

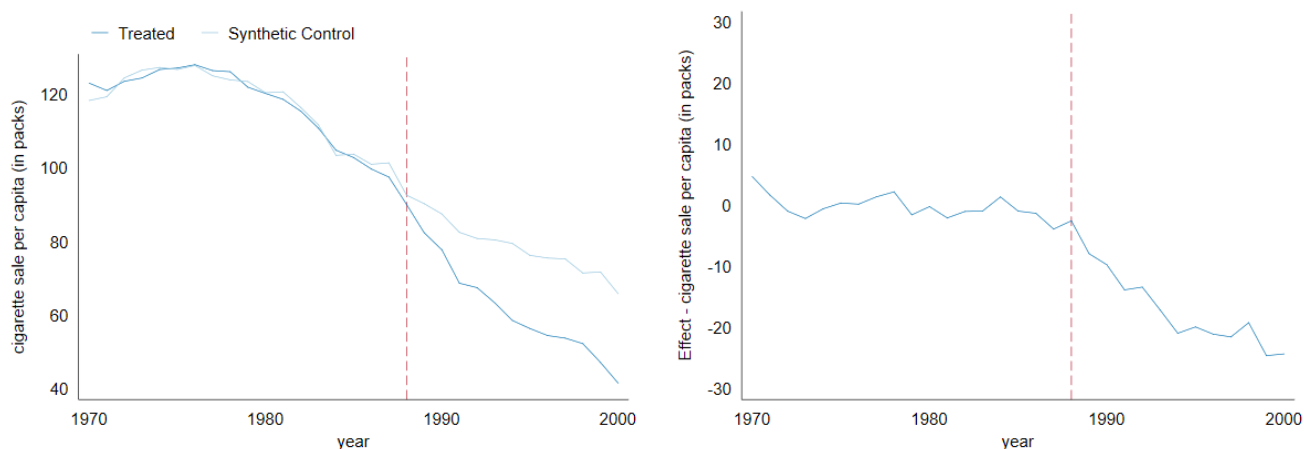
Problem 1.a. Utilizamos los datos de Abadie et al. para construir un control sintético de California entre 1970 y el 2000. Este control es una combinación lineal de los otros 38 estados de la muestra que se comporte lo mas parecido posible a California antes de 1988. La función generada es de

$$Synth = 0,285 * Colorado + 0,101 * Connecticut + 0,245 * Nevada + 0,369 * Utah$$

Para generar este resultados, usamos dos comandos distintos: el comando *Synth* y el comando homologo *synthrunner*

Problem 1.b. Replicamos los gráficos 2 y 3 de Abadie et al. Estos muestran las curvas de California y California Sintéticos, y su diferencia.

Figura 1: Réplica de cuadros 2 y 3 de Abadie et al.

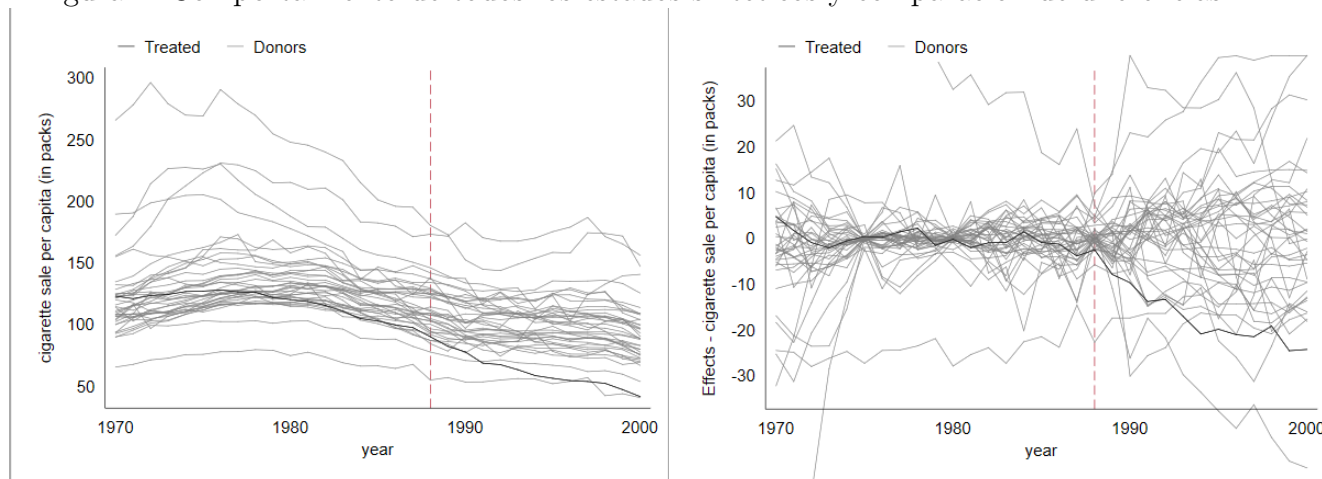


Problem 1.c. Si observamos la figura 2, vemos que California Sintético sigue e comportamiento de la California real hasta el año 1988, año en que ocurre el tratamiento. Esto hace pensar que California y California Sintético se comportarían de manera similar de no ser por el tratamiento y, por lo tanto, es un buen contrafactual. Aún así, este método presenta el problema de no tener pendientes paralelas, en vez teniendo una divergencia, lo que podría complicar un método dif en dif clásico. Por otro lado, al ser el contrafactual California Sintético una construcción matemática, tenemos que creer el supuesto de que, de no ser por el tratamiento, California se hubiese seguido comportando de manera acorde a $0,285 * Colorado + 0,101 * Connecticut + 0,245 * Nevada + 0,369 * Utah$.

Problem 2.a. Al generar los placebos con los otros 38 estados, estamos asegurándonos que la diferencia de California y California Sintética sea causada por el tratamiento en lugar de ser un accidente estadístico inherente del modelo. Entonces, al comparar la diferencia entre cada estado y su contrafactual sintético, debiésemos ver que California sea del 5 % más afectado.

Problem 2.b. Generamos una réplica del gráfico 4 de Abadie et al.. Además, agregamos un gráfico en el que se representa el comportamiento de los otros 38 Estados Sintéticos.

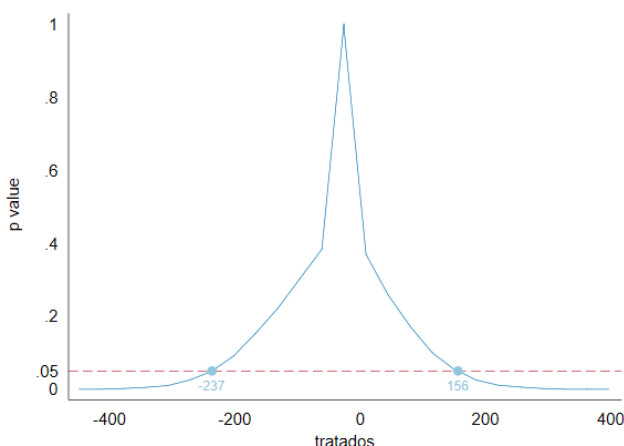
Figura 2: Comportamiento de todos los estados sintéticos y comparación de diferencias.



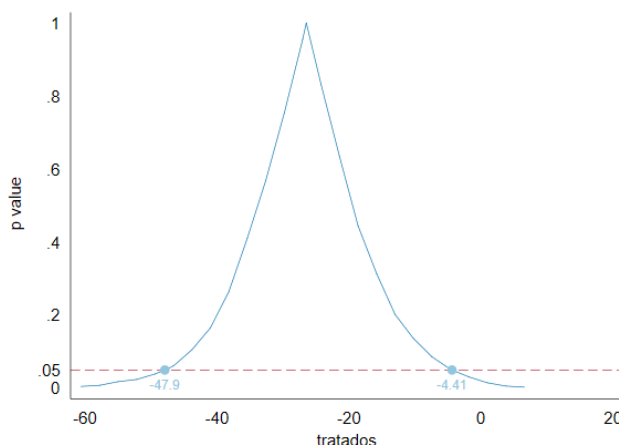
Problem 2.c. Como de los 38 estados sintéticos solo 1 tiene menor efecto respecto a su estados original que California Sintético, podemos creer que $\frac{1}{38}$ es nuestro error. Esta proporción representa la probabilidad de no rechazo cuando se debiese rechazar; la lógica del p-valor. Entonces, podemos creer que California Sintética muestra su diferencia porque esta es real, en lugar de un accidente estadístico con un p-valor de $\frac{1}{39} = 0,025$

Problem 3.a. Si intentamos estimar el efecto de la ley en la cantidad vendida de cajetillas de cigarros en california por medio de un modelo dif-en-dif a travez de una regresión de MCO, obtenemos que este fue de 26.486 cajas menos en promedio.

Problem 3.b. El efecto de una regresión clusterizada por estado es, al igual que la regresión por MCO, de 26.488 cajas menos. La diferencia se muestra en la varianza. La desviación estandar clusterizada es mas de 1.5 puntos menor que la versión MCO. Podemos estimar un efecto a travez de bootstrapping. Este bootstrap intentará remuestrear clusterizado por estado:



este nos dice que el efecto esta entre -237.1 y 155.6 cajas despues de la Propuesta 99. Este efecto no es especialmente creible, siendo un intervalo enorme que incluye el 0. Si en vez generamos el bootstrap clusterizando por estado y por año, obtenemos



donde se muestra un intervalo entre -48.65 y -5.50. Este intervalo no solo es mas compacto sino que se hace consistente con el resultado de las regresiones pues permite con 95 % de confianza afirmar que las regresiones son no-falsas. Como asumimos que esta distribución es simétrica, podemos calcular el valor central como $\frac{(-48,65)-(-5,50)}{2} = -21,575$. Por ultimo, el calculo del bootstrap nos dice que el efecto de la Proposición 99 es distinta de 0 con un p-valor de 1.7 %; altamente significativa.

Efectos estimados			
	MCO	MCO Clusterizado	Bootstrap Clusterizada
Tratamiento	-26.486*** (4.36)	-26.486*** (2.88)	-21.575**

Problem 3.c. Observando los valores tabulados y sus grados de confianza, elegiría la regresión clusterizada. Esta es mas confiable que la regresión sin clusterizacion pues permite atrapar de manera mas confiable el error estandar dentro de estados. Asi mismo, al estar en una serie de tiempo, es altamente probable que los efectos de un periodo esten afectados por periodos anteriores, por lo que un remuestreo aleatorio de bootstrap no necesariamente atraparía todo este efecto temporal, incluso controlando por el año de la muestra.

Problem 4.a. El efecto promedio calculado como el promedio de la diferencia entre California y California Sintético es de -16.582 o 16.58 cajetillas menos por persona en promedio. Esta diferencia esta subestimada pues los efectos de la Proposición 99 se hicieron mucho mas claros luego de 5 años, superando las 24 cajetillas por persona. Por otro lado, el efecto estimado en el la estimación MCO de el apartado 3 tiene un valor de -26.486, mucho mayor que el periodo con mayor diferencia entre las diferencias sintéticas.

Problem 4.b. La razón de esta diferencia nace de la contrafactual ocupada. Como en el apartado A creamos un contrafactual sintetico que se comportaba similar a California, el ATE calculado es respecto a lo que creeríamos que hubiese hecho California de no haber pasado la ley. En efecto, se ve en la figura 2 que California historicamente es uno de los estados en que menos se consumian cigarros. Este efecto no es absorbido perfectamente en el estimador de diferencia en diferencia del apartado 3 pues no se puede cumplir que

$E(\text{tratados}, \text{California}) = 0$, lo que significa que existe una endogeneidad del modelo que no se elimina por utilizar controles y sesga nuestro estimador.