

# Problem Set #3

Profesor: Tomás Rau  
Ayudantes: Valentina Andrade y Nicolás Valle

Fecha de Entrega: 7 de Junio

El Problem Set 3 incluye tanto una sección 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.

Note que cada **inciso tiene un valor de 5 puntos**.

## A. Ejercicio Teórico: Matching and Inverse Probability Weighting (30 Puntos)

Suponga que Ud. desea estimar el ATT de una intervención no aleatoria. Para ello Ud. cuenta con una muestra de  $n_1$  tratados y  $n_0$  controles. Sea  $I_0$  el conjunto de índices de controles e  $I_1$  el conjunto de índices de tratados. Así, Ud. observa solamente  $y_i = y_i^1 d_i + (1 - d_i) y_i^0$  donde  $d_i$  es una dummy de tratamiento. Suponga que Ud. desea implementar una estrategia de Matching. Es decir, por cada tratado  $i$  Ud. quisiera encontrar el contrafactual: cuál hubiese sido el resultado de este tratado  $i$ , si no se hubiese tratado. Una manera muy general es aproximar dicho contrafactual individual simplemente como  $\hat{y}_{i0} = \sum_{j \in I_0} \lambda(i, j) y_j$ , donde  $\lambda(i, j)$  es un ponderador (weight).

a) Demuestre que el ATT puede ser escrito de la siguiente forma

$$\tau_{att} = \frac{1}{n_1} \sum_{i \in I_1} y_i - \frac{1}{n_1} \sum_{j \in I_0} \bar{\lambda}(j) y_j$$

donde  $\bar{\lambda}(j) = \sum_{i \in I_1} \lambda(i, j)$  representa la suma de todos los weights que usan el control “j”.

b) Existen muchos tipos distintos de ponderadores. Refiérase a la estructura de dichos ponderadores  $\lambda(i, j)$  para el caso de Abadie e Imbens (2006) al caso de Inverse Probability Weighting (alguna de sus versiones) y al caso de Díaz, Rau y Rivera (2015).

- c) Una amiga le indica que es mejor usar  $\lambda(i, j) = 1/n_1$ . Escriba el análogo poblacional del ATT que resulta de reemplazar el nuevo  $\bar{\lambda}(j)$  en (1) y discuta sobre la posibilidad de remover el sesgo de selección en este caso.
- d) Suponga Ud. decide implementar Kernel Matching a la Ichimura, Heckman y Todd (1998) con lo que  $\lambda(i, j) = \kappa(|e(X_i) - e(X_j)|/h) / \sum_{j \in I_0} \kappa(|e(X_i) - e(X_j)|/h)$ . Donde  $\kappa(\cdot)$  es una función kernel,  $e(X_m)$  es el propensity score evaluado en  $X_m$  y  $h$  el ancho de banda. Suponga que Ud. está interesada en balancear las variables “post-matching”, es decir que para cada unidad del grupo de tratamiento  $i$ , el promedio ponderado de las  $X_j$  en el grupo de tratamiento sea lo más cerca posible. ¿Cómo elegiría el ancho de banda? Desarrolle. Hint: puede proponer una función objetivo a minimizar que mejore el balance siguiendo a Díaz et al. (2015).
- e) Ud. está interesada en obtener el error estándar de su estimador. Utilice la fórmula de la identidad de la varianza para proponer un estimador de la varianza del ATT. Puede suponer que  $y_i$  es i.i.d.
- f) Un primo de su amiga le propone usar Bootstrap para calcular el error estándar. Discuta sobre la factibilidad de encontrar un estimador apropiado para dicho error estándar. ¿Afecta su respuesta la elección de  $\lambda(i, j)$ ? Ayuda: Abadie e Imbens (2008)

## B. Matching: Ejercicio Empírico (35 puntos)

A partir de la lectura del artículo “Human Capital and Industrialization: Evidence from the Age of Enlightenment” responda las siguientes preguntas:

- a) Resuma los puntos más importantes del artículo y las conclusiones que se pueden obtener. Para su respuesta considere la pregunta económica que se plantea en este trabajo y discuta la relevancia de esta. Extensión máxima: 1 plana.
- b) Replique la figura 3 del artículo. Interprete.
- c) Explique detalladamente la(s) estrategia(s) empírica(s) utilizada(s). En particular, discuta la intuición detrás de la utilización de Matching en este caso. ¿En qué supuestos de identificación descansa la estimación?<sup>1</sup>. Discuta acerca de la factibilidad de que dichas condiciones de identificación se cumplan en este caso.
- d) Replique la tabla 1 del artículo y explique por qué es necesario verificar el balance de la muestra. Además, discuta la conclusión realizada por los autores respecto a esta tabla. ¿Está de acuerdo?
- e) Replique la tabla 2 del artículo e interprete los coeficientes obtenidos.
- f) Replique la tabla 8 del artículo, tanto la estimación por Matching como por OLS. Explique el sentido de realizar estas estimaciones, e interprete los resultados obtenidos.
- g) Presente dos críticas al artículo<sup>2</sup>, que no hayan sido abordadas anteriormente. Sea detallada en estas, y argumente respecto a cómo dichas críticas afectan la validez de los resultados, y por lo tanto las conclusiones del artículo.

---

<sup>1</sup>Sea clara en identificar los supuestos de identificación, no se está preguntando por generalidades.

<sup>2</sup>Al menos una de ellas debe ser asociada a la estrategia empírica utilizada por los autores.

## C. Diferencias en Diferencias: Ejercicio Empírico (35 puntos)

En el artículo “Do Teenagers Respond to HIV Risk Information? Evidence from a Field Experiment in Kenya” se estudia si es posible cambiar el comportamiento sexual de adolescentes en Kenia a través de la entrega de información. A partir de este trabajo, responda las siguientes preguntas:

- a) Lea y resuma brevemente el paper, indicando su aporte a la literatura y conclusiones más importantes. Explique en qué contexto y con qué fin se hace uso de un modelo de diferencias en diferencias en este trabajo. ¿Qué supuesto(s) necesitan la autora para poder identificar<sup>3</sup> correctamente los parámetros de interés? ¿Qué puede decir respecto a la plausibilidad que dicho(s) supuesto(s) se cumpla(n)?
- b) Analice cuidadosamente las tablas 1 y 2 del paper. ¿Hay algo que llame su atención? ¿Qué puede decir respecto al balance pre estimación de la muestra?
- c) Replique la tabla 3 del artículo. Explique qué rol cumple en este trabajo e interprete los coeficientes obtenidos.
- d) Replique la tabla 4 del artículo. Explique qué rol cumple en este trabajo e interprete los coeficientes obtenidos. (*Hint*: Utilice la base *homesurvey* para esta pregunta. Tenga cuidado con la variable *agefather27*).
- e) Investigue acerca de diseños tipo “Event Study” (muchas veces considerados como generalizaciones del método de DiD). ¿Por qué y dónde surge la necesidad de utilizar modelos de este tipo? ¿En qué se parecen y en qué se diferencian estos diseños al modelo de DiD estudiado en clase? Mencione el(los) supuesto(s) de identificación detrás de un diseño tipo “Event Study”. ¿Considera que son supuestos más (o menos) fuertes que los de DiD? ¿Por qué?
- f) Usted desea estimar un “Event Study” en el contexto del artículo en cuestión. ¿Qué pregunta de investigación le gustaría responder con su estimación? ¿Son suficientes los datos con los que cuenta para responderla? En caso afirmativo, ¿cuáles usaría? En caso contrario, ¿qué tipos de datos necesitaría para poder responderla correctamente? Plantee el modelo (ecuación) que estimaría, explicando claramente cada uno de sus componentes y justificando cada una de sus decisiones. ¿Qué resultados esperaría obtener?
- g) Presente dos críticas a este trabajo, que no hayan sido ya abordadas en esta tarea. Argumente detallada y cuidadosamente su respuesta, ayudándose de sus conocimientos econométricos y económicos.

---

<sup>3</sup>Sea clara en identificar los supuestos de identificación, no se está preguntando por generalidades.