PP 421: Problem Set 5

Luke Chen

Acknowledgements: Ibrahim Khan

May 29, 2020

1 Theory

1.1

(a) The random utility model is given by

$$U_{ij} = X'_{ij}\beta + \epsilon_{ij}$$

We are given the assumption that ϵ_{ij} is independent across individuals and choices. Additionally, we assume the errors have Type I extreme value distribution.

We would like to show:

$$\begin{aligned} P_{ij} &= Pr(X'_{ij}\beta + \epsilon_{ij} > X'_{im}\beta + \epsilon_{im}) \quad m \neq j \\ &= Pr(X'_{ij}\beta - X'_{im}\beta > \epsilon_{im} - \epsilon_{ij}) \quad m \neq j \\ &= \frac{exp(X'_{ij}\beta)}{\sum_{m=1}^{J} exp(X'_{im}\beta)} \end{aligned}$$

where J is the total number of choices.

First, through simple algebraic manipulation, we can rewrite the probability as:

$$\begin{aligned} P_{ij} &= Pr(X'_{ij}\beta + \epsilon_{ij} > X'_{im}\beta + \epsilon_{im}) \quad m \neq j \\ &= Pr(\epsilon_{im} - \epsilon_{ij} < X'_{ij}\beta - X'_{im}\beta) \quad m \neq j \\ &= Pr(\epsilon_{im} < X'_{ij}\beta - X'_{im}\beta + \epsilon_{ij}) \quad m \neq j \end{aligned}$$

The last line establishes the probability as a CDF of ϵ_{im} (which we assume the functional form as from the Type I extreme value distribution) as a function of $X'_{ij}\beta, X'_{im}\beta$ and ϵ_{ij}. Since ϵ_{ij} is unobservable, let us put it to the side for now by making the probability conditional on ϵ_{ij}. Moreover, since the error terms for all choices are independent, the CDF for the joint distribution of all $m \neq j$ is simply the product of the individual distributions. Therefore, we can get:

$$P_{ij}|\epsilon_{ij} = \prod_{m \neq j} exp(-exp(-(X'_{ij}\beta - X'_{im}\beta + \epsilon_{ij})))$$

To go from the conditional probability to the unconditional probability, we simply integrate over all values of ϵ_{ij}

$$
\begin{aligned}
P_{ij} &= \int_{-\infty}^{\infty} P_{ij}|\epsilon_{ij} f(\epsilon_{ij}) d\epsilon_{ij} \\
&= \int_{-\infty}^{\infty} \prod_{m \neq j} exp(-exp(-(X'_{ij}\beta - X'_{im}\beta + \epsilon_{ij}))) f(\epsilon_{ij}) d\epsilon_{ij} \\
&= \int_{-\infty}^{\infty} \prod_{m \neq j} exp(-exp(-(X'_{ij}\beta - X'_{im}\beta + \epsilon_{ij}))) exp(-\epsilon_{ij}) exp(-exp(-\epsilon_{ij})) d\epsilon_{ij} \\
&= \int_{-\infty}^{\infty} \prod_{m \neq j} exp(-exp(-(X'_{ij}\beta - X'_{im}\beta + \epsilon_{ij}))) exp(-\epsilon_{ij}) exp(-exp(-(\epsilon_{ij} + (X'_{ij}\beta - X'_{ij}\beta))) d\epsilon_{ij} \\
&= \int_{-\infty}^{\infty} \prod_{m \in \{1,\ldots,J\}} exp(-exp(-(X'_{ij}\beta - X'_{im}\beta + \epsilon_{ij}))) exp(-\epsilon_{ij}) d\epsilon_{ij} \\
&= \int_{-\infty}^{\infty} exp\left(-\sum_m exp(-(X'_{ij}\beta - X'_{im}\beta + \epsilon_{ij}))\right) exp(-\epsilon_{ij}) d\epsilon_{ij} \\
&= \int_{-\infty}^{\infty} exp\left(-exp(-\epsilon_{ij}) \sum_m exp(-(X'_{ij}\beta - X'_{im}\beta + \epsilon_{ij}))\right) exp(-\epsilon_{ij}) d\epsilon_{ij} \\
&= \int_{\infty}^{0} -exp(-T \sum_m exp(-(X'_{ij}\beta - X'_{im}\beta))) dT \\
&= \int_{0}^{\infty} exp(-T \sum_m exp(-(X'_{ij}\beta - X'_{im}\beta))) dT \\
&= \left[\frac{exp(-T \sum_m exp(-(X'_{ij}\beta - X'_{im}\beta))}{-\sum_m exp(-(X'_{ij}\beta - X'_{im}\beta))}\right]_0^{\infty}
\end{aligned}
$$

Note in the last few lines, I used integration by substitution where

$$T = exp(-\epsilon_{ij}) \implies \frac{dT}{d\epsilon_{ij}} = -e^{-\epsilon_{ij}} \implies dT = -e^{-\epsilon_{ij}} d\epsilon_{ij}$$

Moreover, the new bounds of integration for the substituted integration are ∞ to 0 since $\lim_{\epsilon_{ij} \to -\infty} exp(-\epsilon_{ij}) = \infty$ and $\lim_{\epsilon_{ij} \to \infty} exp(-\epsilon_{ij}) = 0$.
Note that when $T \to infty$, $P_{ij} = 0$ and when $T = 0$, $P_{ij} = \frac{-1}{\sum_m exp(-(X'_{ij}\beta - X'_{im}\beta))}$. Therefore, solving the integral from 0 to ∞, and then re-arranging gives:

$$
\begin{aligned}
P_{ij} &= \frac{1}{\sum_m exp(-(X'_{ij}\beta - X'_{im}\beta))} \\
&= \frac{1}{\sum_m exp(-X'_{ij}\beta)exp(X'_{im}\beta)} \\
&= \frac{1}{exp(-X'_{ij}\beta)\sum_m exp(X'_{im}\beta)} \\
&= \frac{exp(X'_{ij}\beta)}{\sum_m exp(X'_{im}\beta)}
\end{aligned}
$$

and we have thus shown what we set out to show.

(b) Now, with the probability function, P_{ij}, and assuming independence between individuals and choices, the likelihood function will simply be given by:

$$
L = \prod_{i=1}^{N}\prod_{j=1}^{J} P_{ij}^{y_{ij}}
$$

where y_{ij} is an indicator variable $= 1$ when individual i chooses j and $= 0$ otherwise. The log-likelihood is thus:

$$
l = lnL = \sum_{i=1}^{N}\sum_{j=1}^{J} y_{ij} ln(P_{ij})
$$

Plugging in the expression for P_{ij} found in the previous part and then simplifying gives:

$$
\begin{aligned}
l &= \sum_{i=1}^{N}\sum_{j=1}^{J} y_{ij} ln(P_{ij}) \\
&= \sum_{i=1}^{N}\sum_{j=1}^{J} y_{ij} ln(\frac{exp(X'_{ij}\beta)}{\sum_m exp(X'_{im}\beta)}) \\
&= \sum_{i=1}^{N}\sum_{j=1}^{J} \left[y_{ij} \left(X'_{ij}\beta - ln(\sum_{m=1}^{J} exp(X'_{im}\beta) \right) \right] \\
&= \sum_{i=1}^{N}\sum_{j=1}^{J} y_{ij} X'_{ij}\beta - \sum_{i=1}^{N}\sum_{j=1}^{J} y_{ij} ln(\sum_{m=1}^{J} exp(X'_{im}\beta))
\end{aligned}
$$

(c) Now, we can take the FOC by taking the first derivative with respect to β of the just derived log-likelihood function, l. The FOC can be written:

$$
\begin{aligned}
\frac{\partial l}{\partial \beta} &= \sum_{i=1}^{N}\sum_{j=1}^{J} y_{ij} X'_{ij} - \sum_{i=1}^{N}\sum_{j=1}^{J} y_{ij} \frac{\sum_{m=1}^{J} exp(X'_{im})X'_{im}}{\sum_{m=1}^{J} exp(X'_{im}\beta)} = 0 \\
&= \sum_{i=1}^{N}\sum_{j=1}^{J} y_{ij} X'_{ij} - \sum_{i=1}^{N}\sum_{j=1}^{J} y_{ij} \sum_{j=1}^{J} P_{ij} X'_{ij} = 0 \qquad (1) \\
&= \sum_{i=1}^{N}\sum_{j=1}^{J} (y_{ij} - P_{ij}) X'_{ij} = 0
\end{aligned}
$$

(d) The Independence of Irrelevant Alternatives (IIA) assumption states that the relative probability with which an individual chooses one alternative over any other alternative is affected *only* by the characteristics of those two alternatives and *not affected by* the characteristics of any other alternatives. This is implied by the expression for the ratio of the two probabilities of choosing between two alternatives, m and k:

$$\frac{P_{im}}{P_{ik}} = \frac{exp(X'_{im}\beta)}{exp(X'_{ik}\beta)}$$

which only contains arguments for X_{im}, X_{ik} (given some β). Note that this result follows directly from the expression for P_{ij} we have derived from the random utility model above (denominators cancel out and so only ratio of numerators remains). Therefore, in the random utility model we have used in this question, we are implicity assuming IIA.

The Red Bus/ Blue Bus problem is a classic example of an intuitive violation of the IIA assumption. Another example could occur in situations of voting between three or more candidates. To illustrate the example, consider a case where there's originally only two candidates, A and B, and there is a voter who is "on the fence" in deciding who to vote for. Then, suppose a third candidate, C, enters the picture who is very similar to B but not very similar to A. In this case, the voter would likely assign some of the probability with which they would have voted for B to C but not change their likelihood of voting for A by very much. Therefore, the relative probability of the individual choosing A versus B would increase, violating IIA.

1.2

(a) The two-step GMM procedure to estimate β first requires that β is a scalar since we need the number of unknowns to be strictly less than the number of moment conditions. If this is true, then we can find an initial estimate of β:

$$\hat{\beta}_{init} = \arg\min_{\beta} \left(\frac{1}{N}\sum_{i=1}^{N}\psi(x,\beta)\right)' W \left(\frac{1}{N}\sum_{i=1}^{N}\psi(x,\beta)\right)$$

where W is an arbitrary, symmetric, positive definite weighting matrix of the form

$$W = \alpha \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix}$$

and

$$\frac{1}{N}\sum_{i=1}^{N}\psi(x,\beta) = \begin{bmatrix} \frac{1}{N}\sum_{i=1}^{N} X_i \cdot (T_i - exp(\beta' X_i)) \\ \frac{1}{N}\sum_{i=1}^{N} X_i \cdot (T_i^2 - exp(2\beta' X_i)) \end{bmatrix}$$

This initial estimate $\hat{\beta}_{init}$ is consistent but not efficient and so we will use this estimate to obtain and estimate of the optimal weighting matrix which will then be used to get $\hat{\beta}_{OGMM}$, the most efficient GMM estimate of β. The estimated optimal weighting matrix is given by:

$$\widehat{S} = \frac{1}{N}\sum_{i=1}^{N} \begin{bmatrix} X_i \cdot (T_i - exp(\hat{\beta}'_{init} X_i)) \\ X_i \cdot (T_i^2 - exp(2\hat{\beta}'_{init} X_i)) \end{bmatrix} \begin{bmatrix} X_i \cdot (T_i - exp(\hat{\beta}'_{init} X_i)) \\ X_i \cdot (T_i^2 - exp(2\hat{\beta}'_{init} X_i)) \end{bmatrix}'$$

Then, plugging in this new estimate of the optimal weighting matrix, we can get the most efficient GMM estimate of β.

$$\hat{\beta}_{OGMM} = \arg\min_{\beta} \left(\frac{1}{N} \sum_{i=1}^{N} \psi(x, \beta) \right)' \widehat{S}^{-1} \left(\frac{1}{N} \sum_{i=1}^{N} \psi(x, \beta) \right)$$

(b) From the lecture, we have that $\hat{\beta}_{OGMM}$ is consistent and is asymptotically normally distributed with a variance of:

$$A.V[\hat{\beta}_{OGMM}] = \frac{1}{N}(\widehat{G}'\tilde{S}^{-1}\widehat{G})^{-1}$$

where

$$\begin{aligned}
\widehat{G} &= \frac{1}{N} \sum_{i=1}^{N} \frac{\partial \psi(x, \hat{\beta})}{\partial \beta} \Bigg|_{\hat{\beta}=\hat{\beta}_{OGMM}} \\
&= \frac{1}{N} \sum_{i=1}^{N} \begin{bmatrix} -X_i^2 exp(\hat{\beta}'_{OGMM} X_i) \\ -2X_i^2 exp(2\hat{\beta}'_{OGMM} X_i) \end{bmatrix}
\end{aligned}$$

and

$$\begin{aligned}
\tilde{S} &= \frac{1}{N} \sum_{i=1}^{N} \psi(x, \hat{\beta})\psi(x, \hat{\beta})' \Bigg|_{\hat{\beta}=\hat{\beta}_{O}GMM} \\
&= \frac{1}{N} \sum_{i=1}^{N} \begin{bmatrix} X_i \cdot (T_i - exp(\hat{\beta}'_{OGMM} X_i)) \\ X_i \cdot (T_i^2 - exp(2\hat{\beta}'_{OGMM} X_i)) \end{bmatrix} \begin{bmatrix} X_i \cdot (T_i - exp(\hat{\beta}'_{OGMM} X_i)) \\ X_i \cdot (T_i^2 - exp(2\hat{\beta}'_{OGMM} X_i)) \end{bmatrix}'
\end{aligned}$$

If we had only used the first moment condition, we would have gotten the Just Identified GMM estimate of β:

$$\hat{\beta}_{JI} = \arg\min_{\beta} \left[\frac{1}{N} \sum_{i=1}^{N} X_i \cdot (T_i - exp(\beta' X_i)) \right]' \left[\frac{1}{N} \sum_{i=1}^{N} X_i \cdot (T_i - exp(\beta' X_i)) \right]$$

Notice that this result could be achievable by manipulating the weights in W in the two-step case such that we only consider the first condition. That is, if the true optimal estimate is in fact $\hat{\beta}_{JI}$ then the two-step estimate would be equivalent. If it is not, then it must be that $\hat{\beta}_{OGMM}$ is actually a better estimate. Therefore, we can know that the variance of the two-step estimate is *at least as* small as the variance from the just identified estimate, if not smaller.

(c) Let us now assume that $T_i|X_i \sim Exponential(\beta' x)$. I.e.,

$$f(T_i = t|X_i = x) = exp(-\beta' x)exp(-t \cdot exp(-\beta' x))$$

Therefore, by probability theory, we can know

$$\mathbb{E}[T_i|X_i] = exp(\beta' X_i)$$

$$Var(T_i|X_i) = exp(2\beta' X_i)$$

Notice that these are the exact same moment conditions that we used for the GMM estimator with both conditions. Also note, the sample first moment condition will be written:

$$\frac{1}{N}\sum_{i=1}^{N} T_i = exp(-\beta' X_i) \implies exp(\beta' X_i) = \frac{N}{\sum_{i=1}^{N} T_i} = \bar{T}$$

Based on prior knowledge of statistics, we can notice that this reminds us of the MLE for the assumed exponential distribution (shown below). Let $\theta = exp(\beta' X_i)$.

$$\begin{aligned} Likelihood = L(\theta|T, X) &= \prod_{i=1}^{N} \theta exp(-T_i\theta) \\ &= \theta^N \cdot exp(-\theta \sum_{i=1}^{N} T_i) \end{aligned}$$

$$Log - Likelihood = l(\theta|T, X) = Nln(\theta) - \theta \sum_{i=1}^{N} T_i$$

We get the MLE by taking the FOC of log-likelihood function and solving:

$$\begin{aligned} \frac{\partial l(\theta|X, T)}{\partial \theta} &= \frac{N}{\theta_{ML}} - \sum_{i=1}^{N} T_i = 0 \\ \implies \theta_{ML} &= \frac{N}{\sum_{i=1}^{N} T_i} = \frac{1}{\bar{T}} \\ \implies exp(\beta' X_i) &= \frac{1}{\bar{T}} \end{aligned}$$

We can now see that the ML and MM procedures both yield the sample estimate $exp(\beta' X_i) = \frac{1}{\bar{T}}$. This is good news because the calculation of the asymptotic variance for the MLE is much easier than it is for the GMM. Note also that the variance of the MLE estimator gives us the following variance of $\theta = exp(\beta' X_i)$ by the Cramer-Rao lower bound:

$$Var(\theta) = \frac{1}{NI(\theta)}$$

where $I(\theta)$ is the Fisher Information, given by:

$$\begin{aligned} I(\theta) &= -\mathbb{E}\left[\frac{\partial^2 ln(f(T_i|X_i, \theta)}{\partial \theta^2}\right] \\ &= -\mathbb{E}\left[\frac{\partial^2 ln(\theta) - \theta T_i}{\partial \theta^2}\right] \\ &= -\mathbb{E}\left[\frac{\partial 1/\theta - T_i}{\partial \theta}\right] \\ &= -\mathbb{E}\left[\frac{-1}{\theta^2}\right] \\ &= \frac{1}{\theta^2} \end{aligned}$$

Variance is therefore given by:

$$Var(\hat{\theta}_{ML}) = Var(\bar{T}) = \frac{1}{N\theta_{ML}^2} = \frac{exp(2\beta' X_i)}{N}$$

It is worth mentioning that we know this is the most efficient estimate of the variance by theorem.

1.3

(a) The sample moment conditions that follow from the population moment conditions are:

$$\frac{1}{N}\sum_{i=1}^{N}\psi(x,y,\theta) = \begin{bmatrix} \frac{1}{N}\sum_{i=1}^{N} 3x_i - \theta \\ \frac{1}{N}\sum_{i=1}^{N} 12y_i - \theta \end{bmatrix} = 0$$

This is equivalent to writing:

$$\frac{1}{N}\sum_{i=1}^{N}\psi(x,y,\theta) = \begin{bmatrix} 3\bar{x} - \theta \\ 12\bar{y} - \theta \end{bmatrix} = 0$$

where $\bar{x}$ and $\bar{y}$ are sample means as conventional.

(b) If we set the weight matrix $W = I_2$ where I_2 is the 2×2 identity matrix, the objective function, $Q_w(\theta)$ is:

$$\begin{aligned} Q_w(\theta) &= \begin{bmatrix} 3\bar{x} - \theta \\ 12\bar{y} - \theta \end{bmatrix}' \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 3\bar{x} - \theta \\ 12\bar{y} - \theta \end{bmatrix} \\ &= (3\bar{x} - \theta)^2 + (12\bar{y} - \theta)^2 \end{aligned}$$

(c) Let us take the FOC of the objective function we have just written in order to get our GMM estimator, $\hat{\theta}_{gmm}$.

$$\begin{aligned} \frac{\partial Q_w(\theta)}{\partial \theta} &= -2(3\bar{x} - \hat{\theta}_{gmm}) - 2(12\bar{y} - \hat{\theta}_{gmm}) = 0 \\ &= -6\bar{x} - 24\bar{y} + 4\hat{\theta}_{gmm} = 0 \\ &\implies \hat{\theta}_{gmm} = \frac{3}{2}\bar{x} + 6\bar{y} \end{aligned}$$

(d) We are given the assumptions that $x \perp\!\!\!\perp y$ and $Var(x) = Var(y) = 1$. Then, the variance of our GMM estimator is:

$$
\begin{aligned}
Var(\hat{\theta}_{gmm}) &= Var(\frac{3}{2}\bar{x} + 6\bar{y}) \\
&= \frac{9}{4}Var(\bar{x}) + 36Var(\bar{y}) \\
&= \frac{9}{4}Var(\frac{1}{N}\sum_{i=1}^{N} x_i) + 36Var(\frac{1}{N}\sum_{i=1}^{N} y_i) \\
&= \frac{9}{4N^2}\sum_{i=1}^{N} Var(x) + \frac{36}{N^2}\sum_{i=1}^{N} Var(y) \\
&= \frac{9}{4N}Var(x) + \frac{36}{N}Var(y) \\
&= \frac{9}{4N} + \frac{36}{N} \\
&= \frac{38.25}{N}
\end{aligned}
$$

(e) Suppose we have the same assumptions except for instead of $W = I_2$, we now have $W = M$ where

$$M = \begin{bmatrix} c & 0 \\ 0 & 1 \end{bmatrix}$$

Then, the new objective function is $Q_w(\theta) = c(3\bar{x} - \theta)^2 + (12\bar{y} - \theta)^2$. The corresponding $\hat{\theta}_{gmm}$ can be solved for through the FOC:

$$\frac{\partial Q_w(\theta)}{\partial \theta} = 2c(3\bar{x} - \hat{\theta}_{gmm}) - 2(12\bar{y} - \hat{\theta}_{gmm}) = 0$$

Solving gives

$$\hat{\theta}_{gmm} = \frac{3c\bar{x} + 12\bar{y}}{c+1}$$

The corresponding variance (by the same calculations as in (d)) is thus

$$
\begin{aligned}
Var(\hat{\theta}_{gmm}) &= Var(\frac{3c\bar{x} + 12\bar{y}}{c+1}) \\
&= \frac{1}{(c+1)^2}(9c^2Var(\bar{x}) + 144Var(\bar{y})) \\
&= \frac{9c^2}{N(c+1)^2} + \frac{144}{N(c+1)^2}
\end{aligned}
$$

To minimize the variance, we take the FOC over c.

$$
\begin{aligned}
\frac{\partial Var(\hat{\theta}_{gmm})}{\partial c} &= 18c^* - \frac{18{c^*}^2 + 288}{c^* + 1} = 0 \\
&\implies 18c^* = 288 \\
&\implies c^* = 16
\end{aligned}
$$

When we evaluate the variance at this optimal c^*, we get:

$$Var(\hat{\theta}_{gmm})|_{c^*=16} = \frac{2304}{289N} + \frac{144}{289N} = \frac{2448}{289N} = \frac{8.47}{N}$$

The denominators in the variance from (d) and the variance from (e) are the same given any N. Meanwhile, the numerator of the variance in (d) is larger than the numerator of the variance in (e). Therefore, the variance from (e) is smaller than the variance from (d).

2 Empirical Analysis

(a) (i) Table 1 and Figure 1 present comparisons of the pre-treatment period Log Traffic Fatalities Per Capita for our treatment unit (TU) and the control states (states that did not have Primary seatbelt laws by 1986). Overall, we can see that TU had much lower traffic fatalities than the averaged control states. Additionally, from the graph, we can see that the yearly trend for TU is very similar to that of the averaged control states. This suggests that there are strong state-specific characteristics at play and also some shared shocks over time that affect all states somewhat equally. However, we do see that in 1985, the last year before treatment, TU experiences a significant increase in fatalities whereas the control states more or less maintain the previous year's level. This could suggest there was some idiosyncratic shock that occurred to the states that comprise TU which caused them to pass the Primary seat belt law which took effect in 1986. If this the case, then directly using a Diff-in-Diff model that uses the average of all control states may give a biased estimate of the treatment effect of the Primary seatbelt law (since the averaged controls no longer satisfy the parallel trend assumption).

(ii) In R, I calculated the MSE between the pre-treatment outcome trends of TU and each control state.

$$MSE = \frac{1}{5}\sum_{i=1981}^{1985} (log(fatalities\ per\ 1000)_{TU} - log(fatalities\ per\ 1000)_{Control_i})^2$$

I then select the single state with the smallest MSE to consider as a potential counterfactual state. This happens to be Florida (with an MSE = 0.00868).[1] The MSE is very low and so at first glance, it seems Florida may be a good candidate as a counterfactual state. However, after creating a balance table (Table 2), we can see that Florida and TU are generally not very balanced at all across covariates. As such, the parallel trend assumption is shaky and Florida is likely not a suitable counterfactual state with which to measure the treatment effect for TU.

(b) (i) Given our concern with directly using Florida as the counterfactual state, we may want to use a synthetic control. The benefits of using a synthetic control are that we would, by construction, obtain a strong pre-treatment trend fit and also obtain mostly balanced covariates. This makes the synthetic control a stronger candidate for being a valid counterfactual state.

[1] The next lowest MSE (=0.0124) belongs to Nevada.

However, the synthetic control method also carries a lot of limitations. Most notably, the synthetic control may fit the pre-trend data well, but that does not necessarily mean it will hold past the fitting period. We can test this to a limited extent by "training" the synthetic control on only a portion of the pre-treatment data but in cases like ours where we only have 5 pre-treatment periods in the first place, this is challenging. Some indicators that this false parallel trend may occur in our data are that there are states that introduce Secondary seat-belt laws but not Primary seat-belt laws and there are states that introduce Primary seat-belt laws later than 1986. If these states get included with positive weight in the synthetic control, then the estimated treatment effect may be biased (especially farther away from 1986). Similarly, if any of the states included in the synthetic control experience spillover effects from the Primary seatbelt laws passed in TU states, then the results may be biased.

Lastly, another drawback is that the synthetic control method cannot directly be used for statistical inference. Instead, we must go through a relatively cumbersome process to use non-parametric means such as the Placebo Test to check for statistical significance.

(ii) Using Abadie et al.'s (2010) method in R, I ran several specifications using different methods of optimization and different sets of predictor variables. Note that the program in R can run multiple methods of optimization in a single command and will return the results with the lowest MSPE (mean squared prediction error). The default methods are "Nelson-Mead" and "BFGS". I run my first specification using these default optimization methods on the full set of predictor variables ("college","beer","secondary","population","unemploy","totalvmt","precip","snow"). The command uses these predictors and the set of control states to search for the optimal combination of weights that minimizes the MSPE. From here, I try running the synthetic control command using the other available optimization methods and then narrowing it down to the method that yields the lowest MSPE ("nlm"); I use this optimization method for the rest of the problem. Next, I begin varying the predictors over which the command optimizes. First, I omit Precipitation because in the full-covariate case, it had a very low weight. However, omitting precipitation actually significantly increases the MSPE and so I avoid omitting Precipitation in my preferred specification. Next, I remove the Secondary seat-belt law predictor since in the period over which I am optimizing (1981 - 1985), almost no states differ from TU and so it is easy to think that this variable may not matter much. The resulting MSPE is slightly higher than the full specification and so I keep this predictor in the optimization process for the rest of the problem. Then, I remove the Population variable from the predictors because the outcome variable is already normalized on population and therefore there may be little correlation between total population size and log fatalities *per capita*. This removal actually lowers the MSPE very slightly and so I keep this change. I then check this predictor specification using the original, default optimization settings out of curiousity and confirm that the "nlm" method still yields a lower MSPE.

I present a summary table (Table 3) detailing the MSPEs of each mentioned specification. Table 4 shows the balance between the TU and final Synthetic Control. It suggests that they are reasonably well balanced on observables. Table 5 gives the weights assigned to each predictor variable. Table 6 shows the positive weights assigned to control states: states with weights < 0.005 are omitted from the table. Finally, Figure 2 shows the actual

TU and Synthetic Control trends in the outcome variable over the whole duration of the data.

Table 1: Averaged Control and TU Log Traffic Fatalities Per Capita (1981-1985)

	Control States	Treatment
Mean Log Fatalities Per Capita	-1.61	-1.34

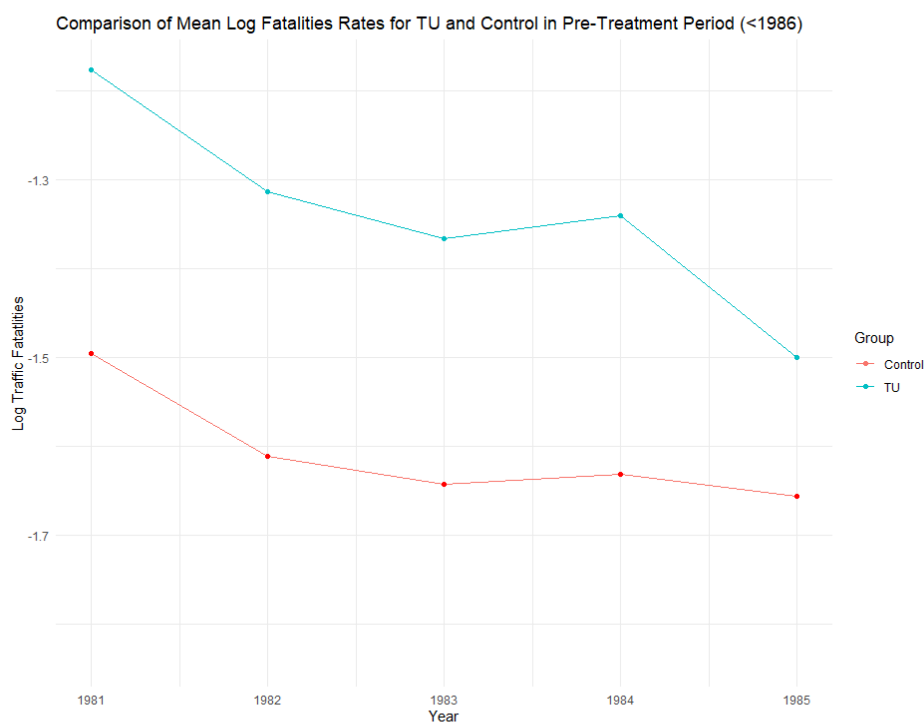

Figure 1: Comparison of Log Traffic Fatalities (Per Capita) in Years 1981 - 1985

Table 2: Balance Table of Covariates for Florida (8) and TU (99)

	8 (N=5)	99 (N=5)	Total (N=10)	p value
Pct College Grads (3 Yr MA)				0.004
Mean (SD)	0.259 (0.023)	0.315 (0.022)	0.287 (0.036)	
Range	0.234 - 0.285	0.290 - 0.343	0.234 - 0.343	
Per Capita Beer Consumption (gal of ethanol)				0.145
Mean (SD)	1.548 (0.136)	1.663 (0.084)	1.606 (0.123)	
Range	1.398 - 1.677	1.564 - 1.762	1.398 - 1.762	
Secondary Seat Belt Law				NaN
Mean (SD)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	
Range	0.000 - 0.000	0.000 - 0.000	0.000 - 0.000	
Population (thousands)				0.091
Mean (SD)	10794.002 (520.447)	11395.581 (466.662)	11094.791 (563.649)	
Range	10017.944 - 11298.433	10737.806 - 12009.348	10017.944 - 12009.348	
Unemployment Rate				0.106
Mean (SD)	7.458 (1.201)	6.218 (0.937)	6.838 (1.207)	
Range	6.300 - 8.900	5.083 - 7.366	5.083 - 8.900	
Vehicle Miles Traveled (VMT)				0.003
Mean (SD)	84224.800 (5089.060)	97499.800 (5021.574)	90862.300 (8465.768)	
Range	76145.000 - 88970.000	90286.000 - 103112.000	76145.000 - 103112.000	
Precipitation (inches)				< 0.001
Mean (SD)	5.163 (0.743)	2.914 (0.154)	4.038 (1.289)	
Range	4.472 - 6.164	2.718 - 3.062	2.718 - 6.164	
Snow (inches)				< 0.001
Mean (SD)	0.000 (0.000)	0.173 (0.060)	0.087 (0.100)	
Range	0.000 - 0.000	0.115 - 0.270	0.000 - 0.270	
Rural Interstate Speed Limit				0.347
Mean (SD)	55.000 (0.000)	55.000 (0.000)	55.000 (0.000)	
Range	55.000 - 55.000	55.000 - 55.000	55.000 - 55.000	
Urban Interstate Speed Limit				0.347
Mean (SD)	55.000 (0.000)	55.000 (0.000)	55.000 (0.000)	
Range	55.000 - 55.000	55.000 - 55.000	55.000 - 55.000	

	Specification	MSPE
A	Basic: Full	0.0011639
B	Alt Optimization: Full	0.00090423
C	Alt Opt: No Precip	0.0020368
D	Alt Opt: No Secondary	0.00090565
E	**Alt Opt: No Population**	**0.00090372**
F	Basic: No Population	0.00104

Table 3: MSPE of Different Specifications. Row E is "Preferred" specification.

	Treated	Synthetic	Sample Mean
College	0.31505	0.32603	0.28633
Beer	1.66350	1.55802	1.37695
Secondary	0.00000	0.00000	0.00476
Unemployment	6.21804	6.53837	8.27235
Vehicle Miles Travelled	97499.80000	77415.69736	32229.91905
Precipitation	2.91424	2.29018	3.60526
Snow	0.17345	0.19145	0.36456

Table 4: Balance Table of Predictor Variables

	Weights
College	0.02513
Beer	0.09276
Secondary	0.05768
Unemployment	0.41875
Vehicle Miles Travelled	0.02972
Precipitation	0.00012
Snow	0.37585

Table 5: Predictor Variable Weights

Weights	State	State Code
0.39929	WY	48
0.38126	CA	4
0.17940	NE	27
0.04002	NH	28

Table 6: Control State Weights (> 0.005)

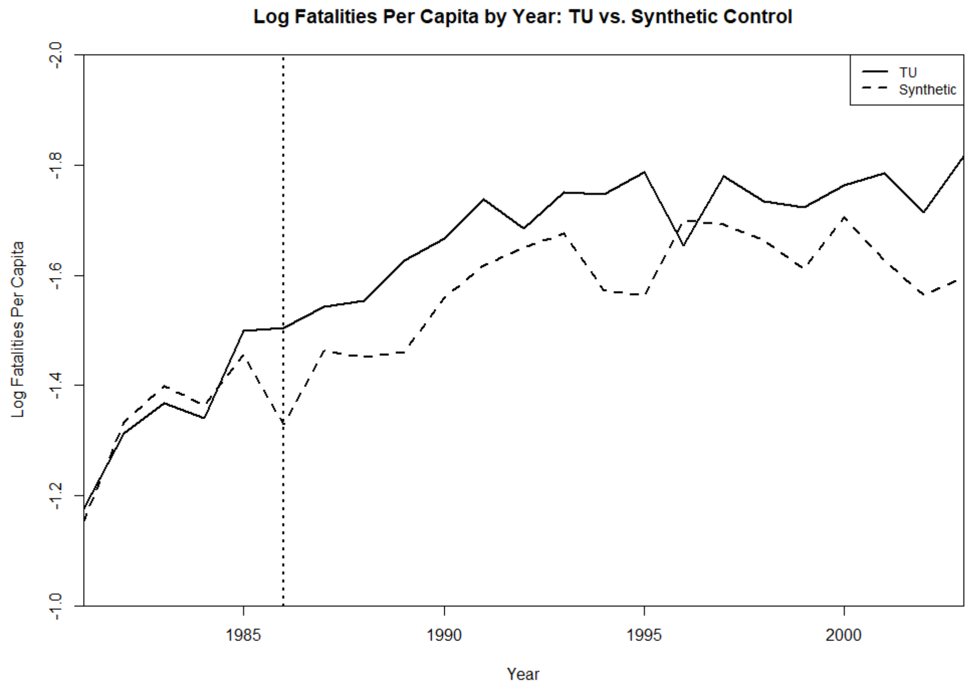

Figure 2: Log Fatalities Per Capita (1981 - 2003) of TU and Synthetic Control

(c) (i) Here, I present several plots of the gap between the TU and Synthetic Control over time using various specifications.

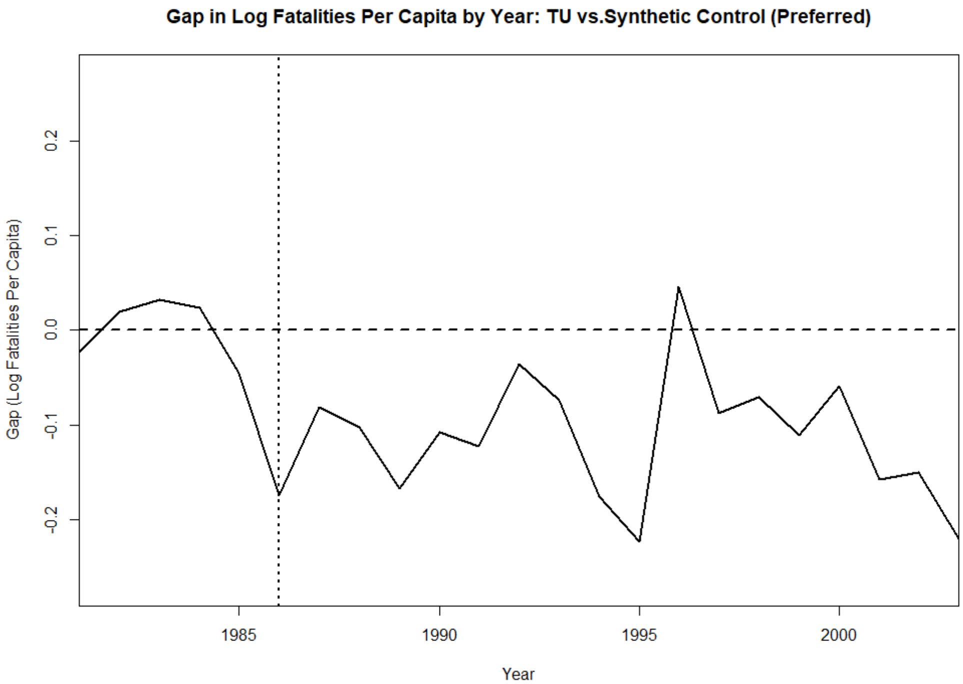

Figure 3: Preferred Specification (Optimization Method = “nlm”; Predictors = All but Population

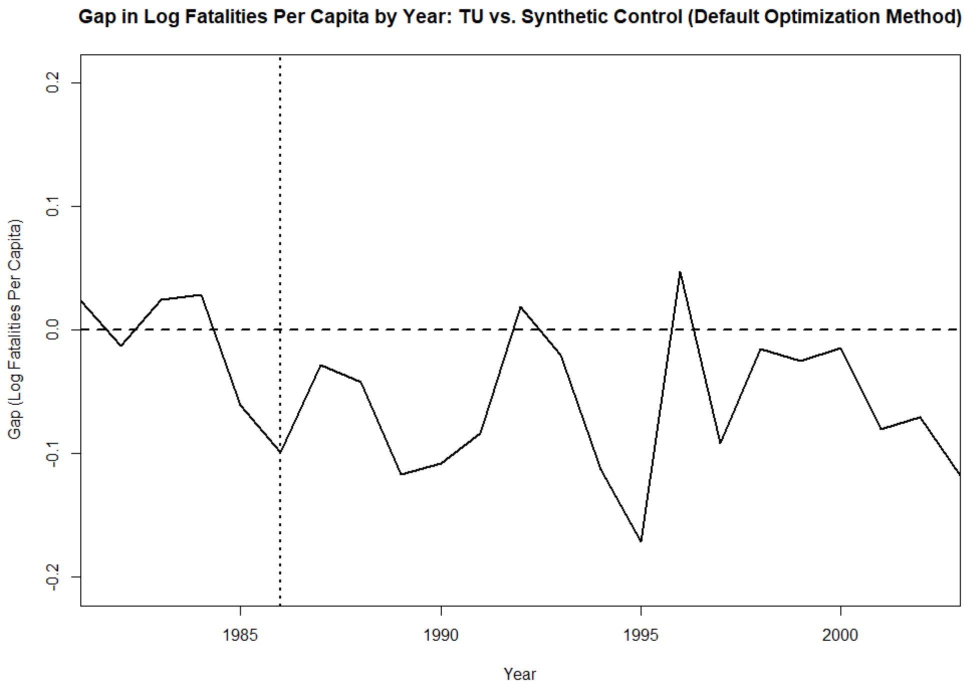

Figure 4: Default Full Specification (Optimization Method = "Nelson-Mead" & "BFGS"; Predictors = All)

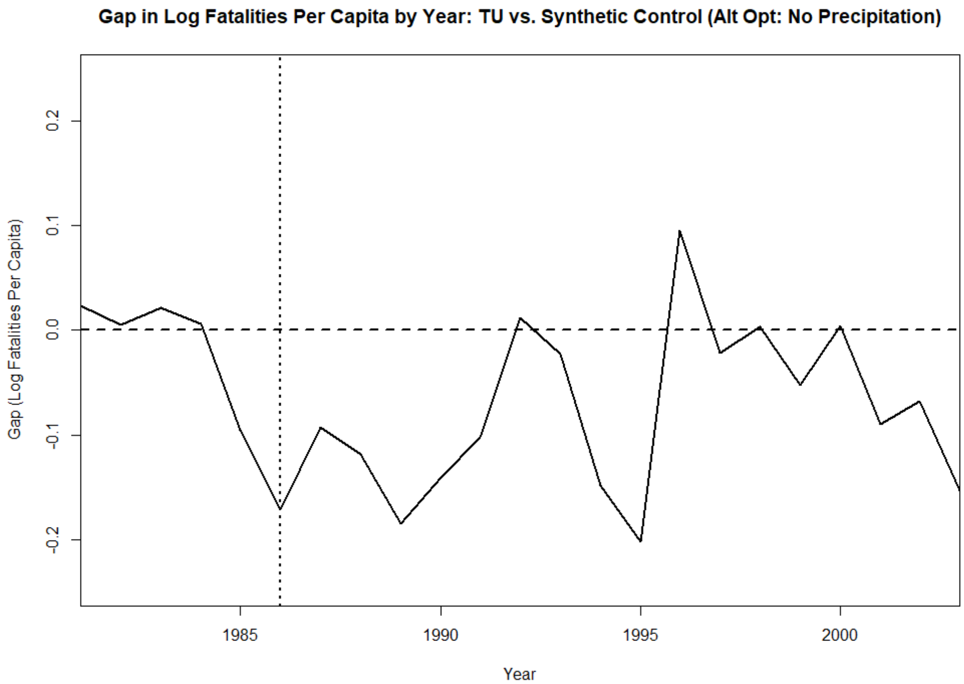

Figure 5: “nlm” Optimization with All Predictors but Precipitation

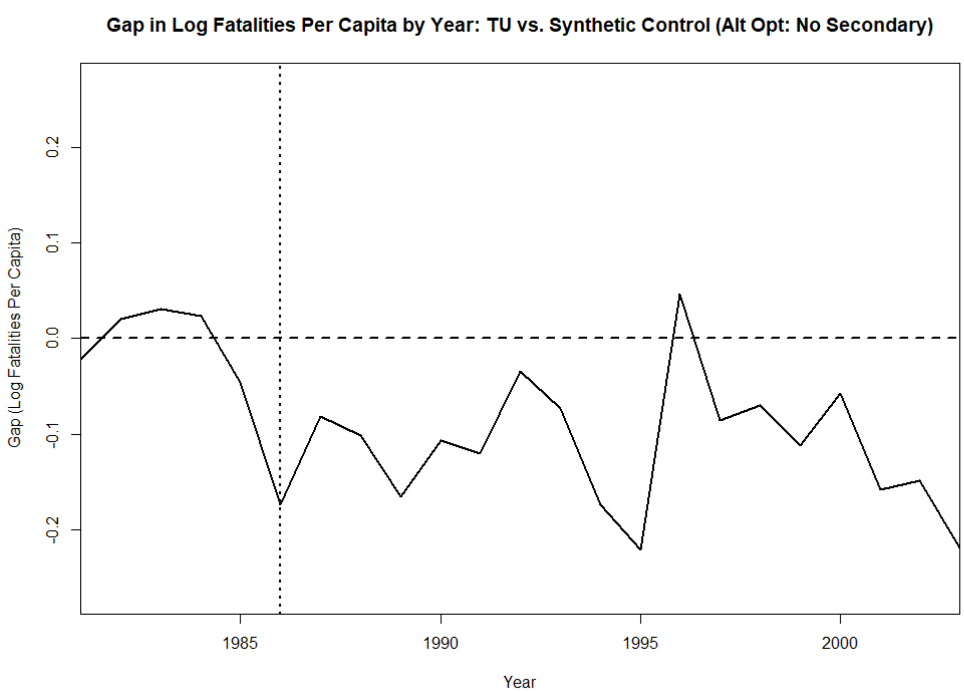

Figure 6: “nlm” Optimization with All Predictors but Secondary Seat-belt Laws

(ii) Finally, Figure 7 visualizes the placebo test conducted by constructing a synthetic control for each control state using the same specification we used for TU and then plotting the gap between the actual values and Synthetic Control values for that state each year. The figure suggests that the treatment was not significant. We can see that there are many control states, that we know *did not* receive treatment, that experienced larger gaps between the actual and predicted values. This suggests that there is a high chance the gaps found for the TU are generated by chance rather than measurement of causal effect. Additionally, our concerns from the discussion in part (b) remain. The general noisiness of the gap plot (i.e. many of the control lines do not have mean zero) implies that we are generally not able to reliably predict the control (counter-factual) trend. This is likely due to some states joining treatment past 1986 and heterogeneity in Secondary seat-belt laws. Therefore, we cannot definitively say whether the treatment had any real effect or not but based on the analysis we have conducted thus far, we absolutely can not reject the null hypothesis that it had no effect.

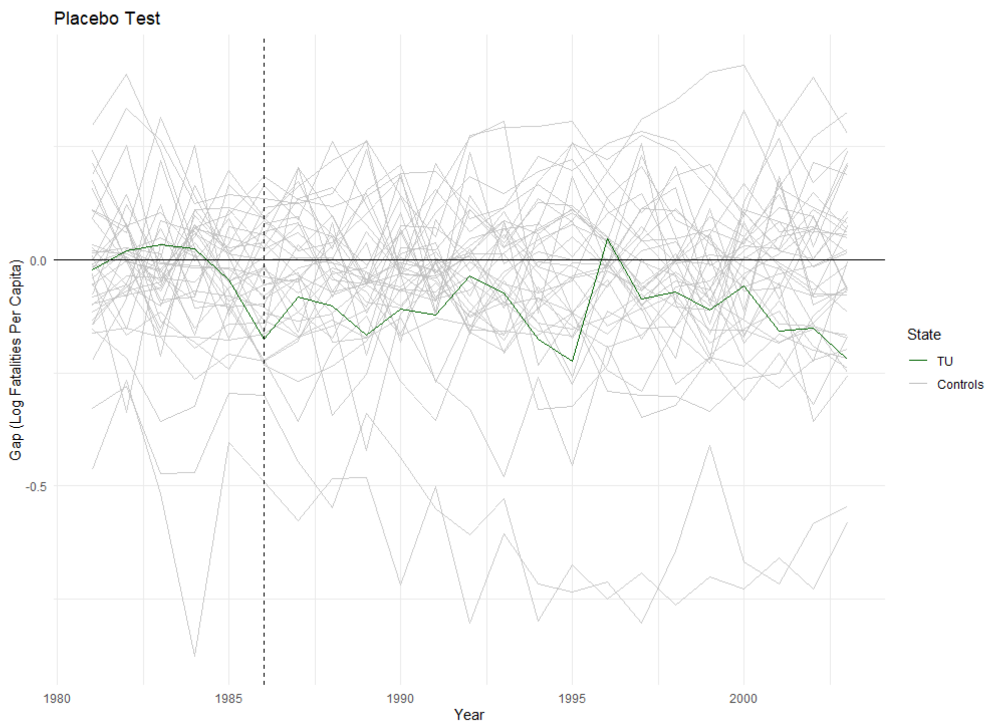

Figure 7: Gaps between actual and synthetically predicted values for TU and most control states. UT, WA, WI, WV, WY are excluded because the selected specification produces errors when running on these states.