

Problem Set #4

Profesor: Tomás Rau
Ayudantes: Valentina Andrade y Nicolás Valle

Fecha de Entrega: 28 de Junio

El Problem Set 4 incluye tanto una sección de teórica y empírica. Todas las secciones se deben desarrollar de manera **individual**. Ud. puede intercambiar ideas, no respuestas ni códigos.

Las entregas deben realizarse exclusivamente a través de CANVAS. Cualquier entrega realizada después de la fecha límite será penalizada con 5 décimas por día de atraso.

En cuanto al formato, se requiere la entrega de un archivo comprimido que contenga: 1) documento pdf (escrito preferentemente en Latex) y 2) códigos para replicación (puede utilizar Matlab, R, STATA, Julia). Respuestas sin códigos tendrán la nota mínima. Los códigos entregados deben replicar los resultados de las respuestas.

A. Aleatorización (25 Puntos)

El ejercicio a continuación se aleja un poco de los ejercicios a los que se ha visto enfrentada en este curso. A diferencia de otras veces, no se le pedirá replicar el trabajo de otros autores, sino que será usted misma la investigadora encargada de estimar los efectos de un programa. Con eso en mente, recuerde la importancia de justificar **cada** decisión que toma y se le sugiere que disfrute el desafío.

Con el objeto de mejorar el aprendizaje y desempeño de los estudiantes, los gobiernos de todo el mundo financian una gran variedad de insumos del proceso educativo: textos, libros, material didáctico, etc. El objetivo de este ejercicio es evaluar el efecto de la entrega de uno de estos insumos en el aprendizaje de los estudiantes, medido mediante pruebas estandarizadas (tipo SIMCE). El programa a evaluar corresponde a la entrega de “flip charts” en Kenya. Los flip charts son carteles grandes con elementos didácticos que el profesor puede ir rotando para usar de apoyo en su clase. Usted cuenta con la base de datos **retro-data-kremer.dta**, base con datos de una Organización No Gubernamental (ONG) que distribuye los flip charts en Kenya. La ONG le cuenta que los flip charts se asignan según una regla que incluye los siguientes elementos: (1) capacitación a los profesores, (2) si la sala de clases es cubierta, (3) si el techo tiene goteras, (4) el número de escritorios por alumno, (5) el número de libros por niño y (6) el tamaño de la sala de clases.

- a) Estime un modelo mediante MCO para calcular la correlación entre tener algún flip chart (any-flip) y el puntaje en pruebas estandarizadas (nmse). ¿Por qué podría esta correlación no representar el impacto de recibir un flip chart? Mencione al menos 2 ejemplos concretos.

- b) Estime el impacto de recibir flip charts mediante un modelo de matching usando las 6 variables mencionadas más arriba. Escoja el método de matching que según usted mejor se ajusta a este escenario. Muestre el balance post matching para agregar robustez a su estimación.
- c) Usted se entera que los flip charts solo incluyen temas de matemáticas (mat), ciencias (sca) y tecnología (hsb) y no temas como inglés (eng), swahili (kis), geografía (ghc) ni artes (acm). ¿Cómo podría usar esta información para medir el impacto de los flip charts de una manera distinta a la realizada en a) y b)? Estime el impacto del programa usando esta metodología y explique los supuestos que deben cumplirse para lograr una correcta identificación. Realice un chequeo de robustez para convencer al lector de su investigación que su estrategia es válida.
- d) Tiempo después, la ONG aleatoriza la entrega de flip charts usando una lotería básica. Los resultados de esta se encuentran en la base **prosp-data-kremer.dta**. Estime el efecto promedio de tratamiento (ATE) de los flip charts sobre el puntaje en pruebas estandarizadas. Estudie además los efectos del programa para cada materia por separado. ¿Qué espera obtener? ¿Qué encuentra? ¿Hay algo que llame su atención? ¿Algo que le preocupe?
- e) Compare los estimadores del impacto obtenidos en las preguntas a), b), c) y d). ¿Qué puede decir respecto al sesgo de selección en este programa?

B. Variables Instrumentales (40 puntos)

Usted cuenta con la base de datos **acemoglu.2001.dta**, correspondiente al paper *The Colonial Origins of Comparative Development: An Empirical Investigation* de Acemoglu et al. (2001). Su primera tarea es **leer el artículo** ya que facilitará el desarrollo del ejercicio. Nota: Los resultados no serán idénticos a los obtenidos en el paper pero si muy parecidos.

- a) Replique la tabla 2 (Columnas 1-6) del paper y discuta (siguiendo los argumentos de los autores) sobre la interpretación de los resultados en términos de causalidad.
- b) Discuta brevemente el(los) instrumento(s) utilizado(s) por los autores, especialmente en relación a las condiciones que este(os) debe(n) cumplir para identificar insesgadamente el parámetro de interés.
- c) Replique la tabla 4 (Columnas 1-8) del paper e interprete los resultados obtenidos con 2SLS vs OLS.
- d) Para saber si estamos o no ante la presencia de instrumentos débiles, testee usando el estadístico de Cragg y Donald junto con las tablas de Stock y Yogo con respecto al tamaño o *size* del test de Wald. ¿Cómo testean los autores la fortaleza del instrumento? Compare ambos resultados.
- e) Replique la tabla 8 (excepto el panel C) del paper y comente sobre la interpretación que hacen los autores. Comente además acerca de los tests de sobreidentificación realizados aún cuando utilizan un solo instrumento.
- f) Utilizando los resultados de replicación de las Tablas 2 y 4, realice el test de Hausman e interprete sus resultados. **Nota:** Esto debe hacerse para aquellas especificaciones que son iguales; es decir: Tabla 2 columnas 2, 5 y 6 vs Tabla 4 columnas 1, 2 y 8 respectivamente.
- g) Lea el paper *The Colonial Origins of Comparative Development: An Empirical Investigation: Comment* de David Albouy (2012). ¿Cuál es el argumento central del paper? ¿Qué hace el autor para cuestionar los resultados de Acemoglu et al. (2001)? ¿Qué puede decir respecto a sus resultados en d) luego de leer el paper?

- h) Utilizando la base de datos **albouy_2012.dta** replique la tabla 2 (Columnas 1-4, paneles A-D) de Albouy (2012) y comente sobre los resultados obtenidos.

C. Regresión Discontinua (35 puntos)

Usted cuenta con el texto y base de datos que corresponden a la investigación *The effects of drinking and driving laws on car crashes, injuries, and deaths: Evidence from Chile* de Otero y Rau (2017). En base a la lectura del paper responda las siguientes preguntas:

- a) Describa brevemente el objetivo del paper, la estrategia de identificación. Refiérase a la “running variable” y al diseño empírico.
- b) Escriba la ecuación (1) del paper, explicando sus componentes, y demuestre que δ corresponde al efecto causal del programa en el contexto de una regresión discontinua. Sea formal y explícita en los supuestos necesarios para poder identificar δ . ¿Son plausibles?
- c) Los autores eligen el método de “regresión lineal local” para estimar el efecto causal del programa. ¿Cuál es la intuición de esta estimación? ¿Cómo eligen el ancho de banda los autores y cuál es la importancia de esto? ¿Qué rol juega la elección del Kernel?
- d) Usando la base **rdd_otero_rau.dta**, replique la Figura 1 completa y el panel (a) de la Figura 2 del paper. Sea cuidadoso con el orden del polinomio de la running variable. ¿Qué puede decir de los efectos del programa en los outcomes de interés? (Puede utilizar el comando *rdplot* en STATA para hacerlo).
- e) Replique las columnas 1, 2 y 6 de la Tabla 5 del paper. ¿Cómo se interpretan los resultados?
- f) Estime los resultados presentados en la primera columna de la Tabla 5 manteniendo el ancho de banda utilizado, pero estimando con Kernel Uniforme, Triangular, y Epanechnikov. Comente sobre sus resultados.
- g) Ahora estime los resultados de la primera columna de la Tabla 5, pero utilizando el ancho de banda conocido como IK. ¿Cambian sus resultados? ¿Cuál es la importancia de la elección del ancho de banda? ¿Cómo se diferencia éste método de calculo del ancho de banda con el utilizado en el paper?