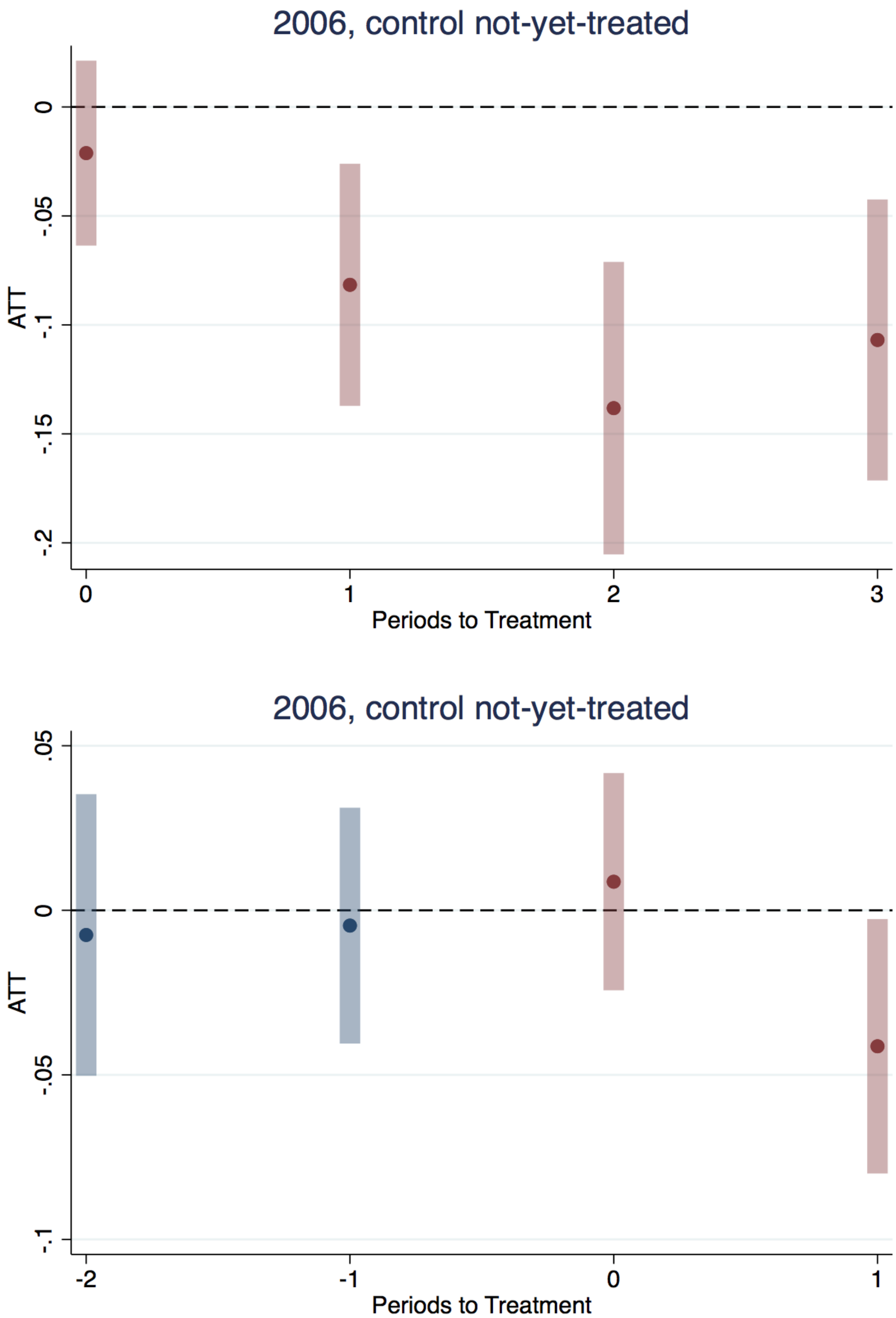

Figure 2: Group-time effects, not-yet-treated as the comparison group.

why.

The sample includes information from 25 states (although this variable is not in the dataset). Given this sample size, indicate how you would calculate the clustered standard errors and explain why.

Answer: Usually it is appropriate to cluster at the level at which the treatment varies, in this case at the level of state. Clustering at the state level would allow to account for common shocks within states. Given that there are only 25 states in the sample, the asymptotic properties of the standard estimators are unlikely to hold. Instead, in a small sample context it would be more convenient to use randomization inference or block bootstrap.

B. The impact of gender quotas in electoral lists

To address the scarcity of women in politics, in recent years many countries have adopted gender quotas in candidate lists requiring the presence of a minimum share of female candidates. By construction, these quotas increase the share of female candidates. However, it remains an empirical question whether they also lead to an increase in the share of women getting elected or reaching top political positions. Let us focus in the case of local elections in Spain. Within a proportional representation electoral system with closed lists, a quota requiring the presence of at least 40% of candidates of each gender on the ballot was implemented in 2007 in municipalities with more than 5,000 inhabitants.

1. One could use a difference-in-differences (DID) strategy to analyze the impact of these quotas on the probability that a woman is elected as mayor, using as treatment group municipalities with population between 5,000 and 10,000 inhabitants, and as control group smaller municipalities, and data for the 2003 and the 2007 local elections (as in Casas-Arce and Saiz (2015)). Explain briefly the potential threats to the validity of the DID strategy in this context and which robustness tests you might want conduct.

   Answer: The DID strategy relies on the assumption that the evolution of the control group provides information on how the treatment group would have evolved in the absence of the quotas. In this context this assumption may not hold for a number of reasons. First, it is possible that small municipalities do not provide a valid counterfactual for the evolution of the outcome variable in large municipalities. For instance, the timing of political and socio-economic shocks might differ depending on municipality size. Checking the evolution of the two groups in the past might provide supportive evidence for this assumption. It might also be useful to conduct some placebo tests. Restricting the analysis to untreated municipalities (e.g. below 5,000 inhabitants), we can run placebo DIDs assuming that quotas were implemented in municipalities with more than 1,000 inhabitants, 1100, 1200 etc. Second, another potential problem is the possibility that the treatment affects also the control group. For instance, the existence of quotas in large municipalities may create role models for women in smaller ones. Finally, it would be a problem if, during this period, there were other policies that had been implemented using the same population threshold.

2. Let us consider now the possibility of using a regression discontinuity design (RDD) to estimate how quotas affect the probability that a woman becomes mayor Bagues and Campa (2021). Explain briefly what are the key requirements that would make the RDD strategy adequate in this particular context.

   Answer: The crucial assumption for the validity of the regression discontinuity design is that there are no discrete changes in any relevant variable at the threshold, other than the treatment (i.e. quotas). A possible threat would be that some municipalities may be able to manipulate their population count. As in DID, other potential threats are the existence of other policies based on this threshold or possible spillovers. Finally, RDD requires a sufficiently large mass of observations around the threshold in order to provide estimates that are precise.

3. Let us now compare the RDD and the DID estimates. Discuss the potential advantages and disadvantages of these two approaches in this particular context, indicating potential differences in terms of (i) consistency, (ii) precision (iii) the locality of the estimates (i.e. ATT vs LATE).

   Answer: Since DID considers typically a larger set of municipalities, it provides a more precise estimate, however the assumptions required for consistency (e.g. parallel trends) tend to be more demanding than in the case of RDD (e.g. no precise manipulation around the threshold).

Another difference is that RDD and DID estimate the impact of the treatment for different populations. While RDD captures the impact for municipalities around the population threshold, DID estimate the average treatment on the treated for a larger set of municipalities (the exact set depends on which municipalities are being considered in the study, in this case municipalities with 5,000 to 10,000 inhabitants). Depending on the context, we might be more interested in the RDD or the DID estimate. For instance, in this particular context, it might be argued that the policy maker used the 5,000 threshold precisely because they believed that gender quotas would not be effective below this threshold, perhaps because in small municipalities the support for gender equality was lower, and the supply of qualified women might be limited. In this respect, a non-significant RDD estimate might be difficult to interpret. Instead, using DID we are learning about the impact of the policy in larger municipalities, which might be a more relevant population.

4. Bonus question:
   It is possible to estimate an RDD that yields identical results as the above DID. Indicate what is the RDD estimation that would be equivalent to the above DID in terms of the (i) bandwidth (i.e. how large?), (ii) kernel (i.e. triangular vs uniform) and (iii) order of the polynomial controlling for the running variable (i.e. 0, 1, 2...)

   Answer: We would obtain identical results if we were estimating a RDD specification where the outcome variable was in differences, bandwidth of 5,000, uniform kernel and polynomial of order 0 at each side of the threshold. In other words, we may thing about the DID as a particular case of RDD where we set an arbitrary bandwidth and we do not take explicitly into account the role of the running variable.

References

Bagues, M. and P. Campa (2021). Can gender quotas in candidate lists empower women? evidence from a regression discontinuity design. *Journal of Public Economics 194*, 104315.

Callaway, B. and P. H. Sant'Anna (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics 225*(2), 200–230.

Casas-Arce, P. and A. Saiz (2015). Women and power: unpopular, unwilling, or held back? *Journal of political Economy 123*(3), 641–669.