Econometría II Tarea Examen I

Benjamín Oliva [1]
Jésica Tapia [2]
Omar Alfaro [3]

Marzo 2022

[1] benjov@ciencias.unam.mx y https://github.com/benjov
[2] jesicatapia@gmail.com y https://github.com/
[3] omarxalpha@gmail.com y https://github.com/omarxalpha

Instrucciones generales

- La entrega podrá ser en equipos de hasta 3 personas.
- Siempre es posible atender dudas por correo o agendar sesiones de dudas por Zoom.
- Fecha de entrega es el 30 de marzo de 2022.

Índice general

1

Problemas Teóricos

1.1. Bases del análisis de regresión (40 %)

Resuelva la sigueinte serie de ejercicios para continuar con la discusión del Efecto Marginal y algunos puntos básicos del análisis de regresión.

1. Sean y, x_1 y x_2 variables aleatorias y una forma funcional de la esperanza condicional de la forma:

   $$\mathbb{E}[y|x_1, x_2] = \beta_1 + \beta_2 x_1 + \beta_3 x_2^2 + \beta_4 x_1 x_2$$

   Al respecto, determine el Efecto Marginal de x_1 y x_2.

2. Sean y, x_1 y $D = \{0, 1\}$ variables aleatorias y una forma funcional de la esperanza condicional de la forma:

   $$\mathbb{E}[y|x_1, D] = e^{\beta_1 + \beta_2 x_1 + \gamma D}$$

   Al respecto, determine el Efecto Marginal de x_1 y $D = \{0, 1\}$.

3. Utilice el material expuesto en el curso para analizar y derterminar si las siguientes afirmaciones son FALSAS o VERDADERAS.

   a) Cuando se realiza un proceso de estimación por el método de MCO, la suma de los erroes resultantes siempre es igual a cero, es decir,

   $$\sum_{i=1}^{n} e_i = 0$$

b) Suponga que realiza una estimación por el método de mínimos cuadrados de la ecuación dada por $y_i = \beta_2 x_{i2} + \beta_3 x_{i3} + ... + +\beta_K x_{iK} + \varepsilon_i$, para $i = 1, 2, \ldots, N$. Asimismo, suponga que como resultado de su estimación encunetra que la estadística R^2 es muy próxima al valor de 0.5, entonces ¿se puede decir que las variables dadas por $x_{i2}, x_{i3}, ..., x_{iK}$ explican aproximadamente la mitad de la variabilidad de la variable y_i?

c) Suponga que estima una ecuación de salario dada por:

$$salario_i = \beta_1 + \beta_2 educa_i + \beta_3 educa_i^2 + \beta_4 DSexo_i + \varepsilon_i$$

Donde todas los estimadores $\hat{\beta}_k$, $k = 1, 2, 3, 4$ son estadísticamente significativos; $i = 1, 2, \ldots, N$; $salario_i$ es el ingresos mesual por sueldos y salarios de una persona; $educa_i$ es la educación de una persona en años, y $DSexo_i$ es una variable dummy que toma el valor de 1 si el individuo es hombre y 0 en cualquier otro caso. Asuma que los resultados indican que: $\hat{\beta}_2 > 0$ y $\hat{\beta}_3 > 0$. Entonces, ¿podemos afirmar qué existe evidencia estadística de la presencia de redimientos marginales decrecientes de la educación en los salarios?

d) Asuma que estima una regresión de los salarios de las personas en una cierta industria y encuenta que sus resultados son:

$$\begin{array}{ccccc} \hat{y}_i & = & 5.15 & + & 1.17\, x_i \\ & & (2.5) & & (0.3) \end{array}$$

Donde y_i es el salario mensual en miles de pesos de las personas, y x_i es una variable dummy que toma el valor de 1 cuando el individuo i es hombre y 0 cuando es mujer. Entre paréntesis se indica los errores estándar de los estimadores (conocidos como término respectivo a $\hat{\sigma^2}(\mathbf{X}'\mathbf{X})^{-1}_{kk}$).

Entonces, existe evidencia estadística y significativa para afirmar que la mejor estimación al salario de los hombres es 5.15 mil pesos y para las mujeres es de 6.31 mil pesos.

4. Suponga que ha decidido estimar una regresión del logarittmo natural del salario en función de un conjunto de variables para un conjunto de individuos en una industria; de tal forma que obtiene los resultados mostrados en el Cuadro 1.1.

Cuadro 1.1: Resultados de la regresión de salarios

Variable	Coeficiente	Error Estándar	Estadística t
Constante	4.216	0.078	
Masculino	0.154	0.005	
Primaria	0.224	0.018	
Secundaria	0.833	0.063	
Preparatoria	0.902	0.063	
Univesidad	0.550	0.065	8.462
ln(Experiencia)	0.907		22.535
Primaria×Masculino	-0.97		-1.242
Secundaria×Masculino	-0.67		-2-272
Preparatoria×Masculino	-0.72		-2.317
Univesidad×Masculino	-0.146	0.076	
ln(Experiencia)×Masculino	0.041		4.891
$R^2 = 0.6032$	F = 89.69	n = 1,000	$\mathbf{e}'\mathbf{e} = 415.37$

En donde la variables son como se describe a continuación:

a) Masculino: es una variable binaria que toma el valor de 1 (uno) cuando el individuo es del sexo masculino y 0 (cero) en cualquier otro caso.

b) Primaria, Secundaria, Preparatoria y Universidad: son un conjunto de variables categóricas (Dummy) que indican el nivel máximo de estudios del individo, es decir, toman el valor de 1 si el individuo cumple con la categoria y cero si su nivel máximo de estudios es algún otro novel. Por lo tanto hemos dejado fuera la dummy que indica nula educación.

c) Experiencia: es una variable que indica el número de años de experiencia del individuo.

Se le solicita:

a) Completar el Cuadro 1.1, así como indicar la hipótesis nula y alternativa asociada. También deberá determinar cuales de las variables incluidas en la regresión resultan estadísticamente significativas.

b) Utilizando la información de Cuadro 1.1 determine la hipótesis nula y alternativa para saber si el logarítmo natural de la experiencia es estadísticamente igual a 1. realice la prueba t asociada y concluya si se puede rechazar la hipótesis nula.

c) Utilizando la información de Cuadro 1.1 determine y explique cuál es el Efecto Marginal de la variable Universidad.

5. Un embotellador de bebidas gaseosas analiza las rutas de servicio de las máquinas expendedoras en su sistema de distribución. Le interesa predecir el tiempo necesario para que el representante de ruta atienda las máquinas expendedoras en una tienda. Esta actividad de servicio consiste en abastecer la máquina con productos embotellados, y algo de mantenimiento o limpieza. El ingeniero industrial responsable del estudio ha sugerido que las dos variables más importantes que afectan el tiempo de entrega, y, son la cantidad de cajas de producto abastecido, x_1, y la distancia caminada por el representante, x_2. El ingeniero ha reunido una muestra aleatoria de algunas observaciones de tiempo de entrega.

   Suponga que desea estimar la ecuación de regresión asumiendo una media condicional dada por:

$$\mu_{y_i|x_{i1},x_{i2},x_{i3}} = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} \tag{1.1}$$

   Derivado de la operación matricial de la muestra suponga que obtiene los siguientes resultados:

$$\begin{bmatrix} \sum y_i^2 & \sum y_i x_{i1} & \sum y_i x_{i2} & \sum y_i x_{i3} \\ \sum x_{i1} y_i & \sum x_{i1}^2 & \sum x_{i1} x_{i2} & \sum x_{i1} x_{i3} \\ \sum x_{i2} y_i & \sum x_{i2} x_{i1} & \sum x_{i2}^2 & \sum x_{i2} x_{i3} \\ \sum x_{i3} y_i & \sum x_{i3} x_{i1} & \sum x_{i3} x_{i2} & \sum x_{i3}^2 \end{bmatrix}$$

$$= \begin{bmatrix} 18310.63 & 559.60 & 7375.44 & 324972.70 \\ 559.60 & 25.00 & 219.00 & 9932.00 \\ 7375.44 & 219.00 & 3055.00 & 129099.00 \\ 324972.69 & 9932.00 & 129099.00 & 6402888.00 \end{bmatrix} \tag{1.2}$$

   Al respecto:

a) Determine los coeficientes $\boldsymbol{\beta}$ de regresión lineal

b) Determine el valor de la varianza estimada de los errores: $\hat{\sigma}^2$

c) Pruebe la hipótesis nula de que $\beta_2 = 0$ con un $\alpha = 0.05$

d) Pruebe la hipótesis nula de que $\beta_3 = 0$ con un $\alpha = 0.05$

e) Escriba una conclusión de las dos pruebas de hipótesis planteadas

f) Determine el valor del R^2 asociado al caso

1.2. Tópicos intermedios del análisis de regresión (20 %)

En esta sección platearemos un caso típico del método de Variables Insttrumentales. Para lo cual le pedimos seguir la discusión y realizar las actividades que se le piden. Consideremos un modelo de equilibrio de mercado (la discusión original se encuentra en el trabajo de *Working (1927). What do statistical demand curves show?. Quaterly Journal of Economics, 41 pp. 212 - 235*). Supongamos que tenemos el siguiente modelo de demanda o modelo en su forma estructural:

$$\begin{aligned} q^d &= \alpha_0 + \alpha_1 p_i + u_i \\ q^s &= \beta_0 + \beta_1 p_i + v_i \end{aligned}$$

Donde q^d es la cantidad demandada para el bien en cuestión, q^s es la cantidad ofertada, p_i es el precio de dicho bien. Los términos de error u_i y v_i representan otros factores observables que afectan o desplazan a la cantidad ofertada y a la cantidad demandada. Podemos asumir que dichos elementos u_i y v_i son factores no observables. De esta forma $\mathbb{E}[u_i] = 0$ y que $\mathbb{E}[v_i] = 0$. Adicionalment, suponga que que $Cov[u_i, v_i] = 0$ y que $q^d = q^s = q$, de tal forma que el mercado está en equilibrio, por lo que tenemos:

$$q = \alpha_0 + \alpha_1 p_i + u_i \tag{1.3}$$

$$q = \beta_0 + \beta_1 p_i + v_i \tag{1.4}$$

Ambas ecuaciones tienen problemas de endogeneidad, ya que el precio p_i es endógeno en cada una de las ecuaciones, ya que se determina simúltaneamente.

Utilizando las ecuaciones (1.4) y (1.4) obtenga la ecuación de demanda:

$$q_i = \frac{\alpha_1\beta_0 - \alpha_0\beta_1}{\alpha_1 - \beta_1} + \frac{\alpha_1 v_i - \beta_1 u_i}{\alpha_1 - \beta_1} \tag{1.5}$$

$$p_i = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1} + \frac{v_i - u_i}{\alpha_1 - \beta_1} \tag{1.6}$$

De esta forma hemos obtendrá dos expresiones (1.5), ecuación de demanda, y (1.6), ecuación de oferta, que representan la solución para el sistema de ecuaciones. No obstante, se puede apreciar que cada una de las variables tiene influencia del término de error de cada una de las ecuaciones, por lo que NO se pueden separar enteramente y se sule decir que tenemos un problema de endogeneidad.

A partir de la ecuación (1.6), determine la covarianza entre p_i y cada uno de los factores de cambio de la demanda (u_i) y de la oferta (v_i). Muestre que es igual a:

$$\begin{aligned} Cov(p_i, u_i) &= -\frac{Var(u_i)}{\alpha_1 - \beta_1} && (1.7) \\ Cov(p_i, v_i) &= \frac{Var(v_i)}{\alpha_1 - \beta_1} && (1.8) \end{aligned}$$

De esta forma, el precio está correlacionado positivamente con los factores que modifican la demanda y negativamente con los facores que modifican la oferta, ello en razón de que la pendiente de la demanda es negativa ($\alpha_1 < 0$) y la de la oferta es positiva ($\beta_1 > 0$). En este caso, decimos que el problema de endogenidad se origina por el proceso de equilibrio en el mercado.

La razón de la endogeniedad referida es que de la curva de demanda o de la curva de oferta no es posible inferir si los cambios en el nivel de precios o en las cantidades en el mercado son por causa de desplazamientos de la curva de oferta o de la curva de demanda. Esto significa que sólo es posible estimar la curva de oferta o de demanda si es posible observar alguno de los factores que determina a la demanda o la oferta.

Supongamos que el efecto del termino de error en la cantidad ofertada, v_i, puede ser dividido en: los fatores observables que están asociados a la demanda x_i y un factor no observable ζ_i que no está correlacionado con x_i. Es decir, que tenemos una formulación de la ecuación de oferta similar a:

$$q_i = \beta_0 + \beta_1 p_i + \beta_2 x_i + \zeta_i \tag{1.9}$$

De esta forma podemos decir que existe una posibilidad de que se elimine la endogeidad, esto es, que se pueda instrumentar mediante el uso de variables asociadas a la demanda (x_i). Lo anterio, para efecto de separar del término de error la información que puede ser endogena al sistema, tal y como se muestra en la ecuación (1.9). La Figura 1.1 ilustra está discusión para cada uno de los casos: de la demanda y de la oferta.

Esta discusión es una aplicación del Método de Variables Instrumentales (IV). Para el cual requerimos un regresor que es endogeno y que requerimos sea instrumentado. Al instrumento se le denomina variable instrumental. En este caso, es la variable x_i quién sera nuestra variable instrumental. Retomemos nuestro sistema de ecuaciones de demanda y oferta. Si realizamos el mismo procedimiento de solución del sistema de ecuaciones pero considerando a la ecuación (1.9).

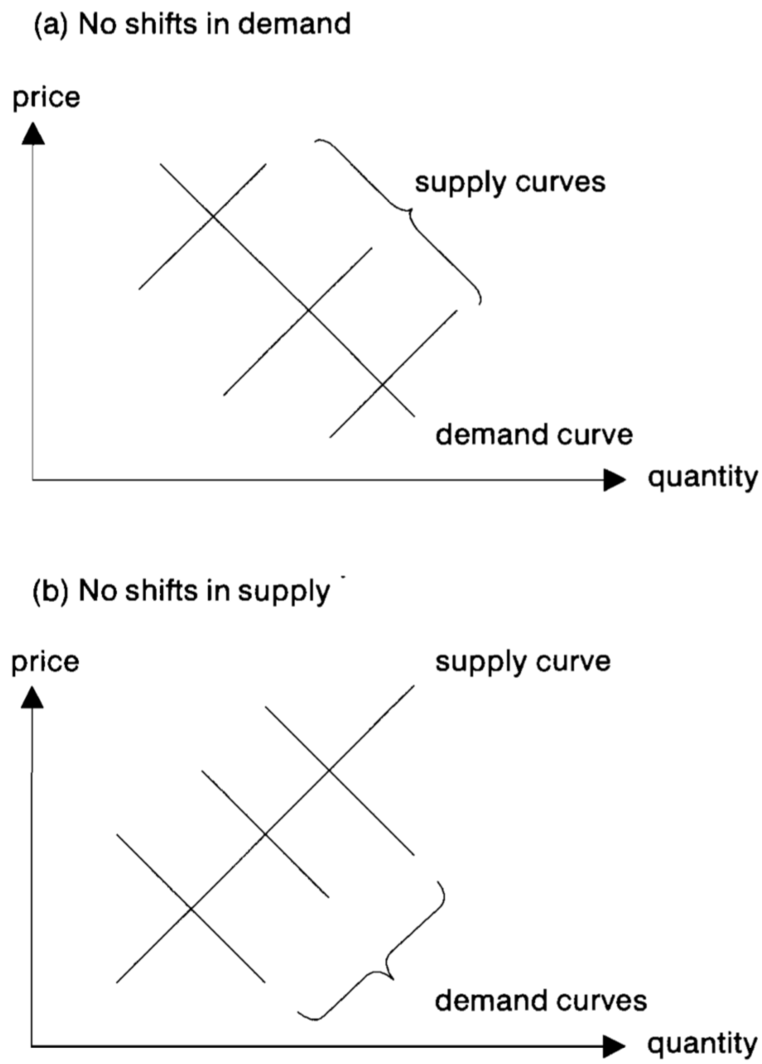
Dicho lo anterior, muestre que obtendríamos un sistema que quedará como sigue:

$$p_i = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1} + \frac{\beta_2}{\alpha_1 - \beta_1} x_i + \frac{\zeta_i - u_i}{\alpha_1 - \beta_1} \tag{1.10}$$

$$q_i = \frac{\alpha_1 \beta_0 - \alpha_0 \beta_1}{\alpha_1 - \beta_1} + \frac{\alpha_1 \beta_2}{\alpha_1 - \beta_1} x_i + \frac{\alpha_1 \zeta_i - \beta_1 u_i}{\alpha_1 - \beta_1} \tag{1.11}$$

En este punto, si fue posible instrumentar correctamente la variable endogena, muestre que deberá pasar que $Cov(x_i, \zeta_i) = 0$. Es decir, es esperado que no exista información endogena. Así aseguraríamos que $Cov(x_i, u_i) = 0$.

Figura 1.1: Instrumentación de la información de la demanda y la oferta, respectivamente



2

Prácticas de Python - Jupyter Notebook (40 %)

2.1. Regresión Lineal MCO

Resolver la práctica de Python ubicado en la carpeta llamada Parte 1. La práctica está relacionada con la replíca que Christensen & Greene (1976) hicieron al trabajo de Nerlove (1963).

2.2. Regresión por el método de Variables Instrumentales

Resolver la práctica de Python ubicado en la carpeta llamada Parte 2. La práctica está relacionada con la implementación de una prueba de Haussman.