Análisis Econométrico

Práctica II

19 July 2025

Regresión Generalizada

1. ¿Cuál es la matriz de convarianzas, $Cov[\hat{\beta}, \hat{\beta} - \mathbf{b}]$, del estimador de MCG $\hat{\beta} = (\mathbf{X'\Omega^{-1}X})^{-1}\mathbf{X'\Omega^{-1}y}$ y la diferencia de esta con la del estimador de MCO $\mathbf{b} = (\mathbf{X'X})^{-1}\mathbf{X'y}$?

2. Suponga que en el modelo de regresión generalizado Ω es conocida.

a) ¿Cuál es la matriz de covarianzas de los estimadores de MCO y MCG de β?

b) ¿Cuál es la matriz de covarianzas del vector de residuos de MCO $\mathbf{e} = \mathbf{y} - \mathbf{Xb}$?

c) ¿Cuál es la matriz de covarianzas del vector de residuos de MCG $\hat{\varepsilon} = \mathbf{y} - \mathbf{X}\hat{\beta}$

d) ¿Cuál es la matriz de covarianzas de los vectores de residuos de MCO y MCG?

3. Suponga que el modelo de regresión es $y_i = \mu + \varepsilon_i$, donde

$E[\varepsilon_i|x_i] = 0,\ Cov[\varepsilon_i, \varepsilon_j|x_i, x_j] = 0$ para $i = j$, pero $Var[\varepsilon_i|x_i] = \sigma^2 x_i^2$, $x_i > 0$.

a) Dada una muestra de observaciones de y_i y x_i, ¿cuál es el estimador más eficiente de μ?¿Cuál es su varianza?

b) ¿Cuál es el estimador de MCO, y cual es la varianza del estimador de MCO?

4) Dos muestras de 50 observaciones producen las siguientes matrices de mommentos (en cada caso, **X** es una constante y una variable).

$$\begin{array}{ccc} & muestra\ 1 & muestra\ 2 \\ X'X & \begin{bmatrix} 50 & 300 \\ 300 & 2100 \end{bmatrix} & \begin{bmatrix} 50 & 300 \\ 300 & 2100 \end{bmatrix} \\ y'X & \begin{bmatrix} 300 & 2000 \end{bmatrix} & \begin{bmatrix} 300 & 2200 \end{bmatrix} \\ y'y & \begin{bmatrix} 2100 \end{bmatrix} & \begin{bmatrix} 2800 \end{bmatrix} \end{array}$$

a) Compute los coeficientes de regresión de minímos cuadrados y las varianzas de los residuos para cada base de datos. Compute el R^2 para cada regresión.

b) Compute la estimación de MCO del vector de coeficientes, asumiendo que los coeficientes y la varianza de las perturbaciones son las mismas en ambas regresiones. Tambien compute la matriz de covarianza asintótica estimada.

c) Contraste la hipótesis de que las varianzas de ambas regresiones es la misma sin asumir que los coeficientes son los mismos en las dos regresiones.

d) Compute el estimador de MCG factibles de los coeficientes en las regresiones, asumiendo que la constante y la pendiente es la misma para ambas regresiones. Compute el estimador de la matriz de varianza y compare con los resultados de b).

5) Para el modelo de regresión simple con heterocedasticidad, $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$ y $var(\varepsilon_i|x_i) = \sigma_i^2$ muestre que la varianza $var(b_2|x_i) = \left[\sum_{i=1}^N (x_i - \bar{x})^2\right]^{-1}\left[\sum_{i=1}^N (x_i - \bar{x})^2\sigma_i^2\right]\left[\sum_{i=1}^N (x_i - \bar{x})^2\right]^{-1}$ se redice a $var(b_2|x_i) = \sigma^2/\sum_{i=1}^N (x_i - \bar{x})^2$ bajo homocedasticidad.

6) Considere el modelo de regresión $y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i$ con dos variables explicativas, pero sin término constante.

a) La función de suma de cuadrados es $S(\beta_1, \beta_2|x_1, x_2) = \sum_{i=}^N (y_i - \beta_1 x_{i1} + \beta_2 x_{i2})^2$. Encuentre las derivadas parciales respecto a los parámetrosn β_1 y β_2. Muestre que el estimador de mínimos cuadrados de β_2 es:

$$b_2 = \frac{(\sum x_{i1}^2)(\sum x_{i2}y_i) - (\sum x_{i1}x_{i2})(\sum x_{i1}y_i)}{(\sum x_{i1}^2)(\sum x_{i2}^2) - (\sum x_{i1}x_{i2})^2}$$

b) Sea $x_{i1} = 1$, muestre que el estimado en (a) puede ser reducido a:

$$b_2 = \frac{\frac{(\sum x_{i2}y_i)}{N} - \frac{\sum x_{i2}}{N}\frac{\sum y_i}{N}}{\frac{\sum x_{i2}^2}{N} - \left(\frac{\sum x_{i2}}{N}\right)^2}$$

c) En el estimador de la parte a), remplace y_i, x_{i1} y x_{i2} por $y_i^* = y_i/\sqrt{h_i}$, $x_{i1}^* = x_{i1}/\sqrt{h_i}$ y $x_{i2}^* = x_{i2}/\sqrt{h_i}$. Estas son variables transformadas por el modelo de heterocedásticidad $\sigma_i^2 = \sigma^2 h(z_i) = \sigma h_i$. Muestre que estimador de MCG resultante puede ser escrito como:

$$\hat{\beta} = \frac{\sum a_i x_{i2} y_i - \sum a_i x_{i2} \sum a_i y_i}{\sum a_i x_{i2}^2 - (\sum a_i x_{i2})^2}$$

donde $a_i = 1/(ch_i)$ y $c = \sum(1/h_i)$. Encuentre $\sum a_i$.

d) Muestre que bajo homocedásticidad $\hat{\beta}_2 = b_2$.

7. Considere el modelo simple de regresión: $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$, donde se cree que existe heterocedasticidad de la forma: $\sigma_i^2 == \sigma^2 x_i^2$. Se tienen 4 observaciones, con $x = (1, 2, 3, 4)$ y $y = (3, 4, 5, 6)$

a) Use la fórmula del estimador de MCO para computar el valor estimado de β_2. En este caso $\sum x_{i2}y_i/N = 10$, $\sum x_{i2}^2/N = 7$

b) Refiriendose al ejercicio 6 (b), cuál es el valor de $c = \sum(1/h_i)$?

c) Refiriendose al ejercicio 6 (c), cuál es el valor de $a_i = 1/(ch_i),\ i = 1, ..., 4$? A qué es igual $\sum a_i$?

d) Use la fórmula para el estimador de mínimos cuadrados generalizados en el ejercicio 6 (c) para computar el valor estimado de β_2.

8. Considere el modelo siguiente para explicar el comportamiento del sueño:

$$sleep = \beta_0 + \beta_1 totwrk + \beta_2 educ + \beta_3 age + \beta_4 yngkid + \beta_6 male + u$$

a) Dé un modelo que permita que la varianza de u difiera entre hombres (male) y mujeres. La varianza no debe depender de otros factores.

b) Emplee los datos en *sleep*75 (librería *wooldridge*) para estimar los parámetros de la ecuación de heterocedasticidad. (Tiene que estimar la ecuación para *sleep* primero mediante MCO para obtener los residuales de MCO.) ¿Es la varianza estimada de u mayor para los hombres o para las mujeres?

c) ¿Es la varianza de u diferente estadísticamente para hombres y para mujeres?

9) Considere el siguiente modelo sobre los determinantes del precio de una vivienda:

$$price = \beta_0 + \beta_1 lotsize + \beta_2 sqrft + \beta_3 bdrms + u$$

donde *price* es el precio en miles de dólares, *lotsize* es el tamaño del lote (solar) en pies cuadrados, *sqrft* es el tamaño de la casa en pies cuadrados y *bdrms*.

a) Usando la base de datos *hprice*1 estime el modelo por MCO. Muestre los resultados en una tabla.

b) Realice la prueba de heterocedásticidad de Breusch-Pagan. ¿Qué concluye acerca de los errores estándar estimados del inciso a)

c) Obtenga los errores estándar robustos y realice las pruebas de significancia individual de los coeficientes estimados.

10) Para este ejercicio emplee el archivo *vote*1.

a) Estime un modelo en el que *voteA* sea la variable dependiente y *prtystrA*, *democA*, $log(expendA)$ y $log(expendB)$ sean las variables independientes. Obtenga los residuales de MCO, $\hat{u}_i$, y regrese éstos sobre todas las variables independientes. Explique por qué obtiene $R^2 = 0$.

b) Ahora calcule la prueba de Breusch-Pagan para heterocedasticidad. Emplee la versión del estadístico F y dé el valor-p.

c) Calcule el caso especial de la prueba de White para heterocedasticidad, usando de nuevo la forma del estadístico F. ¿Qué tan fuerte es ahora la evidencia de heterocedasticidad?

Variables explicativas

11) Para este ejercicio emplee los datos del archivo *wage*2.

a) Estime el modelo

$$log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 tenure + \beta_4 married + \beta_5 black + \beta_6 south + \beta_7 urban + u$$

y de el resultado en una tabla. Manteniendo todos los demás factores constantes, ¿cuál es la diferencia aproximada entre el salario mensual de afroamericanos y no afroamericanos? ¿Es esta diferencia estadísticamente significativa?

b) Agregue a esta ecuación las variables $exper^2$ y $tenure^2$ y muestre que no son conjuntamente significativas al nivel de 20%.

c) Amplíe el modelo original de manera que el rendimiento a la educación dependa de la raza y pruebe si en realidad el rendimiento de la educación depende de la raza.

d) Partiendo nuevamente del modelo original, permita que los salarios difieran entre cuatro grupos: casados afroamericanos, casados no afroaméricanos, solteros afroaméricanos y solteros no afroaméricanos. ¿Cuál es la diferencia de salario entre afroaméricanos casados y no afroaméricanos casados?

Variables Instrumentales

12) Considere el siguiente modelo de regresión:

$$y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + u$$

Donde y_2 una variable explicativa con problemas de endogeneidad y z_1 es una variable exógena, mientras que u es el componente de error del modelo y cumple con los supuestos usuales. A esta ecuación se le conoce como **ecuación estructural** para enfatizar que es una relación estructural.

a) Suponga que se dispone de un instrumento para la variable y_2 denominada z_2. ¿Cuáles son las condiciones que debe poseer ese instrumento para que pueda considerarse válido?

b) La **forma reducida** para y_2 es:

$$y_2 = \pi_0 + \pi_1 z_1 + \pi_2 z_2 + v_2$$

Introduciendo esta expresión en la ecuación original, se obtiene la forma reducida para y_1

$$y_1 = \alpha_0 + \alpha_1 z_1 + \alpha_2 z_2 + v_1$$

Encuentre los α_j en términos de los β_j y π_j.

c) ¿Cómo se puede estimar consistentemente los α_j?

13) Considere el siguiente modelo de regresión de series de tiempo donde la variable explicativa tiene el siguiente error de medición

$$y_t = \beta_0 + \beta_1 x_t^* + u_t$$

$$x_t = x_t^* + e_t$$

donde u tiene una media cero y está correlacionada con x^* y que x^* también tiene media cero.

a) Escriba $x_t^* = x_t - e_t$ y sustituya en el modelo. Muestre que el error en el nuevo modelo (v_t) está negativamente correlacionado con x_t si $\beta_1 > 0$.¿Cuál es la implicancia de este supuesto para el estimador de MCO de β_1 de la regresión entre y_t y x_t?

b) En adición a los supuestos previos, asuma que u_t y e_t están no correlacionados con todos los valores pasados de x_t^* y e_t; en particular con x_{t-1}^* y e_{t-1}. Muestre que $E(x_{t-1}, v_t) = 0$, donde v_t es el término error del modelo de la parte a).

c) Explique si x_t y x_{t-1} están probablemente correlacionados.

d) ¿Qué estrategia para la estimación consistenre de β_0 y β_1 sugieren los resultados de las parte a) y b). anteriores?

14) Los datos en $fertil2$ incluyen información sobre un grupo de mujeres de Botswana durante 1988 de el numero de hijos, años de educación, edad y variables de religión y estatus económico.

a) Estime por MCO el siguiente modelo:

$$children = \beta_0 + \beta_1 educ + \beta_2 age + \beta_3 age^2 + u$$

e interprete los estimados. En particular, manteniendo *age* fijo, ¿cuál es el efecto estimado de otro año de educación sobre la fertilidad? Si 100 mujeres reciben otro año adicional de educación, ¿en cuánto se espera ase reduzca la cantidad de niños?.

b) $frsthalf$ es una variable dummy igual a uno si la mujer nació en los primero seis meses del año. Asumiendo que $frsthalf$ está no correlacionada con $u4, muestre que $frsthalf$ es una candidata razonable de VI para *educ*.

c) Estime el modelo en a. utilizando $frsthalf$ como una VI para *educ*. Compare con el efecto estimado con MCO de la sección a.

d) Añada las variables binarias *electric*, *tv* y *biclycle* al modelo y asuma que son exógenas. Estime la ecuación por MCO y MC2E y compare los coeficientes estimados para *educ*. Interprete el coeficiente asociado a *tv* y explique por qué tener televisión tiene un efecto negativo sobre la fertilidad.

15) Una firma consultora manejada por el Sr. Pedro Chardonnay está investigando la eficiencia relativa de producción de vino de 75 viñedos. Pedro especifica la siguiente función de producción:

$$q = \beta_1 + \beta_2 mgt + \beta_3 cap + \beta_4 lab + e$$

donde q es un índice de producción de vino para un viñedo, que toma en cuenta cantidad y calidad, mgt es una variable que toma en cuenta la eficiencia de la gestión, cap es un índice de insumo de capital y $l$ab es un índice asociado al factor trabajo. Dado que no dispone de datos sobre mgt, Pedro utiliza observaciones del número de años de experiencia (*xper*) de cada administrador de viñedo y utiliza esta variable en lugar de mgt. El archivo *chard.csv* tiene la información de dichas variables.

a) Estime la ecuación revisada usando MCO y comente los resultados.

b) Encuentre los intervalos estimados para la producción de vino de aquellos viñedos que tienen promedio muestrales de trabajo y capital y tienen administradores con:

- 10 años de experiencia
- 20 años de experiencia
- 30 años de experiencia

c) Pedro esta preocupado que la proxy $xper$ puede estar correlacionada con el error del modelo.Decide hacer un contraste de Hausman, utilizando la edad (age) del administrador como un instrumento para $xper$. Regrese $xper$ sobre age, cap y lab, y guarde los residuos. Incluya los residuos como una variable extra en la ecuación estimada en (a) y comente el resultado del constraste de Hausman.

d) Utilice el estimador de VI para estimar la ecuación

$$q = \beta_1 + \beta_2 xper + \beta_3 cap + \beta_4 lab + e$$

con age, cap y lab como las variables instrumentales. Comente los resultados y compare con los de la parte (a).

e) Encuentre los intervalos correspondientes a la producción de vino que tiene los promedio de cap y lab y administradores con:

- 10 años de experiencia
- 20 años de experiencia
- 30 años de experiencia

16) Considere el modelo lineal, $y_i = \alpha + \beta x_i + \varepsilon_i$, en el cual $Cov[x_i, \varepsilon_i] = \gamma \neq 0$. Sea z un variable instrumental exógena y relevante para este modelo. Asuma, también, que z es binaria (toma solo los valores de 1 y 0). Muestre las formas algebraicas de los estimadores de MCO y de Variables Instrumentales de α y β.

17. Muestre que el estimador de *función de control* produce las mismas estimaciones que MC2E.

Bono

16) Para el siguiente ejercicio utilice los datos por hogar disponibles en la Encuesta Nacional de Gastos e Ingresos de los Hogares de 2018 (ENGIH2018).Los datos los encuentra la siguiente dirección: https://www.bancentral.gov.do/a/d/4796-engih-2018, sección bases de datos.Note que tendra que combinar módulos, filtrar observaciones y crear las variables observadas relevantes.

a) Para cada grupo de gasto estime un modelo que relacione el gasto percápita y el ingreso percápita de los hogares dominicanos. Utilice como controles las variables que considere necesarias (usualmente, características de los hogares, entre ellas variables dummies y sus interacciones con las demás variables del modelo). Construya distintas definiciones de ingreso del hogar, incluya en el modelo variables relacionadas a los subsidios y ayuda que reciba el hogar por concepto de transferencias del gobierno.

b) Elabore un cuadro para cada modelo comparar las elasticidades de cada uno de los tipos de gasto respecto al ingreso.

c) Realice pruebas de heterocedasticidad para cada ecuación, resuma los resultados en una tabla y comente.

d) Sustituya los errores estándar de las ecuaciones donde encontró heterocedasticidad por los errores estándar robustos. Presente los resultados en la tabla usual y comente.

e) Finalmente, comente la posibilidad de tener regresores endógenos en las ecuaciones de gasto y proponga (si es posible) alguna solución tipo uso de variables instrumentales o de función de control.