Problem Set 3 ECON8825

Federico Rodari

November 21, 2022

The main references used in this problem set are Abadie (2005), Chernozhukov et al. (2018), Chang (2020), and Colangelo and Lee (2022).

1 Differences-in-differences

(a) Recall that the estimator is defined as

$$\begin{aligned}\beta_{DiD} &= \{\mathbb{E}[Y \mid G=1, T=1] - \mathbb{E}[Y \mid G=0, T=1]\} - \{\mathbb{E}[Y \mid G=1, T=0] - \mathbb{E}[Y \mid G=0, T=0]\} \\ &= \{\mathbb{E}[Y(1) \mid D=1] - \mathbb{E}[Y(1) \mid D=0]\} - \{\mathbb{E}[Y(0) \mid D=1] - \mathbb{E}[Y(0) \mid D=0]\}\end{aligned}$$

where the observed outcome $Y(t) = Y_0(t)(1-D(t)) + Y_1(t)D(t)$. We use the following parallel trend assumption

$$\mathbb{E}[Y_0(1) \mid D=1] - \mathbb{E}[Y_0(0) \mid D=1] = \mathbb{E}[Y_0(1) \mid D=0] - \mathbb{E}[Y_0(0) \mid D=0]$$

in other words, we are assuming that the average effect of the treatment had not they been treated is equal to the average treatment effect of the untreated. From the definition of the estimator we can write

$$\begin{aligned}\beta_{DiD} &= \{\mathbb{E}[Y(1) \mid D=1] - \mathbb{E}[Y(1) \mid D=0]\} - \{\mathbb{E}[Y(0) \mid D=1] - \mathbb{E}[Y(0) \mid D=0]\} \\ &= \{\mathbb{E}[Y_1(1) \mid D=1] - \mathbb{E}[Y_0(1) \mid D=0]\} - \{\mathbb{E}[Y_0(0) \mid D=1] - \mathbb{E}[Y_0(0) \mid D=0]\} \\ &= \mathbb{E}[Y_1(1) \mid D=1] - \mathbb{E}[Y_0(1) \mid D=1] + \mathbb{E}[Y_0(0) \mid D=1] - \mathbb{E}[Y_0(0) \mid D=0] - \\ &\quad \mathbb{E}[Y_0(0) \mid D=1] + \mathbb{E}[Y_0(0) \mid D=0] \\ &= \mathbb{E}[Y_1(1) - Y_0(1) \mid D=1] = \tau_{ATT}\end{aligned}$$

In other words, under the parallel trend assumption we are able to identify the treatment for the treated subsample.

(b) We can estimate the following regression

$$Y = \beta_0 + \beta_1 D + \beta_2 T + \beta_3 D \cdot T + \epsilon$$

Consider the following 4 cases

	T = 0	T = 1
D = 0	β_0	$\beta_0 + \beta_2$
D = 1	$\beta_0 + \beta_1$	$\beta_0 + \beta_1 + \beta_2 + \beta_3$

The two differences are

$$\begin{aligned}\mathbb{E}[Y(1) \mid D=1] - \mathbb{E}[Y(1) \mid D=0] &= \beta_1 + \beta_3 \\ \mathbb{E}[Y(0) \mid D=1] - \mathbb{E}[Y(0) \mid D=0] &= \beta_1\end{aligned}$$

so the differences in differences estimator is $\beta_{DiD} = \beta_3$

(c) To construct the Neyman orthogonal score for the ATT, we will proceed in steps. Start from the standard semi-parametric estimator proposed in Abadie (2005)

$$
\begin{aligned}
\tau_{CATT} &= \mathbb{E}\left[Y_1(1) - Y_0(1) \mid X, D = 1\right] \\
&= \mathbb{E}\left[\frac{D - P(D=1 \mid X)}{P(D=1 \mid X)\left(P(D=1 \mid X)\right)}\left(Y(1) - Y(0)\right) \mid X\right] \\
&= \mathbb{E}\left[\frac{1 - P(D=1 \mid X)}{P(D=1 \mid X)\left(P(D=1 \mid X)\right)}\left(Y(1) - Y(0)\right) \mid X, D = 1\right] P(D=1 \mid X) \\
&+ \mathbb{E}\left[-\frac{P(D=1 \mid X)}{P(D=1 \mid X)\left(P(D=1 \mid X)\right)}\left(Y(1) - Y(0)\right) \mid X, D = 1\right] P(D=1 \mid X) \\
&= \left\{\mathbb{E}\left[Y(1) - Y(0) \mid X, D = 1\right]\right\} - \left\{\mathbb{E}\left[Y(1) - Y(0) \mid X, D = 0\right]\right\} = \tau^{DiD}
\end{aligned}
$$

Therefore, the standard ATT is

$$
\begin{aligned}
\tau_{ATT} &= \mathbb{E}\left[Y_1(1) - Y_0(1) \mid D = 1\right] \\
&= \int \mathbb{E}\left[Y_1(1) - Y_0(1) \mid X, D = 1\right] dP(X \mid D = 1) \\
&= \int \mathbb{E}\left[\frac{D - P(D=1 \mid X)}{P(D=1 \mid X)\left(1 - P(D=1 \mid X)\right)}\left(Y(1) - Y(0)\right) \mid X\right] dP(X \mid D = 1) \\
&= \mathbb{E}\left[\frac{D - P(D=1 \mid X)}{P(D=1 \mid X)\left(1 - P(D=1 \mid X)\right)}\left(Y(1) - Y(0)\right)\right] \frac{P(D=1 \mid X)}{P(D=1)} \\
&= \mathbb{E}\left[\frac{Y(1) - Y(0)}{P(D=1)} \frac{D - P(D=1 \mid X)}{1 - P(D=1 \mid X)}\right]
\end{aligned}
$$

With score

$$
\begin{aligned}
\psi(W; \theta_0, p_0, g_0, m_0) &= \frac{Y(1) - Y(0)}{P(D=1)} \frac{D - P(D=1 \mid X)}{1 - P(D=1 \mid X)} - \tau_0 \\
&= \frac{Y(1) - Y(0)}{p_0} \frac{D - g_0(X)}{1 - g_0(X)} - \tau_0
\end{aligned}
$$

which is not Neyman-orthogonal as shown in Chernozhukov et al. (2018) . To orthogonalize it, we must project the original score onto (D, X) to obtain

$$
\begin{aligned}
\psi(W; \tau_0, p_0, g_0, m_0) &= \frac{Y(1) - Y(0)}{P(D=1)} \frac{D - P(D=1 \mid X)}{1 - P(D=1 \mid X)} - \tau_0 - \frac{D - P(D=1 \mid X)}{P(D=1 \mid X)(1 - P(D=1 \mid X))} \mathbb{E}\left[Y(1) - Y(0) \mid X,\right. \\
&= \frac{Y(1) - Y(0)}{p_0} \frac{D - g_0(X)}{1 - g_0(X)} - \tau_0 - \frac{D - g_0(X)}{g_0(X)(1 - g_0(X))} m_0(X)
\end{aligned}
$$

To show that it is Neyman orthogonal, we must compute the Gateux derivative of the nuisance parameters evaluated at their true value.

This type of derivative was new to me, and I could not really follow the book's notation. If my understanding is correct, the nuisance parameters here are a vector, so the Gateux derivative should be a vector of partial derivatives, but in the book it looks to be a scalar (something like a total derivative). Moreover, I cannot reconcile the concept of the derivative in the book with the definition provided in Wikipedia: here we are actually estimating nuisance function and taking the derivative with respect to the function itself. This clearly means that my reasoning is wrong.

2 Continuous Treatment

1. Following the book, consider the standard interactive linear model (I follow the book notation, but we might just replace $g(\cdot)$ by the standard expression for the expectation)

$$Y = g(D, X) + \xi \ \text{ with } \ \mathbb{E}[\xi \mid D, X] = 0$$
$$D = m(X) + V \ \text{ with } \ \mathbb{E}[V \mid X] = 0$$

where now the treatment takes a set of discrete values.
Use the LIE to write $\tau(d) = \mathbb{E}[Y(d)] = \mathbb{E}[\mathbb{E}[Y(d) \mid X]] = \mathbb{E}[g(D = d, X)]$, which leads to construct the standard Neyman orthogonal score for the ATE as

$$\begin{aligned}\psi(W; \tau(d), m, g) &= \mathbb{E}\left[g(D = d, X) + \frac{I(D = d)\,(Y - g(D = d, X))}{P(D = d \mid X)} - \tau(d)\right] \\ &= \mathbb{E}\left[g(D = d, X) + \frac{I(D = d)\,(Y - g(D = d, X))}{m(X)} - \tau(d)\right]\end{aligned}$$

To estimate the ATE of shifting treatment from d to d' using the following score, we can compute

$$\psi(W; \tau(d, d'), m, g)\mathbb{E}\left[g(d', X) - g(d, X) + \left(\frac{I(d')\,(Y - g(d', X))}{P(d' \mid X)} - \frac{I(d)\,(Y - g(d, X))}{P(d \mid X)}\right) - \tau(d, d')\right]$$

To estimate $\tau(d, d')$ we can adopt the DML procedure proposed by Chernozukhov et al. (2018), which consists of (ML) cross-fitting of the high dimensional nuisance parameters. Concretely,

Step 1: Data Preparation Divide the data in K folds. For each fold $1 \le k \le K$

Step 2: Estimation Two different consecutive steps

- Estimate $g(d, X), m(X)$ using any preferred ML method using $\mid I_c^k \mid$ observations (i.e. the set of observation that does not include I^k)
- Estimate using data in I^k

$$\widehat{\tau}_k(d, d') = \frac{1}{\left|I^k\right|} \sum_{i \in I^k} \widehat{\psi}_k(W_i; , \widehat{m}, \widehat{g})$$

Step 3: Aggregation Average across K-folds with

$$\widehat{\tau}(d, d') = \frac{1}{K} \sum_{1 \le k \le K} \widehat{\tau}_k(d, d')$$

This method nests the standard one adopted in the paper using one single split, but this method might reduce the variance of the estimator.

2. From before, we have seen that

$$\tau(d) = \mathbb{E}[Y(d)] = \mathbb{E}[\mathbb{E}[Y(d) \mid X]] = \mathbb{E}[\mathbb{E}[Y \mid D = d, X]]$$

We can estimate the analogous of the discrete model in the continuous case, but now we must make use of kernel functions to weights treatment values close to the one we are estimating the treatment for. Formally,

$$\begin{aligned}\tau(d) &= \mathbb{E}[\mathbb{E}[Y \mid D = d, X]] \\ &= \int_{\mathcal{X}} \mathbb{E}[Y \mid D = d, X]\, dF_X(X) \\ &\vdots \\ &= \lim_{h \to 0} \int_{\mathcal{X}} \int_{\mathcal{Y}} \int_{\mathcal{D}} \frac{K(D - d)Y}{f(d \mid X)} dF(x, y, d) \\ &= \lim_{h \to 0} \mathbb{E}\left[\frac{K(D - d)Y}{f(d \mid X)}\right] \\ &= \mathbb{E}[\gamma(d, X)]\end{aligned}$$

At this point, the Neyman orthogonal score is simply the continuous analogous of the discrete score

$$\psi(W;\tau(d),\gamma,f,K) = \mathbb{E}\left[\gamma(d,X) + \frac{K(D-d)}{f(d\mid X)}\left(Y-\gamma(d,X)\right) - \tau(d)\right]$$

and the estimation procedure follows the same steps as the discrete analogous.

3 Heterogeneous Treatment Effects

1. I borrow the data from the Chernozhukov's book on the 1991 Survey of Income and Program Participation and estimate the (plausibly) causal effect of eligibility to enroll into a 401(k) plan on the accumulation of financial assets. By signing up to such a plan, the employee agrees to have a percentage of each paycheck paid directly into an investment account. The employer may match part or all of that contribution.
   Following previous literature on the topic, causality can be claimed only after controlling for key determinants of assets accumulation, in particular income. The logic behind the causal channel is that, after controlling for potential confounders, the eligibility can be considered as an exogenous source of variation for asset accumulation. The underlying logic is that (ceteris paribus) employees could not anticipate such an outcome, so they see the eligibility status as conditionally exogenous.
   I use net financial asset as dependent variable. The treatment variable is an indicator for being eligible to enroll in a 401(k) plan. The vector of covariates consists of age, income, family size, years of education, a married indicator, a two-earner status indicator, a defined benefit pension status indicator, an IRA participation indicator, and a home ownership indicator. I include a list of polynomial terms of some regressors to augment the dimension of the covariate space, and in the interactive specification I only interact the treatment with the linear term of the covariates for better interpretability.

2. We estimate the two regression specifications using the two proposet equation and find the following

Table 1: OLS Results

	Dependent variable:	
	Financial Wealth	
	(1)	(2)
e401	$9,002.952^{***}$	$6,467.280^{**}$
e401:age_demean		520.239^{***}
		(122.264)
e401:inc_demean		0.260^{***}
		(0.061)
e401:educ_demean		−236.214
		(475.391)
e401:fsize_demean		−1,208.337
		(909.372)
e401:marr		4,310.774
		(3,758.524)
e401:twoearn		$-11,348.490^{***}$
		(3,150.558)
e401:db		1,784.989
		(2,627.107)
e401:pira		−110.428
		(2,815.261)
e401:hown		4,331.259
		(2,723.976)
Constant	2,264,732.000	1,884,195.000
	(2,736,772.000)	(2,729,517.000)
Observations	9,915	9,915
R^2	0.298	0.303
Interaction	No	Yes
Polynomial	Yes	Yes
Note:		*p<0.1; **p<0.05; ***p<0.01

We can observe that there exists some heterogeneity in treatment effects, in particular age,income and being a family of two earners seem to capture some variation of the overall effect.

3. We estimate the previously written down ATE using the DML methodology in the IRM context. To estimate the nuisance parameters we use LASSO and optimize the tuning parameter through k-fold cross-validation. The result is $\tau_{ATE} = 7540.597(1233.765)$, where the term in parentheses is the standard error.

4. I estimate heterogeneous treatment effects across the dimensions we identified above. For two earners, I estimate the HATE by splitting the sample according to the dummy level, while for the continuous variables I split the data based on quartiles. I here report only the estimates for income

Quantile	τ_{HATE}^{inc}	SE_{HATE}^{inc}
25%	3611.918	1110.332
50%	2859.695	1352.093
75%	6307.041	1736.554
100%	17926.722	3962.183

We observe some non-linearity in the treatment effect: it is increasing in income but for the second quartile we observe a reduction in the estimate