

Tarea 2

Regresión Discontinua, Variables Instrumentales y Diff in Diff

Profesores: Omar Pérez, Daniel Schwartz

Auxiliares: Antonia Aceituno, Camila Galarce

Ayudantes: Guillermo Escobar, Bastián Medina

En esta primera parte de la tarea se les pide estudiar la existencia del denominado “puzzle de consumo tras la jubilación” (“retirement consumption puzzle”) en los hogares italianos entre los años 1993 y 2004.

Para examinar esta temática se les pide utilizar los datos de las encuestas de ingresos y riquezas realizadas por el “Banco de Italia” para las olas 1993, 1995, 1998, 2000, 2002 y 2004, los cuales se encontrarán disponibles en la plataforma U-Cursos.

Esta es una tarea “semi-estructurada”. Por tanto, no existe una única forma de responderla. Se evaluará la coherencia del proceso y del análisis. Se les pide lo siguiente:

1. Elaboren una hipótesis sobre el impacto de la jubilación del jefe o jefa de hogar sobre el consumo del hogar. Justifiquen brevemente su hipótesis utilizando al menos dos referencias de estudios empíricos internacionales que sean relevantes y guarden relación a su hipótesis. Utilicen Google Scholar o algún método de buscar referencias para buscar artículos, y recuerden referenciar su fuente. La referencia tiene que seguir un formato (APA, MLA, el que quieran).(4 puntos)
2. En este trabajo se les pide seguir el procedimiento realizado por Battistin (2008). Para esto, construyan una base de datos colapsada por año ($anno$) y “diferencia en años entre la edad mínima de jubilación y la edad del jefe de hogar” ($esse_m$). Es decir, en la base de datos cada fila debe poseer las variables promedio para todos los hogares que durante un año (ejemplo: $anno = 1995$) su jefe de hogar tenía la misma edad (ejemplo: $esse_m = 3$). Su base de datos final debe incluir las variables: “ $anno$ ”, “ $esse_m$ ”, “ $elig_m$ ”, “ lnc ”, “ $lncn$ ”, “ $lnconsal$ ” y cuatro variables adicionales que estimen relevantes para el estudio. Adicionalmente construyan la variable elegibilidad que tome el valor 1 si el jefe del hogar ya tiene la edad mínima para jubilarse y 0 en otro caso. Finalmente, construyan la variable $porcentaje_retirados$ como promedio de la variable “ qu_m ” (por cada fila [combinación $anno - esse_m$]). Expliquen brevemente el proceso de construcción de su base de datos, identificando claramente su unidad de análisis. (6 puntos)

De ahora en adelante, para los análisis de ambos métodos, sólo incluyan a los hogares cuyos jefes de hogar se encuentren a menos de 10 años de cumplir la edad de jubilación y los que tengan menos de 10 años posteriores a la edad de jubilación.

Diseño de Regresión Discontinua

1. Comprueben (o argumenten) el cumplimiento de los supuestos requeridos para utilizar el diseño de RD nítido (sharp). Para esto, asuman que la running variable es $esse_m$, la variable dependiente es lnc y el tratamiento es la elegibilidad. En caso de realizar regresiones no es necesario

incluir tablas con los modelos completos, basta con mencionar el coeficiente de interés, su error estándar y su significancia estadística. (6 puntos)

2. Presenten en una tabla los resultados de cinco modelos de RD, asumiendo que se cumplen todos los supuestos. Los primeros dos modelos tienen que incluir todas las observaciones (en la ventana es_{se_m} entre -10 y 10). El modelo 1 no debe incluir controles, el modelo 2 debe incluir los controles que estimen convenientes. Los siguientes tres modelos tienen que incluir los controles del modelo 2, pero deben poseer distintos “tamaños de ventana” (8, 6 y 3 respectivamente). (6 puntos)
3. ¿Qué pueden concluir a partir del análisis anterior? Responda a partir de los cinco modelos estimados. (5 puntos)

Diseño de Variables Instrumentales

1. Comprueben (o argumenten) el cumplimiento de los supuestos requeridos para utilizar el diseño de variables instrumentales (IV). Para esto, asuman que la variable endógena es *porcentaje_retirados*, el instrumento elegibilidad y la variable dependiente es Inc. (6 puntos)
2. Presenten en una tabla los resultados de seis modelos, asumiendo que se cumplen todos los supuestos para ocupar IV. Los modelos 1, 3 y 5 no tienen que incluir controles, y los modelos 2, 4 y 6 tienen que incluir los controles que estimen convenientes. Para los modelos 1 y 2 se les pide estimar modelos por mínimos cuadrados ordinarios el efecto promedio de *porcentaje_retirados* sobre la variable dependiente. Luego, en los modelos 3 y 4 se les pide estimar la intención a tratar (ITT) y para los modelos 5 y 6 se les pide estimar el efecto promedio del tratamiento en los compliers mediante mínimos cuadrados en dos etapas (2SLS). (6 puntos)
3. ¿Qué pueden concluir a partir del análisis anterior? Genere una narrativa a partir de los seis modelos estimados. (5 puntos)

Conclusiones

1. Discutan y concluyan sobre su hipótesis inicial a la luz de la evidencia encontrada. En la discusión, brevemente discuta cada tipo de validez (interna, externa y del constructo) de ambos métodos. (6 puntos)

Diff in Diff

En esta segunda parte de la tarea se les pide analizar lo realizado por Kiel y McClain (1995), que estudiaron el efecto que tuvo un nuevo incinerador de basura en el valor de las viviendas en North Andover, Massachusetts. Utilizaron datos de varios años y un análisis econométrico muy complejo. Aunque se emplearán datos de dos años y algunos modelos simplificados, el análisis que sigue es similar.

El rumor de que se construiría un nuevo incinerador de basura en North Andover comenzó después de 1978 y la construcción inició en 1981. Se esperaba que el incinerador entrara en operación poco después del comienzo de su construcción; en realidad comenzó a operar en 1985. Se utilizarán datos sobre los precios de las casas que se vendieron en 1978 y otra muestra de aquellas que se vendieron en 1981. La hipótesis es que el costo de las casas ubicadas cerca del incinerador estaría por debajo del precio de las viviendas más lejanas.

Análisis del Efecto del Incinerador de Basura en el Precio de las Viviendas

A manera de ejemplo, se considera que una casa está cerca del incinerador si se halla en un radio de 3 millas. Utilizando la data subida a Ucurso, “Base 2”, responda las siguientes preguntas:

- Realicen una regresión lineal que les permita estudiar el efecto en el precio real ($rprice$) de estar cerca del incinerador de basura, considerando solo los datos del año 1981. Analice y discuta sus resultados (4 puntos).
- Lleven a cabo la misma regresión, pero ahora considerando solo los datos del año 1978. Comparen sus resultados con la parte (a). (3 puntos)
- A partir de los resultados anteriores discutan si la regresión estimada en el apartado (a) resulta apropiada para estimar el efecto de la ubicación del incinerador en el valor de las casas. (3 puntos)
- Grafiquen el precio real de las casas en el tiempo, diferenciando por si las casas están o no cerca del incinerador. Analicen el gráfico. (2 puntos)
- Utilicen el estimador de diferencia en diferencias (Diff in Diff) (sacando los promedios por año y por si está cerca o no del incinerador), para estimar el efecto del incinerador en los valores de las casas cercanas al mismo. (2 puntos)
- Ahora comprueben los resultados obtenidos en la parte (e) y la significancia del efecto realizando una regresión lineal. Expliquen la construcción de la regresión y concluya. (5 puntos)

Análisis de la Influencia de la Distancia al Incinerador sobre el Precio de las Viviendas

Utilizando la misma base de datos que en la parte anterior evalué el efecto la variable $dist$ sobre el precio de una casa, resolviendo las siguientes preguntas:

- Estimen el siguiente modelo e informen los resultados. Interpreten el coeficiente de $y81\log(dist)$. ¿Qué concluyen a partir de esto? (5 puntos)

$$\log(\text{price}) = \beta_0 + \delta_0 y81 + \beta_1 \log(\text{dist}) + \delta_1 y81 \log(\text{dist}) + u$$

- b) Añadan age , age^2 , $rooms$, $bath$, $\log(inst)$, $\log(land)$, $\log(area)$ a la ecuación. Ahora, ¿Cuál es su conclusión acerca del efecto del incinerador sobre los valores de las casas? (3 puntos)
- c) ¿Por qué el coeficiente de $\log(dist)$ es positivo y estadísticamente significativo en (b) pero no en (c)? ¿Qué dice esto sobre los controles utilizados en (c)? (6 puntos)