Pset 1

Anya Marchenko
Applied Econometrics
Peter Hull (Spring 2024)

March 20, 2024

Question 1

[30 points] Consider a causal model of $Y_i = \beta_i X_i + \epsilon_i$ for some outcome Y_i and treatment X_i. Consider also the regression $Y_i = \beta X_i + W_i'\gamma + e_i$ for some covariate vector W_i.

(a) **[3 points]** Explain why the causal model might be considered "without loss" when X_i is binary. What does the model impose outside of this case?

Answer: The model is without loss when X_i is binary because it is a saturated model. In this case, the CEF is linear and the population regression line function perfectly coincides with the CEF (Theorem 3.1.4, MHE). Outside of this, the model imposes linearity, that is, that the causal effect of X_i on Y_i is linear in β_i.

(b) **[4 points]** Suppose that $E[X_i|\beta_i, \epsilon_i, W_i] = W_i'\lambda$ for some λ. Explain why this condition should hold when X_i is randomly assigned within strata indicated in W_i by a set of group fixed effects. What are two (or maybe three) distinct reasons why this condition might fail in other settings?

Answer: If X_i is randomly assigned within strata W_i, then X_i should be uncorrelated with β_i, ϵ_i conditional on W_i (i.e., $X_i \perp \beta_i, \epsilon_i | W_i$). When we take the expectation of a random variable conditioning on r.v.s uncorrelated with it, the expectation is unchanged so we can remove the conditioning: $E[X_i|\beta_i, \epsilon_i, W_i] = E[X_i|W_i]$. Then, the $E[X_i|W_i]$ is the mean of X_i in each strata, where W_i is the indicator for being in the strata and λ_i is the mean of X_i in that strata, so $E[X_i|\beta_i, \epsilon_i, W_i] = W_i'\lambda$.

This condition might fail if X_i is not randomly assigned within strata, such as if members of a strata select into treatment based on selection on gains (they expect their β_i to be higher), or if they select based on their untreated potential outcomes (their ϵ_i).

(c) **[5 points]** Show that $E[\tilde{X}_i \epsilon_i] = 0$ under the condition in (b), where $\tilde{X}_i$ denote the residuals from the population regression of X_i on W_i.

Answer:

$$E[\tilde{X}_i \epsilon_i] = E[E[\tilde{X}_i | \epsilon_i, W_i] \epsilon_i] = E[E[(X_i - E[X_i|W_i]) | \epsilon_i, W_i] \epsilon_i]$$

$$= E\left[(E[X_i|\epsilon_i, W_i] - E[E[X_i|W_i]|\epsilon_i, W_i])\epsilon_i\right]$$
$$= E\left[(E[X_i|\epsilon_i, W_i] - E[X_i|W_i])\epsilon_i\right]$$
$$= E\left[(E[X_i|W_i] - E[X_i|W_i])\epsilon_i\right]$$
$$= E\left[(0)\epsilon_i\right]$$
$$= 0$$

(d) **[5 points]** Show $E[\tilde{X}_i\beta_i X_i] = E[\sigma_i^2\beta_i]$ under the condition in (b), where $\sigma_i^2 = Var(X_i|\beta_i, W_i)$.

Answer:

$$E[\tilde{X}_i\beta_i X_i] = E[\tilde{X}_i X_i\beta_i] = E[\underbrace{E[\tilde{X}_i X_i|\beta_i, W_i]}_{*}\beta_i]$$

Focus just on (*)

$$* = E[\tilde{X}_i X_i|\beta_i, W_i] = E[(X_i - E[X_i|W_i])X_i|\beta_i, W_i] = E[(X_i X_i - E[X_i|W_i]X_i)|\beta_i, W_i]$$
$$= E[X_i X_i|\beta_i, W_i] - E[E[X_i|W_i]X_i|\beta_i, W_i]$$
$$= E[X_i X_i|\beta_i, W_i] - E[E[X_i|\beta_i, W_i]X_i|\beta_i, W_i]$$
$$= E[X_i X_i|\beta_i, W_i] - E[X_i|\beta_i, W_i]E[X_i|\beta_i, W_i]$$
$$= E[X_i X_i|\beta_i, W_i] - E[X_i|\beta_i, W_i]E[X_i|\beta_i, W_i]$$
$$= E[X_i X_i|\beta_i, W_i] - E[X_i|\beta_i, W_i]^2$$
$$= Var[X_i|\beta_i, W_i]$$

So putting * together with the first line, we get:

$$= E[Var[X_i|\beta_i, W_i]\beta_i] = E[\sigma_i^2\beta_i]$$

(e) **[4 points]** Use your results in (c) and (d) to show that the regression estimand can be written $\beta = E[\sigma_i^2\beta_i]/E[\sigma_i^2]$. Say a few nice things about this result, drawing on what we saw in lecture.
Answer: By FWL,

$$\beta = \frac{E[\tilde{X}_i Y_i]}{E[\tilde{X}_i^2]} = \frac{E[\tilde{X}_i(\beta_i X_i + \epsilon_i)]}{E[\tilde{X}_i^2]} = \frac{E[\tilde{X}_i\beta_i X_i + \tilde{X}_i\epsilon_i]}{E[\tilde{X}_i^2]} = \frac{E[\tilde{X}_i\beta_i X_i] + E[\tilde{X}_i\epsilon_i]}{E[\tilde{X}_i^2]}$$

$$= \frac{E[\tilde{X}_i\beta_i X_i]}{E[\tilde{X}_i^2]} = \frac{E[\sigma_i^2\beta_i]}{E[\sigma_i^2]}$$

This result is nice because it shows that the β in the regression we run is a weighted average of the individual β_i effects, weighted by the conditional variance of X_i. This is good because it mimicks the Angrist 1998 result that shows that the coefficient beta on a regression of Y on a binary treatment with controls X equals the convex weighted average of the CATEs.

(f) **[5 points]** State a condition under which β identifies the average causal effect, $E[\beta_i]$. State a condition under which the σ_i^2 are identified. Propose an estimator for $E[\beta_i]$ under this condition. When is this estimator likely to work well?

Answer: The average causal effect is identified if we assume conditional unconfoundedness, that is, if $X_i \perp (\epsilon_i, \beta_i)|W_i$. In this case, $\sigma_i^2 = Var(X_i|W_i, \beta_i) = Var(X_i|W_i)$, which can be identified from the data. In the case when the weights are simplified in this way, we can identify $E[\beta_i] = \beta$.

(g) **[4 points]** Now suppose we have an instrument Z_i satisfying $E[Z_i|\beta_i, \epsilon_i, W_i] = W_i'\lambda$ for some λ. Show that the IV estimand which controls for W_i can be written $\beta^{IV} = E[w(\beta_i)\beta_i]/E[w(\beta_i)]$ for some $w(\cdot)$. State and interpret a condition under which this weighting scheme is convex.

Answer:

Denote $\tilde{Z}_i$ as the residuals from a regression of Z_i on W_i. Then by FWL we can write:

$$\beta^{IV} = \frac{Cov(Y, \tilde{Z})}{Cov(X, \tilde{Z})} = \frac{E[\tilde{Z}_i Y_i]}{E[\tilde{Z}_i X_i]} = \frac{E[\tilde{Z}_i(\beta_i X_i + \epsilon_i)]}{E[\tilde{Z}_i X_i]} = \frac{E[\tilde{Z}_i \beta_i X_i + \tilde{Z}_i \epsilon_i]}{E[\tilde{Z}_i X_i]} = \frac{E[\tilde{Z}_i \beta_i X_i] + E[\tilde{Z}_i \epsilon_i]}{E[\tilde{Z}_i X_i]}$$

$$= \frac{E[\tilde{Z}_i \beta_i X_i]}{E[\tilde{Z}_i X_i]} = \frac{E[\omega(\beta_i)\beta_i]}{E[\omega(\beta_i)]}$$

We need first stage monotonicity to make the weights convex in the presence of heterogeneous treatments effects.

Question 2

[18 points] Suppose we are interested in the effectiveness of some job training program; we assume the program's earnings effect is some constant, β. We observe earnings Y_{it} both before and after treatment assignment, $t \in \{1, 2\}$, as well as an indicator D_i for treatment receipt at the end of period 1. We also observe an indicator W_{it} for whether individual i has a high-skill job in each period t.

(a) [4 points] Consider the regression of Y_{i2} on D_i and W_{i2}. Argue that the coefficient on D_i is a weighted average of $\Delta(1)$ and $\Delta(0)$ for $\Delta(w) = E[Y_{i2}|D_i = 1, W_{i2} = w] - E[Y_{i2}|D_i = 0, W_{i2} = w]$.

Answer: Use FWL to write β_D:

$$\beta = \frac{Cov(\tilde{D}_i Y_{i2})}{Var(\tilde{D}_i)} = \frac{E[\tilde{D}_i Y_{i2}]}{E[\tilde{D}_i^2]} = \frac{E[(D_i - E[D_i|W_{i2}])Y_{i2}]}{E[(D_i - E[D_i|W_{i2}])^2]}$$

$$= \frac{E[E[(D_i - E[D_i|W_{i2}])Y_{i2}|D_i, W_{i2}]]}{E[(D_i - E[D_i|W_{i2}])^2]} = \frac{E[(D_i - E[D_i|W_{i2}])E[Y_{i2}|D_i, W_{i2}]]}{E[(D_i - E[D_i|W_{i2}])^2]}$$

Where the last step comes from the fact that the residuals are uncorrelated with D_i and W_{i2}. Now, consider the $E[Y_{i2}|D_i, W_{i2}]$ expression:

$$E[Y_{i2}|D_i, W_{i2}] = E[Y_{i2}|D_i = 1, W_{i2}]D_i + E[Y_{i2}|D_i = 0, W_{i2}](1 - D_i)$$
$$= E[Y_{i2}|D_i = 1, W_{i2}]D_i + E[Y_{i2}|D_i = 0, W_{i2}] - E[Y_{i2}|D_i = 0, W_{i2}]D_i$$
$$= \underbrace{(E[Y_{i2}|D_i = 1, W_{i2}] - E[Y_{i2}|D_i = 0, W_{i2}])}_{\Delta(w)} D_i + E[Y_{i2}|D_i = 0, W_{i2}]$$
$$= \Delta(w)D_i + E[Y_{i2}|D_i = 0, W_{i2}]$$

Next, we can substitute this into the numerator of the previous expression:

$$= \frac{E[(D_i - E[D_i|W_{i2}])(\Delta(w)D_i + E[Y_{i2}|D_i = 0, W_{i2}])]}{E[(D_i - E[D_i|W_{i2}])^2]}$$
$$= \frac{E[(D_i - E[D_i|W_{i2}])\Delta(w)D_i]}{E[(D_i - E[D_i|W_{i2}])^2]} + \frac{E[(D_i - E[D_i|W_{i2}])E[Y_{i2}|D_i = 0, W_{i2}]]}{E[(D_i - E[D_i|W_{i2}])^2]}$$

Notice that $D_i\tilde{D}_i = \tilde{D}_i^2$, which allows us to simplify the first term. Second, notice that the second term is zero because $E[Y_{i2}|D_i = 0, W_{i2}$ is a function of W_{i2} but not D_i, since $D_i = 0$ is fixed. This CEF is uncorrelated with the residuals of D_i because those residuals are orthogonal to W_{i2}.

$$= \frac{E[(D_i - E[D_i|W_{i2}])^2\Delta(w)]}{E[(D_i - E[D_i|W_{i2}])^2]} + \underbrace{\frac{E[(D_i - E[D_i|W_{i2}])E[Y_{i2}|D_i = 0, W_{i2}]]}{E[(D_i - E[D_i|W_{i2}])^2]}}_{=0}$$

$$= \frac{E[(D_i - E[D_i|W_{i2}])^2\Delta(w)]}{E[(D_i - E[D_i|W_{i2}])^2]}$$

$$= \frac{E[\omega(W_{i2})\Delta(w)]}{E[Var(\tilde{D}_i)]}$$

So now we have written β_D as a function of some weights and $\Delta(w)$.

(b) [5 points] Suppose D_i is randomly assigned at the end of period 1 but can affect employment in period 2. Specifically, suppose the treatment can shift individuals into (but not out of) a high-skilled job. Show that $\Delta(1)$ can be written as the sum of β and a bias term involving a comparison of individuals who always work in a high-skill job and individuals who switch into a high-skill job. A similar result can be derived for $\Delta(0)$, but you don't have to prove it. Draw a conclusion about the regression in (a).

Answer:

Note that the DiD coefficient can be written as:

$$\beta = E[Y_{2i} - Y_{1i}|D_i = 1] - E[Y_{2i} - Y_{1i}|D_i = 0]$$

$$= E[Y_{2i} - Y_{1i}|D_i = 1, W_{1i} = 1] - E[Y_{2i} - Y_{1i}|D_i = 0, W_{1i} = 1]$$

By random assignment of D_i.

Start with the expression for $\Delta(1)$:

$$\Delta(1) = E[Y_{i2}|D_i = 1, W_{i2} = 1] - E[Y_{i2}|D_i = 0, W_{i2} = 1]$$

Rewrite in terms of W_{i1}:

$$= E[Y_{i2}|D_i = 1, W_{i1} = 1] \underbrace{P(W_{i1} = 1|W_{i2} = 1, D_i = 1)}_{:=a}$$
$$+E[Y_{i2}|D_i = 1, W_{i1} = 0] \underbrace{P(W_{i1} = 0|W_{i2} = 1, D_i = 1)}_{:=b}$$
$$-E[Y_{i2}|D_i = 0, W_{i1} = 1]$$

Since no one can switch into being high-skilled if they weren't assigned treatment, there is no reason to consider cases for $D_i = 0$. Now, add and subtract the same term, $E[Y_{i1}|D_i = 1, W_{i1} = 1](a + b)$, noting that $(a + b) = 1$:

$$= E[Y_{i2}|D_i = 1, W_{i1} = 1]a$$
$$+E[Y_{i2}|D_i = 1, W_{i1} = 0]b$$
$$+E[Y_{i1}|D_i = 1, W_{i1} = 1](a + b) - E[Y_{i1}|D_i = 0, W_{i1} = 1](a + b)$$
$$-E[Y_{i2}|D_i = 0, W_{i1} = 1](a + b)$$

Label the first term ①, the second ②, and so forth, so our expression for $\Delta(1)$ becomes:

$$= ①a + ②b - 3(a + b) + 4(a + b) - 5(a + b)$$
$$= a(① - ③ + ④ - ⑤) + b(② - ③ + ④ - ⑤)$$

$$= a(① - ③ + ④ - ⑤) + \underbrace{b(② - ③ + ④ - ⑤)}_{:=d}$$

Notice that $① - ③ + ④ - ⑤ = E[Y_{2i} - Y_{1i}|D_i = 1, W_{1i} = 1] - E[Y_{2i} - Y_{1i}|D_i = 0, W_{1i} = 1] = \beta$, so we can rewrite as:

$$= a\beta + d$$
$$= a\beta + (1 - a)\beta + (1 - a)\beta + d = \beta + \underbrace{b\beta + d}_{\text{bias}}$$

So now we have rewritten $\Delta(1)$ as a function of β and a bias term, where the bias term is composed of the share of people who switch from low to high skill after being treated (b), the share of people who remain high skilled after being treated, and their conditional differences in means.

(c) [5 points] Now suppose we do not assume D_i is randomly assigned, but instead model untreated potential outcomes as $Y_{it}(0) = \alpha_i + \gamma_t + W_{it}'\gamma + \varepsilon_{it}$ where $E[\varepsilon_{it}|D_i, W_{it}] = 0$. Show that a regression of Y_{it} on D_i, a post-period indicator, and their interaction identifies β plus a different bias term. When is this bias term zero?

Answer: I'm very tired.

(d) [4 points] Argue that dropping the W_{i2} control from the regression in (a) solves the bias problem in (b), while adding a W_{it} control to the regression in (c) solves the bias problem there.

Answer: In part (a) we have random assignment of D_i, so D_i is independent of W_i. In that causal model, $E[\epsilon_i|D_i] = 0$, so running a regression of Y_{i2} on D_i without W_i solves the bias problem and estimates the true causal effects. However, in part (c), we assume that D_i is not unconditionally randomly assigned and $E[\epsilon_{it}|D_i, W_{it}] = 0$, so in order to estimate the causal effect without bias, we need to control for W, which is precisely what we do.

Question 3

[22 points] Economists are often interested in the effects of a student's peers on their academic achievement. Consider a simple "peer-effects" regression of some achievement outcome Y_i (e.g. the test score of student i) on some measure of average classroom peer "quality" $\overline{X}_{j(i)}$ where $j(i)$ denotes student i's classroom. Specifically, $\overline{X}_j = \frac{1}{C}\sum_{i':j(i')=j} X_{i'}$ is the average of some baseline achievement measure (e.g. lagged test scores); here we assume all classrooms are of the same size, C.

(a) **[4 points]** Argue that the OLS coefficient from this regression can be identically obtained by a classroom-level OLS regression. What is the outcome variable? Are there any weights?

Answer: Regression: $Y_i = \beta_0 + \beta_D \bar{X}_{j(i)} + \epsilon_i$. By definition,

$$\beta_D = E[\bar{X}_{j(i)}\bar{X}_{j(i)}]^{-1}E[\bar{X}_{j(i)}Y_i]$$

By LIE,

$$\beta_D = E[\bar{X}_{j(i)}\bar{X}_{j(i)}]^{-1}E[\bar{X}_{j(i)}E[Y_i|j(i) = j]]$$

So the LIE shows that we can estimate the same coefficient by regressing the average outcome in a given classroom on the average classroom peer quality. The weights would normally be the number of students in each classroom, but since the classes are all the same size, the weights don't matter.

(b) **[5 points]** Argue that the OLS coefficient can also be identically obtained by an (individual-level) 2SLS regression. What are the instruments? Does this result depend on C being constant?

Answer: The individual-level regression that we can run is of the form:

$$Y_i = \beta_D \hat{X}_i + \epsilon_i$$

Where $\hat{X}_i$ (baseline achievement of student i) is the fitted value from a first stage regression of that student's baseline achievement on a classroom fixed effect $\alpha_{j(i)}$, which equals 1 if student i belongs to classroom j:

$$X_i = \pi_1 \alpha_{j(i)} + \nu_i$$

This 2SLS approach is numerically identical to the OLS coefficient because $\hat{X}_i$ is just going equal the average of achievement in the student's classroom, so instrumenting with this is equivalent to regressing on average achievement directly.

(c) **[4 points]** Suppose classrooms are randomly assigned to students after X_i is determined but before Y_i is determined. Is this enough to interpret the regression of Y_i on $\bar{X}_{j(i)}$ as the causal effect of being in a classroom with high-quality peers (as measured by $\bar{X}_{j(i)}$)? Explain.

Answer: Yes, this is sufficient. To interpret the β_D in the regression $Y_i = \beta_0 + \beta_D \bar{X}_{j(i)} + \epsilon_i$ as causal, it needs to be the case that $\bar{X}_{j(i)} \perp Y_i(0), Y_i(1)$. This is clearly the case since the random assignment does not depend on X_i (even though X_i is measured).

(d) **[5 points]** Which IV condition is violated under the condition in (c) when $C \to \infty$? What is likely to happen to the OLS coefficient when the classrooms are large?

Answer: As the number of students in a given classroom approaches infinity, we have a weak instrument problem. The relevance assumption requires that the instrument ($\alpha_{j(i)}$, whether you are assigned to a classroom) is sufficiently correlated with the average test scores ($\bar{X}_j$). However, a classroom with infinitely many students means that student i's assignment does not affect the average test scores in the class that much, and we will have a weak first stage.

As classrooms get larger and we regress on the average test scores in that classroom, OLS will give noisier estimates because the regression is getting closer to averaging the population of students (so it's not a peer effects regression anymore).

(e) **[4 points]** Suppose you're not worried about the potential issues in (c) and (d). What standard errors would you use when estimating the OLS coefficient?

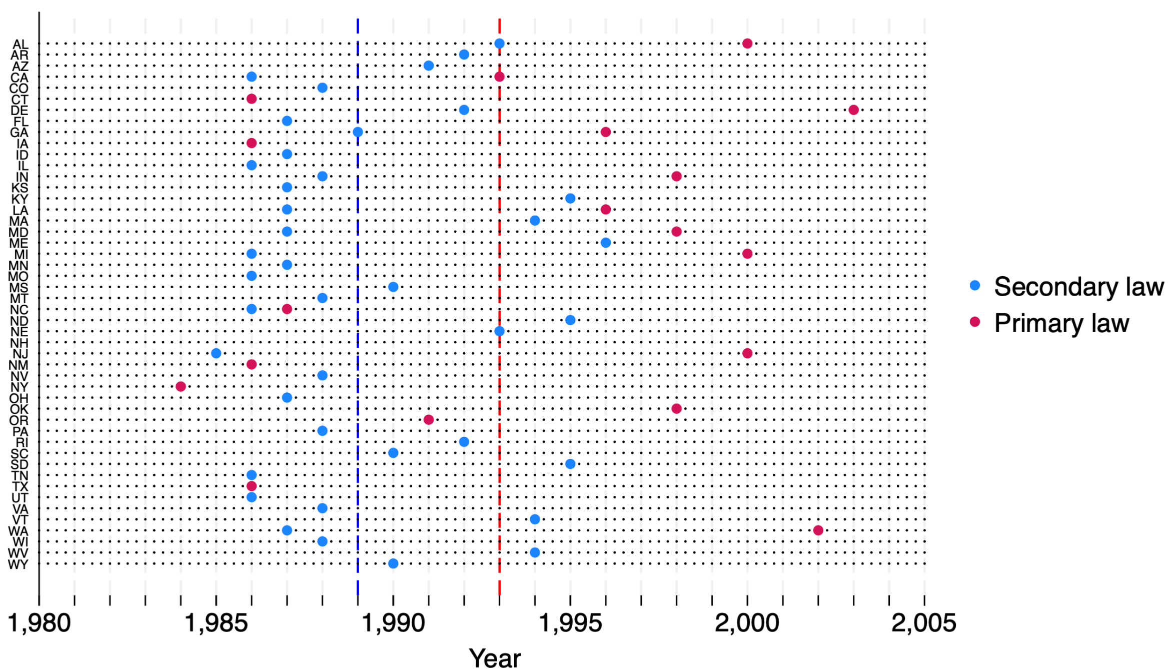
Answer: Abadie et al. (2023) shows that in design-based specifications such as the one given in (c), standard errors should be clustered at the level of treatment assignment. Since treatment was assigned at the classroom level, I would cluster at the classroom level.

Question 4

[30 points] This question asks you to reanalyze data from Anderson (2008)[1]. This paper estimated effects of light trucks on traffic fatalities, but in doing so it also collected data on primary and secondary seat belt laws. A secondary belt law stipulates that law enforcement can only ticket a driver for not wearing a seat belt if the driver has already been pulled over for simultaneously breaking a different traffic law. A primary belt law stipulates that law enforcement can ticket a driver for not wearing a seat belt, regardless of whether he has broken any other laws. This problem set will focus on the question of whether primary seat belt laws save lives. Attached you can find a cleaned dataset for this analysis (both DTA and CSV, with a TXT file giving variable names and labels for the latter).

(a) **[3 points]** Load the dataset and summarize the data. Is the panel balanced (i.e., complete observations for each state-year combination)? Visualize the timing of primary belt laws. Are there any reversals of these laws? Are there "never-treated" states? How do the timing of primary and secondary belt laws relate to each other?

Answer: My summary statistics are in the code. The panel data is balanced, there are no reversals, no missings. 30 states were never treated, 0 states reversed primary laws, and 11 states reversed secondary laws. The timing is visualized here.



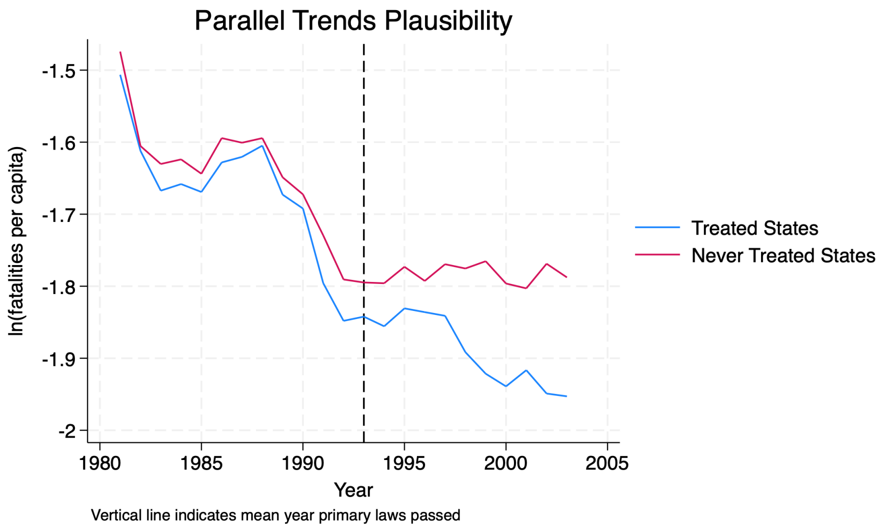
Vertical lines indicate mean enactment year, 1989 for secondary, 1993 for primary

(b) **[4 points]** Define the outcome as log traffic fatalities per capita. Plot this outcome in a way that may be helpful for later difference-in-differences (DiD) analyses. From this graph, can

[1] Link

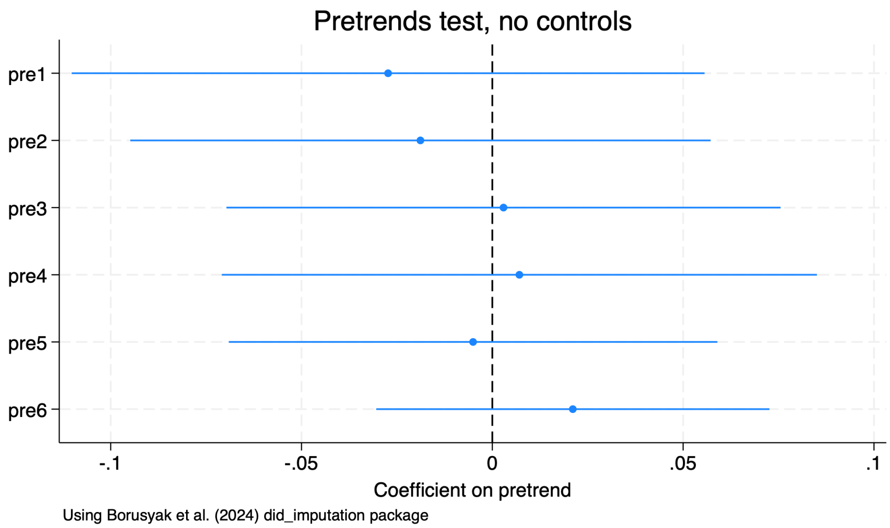
you say anything about the plausibility of the parallel trends assumption, as well as the likely effects of primary belt laws?

Answer: On the graph below I am plotting the trends in fatalities over time among ever treated states (those who ever enacted a primary law) and never treated states. The parallel trends assumption seems to hold - both the never treated and ever treated states are trending downward during this period.



(c) **[4 points]** Test the parallel trends assumption in any manner you find feasible and useful. Summarize the test and your findings. Do secondary belt laws pose a problem for a simple DiD analysis? If so, test whether that problem is likely to be significant. If not, explain why not.

Answer: I used the `did_imputation` package from Borusyak et al. 2024 to test the parallel trends assumption for up to 6 periods before treatment, and I plot the resulting coefficients below. We cannot reject that the coefficients on pretrends are zero, the test seems to be underpowered.



(d) **[6 points]** Using regression, estimate the dynamic effect of primary belt laws on the outcome for each of the horizons where a reasonable sample is available. Then compare your estimates to ones from the Borusyak, Jaravel, and Spiess (2024) imputation estimator. Do the results mostly agree?

Answer: I report the TWFE event-study regression estimates in the table on the next page. Pretrends are insignificant, and post treatment coefficients are negative and insignificant. Since treatment is staggered, we know that classic event-study specifications may be biased. I therefore also created event study graphs showing the dynamic effect of primary laws using the Callaway and Sant'Anna (2021) approach and calculated an overall ATT $= -.068$. The Borusyak, Jaravel, and Spiess (2024) overall ATT $= -.107$, which aligns with the Callaway estimate. The Borusyak dynamic estimates are displayed below the table. The Borusyak approach treatment coefficients are negative and (unlike regression) significant.

	(1)
	Ln(fatalities per capita)
Treat × Lag -4	0.0257
	(0.92)
Treat × Lag -3	0.0311
	(1.84)
Treat × Lag -2	0.00976
	(0.40)
Treat × Lag 0	-0.0566
	(-1.15)
Treat × Lead +1	-0.0398
	(-0.79)
Treat × Lead +2	-0.0479
	(-1.21)
Treat × Lead +3	-0.0454
	(-1.34)
Treat × Lead +4	-0.0330
	(-0.89)
N	414

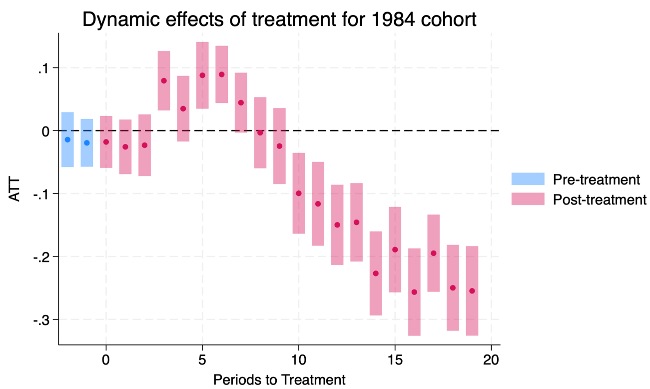
t-stat in parentheses. Table shows coefficients from a TWFE event-study regression, no controls.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

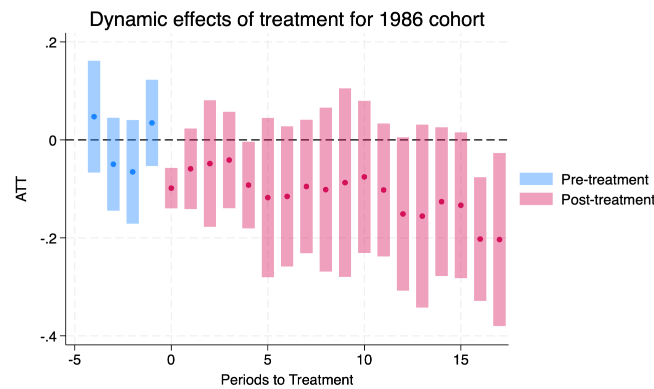
Number of obs = 1,035

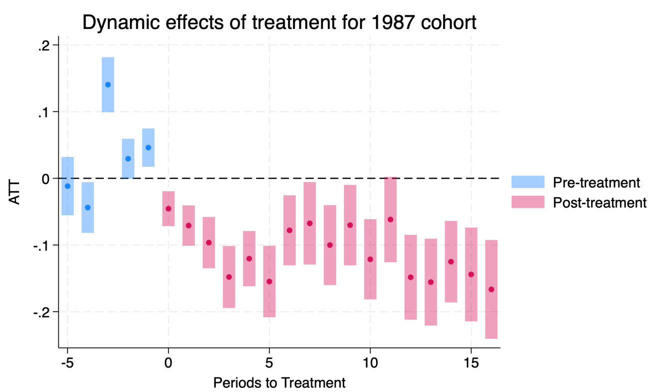
y	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
tau0	-.0535273	.0250533	-2.14	0.033	-.1026309	-.0044236
tau1	-.0650679	.0276325	-2.35	0.019	-.1192265	-.0109092
tau2	-.0675896	.0288414	-2.34	0.019	-.1241176	-.0110616
tau3	-.0637709	.0285417	-2.23	0.025	-.1197116	-.0078303
tau4	-.0852289	.0387245	-2.20	0.028	-.1611276	-.0093302
tau5	-.096785	.0465395	-2.08	0.038	-.1880008	-.0055692
tau6	-.1155178	.0517496	-2.23	0.026	-.2169452	-.0140903
tau7	-.0927107	.0488786	-1.90	0.058	-.1885111	.0030896
pre1	-.0302228	.0399976	-0.76	0.450	-.1086166	.0481711
pre2	-.0216767	.0363079	-0.60	0.550	-.092839	.0494855
pre3	.0002391	.0343866	0.01	0.994	-.0671573	.0676355
pre4	.0040385	.0373376	0.11	0.914	-.069142	.0772189
pre5	-.0075918	.0303079	-0.25	0.802	-.0669941	.0518105

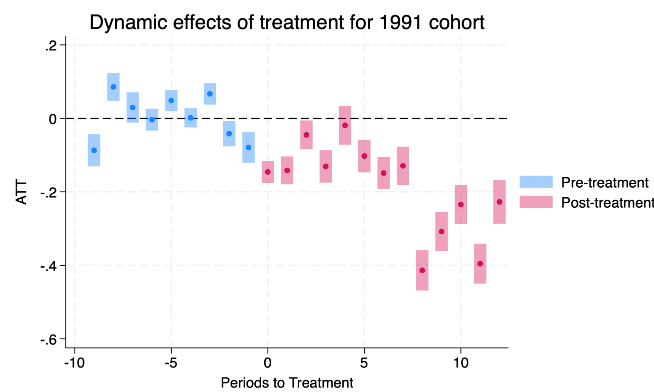
Figure 1: Borusyak et al. (2024) estimates of dynamic effect of primary laws. The pretrends are not significant but the post-treatment coefficients (τ) are significant and negative. The overall ATT is $= -.107$.

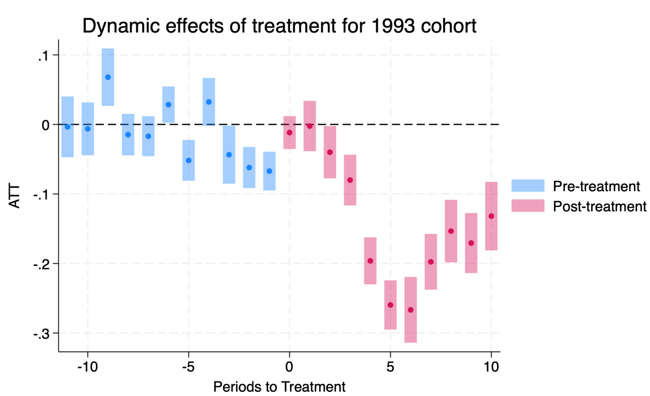
Now, let's use the Callaway approach for staggered treatment to create event-study graphs (for horizons where there is enough of a post-period, I omit 2002 and 2003 cohorts).

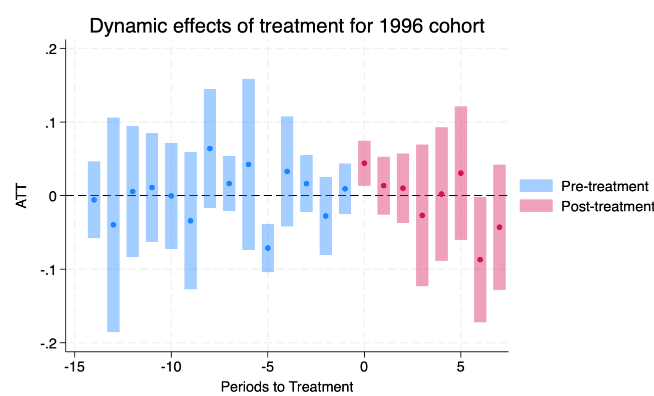


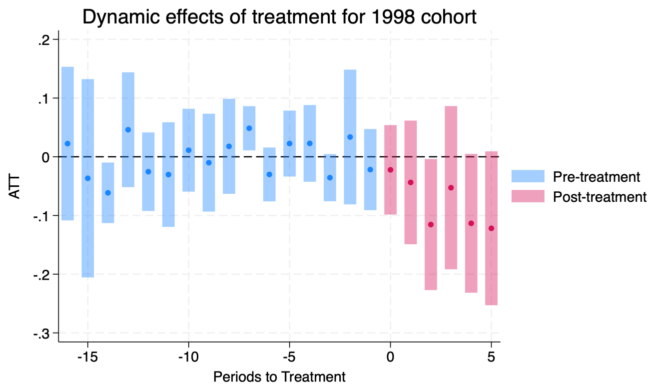


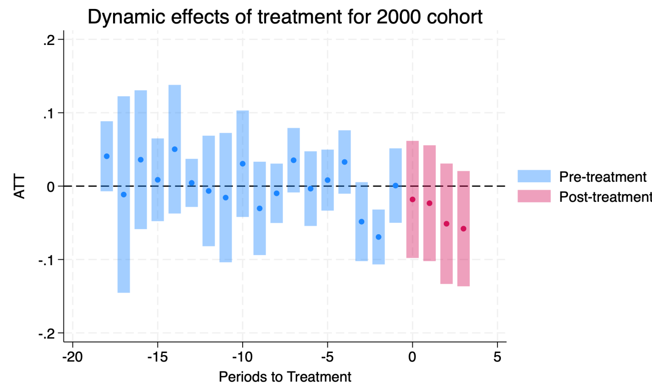












(e) **[4 points]** Check the sensitivity of both estimates in (d) to including state-specific linear trends in your model of untreated potential outcomes.

Answer: The regression estimates are relatively not impacted by the addition of controls, including state-specific trends. As we can see in the table below, the regression estimates shrink slightly, remain the same sign, and remain not significant. In addition to state-specific trends, I added controls for the number of VMT, rain and snow precipitation, and urban and rural speed limits. The Borusyak estimates show a negative and significant treatment effect on fatalities in the periods immediately after the passing of the law, and negative and insignificant treatment effects thereafter.

	(1)
	Ln(fatalities per capita)
Treat × Lag -4	0.0149
	(0.55)
Treat × Lag -3	0.0197
	(1.24)
Treat × Lag -2	0.00521
	(0.27)
Treat × Lag 0	-0.0403
	(-0.95)
Treat × Lead 1	-0.0367
	(-0.93)
Treat × Lead 2	-0.0323
	(-1.27)
Treat × Lead 3	-0.0230
	(-0.82)
Treat × Lead 4	-0.0175
	(-0.45)
Beer consumption	0.748***
	(8.27)
Total VMT	-0.000000198
	(-0.25)
Precipitation (inch)	-0.0252**
	(-2.41)
Snow (inch)	0.0597***
	(3.03)
Rural Speed Limit	0.0000493
	(0.01)
Urban Speed Limit	0.00215
	(0.71)
N	414

t statistics in parentheses. Controls also include state-specific linear trends.
* p¡.1, ** p¡.05, *** p¡.01

Number of obs = 1,035

y	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
tau0	-.045392	.0217749	-2.08	0.037	-.08807	-.002714
tau1	-.0560481	.0254924	-2.20	0.028	-.1060122	-.0060839
tau2	-.0672784	.0323229	-2.08	0.037	-.1306301	-.0039266
tau3	-.0615068	.0405247	-1.52	0.129	-.1409337	.0179202
tau4	-.0672605	.0464469	-1.45	0.148	-.1582947	.0237736
tau5	-.0758719	.0612026	-1.24	0.215	-.1958269	.044083
tau6	-.0851906	.0785872	-1.08	0.278	-.2392186	.0688374
tau7	-.0579566	.0816738	-0.71	0.478	-.2180343	.1021211
pre1	-.0584993	.0640626	-0.91	0.361	-.1840598	.0670611
pre2	-.0468572	.053153	-0.88	0.378	-.1510352	.0573209
pre3	-.0209261	.0397627	-0.53	0.599	-.0988595	.0570074
pre4	-.0133681	.0423248	-0.32	0.752	-.0963232	.069587
pre5	-.0212287	.0300404	-0.71	0.480	-.0801067	.0376494

Figure 2: Borusyak et al. (2024) estimates of dynamic effect of primary laws, with state specific linear trends.

(f) **[6 points]** Estimate the “static” TWFE regression which specifies treatment as only affecting outcomes in the current period. Estimate and plot the total weight this regression places on treated observations at each horizon. In what way are these weights informative? Compare these to the sample weights of each horizon. In your view, does the static regression coefficient provide a useful summary of causal effects in this setting?

Answer: I plot the two sets of weights below. The sampling weights are remarkably similar to the regression weights in relative terms (they are scaled differently). The static regression coefficient provides a pretty good estimate of the causal effects in this setting - the static TWFE coefficient is around -.08, which is similar in magnitude to the dynamic coefficients and the Borusyak ATT estimates, which are around -.07 (see part d).

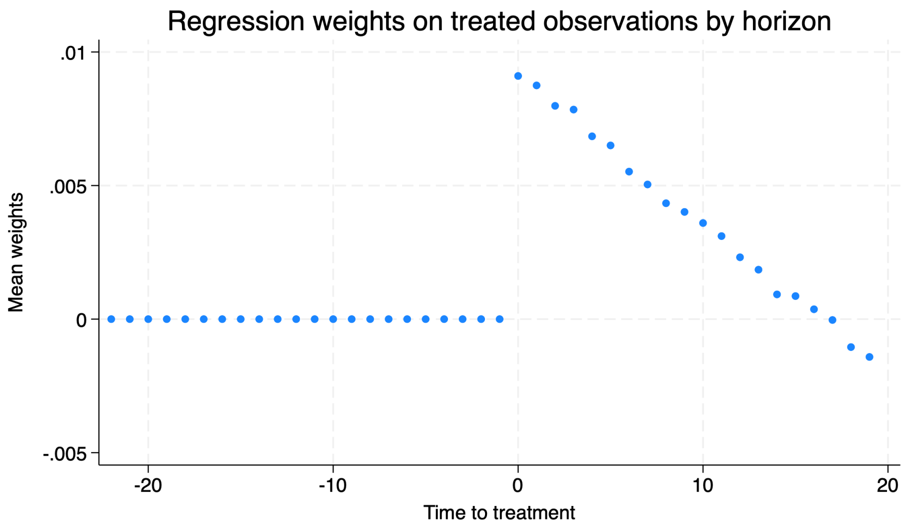


Figure 3: Weights calculated are $\frac{D_{it}\tilde{D}_{it}}{\sum D_{it}\tilde{D}_{it}}$ at each treatment horizon, where $\tilde{D}_{it}$ is the residual from a regression of primary on time and state fixed effects.

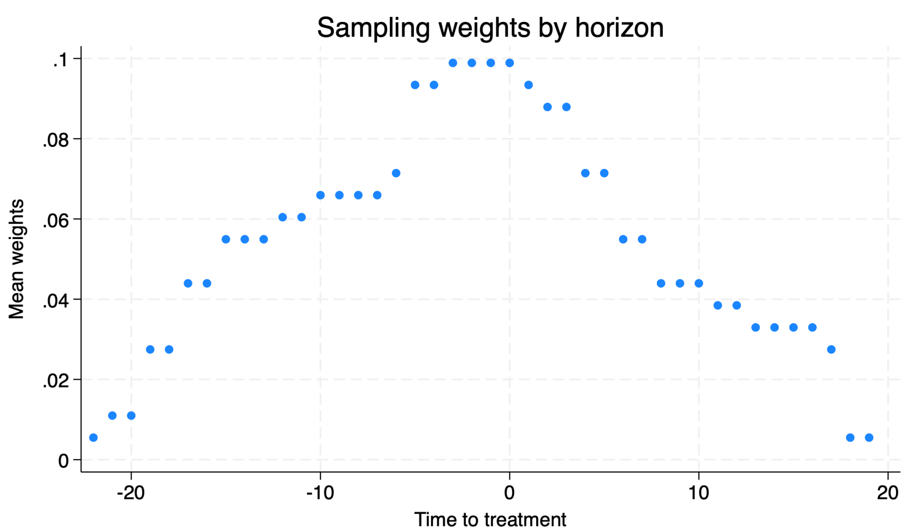


Figure 4: Weights calculated as the number of states in a given time to treat horizon divided by the total number of state-periods.

(g) **[3 points]** Submit the final, well-commented code for this problem.
Answer: Submitted.