PUBLICACIONES DE 4er CURSO

Curso: 4º

Grado: Economía

Asignatura: ECONOMETRÍA III

TRANSPARENCIAS PARTE 2: Regresores estocásticos

TEMA 4: Modelo lineal general

Profesores: Antonio Aznar

Departamento de ANÁLISIS ECONÓMICO

Curso Académico 2015/16

Facultad de Economía y Empresa
Universidad Zaragoza

TEMA 4: MODELO LINEAL GENERAL

1. Introducción: Sesgo de variable omitida
2. Estimadores Mínimo Cuadrático Ordinarios (MCO)
3. Medidas de ajuste
4. Contrastes
5. Variables de control

1. Introducción: Sesgo de Variable Omitida

El error surge por la existencia de factores, o variables, que influyen a *Y* pero que no se incluyen en la función de regresión. Siempre hay variables omitidas.

Algunas veces, pueden provocar que los estimadores MCO sean sesgados.

C.C. 6.1 Sesgo de Variable Omitida en el modelo con un solo regresor

El sesgo en el estimador MCO consecuencia de la omisión de un factor o variable, se llama **sesgo de variable omitida**. Para que se de este sesgo la variable omitida "*Z*" debe satisfacer dos condiciones:

(1) Z es un determinante de Y (i.e. Z es parte de u); y
(2) Z está correlacionada con el regresor X (*i.e.* $\text{corr}(Z,X) \neq 0$)

Ambas condiciones deben de cumplirse para que la omisión de Z produzca el sesgo de variable omitida.

En el **ejemplo de las calificaciones**:

1. El dominio de inglés (si tiene o no el inglés como segunda lengua) con toda seguridad afecta a las calificaciones: *Z* es un determinante de *Y*.
2. Las comunidades de inmigrantes tienden a ser menos ricas por lo que tienen menores presupuestos escolares y mayores REM: *Z* está correlacionada con *X*.

Por lo tanto, $\hat{\beta}_1$ es sesgado. ¿Cuál es el signo del sesgo?

- *¿Qué te dice el sentido común?*
- Si el sentido común falla hay una fórmula…

Una fórmula para el sesgo de variable omitida: recordar la ecuación,

$$\hat{\beta}_1 - \beta_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})u_i}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \frac{\frac{1}{n}\sum_{i=1}^{n} v_i}{\left(\frac{n-1}{n}\right)s_X^2}$$

donde $v_i = (X_i - \bar{X})u_i \approx (X_i - \mu_X)u_i$. Bajo la hipótesis #1,

$$E[(X_i - \mu_X)u_i] = \text{cov}(X_i,u_i) = 0.$$

Pero si $E[(X_i - \mu_X)u_i] = \text{cov}(X_i,u_i) = \sigma_{Xu} \neq 0$?

Bajo las hipótesis #2 y #3 (esto es, incluso si la hipótesis #1 no se cumple),

$$\hat{\beta}_1 - \beta_1 = \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})u_i}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

$$\xrightarrow{p} \frac{\sigma_{Xu}}{\sigma_X^2}$$

$$= \left(\frac{\sigma_u}{\sigma_X}\right) \times \left(\frac{\sigma_{Xu}}{\sigma_X \sigma_u}\right) = \left(\frac{\sigma_u}{\sigma_X}\right)\rho_{Xu},$$

donde ρ_{Xu} = corr(X,u). Si la hipótesis #1 es correcta, entonces $\rho_{Xu} = 0$, pero si no tenemos….

La fórmula del sesgo de variable omitida:

$$\hat{\beta}_1 \xrightarrow{p} \beta_1 + \left(\frac{\sigma_u}{\sigma_X}\right)\rho_{Xu}$$

- Si una variable omitida es, al mismo tiempo:
  (1) Un determinante de *Y*, esto es, está contenida en u;
  (2) Está correlacionada con *X*,

  entonces $\rho_{Xu} \neq 0$ y el estimador MCO $\hat{\beta}_1$ es sesgado y no es consistente.

- Por ejemplo, los distritos con menos estudiantes aprendiendo inglés (1) tienen mejores notas (2) tienen clases con menores tamaños, por lo que omitir la variable de los que aprenden ingles llevará a sobreestimar el efecto del tamaño de la clase. Está ocurriendo esto en los datos de California?

TABLE 6.1 Differences in Test Scores for California School Districts with Low and High Student–Teacher Ratios, by the Percentage of English Learners in the District

	Student–Teacher Ratio < 20		Student–Teacher Ratio ≥ 20		Difference in Test Scores, Low vs. High STR	
	Average Test Score	n	Average Test Score	n	Difference	t-statistic
All districts	657.4	238	650.0	182	7.4	4.04
Percentage of English learners						
< 1.9%	664.5	76	665.4	27	−0.9	−0.30
1.9–8.8%	665.2	64	661.8	44	3.3	1.13
8.8–23.0%	654.9	54	649.7	50	5.2	1.72
> 23.0%	636.7	44	634.8	61	1.9	0.68

- Los distritos con menos PctEL(% de alumnos estudiando inglés) tienen mejores notas
- Los distritos con un PctEL menor tienen clases menores
- Entre los distritos con parecidos *PctEL*, el efecto del tamaño de la clase es pequeño (recordar el "gap de notas" global= 7.4)

Causalidad y análisis de regresión

El ejemplo notas/REM/porcentaje de estudiantes de inglés muestra que, si una variable omitida satisface las dos condiciones para el sesgo de variable omitida, entonces el estimador MCO que se obtiene omitiendo esa variable es sesgado e inconsistente. Por lo tanto, incluso si n es grande, $\hat{\beta}_1$ se alejará de β_1.

Esto genera una cuestión más profunda: ¿cómo definimos β_1? Esto es, ¿qué queremos estimar con exactitud cuando hacemos una regresión?

Al menos hay tres posibles respuestas a esta cuestión:

1. Queremos estimar la pendiente de una línea que resume gráficamente los datos sin asignarle un significado sustantivo.

2. Queremos hacer predicciones de Y en una situación no presente en los datos para la que conocemos el valor de X.

3. Queremos estimar el efecto causal sobre Y consecuencia de un cambio en X.

2. Estimación Mínimo Cuadrático Ordinaria (MCO)

C.C. 6.2 El Modelo de Regresión Múltiple

Considerar el caso de dos regresores:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i, \ \ i = 1,\ldots,n$$

- Y es la *variable dependiente*
- X_1, X_2 son dos *variables independientes* (*regresores*)
- (Y_i, X_{1i}, X_{2i}) denotan la observación i-ésima de Y, X_1, y X_2.
- β_0 =constante poblacional desconocida
- β_1 = efecto sobre Y de un cambio en X_1, manteniendo constante X_2
- β_2 = efecto sobre Y de un cambio en X_2, manteniendo constante X_1

- u_i = error de regresión (factores omitidos)

El Estimador MCO en la Regresión Múltiple

Con dos regresores, los MCO son la solución de:

$$\min_{\beta_0,\beta_1,\beta_2} \sum_{i=1}^{n}[Y_i - (\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i})]^2$$

- Los MCO minimizan esta suma de cuadrados de la diferencia de Y_i y la predicción (valor predicho) basada en la línea estimada.
- Este problema de minimización se resuelve utilizando calculo.

ECUACIONES NORMALES

Derivando respecto a $\beta_0, \beta_1\ y \beta_2$ y reagrupando se obtienen las tres ecuaciones normales

$$n\hat{\beta}_0 + \hat{\beta}_1 \sum_1^n X_{1i} + \hat{\beta}_2 \sum_1^n X_{2i} = \sum_1^n Y_i$$

$$\hat{\beta}_0 \sum_1^n X_{1i} + \hat{\beta}_1 \sum_1^n X_{1i}^2 + \hat{\beta}_2 \sum_1^n X_{1i} X_{2i} = \sum_1^n X_{1i} Y_i$$

$$\hat{\beta}_0 \sum_1^n X_{2i} + \hat{\beta}_1 \sum_1^n X_{1i} X_{2i} + \hat{\beta}_2 \sum_1^n X_{2i}^2 = \sum_1^n X_{2i} Y_i$$

La extensión de tres regresores a *k* es inmediata. Hay una ecuación normal por cada coeficiente.

Ejemplo: Los datos de California

Regresión de las notas sobre REM:

$$\widehat{NotaExamen} = 698.9 - 2.28\times\text{REM}$$

Ahora incluye el porcentaje de estudiantes que aprenden ingles (*PctEL*):

$$\widehat{NotaExamen} = 686.0 - 1.10\times\text{REM} - 0.65 PctEL$$

- ¿Qué le ocurre al coeficiente de REM?
- ¿Por qué? (*Nota*: corr(REM, *PctEL*) = 0.19

Regresión Múltiple en GRETL

```
Modelo 1: MCO, usando las observaciones 1-420
Variable dependiente: NotaExamen
Desviaciones típicas robustas ante heterocedasticidad, variante HC1

             Coeficiente   Desv. Típica   Estadístico t    Valor p
  ----------------------------------------------------------------
  const      686,032        8,72822          78,60       3,98e-252 ***
  REM         -1,10130      0,432847         -2,544      0,0113    **
  PctEI       -0,649777     0,0310318       -20,94       4,70e-067 ***

Media de la vble. dep.  654,1565   D.T. de la vble. dep.   19,05335
Suma de cuad. residuos  87245,29   D.T. de la regresión    14,46448
R-cuadrado              0,426431   R-cuadrado corregido    0,423680
F(2, 417)               223,8229   Valor p (de F)          9,28e-67
Log-verosimilitud      -1716,561   Criterio de Akaike      3439,123
Criterio de Schwarz     3451,243   Crit. de Hannan-Quinn   3443,913
```

$$\widehat{NotaExamen} = 686.0 - 1.10 \times REM - 0.65 PctEL$$

C.C. 6.3 Los Estimadores MCO, valores predichos y Residuos en el modelo de regresión múltiple

Los estimadores MCO $\hat{\beta}_0, \hat{\beta}_1 ......\hat{\beta}_k$ son los valores de $\beta_0, \beta_1 ......\beta_k$ que minimizan la suma de los errores de predicción al cuadrado $\sum_{i=1}^{n}\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - .... - \hat{\beta}_k X_{ki}\right)^2$. Los valores de predicción MCO y los residuos son:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + ...\hat{\beta}_k X_{ki}, i = 1,...,n$$

$$\hat{u}_{\ddot{i}} = Y_i - \hat{Y}_{\ddot{i}}, i = 1,..n$$

3. Medidas de Ajuste

Actual = predicho + residuo: $Y_i = \hat{Y}_i + \hat{u}_i$

ESR = desviación estándar de $\hat{u}_i$ (con corrección de grados de libertad)
$RMSE$ = desviación estándar de $\hat{u}_i$ (sin corrección de grados de libertad)

R^2 = fracción de la varianza de Y explicada por X

$\bar{R}^2$ = " R^2 ajustado " = R^2 con corrección por los grados de libertad que ajusta por la incertidumbre; $\bar{R}^2 < R^2$

ESR y RMSE

Como en la regresión con un solo regresor, el *ESR* y el *RMSE* son medidas de la dispersión de las *Y*s en torno a la línea de regresión:

$$ESR = \sqrt{\frac{1}{n-k-1}\sum_{i=1}^{n}\hat{u}_i^2}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\hat{u}_i^2}$$

$\boldsymbol{R^2}$ y $\bar{R}^2$

El R^2 es la proporción de varianza explicada

$$R^2 = \frac{SE}{ST} = 1 - \frac{SR}{ST},$$

donde SE = $\sum_{i=1}^{n}(\hat{Y}_i - \bar{\hat{Y}})^2$, $SR = \sum_{i=1}^{n}\hat{u}_i^2$, $ST = \sum_{i=1}^{n}(Y_i - \bar{Y})^2$.

- El R^2 siempre aumenta cuando añades otro regresor (porque*?)* – un cierto problema para una medida de ajuste

R^2 y $\bar{R}^2$

El $\bar{R}^2$ corrige este problema penalizando la inclusión de un regresor adicional – el $\bar{R}^2$ no necesariamente crece cuando añades otro regresor.

$$\textit{Ajustado } R^2\text{:}\quad \bar{R}^2 = 1-\left(\frac{n-1}{n-k-1}\right)\frac{SR}{ST}$$

Notar que $\bar{R}^2 < R^2$, aunque si n es grande los dos estarán muy próximos.

Medidas de ajuste: Ejemplo de las notas:

(1) $\widehat{TestScore} = 698.9 - 2.28 \times \text{REM},$

$$R^2 = .05,\ \text{E}SR = 18.6$$

(2) $\widehat{TestScore} = 686.0 - 1.10 \times \text{REM} - 0.65 PctEL,$

$$R^2 = .426,\ \bar{R}^2 = .424,\ \text{E}SR = 14.5$$

- *¿Qué te dicen estos resultados respecto al ajuste de la regresión (2)comparada con el de la regresión (1)?*
- *¿Por qué los valores de* R^2 *y de* $\bar{R}^2$ están tan próximos en **(***2)?*

4. Supuestos y propiedades de los MCO

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + u_i, \quad i = 1,\ldots,n$$

Como en el caso del modelo con un solo regresor, podemos distinguir tres casos:

Caso 1. Regresores no estocásticos.

Caso 2. Regresores Estocásticos. Homocedasticidad.

Caso 3. Regresores Estocásticos. Heterocedasticidad.

Veamos ahora las propiedades de los estimadores MCO que se obtienen a partir de las ecuaciones normales comentadas anteriormente.

Caso 1. Regresores no Estocásticos. Hipótesis

1. $E\left(u_i\right)=0 \quad i=1,2,..n$
2. Esfericidad
   - $E(u_i u_j)=0 \quad i \neq j$
   - $\operatorname{var}\left(u_i\right)=\sigma^2 \quad \vee i$
3. $u \sim N\left(0,\sigma^2 I_n\right)$
4. No hay multicolinealidad perfecta
5. Ausencia de grandes atípicos
6. $\frac{1}{n}\sum_1^n\left(X_i-\bar{X}_i\right)^2 \xrightarrow{n\to\infty} \sigma_{X_i}^2$

   $\frac{1}{n}\sum_1^n\left(X_i-\bar{X}_i\right)\left(Y_j-\bar{Y}_j\right) \xrightarrow{n\to\infty} \sigma_{X_i X_j}^2$

Caso 1. Regresores no Estocásticos. Propiedades de los Estmadores MCO (Resuelto en cursos anteriores)

1. **Todo Tamaño Muestral**
   - Insesgadez: $E\left(\hat{\beta}_i\right) = \beta_i \quad i = 1, 2, ..k$
   - $\operatorname{var}\left(\hat{\beta}_i\right) = \dfrac{\sigma^2}{nS_i^2\left(1 - R_{i.12..k}^2\right)}, \; S_i^2 = \dfrac{1}{n}\sum_1^n \left(X_i - \bar{X}_i\right)^2$ (Ejercicio 2.11)
   - Los MCO son ELIO
2. **Muestras Grandes**
   - Consistencia: $\hat{\beta}_i \xrightarrow{p} \beta_i$
   - Distribución:

$$\sqrt{n}\left(\hat{\beta}_i - \beta_i\right) \sim N\left(0, \sigma_{\hat{\beta}_i}^2\right) \; con \; \sigma_{\hat{\beta}_i}^2 = \frac{\sigma^2}{\sigma_{X_i}^2\left(1 - R_{i.12..k}^2\right)}$$

   - Eficiente asintóticamente.

Caso 2. Regresores Estocásticos. Homocedasticidad

Hipótesis

1. La media de la distribución condicional de u dadas las X's es cero, esto es, $E(u_i|X_{1i} = x_1,\ldots, X_{ki} = x_k) = 0$.
2. $(X_{1i},\ldots,X_{ki}, Y_i)$, $i = 1,\ldots,n$, son i.i.d.
3. Ausencia de grandes atípicos: $X_1,\ldots, X_k$, y Y tienen momentos de cuarto orden finitos $E(X_{1i}^4) < \infty,\ldots,$ $E(X_{ki}^4) < \infty$, $E(Y_i^4) < \infty$.
4. $\operatorname{var}\left(u_i \setminus X_{1i},...X_{ki}\right) = \sigma^2$
5. No hay multicolinealidad perfecta

Caso 2. Regresores Estocásticos. Homocedasticidad. Propiedades de los MCO

1. **Todo tamaño muestral.** Los estimadores MCO siguen siendo insesgados, pero su distribución de probabilidad ya no es la misma que la del caso de regresores no estocásticos. Adopta una forma más compleja.
2. **Muestras Grandes.** Se mantienen todos los resultados vistos para el caso de regresores no estocásticos

Caso 3. Regresores Estocásticos. Heterocedasticidad
Hipótesis
C.C. 6.4 Los supuestos de los MCO en el modelo de regresión múltiple

1. La media de la distribución condicional de u dadas las X's es cero, esto es, $E(u_i|X_{1i} = x_1,\ldots, X_{ki} = x_k) = 0$.
2. $(X_{1i},\ldots,X_{ki}, Y_i)$, i =1,…,n, son i.i.d.
3. Ausencia de grandes atípicos: $X_1,\ldots, X_k$, y Y tienen momentos de cuarto orden finitos $E(X_{1i}^4) < \infty,\ldots,$ $E(X_{ki}^4) < \infty$, $E(Y_i^4) < \infty$.
4. No hay multicolinealidad perfecta.

Caso 3. Regresores Estocásticos. Homocedasticidad. Propiedades de los estimadores MCO

1. **Todo tamaño muestral.** Los estimadores MCO siguen siendo insesgados, pero su distribución de probabilidad ya no es la misma que la del caso de regresores no estocásticos. Adopta una forma más compleja.
2. **Muestras Grandes.** Los estimadores MCO son consistentes pero el resto de los resultados ya no se mantienen.

Hipótesis #1: La media de la distribución condicional de u dadas las *X*s incluidas es cero.

$$E(u|X_1 = x_1,\ldots, X_k = x_k) = 0$$

- Tiene la misma interpretación que en el modelo con un solo regresor.
- El incumplimiento de esta hipótesis lleva al sesgo de variable omitida (VO). Si una VO
  (1) Está incluida en u *y*
  (2) Está correlacionada con una X incluida
  Entonces esta hipótesis no se cumple y se da el sesgo de VO.
- La mejor solución es incluir, si es posible, esa VO en la regresión.
- Una segunda solución es incluir una variable control.

Hipótesis #2: $(X_{1i},\ldots,X_{ki},Y_i)$, $i = 1,\ldots,n$, son i.i.d.

Se satisfice automáticamente si se recogen los datos utilizando el Muestreo Aleatorio Simple (MAS).

Hipótesis #3: Ausencia de grandes atípicos (momentos de cuarto orden finitos)

Como en el modelo con un solo regresor, los MCO pueden ser muy sensibles a grandes atípicos, por lo que hay que chequear los datos (gráficos!) para estar seguros de que no hay valores muy extraños.

Hipótesis #4: No hay multicolinealidad perfecta

La multicolinealidad perfecta se da cuando uno de los regresores es una función lineal exacta del resto de los regresores.

Ejemplo:Suponer que por error incluyes dos veces la variable REM.

- ***Trampa de la Variable Ficticia***
- ***Multicolinealidad imperfecta***

La distribución muestral de los estimadores MCO

C.C. 6.5 Distribución para muestras grandes de $\hat{\beta}_0, \hat{\beta}_1, ..\hat{\beta}_k$

- Si se cumplen los cuatro hipótesis del Concepto Clave 6.4, entonces en muestras grandes los estimadores MCO $\hat{\beta}_0, \hat{\beta}_1, ..\hat{\beta}_k$ están distribuidos normalmente de forma conjunta y cada $\hat{\beta}_j$ se distribuye $N\left(\beta_j, \sigma^2_{\hat{\beta}_j}\right)$, $j = 0,..k$.

¡Conceptualmente, nada nuevo bajo el sol!

5. Contraste de Hipótesis e Intervalos de Confianza

1. Un solo Coeficiente. Sigue la misma lógica vista para el modelo lineal simple

C.C.7.1 Contraste de la hipótesis $\beta_j = \beta_{j,0}$ frente a la alternativa $\beta_j \neq \beta_{j,0}$.

1. **Definir el estadístico de contraste.**

2. **Derivar la distribución de probabilidad de este estadístico bajo la hipótesis nula.**

3. **Determinar la región crítica del contraste.**

1. Definir el estadístico de contraste

Utilizaremos el t-ratio definido como

$$t = \frac{\hat{\beta}_j - \beta_{j,0}}{ES\left(\hat{\beta}_j\right)}$$

En donde ES$\left(\hat{\beta}_j\right)$ es el error estándar del estimador definido como la raíz cuadrada de la varianza. Este error estándar dependerá de si estamos en el caso homocedástico o heterocedástico. Ambos siguen el mismo proceso visto para el modelo con un solo regresor. Solo prestaremos atención al caso homocedástico.

Fórmula para el ES($\hat{\beta}_j$):Homocedasticidad

$$\mathrm{ES}(\hat{\beta}_j) = \sqrt{\frac{1}{n} \times \frac{\frac{1}{n-2}\sum_{i=1}^{n} \hat{u}_j^2}{\frac{1}{n}\sum_{j=1}^{n} (X_j - \bar{X})^2 \left(1 - R_{j.12..k}^2\right)}}.$$

2. Derivar la distribución del estadístico de contraste bajo la hipótesis nula.

Al hablar del estimador MCO hemos obtenido que

$$\hat{\beta}_j \sim N\left(\beta_j, \sigma^2_{\hat{\beta}_j}\right)$$

Por lo tanto, bajo la hipótesis nula tenemos

$$\hat{\beta}_j \sim N\left(\beta_{j,0}, \sigma^2_{\hat{\beta}_j}\right)$$

Por lo que $\dfrac{\hat{\beta}_j - \beta_{j,0}}{\sigma_{\hat{\beta}_j}} \sim N(0,1)$

En muestras grandes el error estándar del estimador será un estimador consistente de la desviación típica del estimador por lo que $t \sim N(0,1)$.

En muestras pequeñas $t \sim t(student)$

3. Determinación de la región crítica

.

Rechazar al nivel de significación del 5% si $|t| > 1.96$, o, equivalentemente, calcular el p-valor, $p = \Pr[|t| > |t^{act}|] =$ probabilidad de las colas de la Normal correspondientes a $|t^{act}|$ y rechazar al nivel del 5% si el p-valor es < 5%. Notar tambien que

$$\text{p-valor} = 2\Phi\left(-\left|t^{act}\right|\right)$$

donde t^{act} es el valor del estadístico t calculado.

Ejemplo

Datos de California

(1) $\widehat{NotaExamen} = 698.9 - 2.28 \times REM$

$\qquad (10.4) \quad (0.52)$

(2) $\widehat{NotaExamen} = 686.0 - 1.10 \times REM - 0.650 PctEL$

$\qquad (8.7) \quad (0.43) \qquad\qquad (0.031)$

- La estimación del coeficiente de REM es el efecto sobre la calificación de una variación unitaria de REM manteniendo constante la proporción que aprende inglés.
- La estimación cae a la mitad.
- El intervalo de confianza es

$$\{-1.10 \pm 1.96 \times 0.43\} = (-1.95, -0.26)$$

- El estadístico-t es $t = -1.10/0.43 = -2.54$, rechazamos

Salida de Gretl (varianzas robustas)

Modelo 1: MCO, usando las observaciones 1-420
Variable dependiente: NotaExamen
Desviaciones típicas robustas ante heterocedasticidad, variante HC1

	Coeficiente	Desv. Típica	Estadístico t	Valor p	
const	686.032	8.72822	78.60	3.98e-252	***
REM	-1.10130	0.432847	-2.544	0.0113	**
PctEL	-0.649777	0.0310318	-20.94	4.70e-067	***

Media de la vble. dep.	654.1565	D.T. de la vble. dep.	19.05335
Suma de cuad. residuos	87245.29	D.T. de la regresión	14.46448
R-cuadrado	0.426431	R-cuadrado corregido	0.423680
F(2, 417)	223.8229	Valor p (de F)	9.28e-67
Log-verosimilitud	-1716.561	Criterio de Akaike	3439.123
Criterio de Schwarz	3451.243	Crit. de Hannan-Quinn	3443.913

Salida de Gretl (homocedasticidad)

Modelo 2: MCO, usando las observaciones 1-420
Variable dependiente: NotaExamen

	Coeficiente	Desv. Típica	Estadístico t	Valor p	
const	686.032	7.41131	92.57	3.87e-280	***
REM	-1.10130	0.380278	-2.896	0.0040	***
PctEL	-0.649777	0.0393425	-16.52	1.66e-047	***

Media de la vble. dep.	654.1565	D.T. de la vble. dep.	19.05335
Suma de cuad. residuos	87245.29	D.T. de la regresión	14.46448
R-cuadrado	0.426431	R-cuadrado corregido	0.423680
F(2, 417)	155.0136	Valor p (de F)	4.62e-51
Log-verosimilitud	-1716.561	Criterio de Akaike	3439.123
Criterio de Schwarz	3451.243	Crit. de Hannan-Quinn	3443.913

2. Contraste de Hipótesis conjuntas

El modelo es

$$NotaExamen_i = \beta_0 + \beta_1 REM_i + \beta_2 Gasto_i + \beta_3 PctEL_i + u_i$$

Expn(Gasto) es el gasto total anual por alumno en el distrito.

La hipótesis nula es que los recursos del colegio no importan

$$H_0\text{: } \beta_1 = 0 \; \boldsymbol{y} \; \beta_2 = 0$$

$$\text{vs. } H_1\text{: } \boldsymbol{\acute{o}} \; \beta_1 \neq 0 \text{ o } \beta_2 \neq 0 \; \boldsymbol{o\ ambos}$$

- ***Una hipótesis conjunta*** *especifica un valor para dos o más coeficientes.*

- En general, una hipótesis conjunta implicará q restricciones. En el ejemplo anterior , $q = 2$, y las dos restricciones son $\beta_1 = 0$ y $\beta_2 = 0$.
- Una primera idea es rechazar la hipótesis conjunta si se rechaza cada hipótesis individual utilizando el estadístico-t..
- Pero esta estrategia "de uno en uno" no es válida porque el porcentaje de rechazos de la hipótesis nula es superior al nivel de significación adoptado.

Suponiendo que los contrastes individuales son independientes la H_0 no se rechaza si tanto $|t_1| \leq 1.96$ como $|t_2| \leq 1.96$. Por ser independientes t_1 y t_2, se tiene:

$$= 1 - \Pr_{H_0}[|t_1| \leq 1.96 \text{ and } |t_2| \leq 1.96]$$

$$= 1 - \Pr_{H_0}[|t_1| \leq 1.96] \times \Pr_{H_0}[|t_2| \leq 1.96]$$

$$= 1 - (0.95)^2$$

$= 0.0975 = 9.75\%$ – que no es el 5% adoptado de partida.

Dos Soluciones:

- Utilizar un punto crítico diferente: Método de Bonferroni.
- Utilizar el estadístico-F. (Este es el más utilizado)

El Contraste -F

$$F = \frac{1}{2}\left(\frac{t_1^2 + t_2^2 - 2\hat{\rho}_{t_1,t_2} t_1 t_2}{1 - \hat{\rho}_{t_1,t_2}^2}\right)$$

- Rechazar si F es grande (¿cuán grande?).

Su distribución para muestras grandes es la chi-cuadrado con q grados de libertad

La distribución chi-cuadrado con q grados de libertad (χ_q^2) es la distribución de la suma de q variables normales estandarizadas elevadas al cuadrado.

Algunos valores críticos de χ_q^2/q

q	5% valor crítico
1	3.84
2	3.00
3	2.60
4	2.37
5	2.21

Calculando el p-valor cuando se utiliza el estadístico-F.

p-valor= probabilidad de la cola de una distribución χ_q^2/q *que corresponde al valor del estadístico-F calculado.*

Modelo 3: MCO, usando las observaciones 1-420
Variable dependiente: NotaExamen
Desviaciones típicas robustas ante heterocedasticidad, variante HC1

	Coeficiente	Desv. Típica	Estadístico t	Valor p	
const	649.578	15.4583	42.02	8.64e-152	***
REM	-0.286399	0.482073	-0.5941	0.5528	
PctEL	-0.656023	0.0317844	-20.64	1.11e-065	***
Gasto	0.00386790	0.00158072	2.447	0.0148	**

Media de la vble. dep.	654.1565	D.T. de la vble. dep.	19.05335
Suma de cuad. residuos	85699.71	D.T. de la regresión	14.35301
R-cuadrado	0.436592	R-cuadrado corregido	0.432529
F(3, 416)	147.2037	Valor p (de F)	5.20e-65
Log-verosimilitud	-1712.808	Criterio de Akaike	3433.615
Criterio de Schwarz	3449.776	Crit. de Hannan-Quinn	3440.003

Calculo del p-valor

A partir de los resultados presentados se puede calcular el valor del estadístico-F. Pero Gretl proporciona el valor calculado a partir de la secuencia

Contrastes/omitir variables

Y se abre una ventana que pide elegir las variables con coeficiente cero. Los resultados son:

F=5,43 p-valor=0,0046

Por lo tanto, la hipótesis nula se rechaza.

Estadístico-F homocedástico

Hay una formula simple para el estadístico-F que sólo es válida bajo homocedasticidad (por lo que no es muy útil) pero puede ayudar a entender lo que el contraste-F está haciendo.

Formula simple para el Estadístico-F homocedástico

$$F = \frac{(R_A^2 - R_R^2)/q}{(1 - R_A^2)/(n - k_A - 1)}$$

donde:

R_R^2= el R^2 para la regresión restringida

R_A^2 = el R^2 para la regresión sin restringir

q = el número de restricciones bajo la nula

k_A = el número de regresores en la regresión sin restringir

Resumen de Contraste Conjunto

- El contraste F homocedástico y la distribución F solo se justifican en situaciones muy especiales poco habituales en la práctica.
- Hay que utilizar el estadístico-F heterocedástico con los valores críticos de χ_q^2/q (esto es, $F_{q,\infty}$).
- Para $n \geq 100$, la distribución *F esencialmente se identifica con la distribución* χ_q^2/q .

Para valores pequeños de n, algunos investigadores utilizan la distribución F porque tiene valores críticos mayores y, en este sentido, es más conservador.

5. Variables de control

Queremos obtener una estimación insesgada del efecto sobre las calificaciones de un cambio en el tamaño de la clase, manteniendo constantes factores que el consejo escolar no controla, como oportunidades de aprendizaje externas (museos, ..), o implicación de los padres en la educación..

Si pudiéramos realizar un experimento, asignaríamos aleatoriamente a los estudiantes y a los maestros a clases de diferente tamaño. Entonces, REM_i sería independiente de todos los factores incluidos en u_i, por lo que $E(u_i / REM_i) = 0$ y el estimador MCO del coeficiente de REM será un estimador insesgado del efecto causal deseado.

Pero con datos observacionales (observación pasiva), u_i depende de factores adicionales (museos, implicación familiar, conocimiento de ingles,…). Algunos de estos factores pueden estar correlacionados con alguno de los regresores y esto nos lleva al **sesgo de variable omitida.**

Las pautas a seguir

- Si puedes observar esos factores entonces se incluyen en la regresión.
- Pero tu no puedes observar todos estos factores (la implicación de los padres). **En este caso, tu puedes incluir "variables control" correlacionadas con estos factores causales omitidos pero que en si mismos no tienen ese carácter causal.**

Una **Variable Control, *W*,** es una variable que está correlacionada con, y que controla a, un factor causal omitido en la regresión de Y sobre X, pero que W no tiene un efecto causal sobre Y.

Variables Control: aplicación al ejemplo.

$$\widehat{NotaExamen} = 700.2 - 1.00REM - 0.122PctEL - 0.547LchPct, \quad \bar{R}^2 = 0.773$$
$$(5.6)\quad (0.27)\qquad (.033)\qquad\qquad (.024)$$

$PctEL$ = % de estudiantes de inglés en el distrito.

$LchPct$ = % de estudiantes que reciben un desayuno libre o subsidiarizado (solamente lo pueden recibir familias de baja renta).

- ¿Qué variable es la variable de interés?
- ¿Qué variables son variables control? ¿Tienen componentes Causales? ¿Qué es lo que controlan?

- REM es la variable de interés.
- *PctEL* es, al mismo tiempo, una variable causal y una variable control. Sin duda tiene un efecto causal ya que se aprende mejor con un buen inglés, pero también está relacionada con factores causales ocultos, por ejemplo oportunidades de aprendizaje fuera del colegio.
- *LchPct* Tiene también este doble carácter, causal y de control

1. **Tres puntos intercambiables que hacen que una variable sea una variable de control de forma efectiva:**
   i. Un variable es realmente una variable de control cuando su inclusión en la regresión logra que el término de error no esté correlacionado con la variable de interés.
   ii. Manteniendo constante la variable de control es “como si” la variable de interés se asignara aleatoriamente..
   iii. Entre los individuos con el mismo valor de la variable control la variable de interés no está correlacionada con los determinantes omitidos de Y.

- **Variables Control. Independencia en media condicional**

Sea X_i la variable de interés y sea W_i la variable de control. Entonces, W es una variable de control efectiva si la independencia en media condicional se cumple; es decir:

$E(u_i|X_i, W_i) = E(u_i|W_i)$ (independencia en media condicional)

Si W es una variable control, entonces el cumplimiento de la independencia en media condicional garantiza el cumplimiento del primer supuesto de los MCO, MCO #1

Variables Control. Independencia en media condicional

Considerar el modelo de regresión,

$$Y = \beta_0 + \beta_1 X + \beta_2 W + u$$

En donde X es la variable de interés y W es una variable de control efectiva de forma que se cumple la independencia en media condicional; Entonces,

$$E(u_i|X_i, W_i) = E(u_i|W_i).$$

Además suponer que se cumplen MCO #2, #3, y #4. Entonces:

1. β_1 tiene una interpretación causal.
2. $\hat{\beta}_1$ es insesgado.
3. El estimador del coeficiente de la variable control, $\hat{\beta}_2$, en general es sesgado.

Reglas para la especificación

1. Identificar la variable de interés.

2. Pensar en los efectos causales omitidos que pueden conducir al sesgo de variable omitida.

3. Si puedes incluye los efectos causales omitidos. Si no puedes incluye variables correlacionadas con ellos que hacen de variables control que sean realmente efectivas. Este es el modelo “base”

4. Especificar también un conjunto de modelos alternativos plausibles.

5. Estimar el modelo base y las especificaciones alternativos plausibles y llevar a cabo “chequeos de sensibilidad”
   - ¿Cambia la inclusión de una variable la estimación del coeficiente de la variable de interés (β_1)?
   - ¿Es una variable candidata estadísticamente significativa?
   - Usa el juicio, no mecánicamente el Gretl…
   - No trates de maximizar solo el R^2!

Análisis de la base de datos del ejemplo

1. Identificar la variable de interés: *REM*

2. Pensar en los efectos causales omitidos que pueden llevar al sesgo de variable omitida

   Si los estudiantes saben inglés; oportunidades de aprendizaje externas al colegio; implicación de los padres; calidad del profesor-hay una larga lista¡.

3. Incluir esos efectos causales omitidos o, si no se puede, incluir las variables correlacionadas con ellos que sirven como variable control. Las variables control son efectivas si la hipótesis de independencia en la media condicional puede aceptarse (si *u* no está correlacionada con REM una vez incluidas las variables control). Este es el modelo "base".

   Muchas de las variables causales omitidas no se pueden medir por lo que necesitamos encontrar variables control. Estas incluyen PctEL (que es al mismo tiempo una variable control y un factor causal omitido)y medidas de riqueza de cada distrito.

4. Estimar también un grupo de modelos alternativos que incluyen otras variables candidatas.

No está claro que variables de renta serán más efectivas como variables control para los muchos factores causales omitidos tales como las oportunidades de aprendizaje externas, por lo que se consideran modelos con variables-renta alternativas. Las especificaciones alternativas consideradas en este ejemplo constituyen solamente un punto de partida, no el final del mundo¡

5. Se estima el modelo base y las alternativas y se lleva a cabo un chequeo de sensibilidad.

Discusión sobre como presentar resultados

- Hemos estimado varios modelos y queremos informar de los resultados. Sería difícil y poco útil presentarlos en formato de ecuación. Creemos que es más fácil y útil presentar los resultados en una Tabla.
- Una tabla con los resultados de una regresión debería de incluir:
    - Las estimaciones de los coeficientes
    - Los errores estándar
    - Medidas de ajuste
    - Número de observaciones
    - Valor de estadísticos-F si los hubiere

Dependent variable: average test score in the district.

Regressor	(1)	(2)	(3)	(4)	(5)
Student–teacher ratio (X_1)	−2.28** (0.52)	−1.10* (0.43)	−1.00** (0.27)	−1.31** (0.34)	−1.01** (0.27)
Percent English learners (X_2)		−0.650** (0.031)	−0.122** (0.033)	−0.488** (0.030)	−0.130** (0.036)
Percent eligible for subsidized lunch (X_3)			−0.547** (0.024)		−0.529** (0.038)
Percent on public income assistance (X_4)				−0.790** (0.068)	0.048 (0.059)
Intercept	698.9** (10.4)	686.0** (8.7)	700.2** (5.6)	698.0** (6.9)	700.4** (5.5)
Summary Statistics					
SER	18.58	14.46	9.08	11.65	9.08
$\overline{R}^2$	0.049	0.424	0.773	0.626	0.773
n	420	420	420	420	420

These regressions were estimated using the data on K-8 school districts in California, described in Appendix 4.1. Heteroskedasticity-robust standard errors are given in parentheses under coefficients. The individual coefficient is statistically significant at the *5% level or **1% significance level using a two-sided test.

Resumen: Regresión Múltiple

- La regresión múltiple te permite estimar el efecto sobre Y de un cambio en X_1, manteniendo constante las otras variables incluidas.
- Si puedes medir una variable, puedes evitar el sesgo de variable omitida incluyéndola en la regresión.
- Si no puedes medir la variable omitida aun puedes neutralizarla incluyendo una variable control.
- No hay ningún procedimiento mecánico simple para saber que variables se incluyen. Debes ser juicioso.
- Un enfoque consiste en especificar un modelo base y después explorar la robustez de las estimaciones clave en especificaciones alternativas.