Microeconometrics - Pset4

Jeanne Sorin*

December 12, 2020

Problem 1

The common trend assumption gives us that:

$$E[Y_i(0)|T_i = 2, G_i = g] - E[Y_i(0)|T_i = 1, G_i = g] \perp g = 1, 2$$

In addition, we suppose that :

- Some individuals in both groups are treated in both periods
- The increase between time periods 1 and 2 of the proportion treated in group 1 is larger than the same increase in group 0

We show that α is point identified.

Let $\tau_{i,t}$ the share of individuals in group i that are treated in period t (thus we have $\tau_{1,2} - \tau_{1,1} > \tau_{0,2} - \tau_{0,1}$ by assumption). Given the constant treatment effect assumption that $E[Y(1)|T = t, G = g] - E[Y(0)|T = t, G = g] = \alpha, \forall t, g$, we can write (potential outcome framework within a group)

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

This holds for both $T = 1, T = 2$, and $G = 1, 0$:

$$\begin{aligned} E[Y|T = 2, G = 1] &= E[Y(0)|T = 2, G = 1] + \alpha\tau_{1,2} \\ E[Y|T = 1, G = 1] &= E[Y(0)|T = 1, G = 1] + \alpha\tau_{1,1} \\ E[Y|T = 2, G = 0] &= E[Y(0)|T = 2, G = 0] + \alpha\tau_{0,2} \\ E[Y|T = 1, G = 0] &= E[Y(0)|T = 1, G = 0] + \alpha\tau_{0,1} \end{aligned}$$

Substracting them by pair

$$\begin{aligned} (E[Y|T = 2, G = 1] - E[Y|T = 1, G = 1]) - (E[Y|T = 2, G = 0] - E[Y|T = 1, G = 0]) = \\ \alpha \underbrace{[(\tau_{1,2} - \tau_{1,1}) - (\tau_{0,2} - \tau_{0,1})]}_{>0} \end{aligned}$$

We observe each element of the LHS, and all τ, so α is point identified.

*I am thankful to Nadia Lucas, Tanya Rajan and Sunny Yixin for helpful comments and discussions

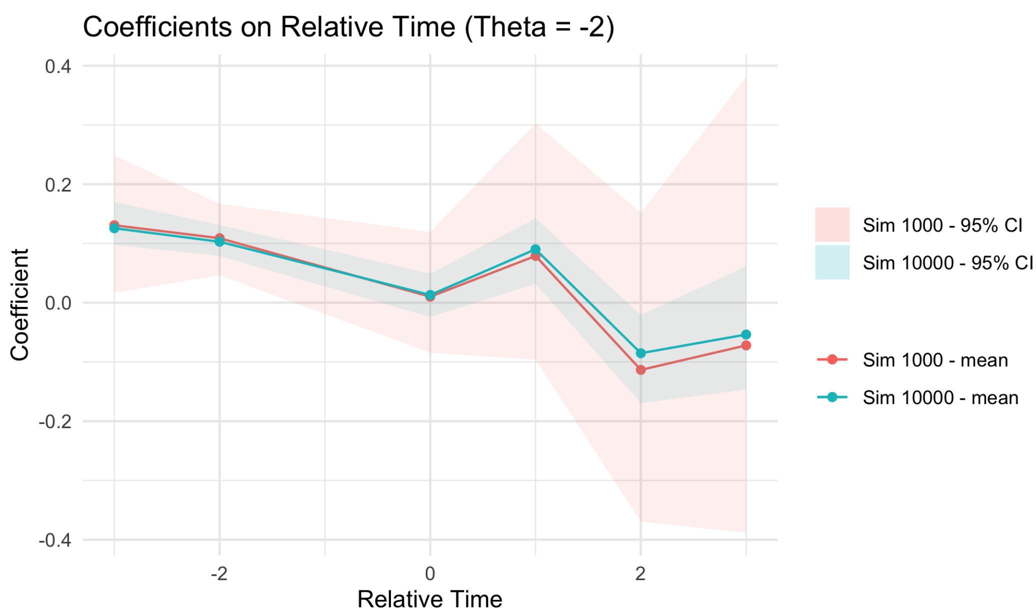
Problem 2

a) The common trend assumption holds here as $\Delta_t E[Y_i(0)], \forall t$ does not depend on the type E_i:

$$E[Y_i(0)|T_i = t] - E[Y_i(0)|T_i = t] = (\rho - 1)U_{i,t-1} + \epsilon_{it}$$

b) For $\theta = -2$

Figure 1: Coefficients on relative time, $\theta = -2$



c) For $\theta = 0$ and $\theta = 1$: As emphasized during the lectures, empirical practice on DiD is often synonym, for better or worse, of "eyeball" test. As shown in question (a), the common trend assumption (common pre-trends) is valid for this DGP. However, it is much easier to see this when $\theta = 0$ than when $\theta = 2$. This is important because this emphasizes we should do the math rather than always trust our eyeball.

d) Let's derive a consistent estimator of $ATE_3(2) \equiv E[Y_{i3}(1) - Y_{i3}(0)|E_i = 2]$.
From the common trend assumption the difference below is not a function of E_i

$$E[Y_i(0)|T_i = t] - E[Y_i(0)|T_i = t] = (\rho - 1)U_{i,t-1} + \epsilon_{it}$$

So, as we are interested in $E[Y_{i,3}|E_i = 2]$:

$$\begin{aligned} E[Y_{i3}(0)|E_i = 2] =& E[Y_{i3}(0) - Y_{i2}(0)|E_i = 2] + \\ & E[Y_{i2}(0)|E_i = 2] - \underbrace{E[Y_{i1(0)|E_i=2}] + E[Y_{i1}(0)|E_i = 2]}_{\text{observed because pre-treatment}} \end{aligned}$$

Notice that because of common trend

$$E[Y_{i2}(0) - Y_{i1}(0)|E_i = 2] = \underbrace{E[Y_{i2}(0) - Y_{i1}(0)|E_i \in \{1,2,3,4,5\}]}_{\text{observed for } E_i \in \{3,4,5\}}$$

$$E[Y_{i3}(0) - Y_{i2}(0)|E_i = 2] = \underbrace{E[Y_{i3}(0) - Y_{i2}(0)|E_i \in \{1,2,3,4,5\}]}_{\text{observed for } E_i \in \{4,5\}}$$

Figure 2: Coefficients on relative time, $\theta = 1$

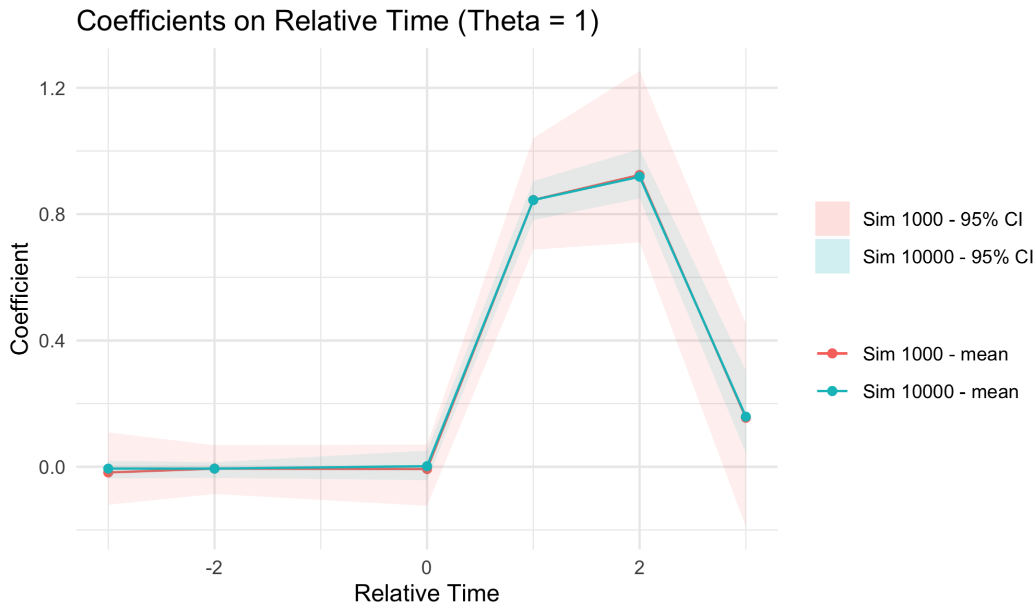
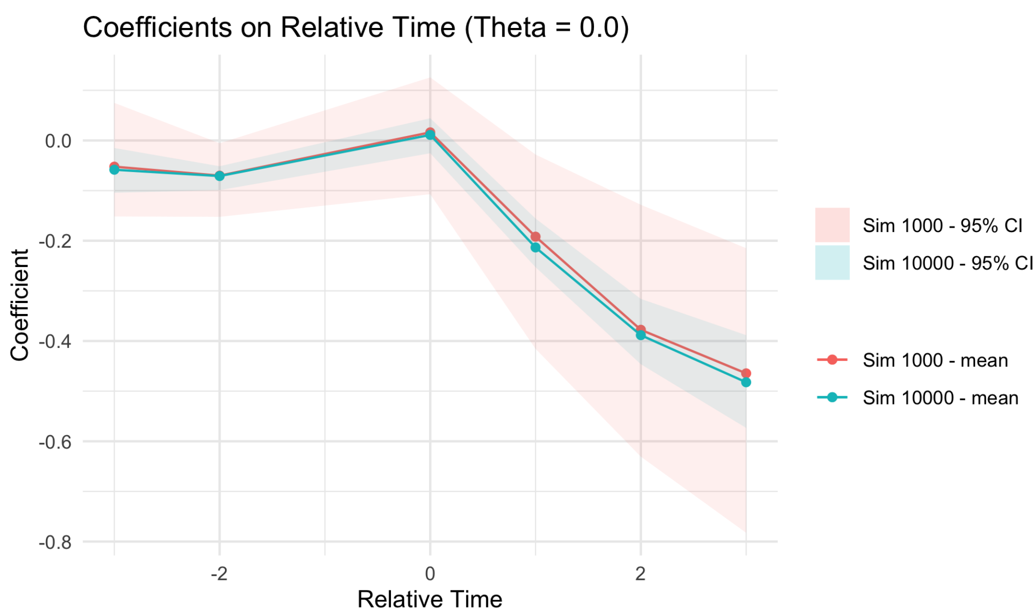


Figure 3: Coefficients on relative time, $\theta = 0$



Because the RHS is observed for $E_i \in \{3, 4, 5\}$, we use their empirical counterpart

$$\begin{aligned}
\hat{tr}_1 &= \frac{1}{M}\sum_{i=1}^{M}[Y\mathbf{1}T_i = 2, E_i \in \{3,4,5\} - Y\mathbf{1}T_i = 1, E_i \in \{3,4,5\})] \\
&\rightarrow E[Y_{i2}(0) - Y_{i1}(0)|E_i = 2] \\
\hat{tr}_2 &= \frac{1}{N}\sum_{j=1}^{N}[Y\mathbf{1}T_j = 2, E_j \in \{4,5\} - Y\mathbf{1}T_j = 1, E_j \in \{4,5\})] \\
&\rightarrow E[Y_{i3}(0) - Y_{i2}(0)|E_i = 2]
\end{aligned}$$

$$\hat{ATE}_3(2) = \frac{1}{K}\sum_{i=1}^{K}(Y\mathbf{1}[T_i = 3, E_i = 2] - Y\mathbf{1}[T_i = 1, E_i = 2]) + \hat{tr}_1 + \hat{tr}_2$$

With

$$M = \sum \mathbf{1}[E_i \in \{3, 4, 5\}]$$
$$N = \sum \mathbf{1}[E_i \in \{4, 5\}]$$
$$K = \sum \mathbf{1}[E_i = 2]$$

Table 1: ATE : simulated vs true for $t = 3$, $E = 2$

Simulation	True
0.627	0.657
0.137	0.141
0.836	0.841

Where the true parameters is $sin(3 - 2 \times \theta)$

e) I use $\theta = 1, n \in \{20, 50, 200\}, M = 200$ the number of Monte Carlo simulations. The "true" hypothesis we are testing against is $D^1_{it} = sin(1)$ (as the omitted dummies are pre-treatment, the ATE should be 0 in any case, so we don't need to deduct them from $sin(1)$).
The table below presents the results of the empirical test with level 0.05 on D^1_{it}.
I believe there might be coding mistakes, especially for the robust approach as the rejection rate is too low (by far). However, we can still see that the cluster and robust methods do not reject enough, which may be because we actually work with a very small number of observations for the clustered approach (N rather than $N * t$). Note, however, that the Wild bootstrap, which also clusters standard errors at the individual level, does much better than its non-wild counterpart. In general, the wild does better than any other method.

Table 2: t - test ($\theta = 1$)

Nb Obs	Homoskedastic	Cluster	Robust	Wild
$\rho = 0$				
20	0.030	0.051	0.010	0.066
50	0.015	0.005	0	0.075
200	0.020	0.005	0	0.040
$\rho = 0.5$				
20	0.010	0.015	0	0.060
50	0.040	0.030	0	0.060
200	0.040	0.005	0	0.100
$\rho = 1$				
20	0.051	0.035	0.005	0.066
50	0.035	0.015	0	0.055
200	0.030	0.010	0	0.055

Problem 3

Following Kellogg, Mogstad, Pouliot & Torgovitsky (2020), as well as the code on M. Kellogg's github, I replicate Figures 10 and 11 from the paper.

In a nutshell, in order to find the MASC estimator I

- Solve for the SC weights separately, using the gurobi solver, where the objective function is given by equation (7) in the paper

$$\hat{\omega}^{sc} \equiv \arg\min_{\omega \in S} ||x_1 - x_0'\omega||^2$$

- Solve for the matching weights separately, by iterating over the matching tuning parameter m, the number of matches.
- Do a cross-validation procedure in order to find ϕ, the weight given to the matching estimator such that

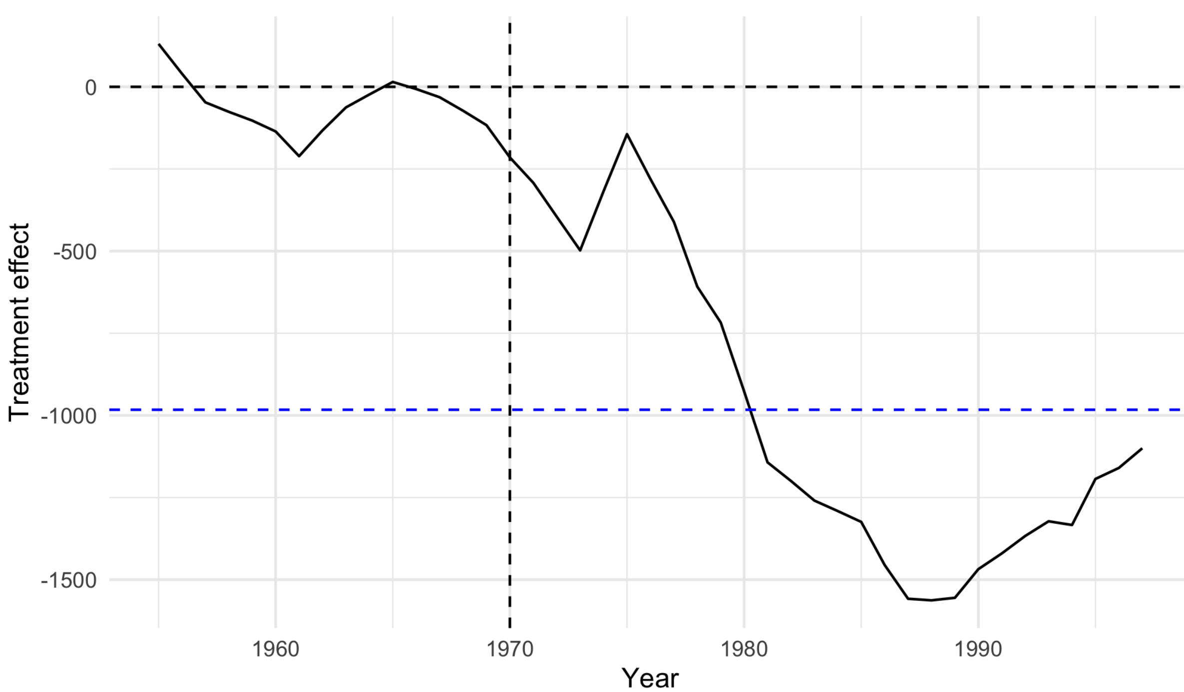
$$\hat{\gamma}_t^{masc} \equiv \phi\hat{\gamma}_t^{ma}(m) + (1-\phi)\hat{\gamma}_t^{sc} = y_{ot}'\hat{\omega}^{masc}$$

  As described in section 3.2, $\hat{\phi}(m)$ as a function of m is given by

$$\hat{\phi}(m) \equiv \frac{\sum_{f-1}^{F}(\hat{\gamma}_f^{ma}(m) - \hat{\gamma}_f^{sc})(y_{1,\bar{t}_f+1} - \hat{\gamma}_f^{sc}}{\sum_{f=1}^{F}(\hat{\gamma}_f^{ma}(m) - \hat{\gamma}_f^{sc})^2}$$

  The optimal $\hat{\phi}(m)$ is found by selecting the one with the lowest sum of squared errors.

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



Mean treatment effect in blue

Next, I find the penalized SC estimator following Abadie and L'Hour (2020). The procedure is quite similar as above, but I now only find one estimator (there is no such thing as a matching estimator anymore) that is both a function of $\omega^{pen,sc}$ and π (or previously ϕ). Note, indeed, that the structure of the penalized SC estimator doesn't allow to solve separately for ω and π as it was the case with the MASC estimator. Instead, I do the following

- For each π on a grid of π I compute the optimal weights given π, where the objective function I optimize over now takes π into account.
- For each $\{\pi, \omega\}$ pair, I compute the prediction error as in MASC

$$\hat{\epsilon}(\pi) \equiv Y_{it} - \sum_{k \neq i} \omega_{i,k}^{pen,sc}(\pi) Y_{jt}$$

  And find the π that minimize the sum of squared prediction errors.

- Given the optimal π, I recover $\omega^{pen,sc}$.

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