Applied Microeconometrics II
Assignment 2 Solutions

Please show your work. For Stata questions, **include the log files of your output.**

(1) (5X6=30 points) Chetty, Raj, Adam Looney and Kory Kroft. (2009). Salience and Taxation: Theory and Evidence, *American Economic Review*, 99(4): 1145-1177.
In Europe, supermarket prices include taxes, whereas in the U.S. price tags on shelves are pre-tax, and the full amount is only "revealed" at the cash register. This paper uses a field experiment to evaluate whether consumers underreact to taxes that are not salient.

(a) Follow the code (and data) posted. Create a table that maps the variables in equation (4) on page 1155 to the names of the variables in the Stata regression used to produce panel C "Third differences" of Table 3. You may need to look in both do files.

TT	treat_time
TS	treat_store
TC	treat_prod
TT*TC	i1=treat_products*treat_time
TT*TS	i3=treat_store*treat_time
TS*TC	i2=treat_products*treat_store
TT*TC*TS	TREATMENT=treat_products*treat_store*treat_time

(b) What is the key identification assumption for the DD estimate?

The common trends assumption implies that the sale of treatment and control products would have evolved similarly absent the experimental intervention.

(c) What is the key identification strategy for the DDD estimate?

There was no shock during the experimental intervention that would differentially affect ONLY the sales of the treatment products in the treatment store.

(d) What is the construct validity concern the authors address in the paper by using a longer observational study?
They are concerned about the presence of Hawthorne effects: namely, that just the presence of additional info on the price tags

(e) What is the level of clustering? How many clusters are there? Do you think this is the correct level? Check out footnote 10.
The authors are clustering at the week level. It appears there are just three clusters. They say they obtain similar standard errors when clustering at the product category level, which has more observations. There is no correct level of clustering, but clustering with a very small number of clusters is bound to

lead to more problems than clustering solves. The fact that standard errors don't change much when clustering at the product category level is to some extent reassuring. Also see footnote 1.

(f) The authors conduct two placebo tests.[1] The first addresses the common trends assumption. What statistical regression evidence do they report to support the assumption?

The coefficients on the before and after dummies are not statistically significant. A statistically significant coefficient would have been cause for concern, indicating a failure of the paralell trends assumption.

(2) (5X8=40 points) Autor, David H. (2003). "Outsourcing at Will: The Contribution of Unjust Dismissal Doctrine to the Growth of Employment Outsourcing." *Journal of Labor Economics,* 21(1): 1-42.

The article examines the effect of three classes of common law exceptions to the employment-at-will doctrine on the use of temporary help services (THS): breach of an implied contractual right to continued employment ("implied contract"), terminations contrary to public policy ("public policy exception") and violations of an implied covenant of good faith and fair dealing ("good faith exceptions").

(a) Reproduce the analysis in column 1, Table 5. Note the main independent variable is **mico**, showing the implied contract exception. Note the regression also controls for any public policy exception, **mppa**, or good faith exception (**mgfa**). Interpret the coefficient on the implied contract exception in column 1, Table 5.

See the attached log file. The coefficient suggests the passage of the implied contract exception increased temporary contract employment by 14.8 per cent in treated states relative to control states. The increase is more exactly 14.8 log points, but since it is a relatively small log point increase, we can approximate it as a percentage change.

(b) Explain why the difference-in-differences coefficient on the implied contract exception in column 1, Table 5, may be biased.

The passage of implied contract exception may not be a random policy decision, therefore it may be triggered by previous trends in employment and other

[1] The second placebo tests they conduct evaluates how likely a treatment effect of the same size as they found is, in relation to the distribution of all the possible treatment effects. The fact that very few treatment effects were observed to be lower than the actual estimated treatment effect is a sign that such a value is unlikely to be observed simply by chance- this is an analogous interpretation to the p-value in regular t-tests. They likely had to implement this test given the small number of clusters and unclear level at which clustering should occur. The code they use for the placebo permutations test is available on the AER website if you're interested.

economic factors. If such is the case, the parallel trends assumption that ensures difference in difference estimates provide causal effects would be violated, and the coefficient would be biased.

(c) How do the various specifications the author tries in Table 5 address the potential bias you identified in question (b)?

To address the failure of the parallel trend assumption, the author includes state time trends, which would capture a differential trend in the treatment states compared to the control states. Since there is no reason to assume such trends are purely linear, we include quadratic time trends. Since there might be trends at the regional level, the author also includes region-year indicators. Finally, some specifications also include time-varying covariates that may specifically bias our results, such as demographic and education trends (Black, female, young, college educated) that could influence employment in the temporary help services industry.

(d) How does the analysis in Table 7 address the risk of reverse causality?

We are specifically checking to see if there is any evidence of "treatment effects" in the years prior to the actual implementation of the treatment. If we observe that the treatment coefficients in years prior to the actual treatment implementation are statistically significant, we would have a reverse causality problem: the significance of the treatment indicators in the years prior to the treatment would represent an indication that pre-existing trends are biasing our results.

(e) Reproduce the analysis in Table 7. Interpret the "Law change t+2", "Law change t0", and "Law change t-4 forward" coefficients.

To reproduce results, you will need to create:
– a continuous time variable in order to create a time trend and a quadratic time trend.
– state time trends and region-by-year indicator variables (check out the Stata **xi** command and help file for an easy way to create these dummy variables).
– the leads and lags variables are already in the dataset provided to you: for the implied contract exception, the **admico** variables; for public policy exceptions, **admppa** variables; for good faith exception, **admgfa** variables.
– cluster standard errors at the state level

See attached log file.

(f) Run a simple difference in differences model using only the **mico** variable (the implied contract exception), continuing to cluster errors at the state level. Report the coefficient on **mico**. Use the **twowayfeweights** command (in Stata, run **ssc install twowayfeweights** first), with the **feTR** option, to identify whether the estimate contains any negative weights. What do you find? What does the pattern of weights suggest about the treatment-control comparisons you are aggregating? Is your estimate on mico an average treatment on the treated (ATT) effect?

You will notice quite a few of the weights are negative. We are in a situation where our aggregate coefficient on mico is obtained through a series of comparisons that include comparisons between already treated states to newly treated states. This implies that the coefficient on mico does not capture an average ATT across the policies we are analyzing.

(g) Use the **csdid** command and the **estat all** postestimation option to estimate the ATT only for the **mico** policy using the methods in Callaway and Sant'Anna (2020). Use the variable **max_first_treat** in the dataset for the **gvar** option of the csdid command. What is the ATT they report? Does their method suggest we were overestimating or underestimating the ATT with the specification in part f)

They find a lower ATT, suggesting our initial estimate of the ATT was upwardly biased.

(h) Run the **eventstudyinteract** command (Abraham and Sun, 2020) to estimate the specification in Table 7 column 1, only for the **mico** treatment (implied contract exception policy). Compare to your estimate using the specification in Table 7 column 1, only for the implied exception policy. Use the **treated_first** variable for the cohort option in the **eventstudyinteract** command, and the **nevertreated** variable as the **control_cohort** option. What do the **eventstudyinteract** results suggest about the persistence of the treatment effect?

The coefficients on the later lags of the policy, the three period lag and t+ periods lag, are much smaller than initially estimated. This suggests the policy effects were not persistent, and that the initial estimates in Table 7 were upwardly biased - a finding similar to the finding in question g).

(3) (5X6=30 points) Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. (2010). "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program." *Journal of the American Statistical Association,* 105(490): 493-505.

Proposition 99 was a Californian anti-tobacco policy initiative which resulted in an increase in the cigarette excise tax of 25 cents per pack. Tax revenues were initially earmarked to health and anti-smoking education budgets.

(a) Use the *smoking.dta* file and declare the dataset as a panel using the *tsset* command. Why do the authors only use panel data until the year 2000? Why does the panel not include all 50 states?

    After the year 2000, all states started implementing anti-tobacco measures. This means there are no more states to serve as a "no policy" control. Similarly, there are states that had similar anti-tobacco measures at or around the time California passed its Proposition 99.

(b) Looking at Table 1 and Figure 1, if all 38 states were used as controls in a difference-in-differences estimation of the effect of the passage of Proposition 99 in California, would we overestimate or underestimate the effect of Proposition 99?

    You can observe that California was registering a more pronounced decrease in cigarette sales compared to other states. As such, the decrease would have probably continued even in the absence of Proposition 99. As such, we would be overestimating the treatment effect.

(c) Implement the synthetic control method to find the weights for states which will generate the synthetic control for California. In order to do this, you will need to install the **synth** command and follow the syntax, as explained in the help file associated with the command. Adapt the syntax to follow the notes to Table 1 (e.g., all variables except lagged cigarette sales are averaged for the 1980 to 1988 period). Your weights may not be exactly the same as those in Table 2 because of differences in the maximization routines used to compute the weights, but they should be close.

    See attached log file.

(d) Using the **fig** option with the synth command and your synth command syntax in part (c) to produce a figure similar to Figure 2 in the paper.

    See attached log file.

(e) Using the weights reported in Table 2 in the paper and the data available to you in the *smoking.dta* dataset, calculate the treatment effect of Proposition 99 for the year 1995.

85.4*0.164+79.3*0.069+90.5*0.199+100.9*0.234+52*0.334-56.4= 22.0654, where we have first taken a weighted average of the cigarette sales in the synthetic control states, using their weights from Table 2, and then subtracted the Californian level of cigarette sales for 1995. You can observe this treatment effect in Figure 2.

(f) The synthetic control method does not provide standard errors for the treatment effect you calculated in part (e). Explain how the authors evaluate whether the treatment effect is obtained solely by chance.

The authors conduct placebo tests for every control state and then plot the resulting treatment effect graphs. Then they compare these graphs with the Californian graph. The authors' explanation: "In each iteration we reassign in our data the tobacco control intervention to one of the 38 control states, shifting California to the donor pool. That is, we proceed as if one of the states in the donor pool would have passed a large-scale tobacco control program in 1988, instead of California. We then compute the estimated effect associated with each placebo run. This iterative procedure provides us with a distribution of estimated gaps for the states where no intervention took place. Figure 4 displays the results for the placebo test. The gray lines represent the gap associated with each of the 38 runs of the test. That is, the gray lines show the difference in per capita cigarette sales between each state in the donor pool and its respective synthetic version. The superimposed black line denotes he gap estimated for California. As the figure makes apparent, the estimated gap for California during the 1989–2000 period is unusually large relative to the distribution of the gaps for the states in the donor pool."