



MAESTRÍA EN ECONOMÍA

Economía Aplicada

PROF. MARTIN A. ROSSI

TUTORES: PAOLA LLAMAS Y TOMÁS
PACHECO

Problem Set 8: Control Sintético

Garcia Ojeda, Juan
Hausvirth, Martina
Hayduk, Gaspar
Salvatierra, Elias Lucas D.

Fecha de entrega: 8 de Noviembre de 2024

PROBLEM SET 8: CONTROL SINTÉTICO

GARCÍA OJEDA - HAUSVIRTH - HAYDUK - SALVATIERRA

Este problem set se basa en el artículo “Evaluating the Effect of Homicide Prevention Strategies in São Paulo, Brazil: A Synthetic Control Approach” de Danilo Freire (Latin America Research Review, 2018). En la carpeta tienen los códigos de R para reproducir los resultados del paper. La consigna es replicar todos los gráficos (excepto el 8) en Stata con los comandos vistos en clase. Nota: no se preocupen si los resultados de los sintéticos no son iguales a los del paper. Las diferencias se dan porque los algoritmos de optimización son ligeramente distintos.

INFORME

En su trabajo, Freire evalúa el impacto de las políticas anti-delincuencia (mayores intensidades de penas, control de armas, aumento de las tasas de encarcelamiento, sentencias más severas a condenados, etc.) implementadas a partir de 1999 en el estado de São Paulo sobre las tasas de homicidio cada 100.000 habitantes utilizando un enfoque de control sintético. Este método es una técnica econométrica que permite comparar el resultado de una unidad (en este caso, São Paulo) con una combinación ponderada de otras unidades que no fueron afectadas por la política, creando un “grupo de control” sintético. El método de control sintético busca crear un contrafactual; compara la evolución de las tasas de homicidio en São Paulo con la evolución de un grupo de control sintético compuesto por otras regiones del país que no adoptaron políticas similares. El control sintético busca generar una serie temporal de homicidios en un “Estado contrafáctico” que represente lo que podría haber ocurrido si no se hubiera implementado la política. El autor parte desde el hecho de que São Paulo tuvo una reducción del 70% en la tasa de homicidio entre 1999 y 2009.

La Figura 1 presenta la evolución de la tasa de homicidio de São Paulo en comparación al promedio brasileño (excluyendo a SP). Podemos observar que la tasas de homicidio en São Paulo presentan una reducción a partir de 1999, mientras que en el resto de Brasil se sigue una tendencia creciente.

Para la construcción del São Paulo sintético, el autor utiliza los siguientes predictores: State GDP per Capita, State GDP Growth, Years of Schooling, Gini Index, Natural Logarithm of Population y Population Living in Extreme Poverty. La Figura 2 presenta la evolución de la tasa de homicidio para São Paulo y el São Paulo Sintético. Podemos observar que ambas trayectorias tienen una tendencia similar a la alza entre 1990 y 1999; pero que a partir de ese momento empiezan a diverger, con São

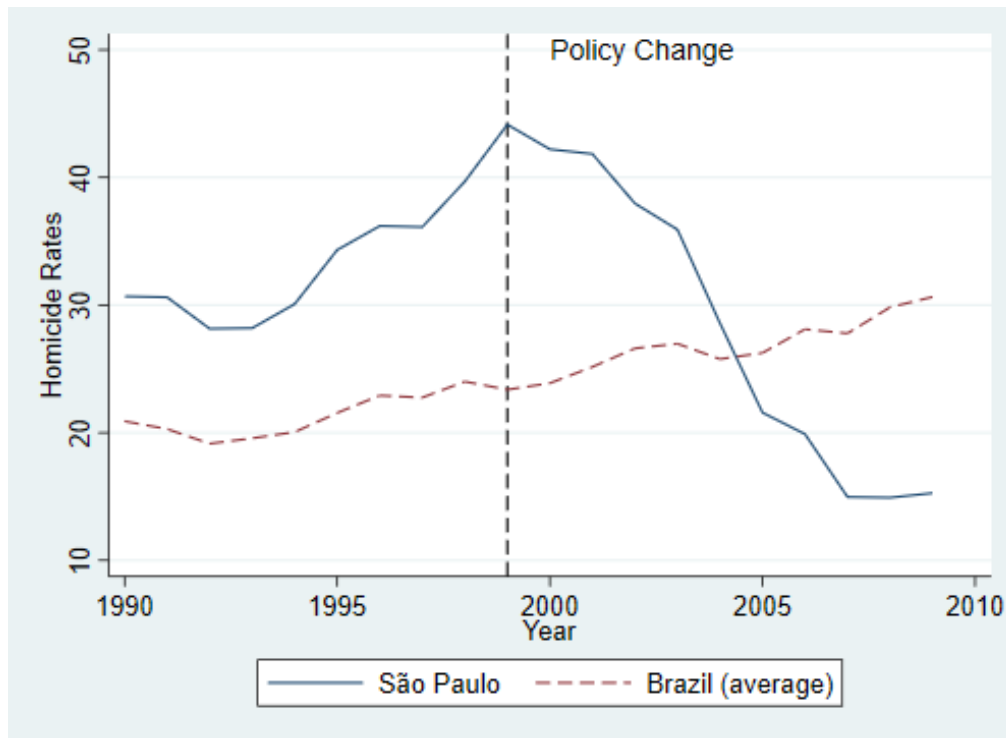


FIGURE 1. Homicide rates per 100,000 population: São Paulo and Brazil (excluding the state of São Paulo).

Paulo experimentando una fuerte reducción en la tendencia de su tasa de homicidio en comparación al sintético.

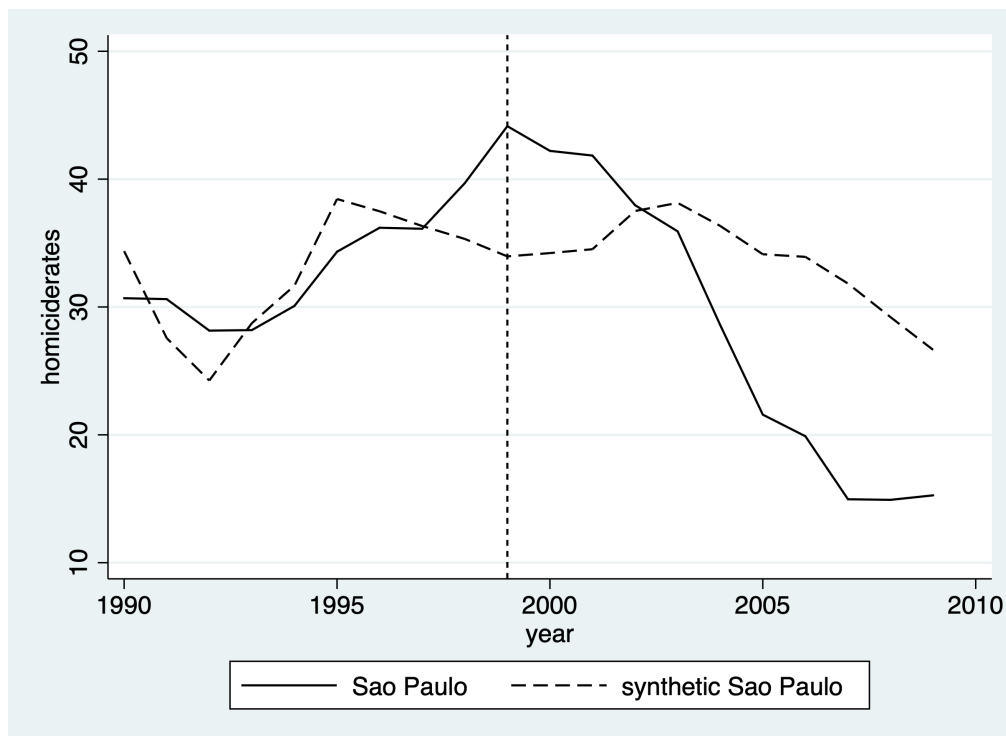


FIGURE 2. Trends in homicide rates: São Paulo versus synthetic São Paulo.

La Figura 3 nos permite hacer zoom sobre el gráfico anterior ya que ilustra el gap en las tasas de homicidio entre São Paulo y el São Paulo Sintético. Podemos ver que el gap aumenta considerablemente a partir de 1999.

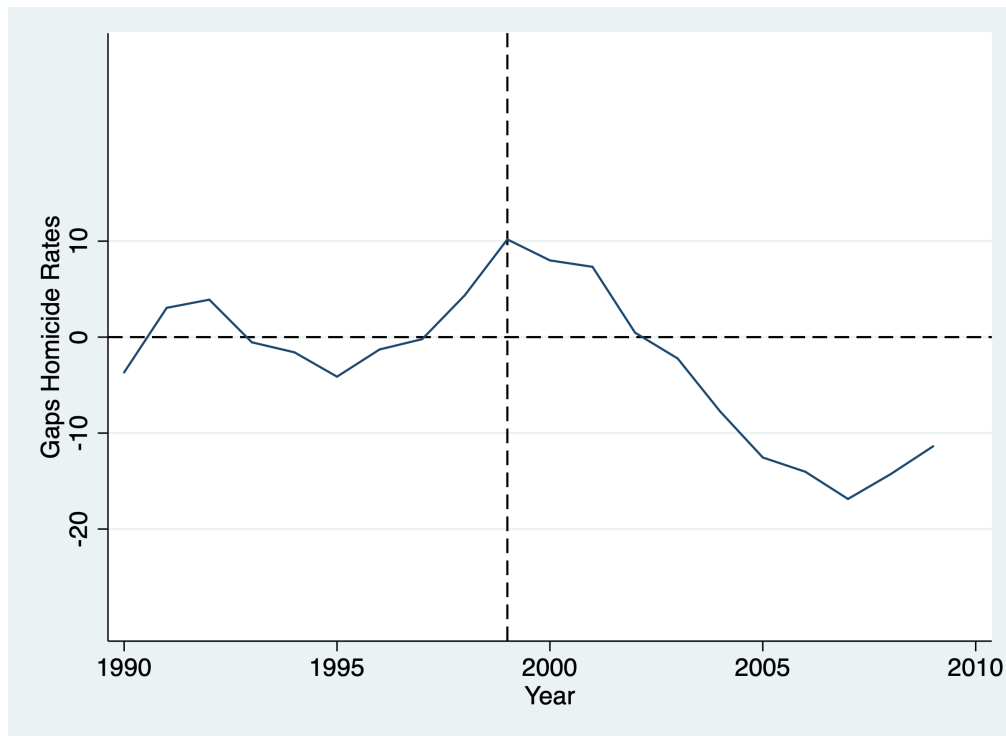


FIGURE 3. Homicide rates gap between São Paulo and synthetic São Paulo.

Para que los resultados presentados sean creíbles, el autor realiza una serie de Tests de Robustez. En primer lugar, realiza el "in-time placebo" synthetic control. En esta prueba, se crea una falsa fecha de intervención para la unidad tratada para ver si hay falsos efectos de la intervención en los años pre-tratamiento. La Figura 4 ilustra los resultados de esta prueba, donde se selecciona al año 1994 como año donde comienzan las políticas anti-delincuencia en São Paulo. Podemos observar que el método no genera quiebres entre la evolución de São Paulo y el São Paulo Sintético a partir de 1994.

En segundo lugar, el autor también realiza el leave-one-out test. En este test, el autor estima nuevos São Paulo Sintéticos pero sacando de a uno a los estados que componen el São Paulo Sintético original. El objetivo de este test es determinar si hay un estado driving the results; si esto llega a ocurrir es porque probablemente el São Paulo Sintético original no es un buen contrafactual de São Paulo. La Figura 5 muestra los resultados de este test. Podemos observar que ningún estado del donor pool está sesgando los resultados y la mayor diferencia entre São Paulo y São Paulo Sintético se da cuando no se restringe ningún estado.

Por último, para verificar si existe algún factor nacional o regional no observado que pueda explicar los resultados de las Figuras 2 y 3, el autor realiza dos test de permutación donde le aplica al tratamientos a los demás estados y para cada estado estima un correspondiente sintético. La

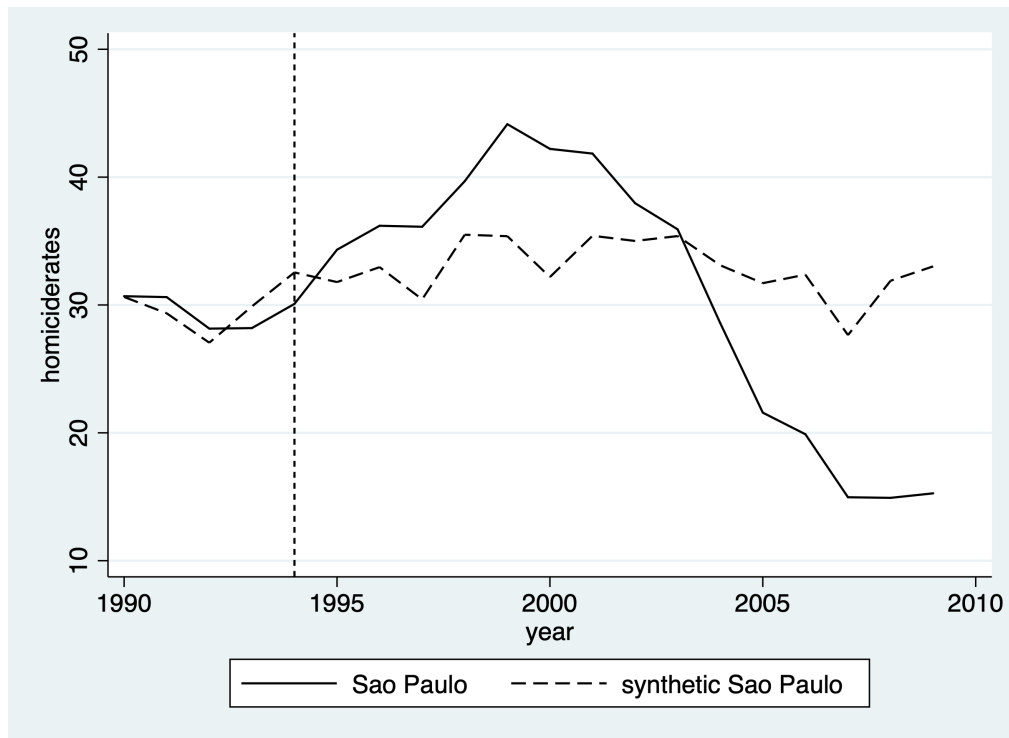


FIGURE 4. Placebo policy implementation in 1994: São Paulo versus synthetic São Paulo.

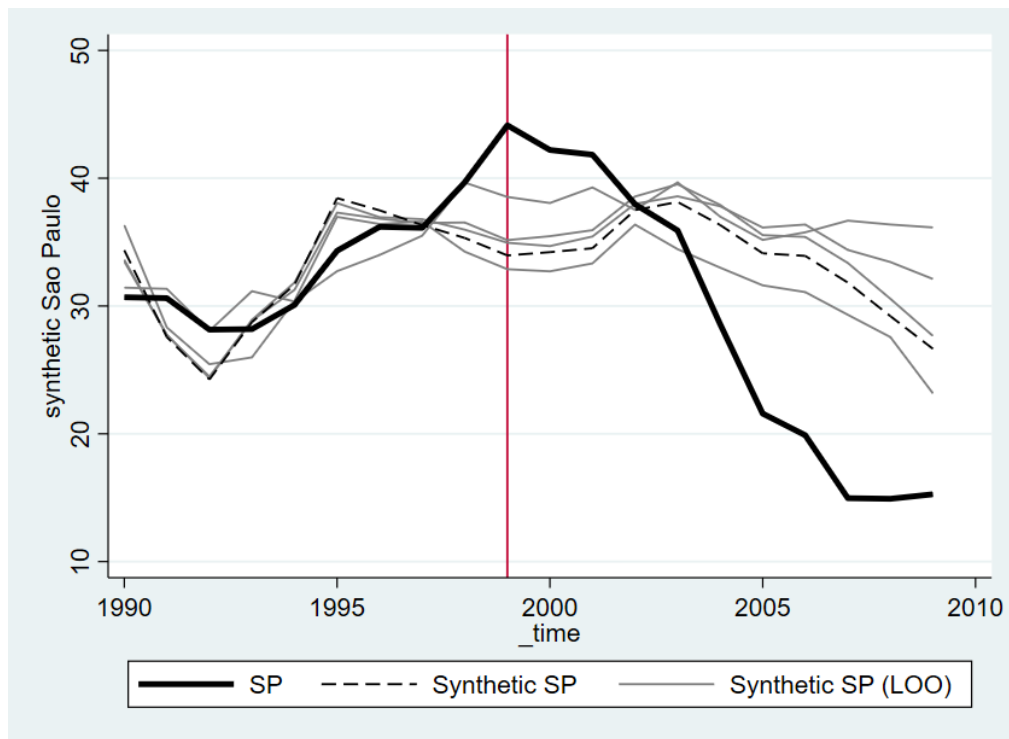


FIGURE 5. Leave-one-out distribution of the synthetic control for São Paulo.

Figura 6 muestra el gap de la tasa de homicidio para cada estado y su correspondiente sintético; la línea roja representa el gap entre São Paulo y el São Paulo Sintético. Si no existen factores

nacionales o regionales no observados que puedan explicar los resultados, entonces el gap para São Paulo y el São Paulo Sintético debería ser el mayor y los demás gaps deberían moverse en direcciones aleatorias (algunos para arriba y otros para abajo). Podemos observar que solo existe un gap que supera al de São Paulo y que sí ocurre que los demás gaps se mueven en direcciones aleatorias.



FIGURE 6. Permutation test: Homicide rate gaps in São Paulo and twenty-six control states.

La Figura 7 presenta lo mismo que la Figura 6, pero solo muestra aquellos sintéticos cuyo Mean Squared Prediction Error es no mayor a 2 veces el Mean Squared Prediction Error de São Paulo; es decir, solo se muestran placebos que tienen un buen sintético en términos de predicción. Podemos observar que el gap entre São Paulo y el São Paulo Sintético (línea roja) es el más relevante, proporcionándonos mejor evidencia sobre los resultados previos.



FIGURE 7. Permutation test: Homicide rate gaps in São Paulo and selected control states.

```

1  /*****
   *****/
2
3          Problem Set 8
4          Control sintético
5          Universidad de San Andrés
6          Economía Aplicada
7  *****/
8  * Gaspar Hayduk; Juan Gabriel García Ojeda; Elias Lucas
9  Salvatierra; Martina Hausvirth
10 *****/
11
12 * 0) Set up environment
13
14 *=====
15 =====*
16
17 global main "/Users/gasparhayduk/Desktop/Economía
18 Aplicada/ConsignasPS8"
19 global output "$main/output"
20 global input "$main/input"
21
22 cd "$output"
23
24 * Importo el dataset. Usamos la base df.csv
25 import delimited "$input/df.csv", clear
26
27 * Las variable state es categórica string y encima con errores
28 de escritura, la pasamos a código
29 encode state, generate(state2)
30 drop state
31 rename state2 state
32
33 * Arreglo los labels que se importaron mal
34 label define state2 3 "Amapá" 6 "Ceará" 8 "Espírito Santo" 9
35 "Goiás" 10 "Maranhao" 15 "Paraíba" 16 "Pará" 14 "Paraná" 18
36 "Piauí" 22 "Rondonia" 26 "Sao Paulo", modify
37 label values state state2
38
39 * Aclarar a Stata que es un panel
40 tsset state year
41
42 * Guardamos la base
43 save "df.dta", replace
44
45 * 1) Réplica graficos del paper de Freire
46
47 * ---- FIGURA 1: Homicide rates per 100,000 population: São
48 Paulo and Brazil (excluding the state of São Paulo) ---- >

```



```

r paulo and Brazil (excluding the state of Sao Paulo)---- ↑
42 * La primer figura es simplemente la evolución histórica de la
    tasa de homicidios por 100k habitantes en San Pablo en relación
    con el promedio. La aplicacion de la politica fue en 1999
43
44 * Obtenemos el promedio de homicidios para el resto de Estados
    de Brasil
45 bysort year: egen hr_promedio_resto = mean(homiciderates) if
    state != 26
46
47 * Me quedo con los datos solo de Sao Paulo
48 gen hr_sp = .
49 replace hr_sp = homiciderates if state == 26
50
51 * Armamos el gráfico
52 twoway (line hr_sp year, lwidth(medium) lpattern(solid)) ///
53       (line hr_promedio_resto year, lwidth(medium) lpattern(dash
    )), ///
54       ytitle("Homicide Rates") xtitle("Year") ///
55       legend(order(1 "São Paulo" 2 "Brazil (average)") pos(6))
    ///
56       xline(1999, lpattern(dash) lcolor(black)) ///
57       text(50 2000 "Policy Change", place(e) size(medium))
58
59 * Exporto el grafico
60 graph export "$output/figura1.png", replace
61
62
63 * ---- FIGURA 2: Trends in homicide rates: São Paulo versus
    synthetic São Paulo.---- *
64 /* La segunda figura surge luego de la construcción del Sao
    Paulo sintético. Lo que realiza es la evolución del sintético y
    de Sao Paulo.
65 Las variables de control que utiliza son: stategdpcapita
    yearsschoolingimp populationprojectionln giniimp
    populationextremepovertyimp stategdpgrowthpercent */
66
67 * Instalar el comando
68 *ssc install synth
69
70 * Estimamos el sintético
71 synth homiciderates stategdpcapita yearsschoolingimp
    populationprojectionln giniimp populationextremepovertyimp
    stategdpgrowthpercent homiciderates(1990(1)1998), trunit(26)
    trperiod(1999) nested fig
72
73 * Exporto el grafico
74 graph export "$output/figura2.png", replace
75
76 * ---- FIGURA 3: Homicide rates gap between São Paulo and
    synthetic São Paulo.---- *
77 /* Grafico con la misma info del 2, sólo que considera el gap

```

```

pretratamiento entre Sao Paulo y el sintético.
78 Hay que volver a cargar la base de datos y correr de nuevo el
synth porque sino se grafica mal (no termino de entender bien
por que pasa).*/
79
80 * Cargamos la base de datos
81 use "$output/df.dta", clear
82 tsset state year
83
84 * Estimamos el sintético
85 synth homiciderates stategdpcapita yearsschoolingimp
populationprojectionln giniimp populationextremepovertyimp
stategdpgrowthpercent homiciderates(1990(1)1998), trunit(26)
trperiod(1999) nested
86
87 * Obtenemos el gap
88 matrix gaps=e(Y_treated) -e(Y_synthetic)
89 matrix Y_treated=e(Y_treated)
90 matrix Y_synthetic=e(Y_synthetic)
91 keep year
92 svmat gaps
93 svmat Y_treated
94 svmat Y_synthetic
95 *svmat hace que las columnas de las matrices se vuelvan
variables.
96
97 * Graficamos
98 twoway (line gaps1 year, lwidth(medium) lpattern(solid)), ///
99 ytitle("Gaps Homicide Rates") xtitle("Year") ///
100 xline(1999, lpattern(dash) lcolor(black)) ///
101 yline(0, lpattern(dash) lcolor(black)) ///
102 text(50 2000 "Policy Change", place(e) size(medium)) ///
103 yscale(range(-30 30))
104
105 * Exporto el grafico
106 graph export "$output/figura3.png", replace
107
108
109 * ---- FIGURA 4: Placebo policy implementation in 1994: São
Paulo versus synthetic São Paulo.---- *
110 /* En este gráfico se incluye un placebo en 1994, es decir, se
corre el modelo como si el tratamiento hubiera sido hecho en
1994 (cuando en realidad fue en 1999)*/
111
112 * Cargamos la base de datos
113 use "$output/df.dta", clear
114 tsset state year
115
116 * Estimamos el sintético cambiando el año de tratamiento (1994)
117 synth homiciderates stategdpcapita yearsschoolingimp
populationprojectionln giniimp populationextremepovertyimp
stategdpgrowthpercent homiciderates(1990(1)1993), trunit(26)

```

```

stategdpgrowthpercent homiciderates(1990(1)1995), trunit(20)
trperiod(1994) nested fig
118
119 * Exporto el grafico
120 graph export "$output/figura4.png", replace
121
122
123 * ---- FIGURA 5: Leave-one-out distribution of the synthetic
    control for São Paulo---- *
124 * Leave one out. En esta figura lo que hace es estimar el
    sintético para Sao Paulo con los n-1 estados restantes (siendo n
    los estados con los que estimaba antes). En cada estimación saca
    un Estado distinto. La idea es mostrar que no importa con que
    Estados lo hacíamos igual nos daba diferencias post tratamiento
    entre Sao Paulo y su sintetico
125
126 * Cargamos la base de datos
127 use "$output/df.dta", clear
128 tsset state year
129
130 * Estimamos el sintético
131 synth homiciderates stategdpcapita yearsschoolingimp
    populationprojectionln giniimp populationextremepovertyimp
    stategdpgrowthpercent homiciderates(1990(1)1998), trunit(26)
    trperiod(1999) nested
132
133 * Queremos saber que Estados eran utilizados para contruir el
    sintetico original (es decir, aquellos para los cuales el peso
    era >0)
134 * En e(W_weights) se guardan los pesos de cada unidad
135 mat list e(W_weights)
136
137 * El sintetico se construye usando a Distrito Federal (7),
    Maranhao (10), Minas Gerais (13) y Rio de Janeiro (21)
138 * Vamos a correr nuevos sinteticos sacando de a uno los estados
    que forman del Sao Paulo sintetico para ver si no hay ningun
    estado driving the results
139
140 * Cargamos la base de datos
141 use "$output/df.dta", clear
142 tsset state year
143
144 tempname resmat
145         local i 26
146         qui synth homiciderates stategdpcapita yearsschoolingimp
    populationprojectionln giniimp populationextremepovertyimp
    stategdpgrowthpercent homiciderates(1990(1)1998), trunit(`i')
    trperiod(1999) nested keep(loo-resout`i', replace)
147         *primero estimo para sao paulo. la base loo-resout26 es
    el sintetico para SP usando TODOS los estados del donor pool
148
149

```

```

150      *ahora construyo todos los sintetiticos posibles para
sao paulo sacando de a uno a los estados del donor pool. siempre
le aplico el tratamiento a sao paulo
151      *quiero mostrar que el efecto se mantiene aunque vaya
sacando de uno del donor pool. quiero decir que no hay ningun
estado driving the results.
152      local i 26
153      forvalues j=1/27 {
154      if `j'== 26 {
155      continue
156      }
157      use "$output/df.dta", clear
158      tsset state year
159      * sacamos un estado
160      drop if state==`j'
161      * corremos el sintetico
162      qui synth homiciderates stategdpcapita yearsschoolingimp
populationprojectionln giniimp populationextremepovertyimp
stategdpgrowthpercent homiciderates(1990(1)1998), trunit(`i')
trperiod(1999) nested keep(loo-resout`j', replace)
163      * la base loo-resoutj me dice el sintetico para SP
excluyendo al estado j
164      }
165
166      * volvemos a mergear las bases individuales
167      forvalues i = 1/27 {
168      use "$output/loo-resout`i'.dta", clear
169      ren _Y_synthetic _Y_synthetic_`i'
170      ren _Y_treated _Y_treated_`i'
171      gen _Y_gap_`i'=_Y_treated_`i'-_Y_synthetic_`i'
172      save "$output/loo-resout`i'.dta", replace
173      }
174
175      use "$output/loo-resout1.dta", clear
176      forvalues i = 2/27 {
177      merge 1:1 _Co_Number _time using "$output/loo-resout`i'.dta",
nogen
178      }
179
180      *twoway (line _Y_synthetic_20 _time, lcolor(grey)) (line
_Y_synthetic_21 _time, lcolor(grey)) (line _Y_treated_26 _time,
lcolor(black) lwidth(thick)) (line _Y_synthetic_26 _time,
lcolor(black) lpattern(dash)), xline(1999) legend(off)
181
182      *twoway (line _Y_synthetic_20 _time, lcolor(grey) lwidth(medium)
lpattern(solid)) ///
183      *      (line _Y_synthetic_21 _time, lcolor(grey) lwidth(medium)
lpattern(solid)) ///
184      *      (line _Y_treated_26 _time, lcolor(black) lwidth(thick))
///
185      *      (line _Y_synthetic_26 _time, lcolor(black)
lpattern(dash)) ///

```

```

186 *      xline(1999) ///
187 *      legend(order(3 "SP" 4 "Synthetic SP" 1 "Synthetic SP
(leave-one-out)") ///
188 *      rows(1) position(6) colgap(5))
189
190 twoway (line _Y_synthetic_7 _time, lcolor(gray) lwidth(medium)
lpattern(solid)) ///
191 (line _Y_synthetic_10 _time, lcolor(gray) lwidth(medium)
lpattern(solid)) ///
192 (line _Y_synthetic_13 _time, lcolor(gray) lwidth(medium)
lpattern(solid)) ///
193 (line _Y_synthetic_21 _time, lcolor(gray) lwidth(medium)
lpattern(solid)) ///
194 (line _Y_treated_26 _time, lcolor(black) lwidth(thick)) ///
195 (line _Y_synthetic_26 _time, lcolor(black) lpattern(dash
)), ///
196 xline(1999) ///
197 legend(order(3 "SP" 4 "Synthetic SP" 1 "Synthetic SP
(leave-one-out)") ///
198 rows(1) position(6) colgap(5))
199
200
201
202 * Exporto el grafico
203 graph export "$output/figura5.png", replace
204
205
206 * ---- FIGURA 6: Permutation test: Homicide rate gaps in São
Paulo and twenty-six control states---- *
207 * Tengo que aplicarle el tratamiento a los demas estados y ver
qué onda.
208 * El gap para SP deberia hacerse negativo desp de 1999 mientras
que el gap para los demas estados deberia tener movimientos
aleatorios (algunos para arriba y otros para abajo)
209
210 * Cargamos la base de datos
211 use "$output/df.dta", clear
212 tsset state year
213
214 tempname resmat
215 local i 26
216 qui synth homiciderates stategdpcapita yearsschoolingimp
populationprojectionln giniimp populationextremepovertyimp
stategdpgrowthpercent homiciderates(1990(1)1998), trunit(`i')
trperiod(1999) nested keep(resout`i', replace)
217
218 * Resout es una matriz que nos guarda para cada "year": outcome
sintetico, outcome treated, y weights.
219 matrix `resmat' = nullmat(`resmat') \ e(RMSPE)
220 local names ``names' ``i''''''
221 mat colnames `resmat' = "RMSPE"

```

```

222     mat rownames `resmat' = `names'
223     matlist `resmat' , row("Treated Unit")
224
225     *saco a SP y aplico el tratamiento a estados fakes
226     drop if state == 26
227
228     forvalues i = 1/27 {
229         if `i'==26 {
230             continue
231         }
232         qui synth homiciderates stategdpcapita yearsschoolingimp
populationprojectionln giniimp populationextremepovertyimp
stategdpgrowthpercent homiciderates(1990(1)1998), trunit(`i')
trperiod(1999) keep(resout`i', replace)
233         matrix `resmat' = nullmat(`resmat') \ e(RMSPE)
234         * Nos guardamos el MSPE. mean squared prediction error
235         local names ``names' ``i'""'
236     }
237
238     mat colnames `resmat' = "RMSPE"
239     mat rownames `resmat' = `names'
240     matlist `resmat' , row("Treated Unit")
241
242     * Si corremos hasta aca, podemos notar que las unidades tratadas
cuyo RMSPE supera al de Sao Paulo en mas de 2 veces son las
unidades 1, 3, 12, 21, 8, 23 y 27. Estas unidades se excluiran
en la figura 7
243
244     * fijemonos que contiene cada base resout`i'
245     use "$output/resout26.dta", clear
246     * el 26 es el tratamiento aplicado a SP, el 1 aplicado a Acre,
el 2 aplicado a Alagoas, etc..
247     * tengo una base por estado, en la que se dice qué peso se le da
a cada estado, el outcome del estado y el outcome sintetico para
cada año.
248
249     * mergeo todas las bases.
250     * renombro variables de cada base individual para despues hacer
un merge.
251     forvalues i = 1/27 {
252         use "$output/resout`i'.dta", clear
253         ren _Y_synthetic _Y_synthetic_`i'
254         ren _Y_treated _Y_treated_`i'
255         gen _Y_gap_`i'=_Y_treated_`i'-_Y_synthetic_`i'
256         save "$output/resout`i'.dta", replace
257     }
258
259     use "$output/resout1.dta", clear
260     forvalues i = 2/27 {
261         merge 1:1 _Co_Number _time using "$output/resout`i'.dta", nogen
262     }
263     * hasta aca tengo todas las merges

```

```

263 * hasta aca tengo todos los merges.
264
265 * Grafico los gaps:
266 twoway (line _Y_gap_1 _time, lcolor(gray)) (line _Y_gap_2 _time,
267         lcolor(gray)) (line _Y_gap_3 _time, lcolor(gray)) ///
268         (line _Y_gap_4 _time, lcolor(gray)) (line _Y_gap_5 _time,
269         lcolor(gray)) (line _Y_gap_6 _time, lcolor(gray)) ///
270         (line _Y_gap_7 _time, lcolor(gray)) (line _Y_gap_8 _time,
271         lcolor(gray)) (line _Y_gap_9 _time, lcolor(gray)) ///
272         (line _Y_gap_10 _time, lcolor(gray)) (line _Y_gap_11 _time
273         , lcolor(gray)) (line _Y_gap_12 _time, lcolor(gray)) ///
274         (line _Y_gap_13 _time, lcolor(gray)) (line _Y_gap_14 _time
275         , lcolor(gray)) (line _Y_gap_15 _time, lcolor(gray)) ///
276         (line _Y_gap_16 _time, lcolor(gray)) (line _Y_gap_17 _time
277         , lcolor(gray)) (line _Y_gap_18 _time, lcolor(gray)) ///
278         (line _Y_gap_19 _time, lcolor(gray)) (line _Y_gap_20 _time
279         , lcolor(gray)) (line _Y_gap_21 _time, lcolor(gray)) ///
280         (line _Y_gap_22 _time, lcolor(maroon)) (line _Y_gap_23
281         _time, lcolor(gray)) (line _Y_gap_24 _time, lcolor(gray)) ///
282         (line _Y_gap_25 _time, lcolor(gray)) (line _Y_gap_27 _time
283         , lcolor(gray)) (line _Y_gap_26 _time, lcolor(red) lwidth(thick
284         )), ///
285         xline(1999) legend(off) xtitle("Year") name(gg2, replace)
286
287 graph export "$output/figura6.png", replace
288
289
290 *----- Figura 7: Permutation test: Homicide rate gaps in
291 Sao Paulo and selected control states -----
292 * las unidades tratadas cuyo RMSPE supera al de Sao Paulo en mas
293 de 2 veces son las unidades 1, 3, 12, 21, 8, 23 y 27. Estas
294 unidades se excluiran en la figura 7
295 * Grafico los gaps:
296 twoway (line _Y_gap_2 _time, lcolor(gray)) ///
297         (line _Y_gap_4 _time, lcolor(gray)) (line _Y_gap_5 _time,
298         lcolor(gray)) (line _Y_gap_6 _time, lcolor(gray)) ///
299         (line _Y_gap_7 _time, lcolor(gray)) (line _Y_gap_9 _time,
300         lcolor(gray)) ///
301         (line _Y_gap_10 _time, lcolor(gray)) (line _Y_gap_11 _time
302         , lcolor(gray)) ///
303         (line _Y_gap_13 _time, lcolor(gray)) (line _Y_gap_14 _time
304         , lcolor(gray)) (line _Y_gap_15 _time, lcolor(gray)) ///
305         (line _Y_gap_16 _time, lcolor(gray)) (line _Y_gap_17 _time
306         , lcolor(gray)) (line _Y_gap_18 _time, lcolor(gray)) ///
307         (line _Y_gap_19 _time, lcolor(gray)) (line _Y_gap_20 _time
308         , lcolor(gray)) ///
309         (line _Y_gap_22 _time, lcolor(maroon)) (line _Y_gap_24
310         _time, lcolor(gray)) ///
311         (line _Y_gap_25 _time, lcolor(gray)) (line _Y_gap_26 _time
312         , lcolor(red) lwidth(thick)), ///
313         xline(1999) legend(off) xtitle("Year") name(gg2, replace)
314
315

```

294

295 graph export "\$output/figura7.png", replace