Applied Microeconometrics Problem Set 4

Yixin Sun

December 12, 2020

Problem 1

Let

$$p_{gt} = P(D = 1|G = g, T = t)$$

For each t, g pair, we can write

$$\begin{aligned}
E[Y|T = t, G = g] &= E[Y(1)D + Y(0)(1 - D)|G = g, T = t] \\
&= p_{gt}E[Y(1)D + Y(0)(1 - D)|G = g, T = t, D = 1] \\
&\quad + (1 - p_{gt})E[Y(1)D + Y(0)(1 - D)|G = g, T = t, D = 0] \\
&= E[Y(0)|T = t, G = g] + p_{gt}\alpha
\end{aligned}$$

Taking the difference between two periods, we have

$$\begin{aligned}
\Delta_g &= E[Y \mid T = 2, G = g] - E[Y \mid T = 1, G = g] \\
&= E[Y(0)|T = 2, G = g] - E[Y(0)|T = 1, G = g] + \alpha(p_{g2} - p_{g1})
\end{aligned}$$

Now we can use the common trends assumption to write a difference in difference equation, where $E[Y(0)|T = 2, G = g] - E[Y(0)|T = 1, G = g]$ is the same between two groups, so the terms difference out, and we have

$$\Delta_2 - \Delta_1 = \alpha(p_{12} - p_{11} - p_{02} + p_{01})$$

Since we have that $p_{12} - p_{11} > p_{02} - p_{01}$, the right hand side will not be 0, $\Delta_2, \Delta_1, p_{12}, p_{11}, p_{02}, p_{01}$ are all observable in the data, so we have all the ingredients we need to point identify α.

Problem 2

(a) The common trends assumption is

$$E\left[Y_{i,t}(0) - Y_{i,1}(0) \mid E_i = e\right] = E\left[Y_{i,t}(0) - Y_{i,1}(0) \mid E_i = e'\right]$$
$$e' > t, \quad t \geq e$$

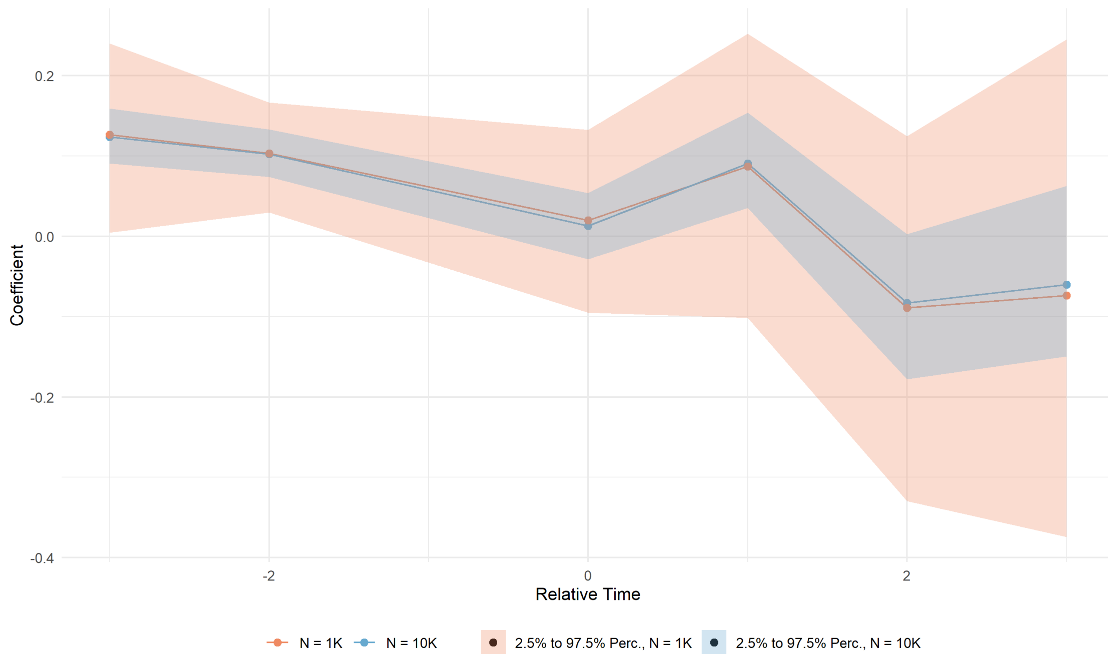
From the setup, we have

$$\begin{aligned}
E\left[Y_{i,t}(0) - Y_{i,1}(0) \mid E_i = e\right] &= -.2 + .5E_i + U_{it} - (-.2 + .5E_i + U_{i1}) \\
&= U_{it} - U_{i1} \\
&= \sum_{t'=0}^{t} \rho^{t'} \epsilon_{t-t'}
\end{aligned}$$

which is independent of E_i and therefore the common trends assumption holds.

(b) .

Figure 1: Monte Carlo, $\theta = -2$



(c) .

Figure 2: Monte Carlo, $\theta = 1$

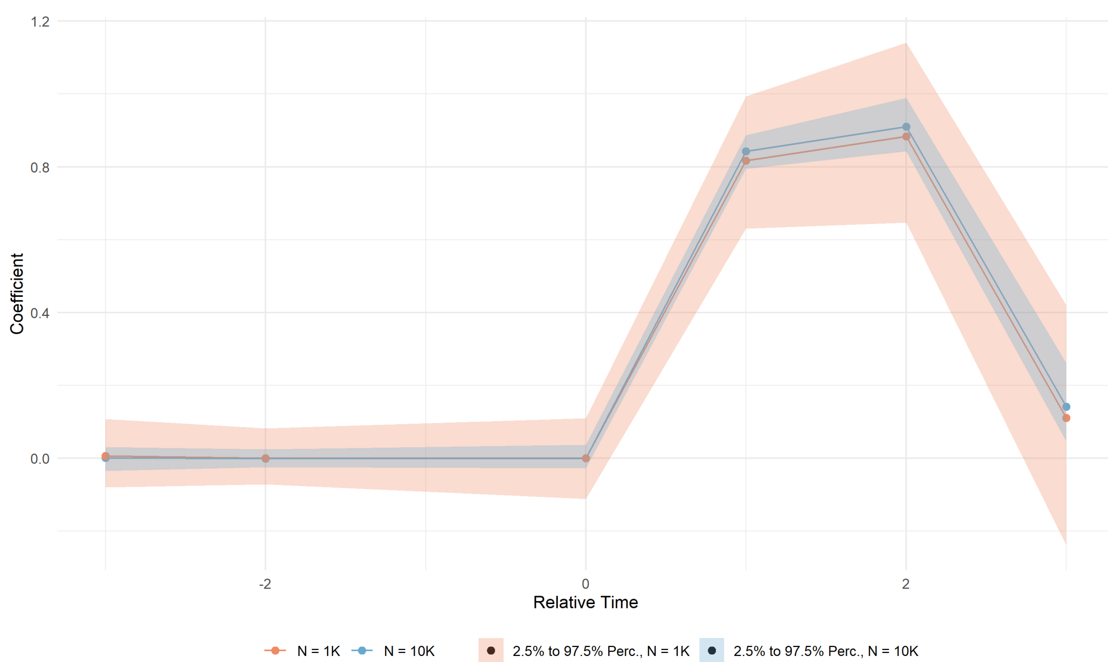
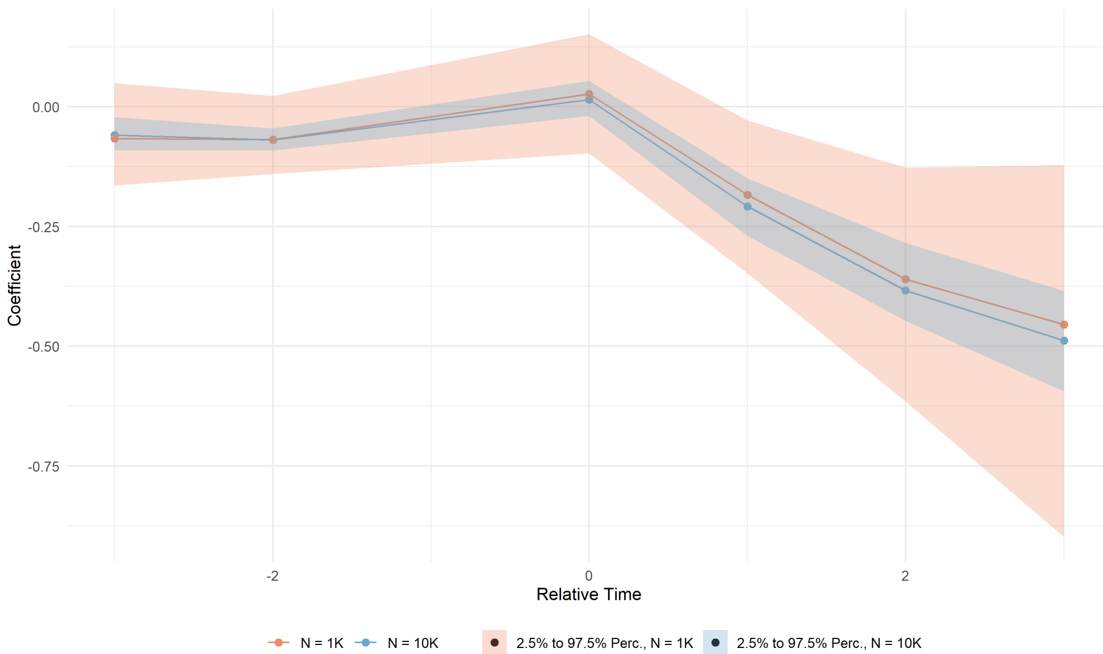


Figure 3: Monte Carlo, $\theta = 0$



(d)

$$
\begin{aligned}
ATE_3(2) &\equiv E\left[Y_{i3}(1) - Y_{i3}(0) \mid E_i = 2\right] \\
&= E\left[Y_{i3}(1) \mid E_i = 2\right] - E\left[Y_{i1}(0) \mid E_i = 2\right] + E\left[Y_{i1}(0) \mid E_i = 2\right] - E\left[Y_{i3}(0) \mid E_i = 2\right] \\
&= E\left[Y_{i3}(1) \mid E_i = 2\right] - E\left[Y_{i1}(0) \mid E_i = 2\right] + E\left[Y_{i3}(0) - Y_{i1}(0) \mid E_i = 2\right] \\
&= E\left[Y_{i3}(1) \mid E_i = 2\right] - E\left[Y_{i1}(0) \mid E_i = 2\right] + \underbrace{E\left[Y_{i3}(0) - Y_{i1}(0) \mid E_i = e'\right]}_{\text{common trends, } e' > t}
\end{aligned}
$$

All three components are observable in the data, so we can construct a consistent estimator by taking means from the sample. We see in the table below that this nonparametric estimator does yield results very close to the true $ATE_3(2)$ values. We also see that these values do not coincide with the coefficient on D^1_{it} in the previous graphs, demonstrating that we have to be careful with how we interpret the regression estimates.

Table 1: $ATE_3(2)$

θ	Estimate	True ATE
-2	0.677	0.657
0	0.130	0.141
1	0.852	0.841

Results from a Monte Carlo simulation with $N = 10,000$ and varying θ values.

(e) As we saw in class, the regression is pulling variation from everywhere, so the coefficient on D^1_{it} depends on every DID contrast from all periods and all cohorts. However, from Abraham and Sun (2020), we know that if parallel trends and treatment effect homogeneity (each cohort experiences the same path of treatment effects) hold, then the regression coefficient on D_{it}^1 represents the treatment effect in the first relative time period bin. This is the case here, and thus the coefficient of interest should coincide with the true value of the coefficient, which is the treatment effect relative to the treatment effects of the omitted relative time indicators, giving us that the null hypothesis should be $sin(1) - sin(-1) - sin(-4)$.

Below are the empirical size of a level .05 t-test of the coefficient on D_{it}^1. We see that when there is no serial autocorrelation in the errors, ($\rho = 0$), the different estimation methods all do fairly well, although the homoscedastic errors are a bit overconfident. When ρ gets large, clustering at the individual level accounts for the within-individual correlation, os we see that the cluster-robust does fairly well, while homoscedastic errors are way overconfident. The cluster-robust error does not do well when $N = 20$, which makes sense since for the cluster-robust method, there is essentially only 20 data points, while the homoescedastic and robust standard errors have $N \times t$ data points. The wild bootstrapping is way overconfident in all cases. This is likely due to coding error, but could possibly be due to the fact that the assumptions for wild bootstrapping rely on a large number of within-in cluster observations, while in our case, we only have five observations within.

Table 2: t-test

	ρ		
N	0	0.5	1
Homoscedastic			
20	0.045	0.026	0.006
50	0.026	0.022	0.008
200	0.028	0.038	0.014
Robust			
20	0.091	0.085	0.048
50	0.048	0.056	0.030
200	0.046	0.070	0.030
Cluster-Robust			
20	0.101	0.119	0.109
50	0.056	0.064	0.074
200	0.040	0.090	0.074
Wild Bootstrap			
20	0.006	0.000	0.012
50	0.000	0.000	0.000
200	0.000	0.000	0.000

Rejection probability for D_{it}^1 when null is the true value, $sin(1) - sin(-1) - sin(-4)$, using 1,000 Monte Carlo simulations.

Problem 3

Set Up:
In this problem, we are replicating a re-analysis the effect of Spanish terrorism on GDP. The economic costs of conflict are calculated by taking the difference between Basque Country's actual and forecasted outcome path over the post period. We have four ways of forecasting outcome path: matching, synthetic control, penalized synthetic control, and MASC.

Matching:

$$\hat{\gamma}_t^{\mathrm{ma}}(m) \equiv \boldsymbol{y}_{0t}' \hat{\boldsymbol{w}}^{\mathrm{ma}}(m)$$

where the weights $\hat{w}_i^{\mathrm{ma}}(m)$ are $1/m$ for the m units with smallest $\|x_i - x_1\|$ and 0 for all other units.

This can be written as the solution to the optimization problem

$$\hat{\boldsymbol{w}}^{\mathrm{ma}}(m) = \underset{\boldsymbol{w}\in\mathcal{S}}{\arg\min} \underbrace{\sum_{i\geq 2} w_i \|\boldsymbol{x}_1 - \boldsymbol{x}_i\|}_{\equiv \mathrm{Int}(\boldsymbol{w})} \quad \text{s.t.} \quad w_i \leq \frac{1}{m} \quad \text{for all } i \geq 2$$

Synthetic Control:

$$\hat{\gamma}_t^{\mathrm{sc}} \equiv \boldsymbol{y}_{0t}' \hat{\boldsymbol{w}}^{\mathrm{sc}} \quad \text{where} \quad \hat{\boldsymbol{w}}^{\mathrm{sc}} \equiv \underset{\boldsymbol{w}\in\mathcal{S}}{\arg\min} \left\|\boldsymbol{x}_1 - \boldsymbol{x}_0'\boldsymbol{w}\right\|^2$$

MASC:

$$\begin{aligned} \hat{\gamma}_t^{\mathrm{masc}} &\equiv \phi \hat{\gamma}_t^{\mathrm{ma}}(m) + (1-\phi)\hat{\gamma}_t^{\mathrm{sc}} \equiv \boldsymbol{y}_{0t}' \hat{\boldsymbol{w}}^{\mathrm{masc}} \\ \hat{\boldsymbol{w}}^{\mathrm{masc}} &\equiv \phi \hat{\boldsymbol{w}}^{\mathrm{ma}}(m) + (1-\phi)\hat{\boldsymbol{w}}^{\mathrm{sc}} \end{aligned}$$

Penalized Synthetic Control:

$$\begin{aligned} \hat{\gamma}_t^{\mathrm{pen}} &\equiv y_{0t}' \hat{w}^{\mathrm{pen}} \\ \text{with } \hat{\boldsymbol{w}}^{\mathrm{pen}} &\equiv \underset{w\in\mathcal{S}}{\arg\min} (1-\pi) \left\|\boldsymbol{x}_1 - \boldsymbol{x}_0'\boldsymbol{w}\right\|^2 + \pi \left(\sum_{i\geq 2} w_i \|\boldsymbol{x}_i - \boldsymbol{x}_1\|^2 \right) \end{aligned}$$

Cross Validation:
To find the optimal tuning parameters, m for matching, m and ϕ for MASC, and π for the penalized synthetic control, we use the paper's cross-validation procedure. The folds here are the data running between $\underline{t}_f = 1955$ and $\bar{t}_f \in \{1962, 1963, \ldots, 1968\}$. The cross validation procedure chooses, τ, the tuning parameters, to minimize

$$\mathrm{cv}(\boldsymbol{\tau}) \equiv \frac{1}{F} \sum_{f=1}^{F} \left(y_{1(\bar{t}_f+1)} - \hat{\gamma}_f(\boldsymbol{\tau}) \right)^2$$

Below are replications of Figure 10 and Figure 11 from Kellogg, Mogstad, Pouliot, and Torgovitsky (2020).

Figure 4: Pre-terrorism fit and post-period estimated annual costs of terrorism for MASC estimator

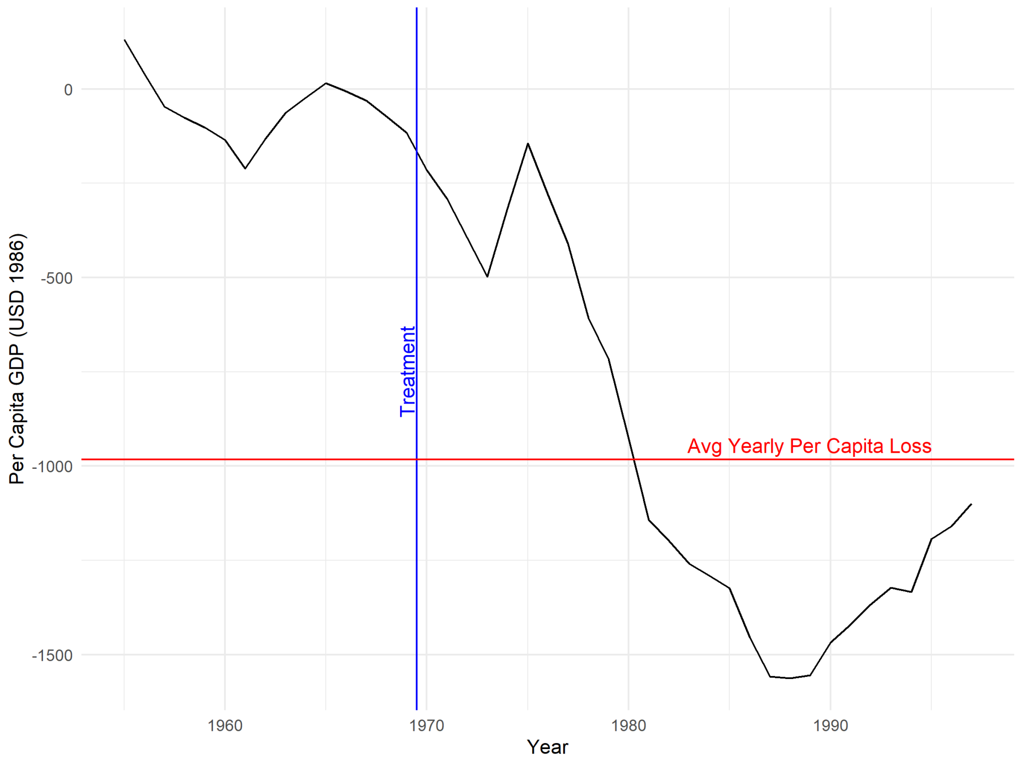


Figure 5: Difference from MASC in cost of terrorism produced by alternative estimators