émicos/Cursos
Dic-
ta-
dos/PUCP/QLab/Clase/"pd1enc.dfu

émicos/Cursos
Dic-
ta-
dos/PUCP/QLab/Clase/"rerunfilecheck.cfg

Synthetic Control Methods: Theory and application

Juan Manuel del Pozo Segura

The University of Sussex

j.del-pozo@sussex.ac.uk

PUCP Q Lab - Jan. 2020

Outline

1. Current developments
   - How reasonable is the Convex Hull condition?
   - The problem with the Cross Validation method for V^*
   - How to correctly specify the model and avoid cherry picking

2. Empirical application
   - The intervention in Abadie et al . (2010)
   - The command synth and examples
   - Replication of the paper

Outline

1. Current developments
   - How reasonable is the Convex Hull condition?
   - The problem with the Cross Validation method for V^*
   - How to correctly specify the model and avoid cherry picking

2. Empirical application
   - The intervention in Abadie et al . (2010)
   - The command synth and examples
   - Replication of the paper

Ferman & Pinto (2019): only stationary common factors

- Analyze the properties of the SC when the pre-treatment fit is imperfect, i.e. when *the CH assumption is violated* CH assump.
- When $T_0 \to \infty$, when the pre-treatment fit is imperfect in a model with $I(0)$ $\boldsymbol{\lambda}_t$, SC weights converge in probability to weights that *do not match the factor loadings* $\boldsymbol{\mu}$ *of* $j = 1$
- Intuition: if treatment assignment is correlated with **common factors** $\boldsymbol{\lambda}_t$ *in the post-treatment periods*, then *we would need a SC unit which*
  - *did not receive the treatment*
  - *is affected in exactly the same way by these* **common factors** $\boldsymbol{\lambda}_t$ *as the* $j = 1$

  this would be attained with *weights that reconstruct the factor loadings* $(\boldsymbol{\mu}_j)$ *of* $j = 1$. However, when the pre-treatment fit is imperfect
  - SC weights do *not* converge to weights that satisfy this condition
  - the distribution of the SC estimator will still depend on the $\boldsymbol{\lambda}_t$, *implying a biased estimator when selection depends on these* $\boldsymbol{\lambda}_t$
- They propose a **modified SC estimator** which demeans the data using information from the preintervention period

$$\hat{\alpha}_{0t}^{SC'} = y_{0t} - y_t' \hat{w}^{SC'} - \left(\overline{y}_0 - \overline{y}' \hat{w}^{SC'} \right)$$

under $E(\boldsymbol{\lambda}_t | D(0,0) = 1) = \omega_0$ for $t > 0$ and stability conditions in the pre- and

Ferman & Pinto (2019): $I(0)$ and $I(1)$ common factors

- So far, λ_t is only stationary; what if we also have $\gamma_t \sim I(1)$ such that treatment assignment can be correlated with $\boldsymbol{\lambda}_t$ *and* $\boldsymbol{\gamma}_t$ in the post-intervention?
- Under the *additional* assumption on the **existence of weights** that reconstruct the $\boldsymbol{\mu}$ of unit 1 associated with the $\gamma_t \sim I(1)$, then *the asymptotic distribution $T_0 \to \infty$ of the **demeaned** SC estimator does not depend on the $I(1)$ common trends*
- Intuition: demeaned SC weights will converge to weights that reconstruct the $\boldsymbol{\mu}$s of $j = 1$ associated with the $I(1)$ $\boldsymbol{\lambda}_t$s. Then
  - $I(1)$ common factors will *not* lead to asymptotic bias
  - We only need caring about correlation between treatment and $I(0)$ $\boldsymbol{\lambda}_t$s
- We need checking **detrended** pre-treatment fit (eliminating $I(1)$ trends)
  - This implies subtracting from the outcome of the treated *and* control units the average of the *control units* only at t, $a_t = \frac{1}{J}\sum_{j\neq 1} y_{jt}$
  - This is indicative of potential bias from possible correlation between treatment assignment and $I(0)$ common factor
  - If pre-treatment fit with these series is *not as good as the original*, for a finite T_0, *a mismatch in $\boldsymbol{\mu}$s associated with $I(0)$ common factors might be relevant*

Outline

1. Current developments
   - How reasonable is the Convex Hull condition?
   - The problem with the Cross Validation method for V^*
   - How to correctly specify the model and avoid cherry picking

2. Empirical application
   - The intervention in Abadie et al . (2010)
   - The command synth and examples
   - Replication of the paper

Klößner, S. et al. (2015)

- Abadie et al. 2015 state that a way to find V^* consists in a Cross Validation approach.
  - We separate the pre-treatment period in a training and validation period
  - Using data from the training period we take any given V and find a preliminar $\check{W}$
  - Using data from the validation period we use this $\check{W}$ to find the V^*
  - Using data from the validation periood we use this last V^* to find the W^*
- Yet, Klößner, S. et al. (2015) show that this method is not recommended for empirical practice
  - The reason for this is the non injectivity of the $\check{W}$ in the training period : the same $\check{W}$ which leads to very different V^*s
  - This implies that we end up with different W^* depending on factors such as the ordering of the data!!
- Hence, when finding the V we need to rely on the regression based method or the nested approach

Outline

1. Current developments
   - How reasonable is the Convex Hull condition?
   - The problem with the Cross Validation method for V^*
   - How to correctly specify the model and avoid cherry picking

2. Empirical application
   - The intervention in Abadie et al . (2010)
   - The command synth and examples
   - Replication of the paper

Kaul et al. (2018)

1. If in$\boldsymbol{X}$ we use *all pre-treatment* Y along with Zs, then these Zs become irrelevant and the W^*will be calculated as if optimizes considering only Y
   - I.e. we get a SC based on
     - using economically meaningless covariates, or
     - not using covariates
   - This hold true regardless of using the RM or the nested approach to find V^*
2. Theoretically, doing this introduces a trade off
   2.1 we get an additional **small-sample bias** to the estimation***:*** this is likely to be significant, especially when effect of the $\boldsymbol{Z}$s is large
   2.2 we take more care of unobserved confounders by fitting with respect to lagged outcomes alone
3. In MC experiment, they find that using all lags of the outcome in $\boldsymbol{X}$
   3.1 effectively ignores the covariates which leads to small-sample bias when estimating the outcome's counterfactual development
   3.2 makes estimates less precise in terms of RMSE compared to those which effectively use the covariates

Ferman et al. (2019)

- Still, theory rarely tells us about what covariates to include in X
  - If different models result in different choices of the SC unit, *then a researcher would have relevant opportunities to select "statistically significant" specifications even when there is no effect.*
  - This flexibility may undermine the main advantage of the SC method, as it *implies some discretionary power for the researcher to construct the counterfactual for the treated unit*
- Ferman et al. show that the SC method *and* results using p-values in Abadie et al. (2010) are robust to specification searching only if
  1. T_0 is large
  2. we use models whose number of pre-treatment outcome lags go to ∞ with T_0

  otherwise, specification searching is a problem.
- They recommend
  1. Focusing on the specification that uses **all** the pre-treatment outcome lags unless there is a strong belief that it is *crucial* to *also* balance on specific Zs
  2. Presenting results for different specifications; e.g. $X_j = \left(Y_{j,1}, ..., Y_{j,\frac{3T_0}{4}}\right)'$ or

     $X_j = \left(Y_{j,1}, ..., Y_{j,\frac{T_0}{2}}\right)'$

Outline

1. Current developments
   - How reasonable is the Convex Hull condition?
   - The problem with the Cross Validation method for V^*
   - How to correctly specify the model and avoid cherry picking

2. Empirical application
   - The intervention in Abadie et al . (2010)
   - The command synth and examples
   - Replication of the paper

The Proposition 99

- In 1988, the **Proposition 99 in California** was the first modern-time large-scale tobacco control program in USA. This
  - Triggered a wave of local clean-air ordinances in California
  - Launched a new wave of state and federal anti-tobacco laws
  - Resulted in the tobacco industry response increase in its political activity in California at both the state and local levels.
- *Proposition 99 was widely perceived to have successfully cut smoking in California*. From the passage of Proposition 99 through 1999
  1. adult smoking prevalence fell in California by more than 30%,
  2. youth smoking levels dropped to the lowest in the country
  3. per capita cigarette consumption more than halved (California Department of Health Services 2006).
- Following early reports of California's success with Proposition 99, other states adopted similar policies.
  - As of April 20, 2009, 30 states, the DC, and 792 municipalities had laws requiring 100% smoke-free workplaces, bars, or restaurants (ANRF 2009).

Data and variables setting

- Proposition 99 went into effect in January 1989

$$\underbrace{1970, \dots, 1988}_{\text{Pre-treatment}}; \underbrace{1989, \dots, 2000}_{\text{Post-treatment}}$$

  and so
  - we have 19 years of pre-intervention data.
  - we only take 10 year of post-intervention because at about this time anti-tobacco measures were implemented across many states, invalidating them as potential control units.
- Outcome is *cigsale*, annual per capita cigarette consumption at the state level, measured as per capita cigarette *sales* (in packs)
- Predictors of *cigsale* for California and the *j*s in DP take information *only on preintervention period* Matrices
  1. $\overline{Y}^{K_1}, \dots, \overline{Y}^{K_M}$: lagged (observed) smoking consumption in 1975, 1980, and 1988
  2. $\boldsymbol{Z}$: averaged over the 1980–1988 period:
     - 2.1 average retail price of cigarettes
     - 2.2 per capita state personal income (logged)
     - 2.3 percentage of the population age 15–24
     - 2.4 per capita beer consumption

Choosing the Donor Pool

- The SC (California) is constructed as a *weighted average* of js in the DP with weights chosen so that this best reproduces
  1. the values of a set of predictors of cigarette consumption in California in the pre-treatment period (before passage of Proposition 99)
  2. the outcome (*cigsale*) that would have been observed for California in absence of Proposition 99
- Hence, we need to *discard* from the DP those states that
  1. adopted some other large-scale anti tobacco program during any year: MA, AZ, OR, and FL
  2. raised cigarette taxes by 50c. or more over post-treatment period: AK, HI, MD, MI, NJ, NY, WA
  3. DC

  So, the DP includes the remaining 38 states ($J+1=38+1$)
- The treatment effect estimate obtained for California would be **attenuated** if *any* of the states in the DP that gets a weight in the SC increased unilaterally taxes that reduced smoking

The work plan

- Using the techniques described in Section 2, we
  1. Construct a SC for California that mirrors the values of the predictors of cigsale in California before the passage of Proposition 99.
  2. Estimate the effect of Proposition 99 on cigsale as the difference in cigarette consumption levels between California and its synthetic versions in the post-intervention years (after Proposition 99 was passed) following

$$\hat{\alpha}_{1t} = Y_{1t} - \sum_{j=2}^{J+1} w_j^* Y_{jt}; t \in \{T_0 + 1, ..., T\} \text{ where } j = 2, ..., 38 \text{ and } 1\text{=California}$$

  3. Perform **placebo studies** that confirm that our estimated effects for California are unusually large relative to the distribution of the estimate that we obtain when we apply the same analysis to the states in the DP

Outline

1. Current developments
   - How reasonable is the Convex Hull condition?
   - The problem with the Cross Validation method for V^*
   - How to correctly specify the model and avoid cherry picking

2. Empirical application
   - The intervention in Abadie et al . (2010)
   - The command synth and examples
   - Replication of the paper

synth: basic syntax and mandatory options

synth *depvar predictorvars*, ***trunit(#) trperiod(#)*** *[counit(numlist) xperiod(numlist) mspeperiod() resultsperiod() nested allopt unitnames(varname) figure keep(file) customV(numlist) optsettings]*

- Dataset must be a **balanced** panel after `xtset panelvar timevar`
  - *depvar* is the outcome variable.
  - *predictorvars* is the list of predictor variables.
- Mandatory options
  - *trunit(#)*: the unit *number* of the $j = 1$ as given in *panel id variable specified in* `xtset`
    - Only 1 number can be specified.
    - If the intervention affected several *js* we *combine these* and *then* treat them as 1
  - *trperiod(#)*: the time *period* when intervention occurred $(T_0 + 1)$ as given in *panel time variable* specified in `xtset`. Only 1 number can be specified.
- By default, *all predictor variables are averaged over the entire pre-intervention period*, which ranges
  - from the period earliest available in the panel time variable
  - to the period ***immediately prior*** to the beginning of the intervention (T_0)

`synth`: optional options for handling Xs

- *counit(numlist):* a list of *j*s in the DP the control units as given in the panel id variable
  - Should contain at least 2 *j*s in the DP
  - If *no* counit() is specified, the DP defaults to *all units available in the panel id variable* excluding the unit specified in *trunit()*
- *xperiod(numlist)*: a list of periods over which the predictor variables specified in predictorvars are averaged as given in the panel time variable
  - E.g. *xperiod(1980(1)1988)* indicates that predictor variables are averaged over all years from 1980, 1981,...,1988.
  - If *no* xperiod() is specified, *xperiod()* defaults to the entire pre-intervention period
  - Missing data points in the variables are ignored in the X matrix
- *figure*: produces a line plot with outcome trends for $j = 1$ and the SC for the years specified in *resultsperiod()*

`synth`: optional options for optimization

- *nested:*
  - `synth` uses a **data-driven regression based method** to obtain the variable weights contained in the V-matrix, which is *not* covered in Abadie and Gardeazabal [2003], Abadie et al. [2010] and Abadie et al. [2015].
    - This relies on a **constrained quadratic programming routine** that finds the best fitting W-weights conditional on the regression based V-matrix.
    - This procedure is fast and yields good results in terms of minimizing the MSPE
  - *nested* will lead to better performance using additional computing time
    - `synth` will embark on an optimization procedure that *searches among all (diagonal) positive semidefinite V-matrices and sets of W-weights for the best fitting convex combination of the control units*.
    - This takes as starting point the regression based V and produces convex combinations that achieve even lower
- *allopt*:
  - If nested is specified we can *also* specify *allopt* if we want to check that the minimum found is not a local one
  - This is done by running nested optimization 3 times using 3 different starting points (regression based V, equal V-weights, and using Stata's `ml search`) and returning the best of these 3.
  - This will take 3 times the amount of computing time compared to *nested*

Example 1

```
synth cigsale cigsale(1988) cigsale(1980) cigsale(1975)
beer(1984(1)1988) lnincome retprice age15to24, trunit(3)
trperiod(1989)
```

- Parameters
  - *trunit(3)* : the unit affected by the intervention ($j = 1$) is unit 3 (California). Since no *counit()* is specified, the DP defaults to the other 38 states in the dataset: $j = 1, 2, 4, ..., 39$
  - *trperiod(1989)* : the *first* year of the treatment ($T_0 + 1 = 1989$)
- Model: what we include in X ▸ X matrices
  - *cigsale(1988) cigsale(1980) cigsale(1975)* : the observed values of cigsale in those 3 years (Y)
  - *beer(1984(1)1988)*: the average valueof beer taking *only* 1984, 1985, 1986, 1987 and 1988
  - *lnincome retprice age15to24*: the average value of lnincome, retprice and age15to24 for the whole pre-treatment period (1970 to 1988) since no *xperiod()* is provided

Example 2

```
synth cigsale cigsale(1988 1980 1975) lnincome(1980 1985)
beer retprice age15to24, trunit(3) trperiod(1989) fig
```

- Parameters
  - *trunit(3)* : the unit affected by the intervention ($j = 1$) is unit 3 (California). Since no *counit()* is specified, the DP defaults to the other 38 states in the dataset: $j = 1, 2, 4, ..., 39$
  - *trperiod(1989)* : the *first* year of the treatment ($T_0 + 1 = 1989$)
- Model: what we include in X ▸ X matrices
  - *cigsale(1988 1980 1975)*: the *average value* of *cigsale* taking only 1998, 1980 and 1975. This is DIFFERENT from *cigsale(1988) cigsale(1980) cigsale(1975)*
  - *lnincome(1980 1985)*: the *average value* of *lnincome* taking only 1980 and 1985. This is DIFFERENT from *lnincome(1980) lnincome(1985)*
  - *beer*: the *average value* of *beer* for the whole pre-treatment period (1970 to 1988). But since *beer* has missings pre-1984, synth will inform about this for this variable and these missings are ignored in the averaging
  - *retprice age15to24*: the *average value* of *retprice* and *age15to24* for the whole pre-treatment period (1970 to 1988) since no *xperiod()* is provided

Example 3

```
synth cigsale cigsale(1970 1979) retprice age15to24,
trunit(33) counit(1(1)20) trperiod(1980)
resultsperiod(1970(1)1990) fig
```

- Parameters
  - *trunit(33)*: the unit affected by the intervention ($j = 1$) is unit no 33
  - *counit(1(1)20)*: since this is specified, the DP is restricted to *only* states no 1,2,...,20. Note how the treated state does not appear here
  - *trperiod(1989)*: the first year of the treatment
  - *resultsperiod(1970(1)1990)*: results are obtained for 1970,1971,...,1990
- Model: what we include in X ▸ X matrices
  - *cigsale(1970 1979)*: the *average value* of cigsale taking only 1970 and 1979. This is DIFFERENT from *cigsale(1970) cigsale(1979)*
  - *retprice age15to24*: the *average value* of retprice and age15to24 for the whole pre-treatment period (1970 to 1988) since no *xperiod()* is provided
- Note the equivalence of this with

```
keep if inrange(state, 1, 20) | state==33
synth cigsale cigsale(1970 1979) retprice age15to24, trunit(33) trperiod(1980)
resultsperiod(1970(1)1990) fig
```

Example 4

```
ssynth cigsale cigsale(1975 1988 1980) beer lnincome
retprice age15to24, trunit(3) trperiod(1989)
xperiod(1980(1)1988) nested
```

- Parameters
  - *trunit(3)* : the unit affected by the intervention ($j = 1$) is unit 3 (California). Since no *counit()* is specified, the DP defaults to the other 38 states in the dataset: $j = 1, 2, 4, ..., 39$
  - *trperiod(1989)* : the year of the treatment ($T_0 + 1 = 1989$)
- Model: what we include in X
  - *cigsale(1975 1988 1980)*: the *average value* of *cigsale* taking only 1998, 1980 and 1975. This is DIFFERENT from *cigsale(1988) cigsale(1980) cigsale(1975)*
  - *beer lnincome retprice age15to24*: average of beer (obviating its missings) lnincome, retprice and age15to24 for the whole pre-treatment period (1970 to 1988) since NO *xperiod()* is provided.
- By specifying *nested*, synth will do an optimize searching *among all* (diagonal) positive semidefinite V-matrices and sets of W-weights for the best fitting convex combination of the control units.

Outline

1. Current developments
   - How reasonable is the Convex Hull condition?
   - The problem with the Cross Validation method for V^*
   - How to correctly specify the model and avoid cherry picking

2. Empirical application
   - The intervention in Abadie et al . (2010)
   - The command synth and examples
   - Replication of the paper

Figure 1

- *Figure 1. Trends in per-capita cigarette sales: California vs. the rest of the United States.*
  - This shows how *bad* a simple average of cigsales for the 38 js in the DP approximates cigsales for California
- So we need
  1. for *every year*, collapse into 1 the cigsales for the 38 js in the DP
  2. plot it along the observed series for California

The command `synth` for this study

- To do what follows, we need to run the `synth` command to find $\boldsymbol{W}^*$

```
synth cigsale lnincome age15to24 retprice beer(1984(1)1988)
cigsale(1988) cigsale(1980) cigsale(1975), trunit(3)
trperiod(1989) xperiod(1980(1)1988)
resultsperiod(1970(1)2000) nested
```

where

- Zs
  - *beer(1984(1)1988)*: takes the average value of beer only between those years because this *has missings* before
  - *lnincome age15to24 retprice*: takes the average value of these variables over *only* 1980-1988 because we specified *xperiod(1980(1)1988)* which means that we no longer take the *whole pre-treatment period* (1970-1988).
- $\overline{Y}^{K_1}, \ldots, \overline{Y}^{K_M}$*: cigsale(1975) cigsale(1980) cigsale(1988)*: takes the values of cigsale for onlu those years
- *resultsperiod(1970(1)2000)*: the latest year is because by 2001 many states had adopted similar policies

Table 2

- *Table 2. State weights in the synthetic California*
  - This is the main result of the SCM estimator: the SC weights
  - This is conveniently stored as a return matrix after the estimation of the command
    `matrix list e(W_weights)`
- Note how there are 38 entities in the DP
  - Yet, there are only 5 entities (states) who receive (positive) weights
  - This is a consequence of the 2 restrictions imposed on W, which yields **sparse synthetic controls**
  - In fact, this is a powerful feature of SCM since, precisely, this sparsity allows for easiness of interpretation and analysis

Table 1

- *Table 1. Cigarette sales predictor means*
  - This shows how *good* the weighted average of the $\boldsymbol{X}s$ using the $w_1^*, w_2^*, w_4^* ..., w_{39}^*$ for the 38 js in the DP approximates $\boldsymbol{X}s$ for California
  - This is important to do before we calculate the SC (using $\sum_{j=2}^{J+1} w_j^* Y_{jt}$)
- So

1. for column 1: for California only (so only 1 entity)
   1.1 simple average over 1980-1988 for *lnincome*, *age15to24*, *retprice*
   1.2 simple average over 1984-1988 for *beer*
   1.3 observed values over 1975, 1980 and 1988 for *cigsales*
2. for column 2: for the DP only (so 38 entities)
   2.1 weighted average (using the $\boldsymbol{W}^*$) over 1980-1988 for *lnincome*, *age15to24*, *retprice*
   2.2 weighted average (using the $\boldsymbol{W}^*$) over 1984-1988 for *beer*
   2.3 weighted average (using the $\boldsymbol{W}^*$) of observed values over 1975, 1980 and 1988 for *cigsales*
3. for column 3: same as column 2 (for the DP only so 38 entities) but taking simple averages

Figure 2

- *Figure 2. Trends in per-capita cigarette sales: California vs. synthetic California*
  - This is where we plot the observed series for California (between 1970 and 2000) and the calculated SC using, for any $t = 1970, \ldots, 2000$:

$$\underbrace{\sum_{j=2}^{J+1} w_j^* Y_{jt}}_{\text{SC for California}} = \begin{pmatrix} 0 & 0 & \underbrace{0.160}_{\text{CO}} & \underbrace{0.068}_{\text{CT}} & 0 & \ldots & \underbrace{0.202}_{\text{MT}} & 0 & \underbrace{0.236}_{\text{NV}} & \ldots & 0 & \underbrace{0.335}_{\text{UT}} & \ldots & 0 \end{pmatrix}_{1\times 38}$$

$$\times \begin{pmatrix} Y_{1t} & Y_{2t} & Y_{4t} & Y_{5t} & Y_{6t} & \ldots & Y_{19t} & Y_{20t} & Y_{21t} & \ldots & Y_{33t} & Y_{34t} & \ldots & Y_{38t} \end{pmatrix}'_{38\times 1}$$

  - These are stored in

```
matrix list e(Y_treated)
matrix list e(Y_synthetic)
```

- We can calculate `e(Y_synthetic)`manually
  - Let's keep Y_{j1970} for all the js in the DP

```
keep if state!=3 & year==1970
br state year cigsale
```

  - Now let's generate a variable *weights*, with 0 for all the states other than those 5 which have positive weights in W^*
  - Multiply this variable *weights* with *cigsale*
  - Sum the values: this is the value to e(Y_synthetic) for 1970!

Figure 3

- *Figure 3. Per-capita cigarette sales gap between California and synthetic California*
  - This is where we use our TE estimator $\hat{\alpha}_{1t}$ defined above, which equals the observed series for California minus the calculated SC using, for any $t = 1970, ..., 2000$:

$$\boxed{\hat{\alpha}_{1t} = \underbrace{Y_{1t}}_{\text{observed California}} - \underbrace{\sum_{j=2}^{J+1} w_j^* Y_{jt}}_{\text{SC for California}} \; ; t \in \{T_0 + 1, ..., T\}}$$

  where $1 = \textit{California}$

  - We have to create this manually (no big deal)
- We can calculate the TE manually

```
gen te=cigsale_cal - cigsale_scm
```

and then we plot this

Figure 4

- *Figure x. Per-capita cigarette sales gaps in California and placebo gaps in all 38 control states*
  - This is where we take the permutation tests (in-space placebos) to see the exact significance of our estimates
  - Figure 4 plots *all* the 38 permutations + the estimate for California (so there are 38 + 1 lines). This provides the "raw material" for the next 3 graphs
- This requires looping. For a given *j* (to simplify the coding we also include here California even though we already found its SC estimate)

  1. re-run the `synth` command we ran at the beginning but specifying `trunit(`*j*`)`
  2. save separately `e(Y_treated)`, `e(Y_synthetic)` and `e(RMSPE)`for this *j*
  3. accumulate these along with those of the previous *j*

  We will end up with a matrix, we save into stata via `svmat` and then we graph it

Figure 5, 6 and 7

- *Figure x. Per-capita cigarette sales gaps in California and placebo gaps in Z control states (discards states with pre-Proposition 99 MSPE X times higher than California's).*
  - Figure 5, 6 and 7 drop lines (states) with increasingly large MSPEs relative to that of California
  - This is nothing but figure 4 without some lines, those above a given threshold
- So,

1. We need the `e(RMSPE)`$=RMSPE_j^{\boxed{pre}}$ stored before to create

$$ratio = \frac{RMSPE_j^{pre}}{RMSPE_{\boxed{1}}^{pre}} = \frac{\sqrt{\left(\frac{1}{T_0}\sum_{t=1}^{T_0}\left(Y_{jt} - \sum_{j=1}^{J+1} w_k^* Y_{kt}\right)^2\right)}}{\sqrt{\left(\frac{1}{T_0}\sum_{t=1}^{T_0}\left(Y_{1t} - \sum_{j=2}^{J+1} w_j^* Y_{jt}\right)^2\right)}} \texttt{ with } k = 1, ..., J+1 \texttt{ without the current } j$$

   where $1 = California$. This will provide with 1 unique value of the ratio for the series of every state (38+1 unique values), equal to 1 *only* for California
2. Based on this we filter the dataset: we drop those series with a value of *ratio* larger than a threshold
   - 2.1 $ratio > 20$
   - 2.2 $ratio > 5$
   - 2.3 $ratio > 2$

Figure 8

- *Figure 8. Ratio of post-Proposition 99 MSPE and pre-Proposition 99 MSPE: California and 38 control states*
  - A final way to evaluate the California TE relative to the TEs of the placebo runs is looking at the distribution of *ratios* of post/pre- RMSPE.
  - The main advantage of looking at **ratios** is that it obviates choosing a cutoff for the exclusion of ill-fitting placebo runs.
- So,
  1. We use `e(RMSPE)`$=RMSPE_j^{pre}$ and *also* to calculate $RMSPE_j^{post}$

$$\frac{RMSPE_j^{post}}{RMSPE_j^{pre}} = \frac{\sqrt{\left(\frac{1}{T-T_0}\sum_{t=T_0+t}^{T}\left(Y_{jt}-\sum_{k=1}^{J+1}w_k^* Y_{kt}\right)^2\right)}}{\sqrt{\left(\frac{1}{T_0}\sum_{t=1}^{T_0}\left(Y_{jt}-\sum_{k=1}^{J+1}w_k^* Y_{kt}\right)^2\right)}} \texttt{ with } k=1,...,J+1 \texttt{ without the current } j$$

     This will provide with 1 unique value of the ratio for the series of every state (39 unique values), equal to 1 only for California
  2. Based on this we graph the distribution of the post/pre-Proposition 99 ratios of the MSPE for California and all 38 control states.
  3. If we were to assign the intervention at random in the data, the probability of obtaining a post/pre-Proposition 99 MSPE ratio as large as California's is

$$\frac{1}{39} = 0.026 < 0.05$$

END

- Thanks for your time :)

Example 1 matrices

- The matrix for $(X_1)_{(r+M=)7\times 1}$ has both $(Z_1)_{(r=)4\times 1}$ and $\left(\overline{Y}_1^{K_1},\ldots,\overline{Y}_1^{K_m}\right)_{(M=)3\times 1}$

	3	
lnincome	10.03	Average over 1970-88
age15to24	0.18	Average over 1970-88
retprice	66.64	Average over 1970-88
beer(1984(1)1988)	24.28	Average over 1984-88
cigsale(1975)	127.10	Observed
cigsale(1980)	120.20	Observed
cigsale(1988)	90.10	Observed

$$X_1 = \begin{pmatrix} Z_1 \\ \hline \overline{Y}_1^{K_1} \\ \ldots \\ \overline{Y}_1^{K_M} \end{pmatrix}_{k\times 1=(r+M)\times 1}$$

- The matrix for $(X_0)_{(r+M=)7\times 38}$ has both $(Z_0)_{(r=)4\times 38}$ and $\left(\overline{Y}_0^{K_1},\ldots,\overline{Y}_0^{K_m}\right)_{(M=)3\times 38}$

	1	2	4		38	39	
lnincome	9.63	9.61	9.93	...	9.85	9.90	Average over 1970-88
age15to24	0.18	0.17	0.18	...	0.18	0.18	Average over 1970-88
retprice	66.99	67.69	60.39	...	69.88	59.38	Average over 1970-88
beer(1984(1)1988)	18.96	18.52	25.08	...	32.04	24.98	Average over 1984-88
cigsale(1975)	111.70	114.80	131.00	...	113.50	160.70	Observed
cigsale(1980)	123.20	131.80	131.00	...	117.60	158.10	Observed
cigsale(1988)	112.10	121.50	94.60	...	102.60	114.30	Observed

$$X_0 = \begin{pmatrix} Z_2 & Z_3 & \ldots & Z_{J+1} \\ \hline \overline{Y}_2^{K_1} & \overline{Y}_3^{K_1} & \ldots & \overline{Y}_{J+1}^{K_1} \\ \ldots & \ldots & & \ldots \\ \overline{Y}_2^{K_M} & \overline{Y}_3^{K_M} & \ldots & \overline{Y}_{J+1}^{K_M} \end{pmatrix}_{k\times J=(r+M}$$

Back

Example 2 matrices

- The matrix for $(X_1)_{(r+M=)5\times 1}$ has *only* $(Z_1)_{(r=)5\times 1}$

	3	
cigsale(1988 1980 1975)	112.27	Average over 1988, 1980 and 1975
lnincome	10.03	Average over 1970-88
age15to24	0.18	Average over 1970-88
retprice	66.64	Average over 1970-88
beer(1984(1)1988)	24.28	Average over 1984-88

$$\boldsymbol{X}_1 = (\boldsymbol{Z}_1)_{k\times 1 = r\times 1}$$

- The matrix for $(X_0)_{(r+M=)5\times 38}$ has *only* $(Z_0)_{(r=)5\times 38}$

	1	2	4		38	39	
cigsale(1988 1980 1975)	115.67	122.70	118.87		111.23	144.37	Average over 1988, 1980 and 1975
lnincome	9.63	9.61	9.93	...	9.85	9.90	Average over 1970-88
age15to24	0.18	0.17	0.18	...	0.18	0.18	Average over 1970-88
retprice	66.99	67.69	60.39	...	69.88	59.38	Average over 1970-88
beer(1984(1)1988)	18.96	18.52	25.08	...	32.04	24.98	Average over 1984-88

$$\boldsymbol{X}_0 = (\boldsymbol{Z}_0)_{k\times 1 = r\times 1}$$

Back

Example 3 matrices

- The matrix for $(X_1)_{(r+M=)3\times 1}$ has *only* $(Z_1)_{(r=)3\times 1}$

	33	
cigsale(1970 1979)	129.10	Average over 1970 and 1979
age15to24	0.18	Average over 1970-88
retprice	66.64	Average over 1970-88

$$\boldsymbol{X}_1 = (\boldsymbol{Z}_1)_{k\times 1 = r\times 1}$$

- The matrix for $(X_0)_{(r+M=)3\times 20}$ has *only* $(Z_0)_{(r=)3\times 20}$

	1	2	3		19	20	
cigsale(1970 1979)	113.80	111.55	135.81	...	131.50	125.90	Average over 1970 and 1979 and 1975
age15to24	0.18	0.17	0.18	...	0.17	0.17	Average over 1970-88
retprice	66.99	67.69	60.39	...	63.36	67.92	Average over 1970-88

$$\boldsymbol{X}_0 = (\boldsymbol{Z}_0)_{k\times 1 = r\times 1}$$

Back