Variables Instrumentales (II)

1. Vamos a replicar algunos resultados del artículo de Angrist y Lavy: "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement", *Quarterly Journal of Economics*, 114, 533-575.

   Usaremos los datos para los estudiantes de 5to grado: **AngristLavy5.gdt**. El archivo contiene información de los resultados promedio en cada clase en matemáticas (`avgmath`) y lengua (`avgverb`), el tamaño de la clase en primavera (`classize`), el número de alumnos matriculados al comienzo del año en la escuela (`enrol`), un identificador de la ciudad (`towncode`), y un índice del nivel socioeconómico de los estudiantes que los autores llaman porcentaje de desfavorecidos (`tipuach`). Este estudio se limita a estudiantes en el sistema público Judío e incluye escuelas religiosas y laicas. Los autores eliminan de la muestra las observaciones con un tamaño de clase mayor a 44, las que tienen menos de 5 alumnos matriculados y las que no tienen alumnos tomando el examen de matemáticas (`mathsize`=0).

   El artículo explota el hecho de que el tamaño de las clases en las escuelas en Israel tiene un máximo de 40 estudiantes (lo que se conoce como la regla de Maimonides), para identificar el efecto causal del tamaño de las clases en los resultados de los estudiantes. Se utiliza la regla de Maimonides porque esta regla provoca un cambio (discontinuidad) en la relación entre matriculación y tamaño de la clase para ciertos valores de matriculación (múltiplos de 40). Por ejemplo una escuela con 40 matriculados en quinto grado tendrá una clase de 40 estudiantes y una escuela con 41 estudiantes matriculados tendrá dos clases más pequeñas.

   Dado que este cambio (discontinuidad) es la fuente de indentificación, los autores definen una "muestra con discontinuidad" que incluye las escuelas con matriculaciones en un intervalo cercano a los puntos de discontinuidad (40,80,120). Esta submuestra incluye solamente a las escuelas con matriculaciones en los siguientes intervalos: [36,45], [76,85], [116,125].

   En este ejercicio nos centraremos en los resultados **en matemáticas**.

   i) Restrinja la muestra siguiendo los criterios de los autores.

   ii) Defina una variable binaria que valga uno para las observaciones en la muestra con discontinuidad. ¿Cuántas clases pertenecen a dicha muestra?

   iii) (Extra) Reporte la siguiente información respecto al tamaño de las clases en la muestra total: la media, la desviación estándar, y los percentiles 10, 25, 50, 75 y 90.

      Además reporte los valores medios y los desvíos estándar para las siguientes variables: matriculación, porcentaje de estudiantes desfavorecidos, notas en examen de lengua y notas en examen de matemáticas. Realice los cálculos para la muestra total y la submuestra con discontinuidad. Compare los resultados para ambos grupos.

   iv) En la Tabla II del artículo los autores comparan las estimaciones MCO de tres modelos diferentes. Estime los tres modelos solamente para los resultados **en matemáticas**:

1) un modelo simple con el tamaño de la clase como único regresor, 2) un modelo que incluya también el porcentaje de estudiantes desfavorecidos, y 3) un modelo que controle además por el número de matriculados. Comente los resultados.

v) Compute el tamaño de clase medio predicho en cada escuela s de acuerdo a la "regla de Maimonides":

$$f_s = \frac{enrol_s}{int\left[\frac{enrol_s - 1}{40}\right] + 1}$$

donde $int(n)$ es la parte entera de n.

vi) En la Tabla III los autores reportan la forma reducida de la relación entre el tamaño de clase predicho y el tamaño real así como para los resultados de los tests. Replicar las columnas (2) y (6) y comentar los resultados. La columna(2) corresponde a un modelo con el tamaño de la clase como variable dependiente y f_s, el porcentaje de estudiantes desfavorecidos y el número de matriculados como controles. La columna(6) corresponde a un modelo con el resultado de matemáticas como variable dependiente y f_s, el porcentaje de estudiantes desfavorecidos y el número de matriculados como controles.

vii) En la Tabla IV, los autores comparan estimaciones de MC2E para diferentes modelos utiliando f_s como instrumento para el tamaño de la clase. Compute estimaciones de MC2E para un modelo para los resultados en matemáticas controlando por el tamaño de la clase, el porcentaje de estudiantes desfavorecidos y el número de matriculados (columna(8) en la Tabla IV). Comente las diferencias entre estos resultados y los obtenidos con MCO.

viii) Un aspecto clave en el problema, es capturar la relación en este caso entre matriculación y resultados. Si la relación es no lineal el modelo anterior no sería correcto para capturar el efecto del tamaño de la clase. Por ello los autores incluyen el cuadrado de la variable de matriculación como especificación alternativa. Estime nuevamente el modelo incluyendo esta variable como regresor adicional.

ix) Estime usando MC2E el modelo para los resultados en matemáticas controlando por el tamaño de la clase, el porcentaje de estudiantes desfavorecidos y el número de matriculados para la sumbuestra con discontinuidad (columna(12) en la Tabla IV). Compare los resultados con los obtenidos para la muestra completa.

2. (extraído de Stock and Watson) Durante la década de 1880, un cártel conocido como el Comité Ejecutivo Conjunto (JEC, en sus siglas originales) controlaba el transporte ferroviario de cereales desde las ciudades del medio oeste hacia las del oeste de los Estados Unidos. El cártel precedió a la Ley Antimonopolio Sherman de 1890, y operó legalmente para aumentar el precio del grano por encima de lo que habría sido el precio competitivo. De vez en cuando, los engaños de los miembros del cártel provocaban un colapso temporal en el acuerdo de fijación de precio colusivo. En este ejercicio, se utilizarán las variaciones de la oferta asociadas con los colapsos del cártel para estimar la elasticidad de la demanda de transporte de grano por ferrocarril. La base de datos `JEC` contiene las observaciones semanales sobre el precio del transporte en tren y de otros factores desde 1880 hasta 1886.

Supóngase que la curva de demanda de transporte ferroviario de cereales se especifica como

$$ln(Q_i) = \beta_0 + \beta_1 * ln(P_i) + \beta_2 Ice_i + \sum_{j=1}^{12} \beta_{2+j} Seas_{j,t} + u_i$$

donde Q_i es el total de toneladas de grano enviado en la semana i; P_i es el precio del envío de una tonelada de grano por ferrocarril. La variable Ice_i es una variable binaria igual a 1 si los *Grandes Lagos* no son navegables a causa del hielo, y la variable $Seas_j$ es una variable binaria que capta la variación estacional de la demanda. La variable Ice se incluye porque el cereal podría igualmente transportarse por barco, cuando los *Grandes Lagos* son navegables.

i) Estime la ecuación de demanda por MCO. ¿Cuál es el valor estimado de la elasticidad de la demanda y su error estándar?

ii) Explique por qué la interacción entre la oferta y la demanda podría hacer que el estimador MCO de la elasticidad fuera sesgado

iii) Considérese la utilización de la variable cartel como variable instrumental para $ln(P_i)$. Utilice un razonamiento económico para analizar si resulta verosímil que la variable *cartel* satisfaga las dos condiciones para que un instrumento sea válido.

iv) Estima la regresión de la primera etapa. ¿Es la variable *cartel* un instrumento débil?

v) Estime la ecuación de la demanda mediante Variables Instrumentales. ¿Cuál es la elasticidad de la demanda estimada y su error estándar?

vi) ¿Sugiere la evidencia que el cártel estuviera fijando el precio que maximizaba el beneficio de monopolio? Explíquelo (Pista: ¿Qué debería hacer un monopolista si la elasticidad precio fuese menor que 1?).