1Teoria.pdf

jmgg94

Econometría

3º Doble Grado en Administración y Dirección de Empresas e Informática

Facultad de Ciencias Económicas y Empresariales Universidad de Granada

Descarga la APP de Wuolah.
Ya disponible para el móvil y la tablet.

Available on the App Store
GET IT ON Google Play

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

KEEP CALM AND ESTUDIA UN POQUITO

WUOLAH + #QuédateEnCasa

Descarga la APP de Wuolah.
Ya disponible para el móvil y la tablet.
Available on the App Store
GET IT ON Google Play

TEMA 1. Introducción a la Econometría

1.1. Econometría y modelos econométricos.

- Concepto de Econometría.

A continuación se describen definiciones destacables de Econometría realizadas por algunos autores:

- **Frisch (1933):** "La Econometría se desmarca de campos como la Estadística Económica o la Teoría Económica General, ni es considerada como un sinónimo de la aplicación de las matemáticas a la economía. Estos tres campos comentados son necesarios para la Econometría, pero ninguno por si mismo permite entender las relaciones cuantitativas de la vida económica moderna".
- **Tintner (1952):** "La Econometría consiste en la aplicación de la Teoría Económica, matemática y de los métodos estadísticos a los datos económicos, para establecer resultados numéricos en el campo de la Economía y verificar los teoremas económicos".
- **Gollnick (1973):** "La Econometría combina la teoría económica, las matemáticas y la estadística para cuantificar y verificar las relaciones entre las variables económicas".
- **Pindick y Rubinfeld (1980). "**La ciencia de la elaboración de modelos consiste en un conjunto de instrumentos, en su mayoría cuantitativos que se utilizan para construir y contrastar representaciones matemáticas del mundo real. El desarrollo y la utilización de dichos instrumentos constituyen el objeto de la Econometría".

En resumen, se puede decir que la **Econometría** es la ciencia que utiliza la *Teoría Económica, Matemática y la Inferencia Estadística de forma conjunta, y utilizando como herramienta la informática* (véase la Figura 1.1).

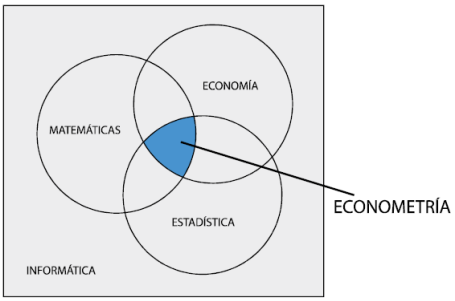


Figura 1.1. *Diagrama de Venn para describir el concepto de Econometría.*

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.
WUOLAH

Modelos económicos y modelos econométricos

- **Modelo Económico:** modelo basado en la representación algebraica simplificada de una realidad económica. En general, los modelos económicos se describen a través de ecuaciones matemáticas que relacionan distintos tipos de variables.
- **Modelo Econométrico:** modelo económico que contienen todos los elementos necesarios para realizar su aplicación práctica. Para ello, también se requiere describir matemáticamente la relación entre una o más variables explicadas en función de una o varias variables explicativas, pero además, se debe añadir un término conocido como término aleatorio o perturbación aleatoria.

Ejemplo 1.1. ***Modelo Económico y Modelo Econométrico: Concesionario de vehículos.***

Una conocida marca de vehículos quiere conocer el efecto que varios factores (tal como los gastos en publicidad, la situación económica del país, el precio de la competencia o el precio de los vehículos de la propia marca, etc) tendrán sobre sus ventas de vehículos.

- **Definición de variables:**

Variable	Definición
y	Ventas de vehículos (en miles de euros)
x_2	Gastos en publicidad (en miles de euros)
x_3	Precio (en miles de euros)
x_4	Precio de la competencia (en miles de euros)
x_5	Situación económica

- **Modelo Económico**

Un modelo económico podría ser el siguiente:

$$y = f(x_2, x_3, x_4, x_5, \cdots)$$

Se observa que en dicho modelo se reflejan las relaciones entre las variables, pero sería más interesante cuantificar la relación entre estas variables. Por ejemplo, resultaría relevante para la empresa conocer cuánto aumentarían las ventas en el caso de aumentar los gastos en publicidad en 1000 euros. Además, es posible que existan otros factores desconocidos que de alguna forma incidan en las ventas de la empresa, y este modelo económico no recoge esos otros factores. Por éstas y otras razones, el modelo económico planteado no resulta útil.

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

WUOLAH

Descarga la app de Wuolah desde tu store favorita

- **Modelo Econométrico**

La solución consiste en transformar el modelo económico en un modelo econométrico, en el cual si resulta posible conocer la variación en las ventas ante un aumento o una disminución de alguno de los factores planteados factores (gastos en publicidad, precio de la competencia, precio del propio concesionario de vehículos, etc). Además, en el modelo econométrico se podrán recoger otros factores que puedan influir sobre las ventas y que por alguna razón no están recogidos de forma explícita en el modelo. Un ejemplo muy común de modelo econométrico es el modelo de regresión lineal múltiple, el cual vendría dado para este ejemplo a partir de la siguiente expresión:

$$y = \beta_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + u$$

A continuación se realizan algunas observaciones sobre este modelo econométrico:

- A la variable de interés de la empresa (ventas) se le puede denominar ***variable dependiente, regresada, endógena, explicada,*** etc.
- Al resto de variables que se incluyen en el modelo de regresión lineal y que se supone que pueden tener algún efecto sobre la variable endógena se las denominan ***variables independientes, regresoras, exógenas, explicativas***, etc. En la Tabla 1.1 se resumen el nombre que suelen tener las variables de un modelo econométrico.

y	***x***
Variable explicada	Variable explicativa
Variable dependiente	Variable independiente
Regresada	Regresora
Endógena	Exógena
Variable de respuesta	Variable de control
Predicha	Predictora

Tabla 1.1. *Denominación de las variables de un modelo econométrico.*

- Los coeficientes que acompañan a las variables exógenas $(\beta_1, \beta_2, \beta_3, \ldots)$ se denominan ***coeficientes del modelo***. Los coeficientes del modelo son ***constantes y desconocidos***, y su objetivo es cuantificar la intensidad con la que su correspondiente variable exógena influye sobre la variable endógena.

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

WUOLAH

Descarga la app de Wuolah desde tu store favorita

- **Uno de los objetivos principales** es la estimación de los coeficientes del modelo a partir de una muestra de observaciones para todas las variables especificadas en el modelo de regresión lineal (endógena y exógenas). Asumiremos que la muestra tendrá un tamaño *n*. A las estimaciones de los coeficientes del modelo se les denotará como $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3$, etc. A lo largo del curso se explicará el método más común para la estimación de los coeficientes del modelo. Es realmente importante entender las diferencias entre los coeficientes del modelo y sus estimaciones. A diferencia de los valores verdaderos, ***las estimaciones no son constantes***, puesto que dependen de las observaciones recogidas en la muestra, y ***además son valores conocidos*** que se obtienen a partir de la información muestral.
- El modelo econométrico que hemos formulado para explicar las ventas de vehículos ***no es el verdadero***, es decir, se está planteando un modelo alternativo al verdadero modelo que pretende explicar las ventas del concesionario de vehículos. En otras palabras, existen otros factores o variables que pueden tener una influencia sobre las ventas de vehículos, bien por la imposibilidad de recoger dichos datos, o bien, por el desconocimiento de la existencia de dichos factores. Estas variables que influyen en las ventas no están recogidas de forma directa en el modelo de regresión lineal. Por este motivo se incluye en el modelo en término *u*, el cual se denomina ***término del error o perturbación aleatoria***. En resumen, el termino *u* representa a todos los demás factores que pueden influir en la ventas de la empresa, y los cuales no estamos incluyendo explícitamente en el modelo. Por ejemplo, supongamos que otro investigador plantea otro modelo de regresión lineal simple y no incluye en el modelo el precio de los vehículos de la competencia, por lo que su modelo quedaría del siguiente modo:

$$y = \beta_1^* + \beta_2^* x_2 + \beta_3^* x_3 + \beta_5^* x_5 + u^*$$

En este modelo la variable x_4 (precio de los vehículos de la competencia) no está incluida de forma directa en el modelo, pero este factor ignorado por el investigador quedará recogido a través del término error u^*.

□

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

WUOLAH

Descarga la app de Wuolah desde tu store favorita

Descarga la APP de Wuolah.
Ya disponible para el móvil y la tablet.
Available on the App Store
GET IT ON Google Play

<u>Ejemplo 1.2.</u> ***Modelo Econométrico para explicar la calificación final en la asignatura de Econometría.***

Variable	Definición
y	Calificación final en la asignatura de Econometria
x_2	Número de horas de estudio a la semana
x_3	Grado de asistencia a clase
x_4	Grado de participación en clase
x_5	Calidad del material disponible para prepararse la asignatura

Modelo Econométrico 1:

$$y = \beta_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + u$$

Modelo Econométrico 2:

$$y = \beta_1^* + \beta_2^* x_2 + \beta_5^* x_5 + u^*$$

□

De una forma más general, un modelo de regresión lineal basado en $k-1$ variables exógenas se expresa del siguiente modo:

$$y = \beta_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots + \beta_k x_k + u,$$

donde

- y *(variable endógena)*: es la variable de interés, y viene explicada en el modelo en función de una serie de variables exógenas. Uno de los objetivos que se persiguen es realizar predicciones para esta variable endógena.
- $x_2, \dots, x_k$ (*variables exógenas*): son las $k-1$ variables que consideramos relevantes para explicar la variable endógena.
- $\beta_1, \beta_2, \beta_3, \dots, \beta_k$ *(coeficientes del modelo o parámetros de posición)*: son constantes y desconocidas, y en total se tienen un total de k parámetros de posición. Al ser desconocidos, estos coeficientes deben estimarse a partir de los datos recogidos en una muestra (de tamaño n) para todas las variables, es decir, tanto las variables exógenas como la endógena son observadas y su información se recoge a partir de la muestra realizada. Los coeficientes $\beta_2, \beta_3, \dots, \beta_k$ miden el efecto de la correspondiente variable exógena sobre la variable endógena, manteniendo fijas el resto de variables exógenas. Al coeficiente β_1 se le denomina constante o intercepto y no tiene una interpretación económica. Se incluye en el modelo con el objetivo de ajustar o igualar las medias de las variables.

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.
WUOLAH

- u *(término de error o perturbación aleatoria)*: en este término se representan todos los factores que no han sido incluidos en el modelo de regresión lineal. Además, el término error se incluye por el hecho de que el modelo de regresión lineal no es el verdadero modelo, dado que en realidad se está planteando una aproximación lineal del modelo verdadero que permite explicar correctamente la variable endógena. El termino error no se puede observar.
- $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_{3,} \dots \hat{\beta}_k$ *(estimaciones de los coeficientes del modelo).* Una vez que se obtienen las estimaciones de los coeficientes del modelo se puede formular el modelo estimado de la siguiente forma:

$$\hat{y} = \hat{\beta}_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3 + \cdots + \hat{\beta}_k x_k.$$

Es conveniente indicar en el modelo que estamos haciendo referencia a un instante "*t*" en el tiempo (en este caso asumiremos un total de *T* observaciones), o bien a la observación "*i*" de una muestra, la cual asumiremos que tiene tamaño *n*. Esta etiqueta (*t* o *i*) se deberá incluir en el modelo de forma de subíndice en cada variable del modelo, es decir, el Modelo de Regresión Lineal General (MRLG) se puede expresar como:

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \cdots + \beta_k x_{kt} + u_t, \quad t = 1, \dots, T$$

o bien

$$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \cdots + \beta_k x_{ki} + u_i, \quad i = 1, \dots, n$$

Motivos para incluir en el modelo econométrico el término error:

1) El término error es la forma de incluir en el modelo propuesto el resto de factores no incluidos expresamente en dicho modelo y que influyen en la variable endógena. La inclusión del término error se debe, entre otras razones, a que:
    - Nunca resultará posible disponer de un listado completo de los factores que afectan a la variable endógena *y*.
    - No es posible disponer de la información de las variables exógenas que afectan a la variable endógena.

2) Principio de simplicidad. Es posible que interese dejar fuera del modelo determinados factores o variables exógenas por ser poco relevantes, y por tanto, estos factores excluidos estarán representados a través del término error.

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

WUOLAH

Descarga la app de Wuolah desde tu store favorita

3) El término error explica el carácter aleatorio de la conducta humana.

4) En numerosas ocasiones sólo tenemos acceso a una variable cercana o similar (la cual denominaremos variable proxy) a la variable exógena que realmente se debe incluir en el modelo. En este caso, el término error también incluye la diferencia entre la variable exógena verdadera y la variable proxy. Por ejemplo, la variable "habilidad innata del trabajador" es un factor relevante que puede afectar en un modelo donde la variable endógena sea el salario. Sin embargo, esta variable exógena no es observable, y en su lugar se puede utilizar una variable exógena similar, tal como el "resultado del trabajador en un test de cociente intelectual".

Nota para el profesor: Ver Tema 1 de la App: Histograma del término error con distintas varianzas, nube de puntos, y ejemplo: Horas de estudio a la semana y calificación en Econometría.

Ejemplo 1.3. ***Efecto del número de profesores y estudiantes en el rendimiento académico de estudiantes de Ciencias Económicas y Empresariales.***

Se desea analizar si variaciones del número de profesores o estudiantes en distintas Facultades en Ciencias Económicas y Empresariales pueden tener un efecto en el rendimiento académico de estos centros. Para responder a esta cuestión, un investigador selecciona de forma aleatoria a 50 facultades de todo el país, y decide considerar el siguiente modelo econométrico:

$$y_i = \beta_1 + \beta_2 \ x_{2i} + \beta_3 \ x_{3i} + u_i$$

donde:

- y_i: es la nota media de los estudiantes de la *i*-ésima Facultad en el curso académico anterior, con $i = 1, \dots, 50$.
- x_{2i}: es el número de profesores de la *i*-ésima Facultad en el curso académico anterior, con $i = 1, \dots, 50$.
- x_{3i}: es el número de estudiantes de la *i*-ésima Facultad en el curso académico anterior, con $i = 1, \dots, 50$.

□

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

WUOLAH

Descarga la app de Wuolah desde tu store favorita

Un objetivo adicional en Econometría es la interpretación de resultados, y en particular, analizar el efecto de cada una de las variables exógenas incluidas en el modelo sobre la variable endógena, es decir, cuantificar el efecto que causa una variación del valor de una determinada variable exógena sobre la variable endógena. A este tipo de análisis se le denomina análisis marginal, dado que se debe realizar de forma individual para cada variable exógena, y asumiendo que el resto de variables exógenas permanecen fijas o no sufren variaciones.

- En el Ejemplo 1.1 (concesionario de vehículos) se podría analizar: ¿qué efecto producirá un aumento en los gastos de publicidad en las ventas de vehículos, y manteniendo los mismos precios y para una misma situación económica?, o bien, ¿qué efecto producirá una disminución del precio de los vehículos en las ventas del concesionario, y manteniendo los mismos gastos en publicidad y para una misma situación económica?

- En el Ejemplo 1.3 (rendimiento académico de Facultades en Ciencias Económicas y Empresariales) se podría analizar: ¿qué efecto se producirá en la nota media de los estudiantes cuando haya un aumento del número de profesores en estos centros y con el mismo número de estudiantes?

1.2. Fases del método econométrico y componentes de un modelo econométrico.

- Fases del método econométrico

1. **Especificación.** En esta fase se establece en forma matemática, mediante una o varias ecuaciones (una ecuación para cada variable endógena), la relación entre las variables en estudio, incluyendo la/s perturbación/es aleatoria/s. Además deben establecerse las hipótesis estadísticas sobre el comportamiento de la/s perturbación/es aleatoria/s.

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

WUOLAH

Descarga la app de Wuolah desde tu store favorita

Descarga la APP de Wuolah.
Ya disponible para el móvil y la tablet.
Available on the App Store
GET IT ON Google Play

2. **Estimación.** Una vez especificado el modelo, se procede a la estimación de cada uno de los parámetros que lo componen. Para ello se precisa de datos de las variables endógenas y exógenas recogidos en una muestra aleatoria, y de la selección de un método de estimación apropiado.

3. **Verificación o validación.** Consiste en el diagnóstico del modelo para detectar errores de especificación del modelo, tal como:
   - Que los estimadores sean aceptables desde el punto de vista teórico y estadístico.
   - Que el modelo sea aceptable o razonable
   - Que se verifiquen las hipótesis estadísticas de partida

   En el caso de que el modelo no supere satisfactoriamente esta fase, será necesario realizar de nuevo la Fase 1 con el objetivo de proporcionar una nueva especificación del modelo que se ajuste mejor al estudio que se desea realizar.

4. **Explotación.** En esta fase se realizará el denominado análisis estructural (para cuantificar la relación entre las variables y poder entender mejor el fenómeno económico), predicción condicionada (obtener predicciones de las variables endógenas a partir de los valores conocidos de las variables exógenas) y/o evaluación de medidas de política económica (simulación).

- Componentes de un modelo econométrico

Recordemos que el Modelo de Regresión Lineal General (MRLG) basado en $k-1$ variables exógenas que se estudia en esta asignatura se ha definido como:

$$y = \beta_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots + \beta_k x_k + u.$$

<u>Ejemplo 1.4.</u> ***Modelo econométrico de renta nacional.***

$$\begin{aligned} c_t &= \alpha_1 + \alpha_2(1-\tau)y_t + \alpha_3 r_t + u_t \\ i_t &= \beta_1 + \beta_2(y_{t-1} - y_{t-2}) + \beta_3 r_{t-1} + v_t \\ y_t &= c_t + i_t + g_t \end{aligned}$$

donde:

- *y:* denota el producto interior bruto del país.
- *r:* es el tipo de interés.
- *c*: son los gastos de consumo.
- *i*: son los gastos de inversión.
- *g*: son los gastos del gobierno.
- τ: es el tipo impositivo. □

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.
WUOLAH

A) VARIABLES.

Las variables que intervienen en un modelo econométrico se pueden clasificar en dos grandes grupos:

1. **Variables observables.** Son aquellas variables de las que se disponen de datos muestrales. A su vez, las variables observables se dividen en:

   - **Endógenas corrientes o variables endógenas.** Son las variables explicadas por el modelo econométrico.

   - **Predeterminadas o variables explicativas o regresores.** Son las variables que describen el comportamiento de la/s variable/s endógenas. Según la naturaleza de las variables predeterminada se puede realizar la siguiente clasificación de éstas:
     - **Endógenas retardadas.**
     - **Exógenas corrientes.**
     - **Exógenas retardadas.**

2. **Variables no observables**. Son variables para las cuales no se dispone de datos. Dentro de este grupo se incluyen las perturbaciones aleatorias.

En el Modelo de Regresión Lineal General (MRLG), y es una variable endógena corriente, $x_2, \dots, x_k$ son variables exógenas corrientes, y u es por tanto la perturbación aleatoria, la cual es una variable no observable.

Por otra parte, en el modelo de renta nacional (Ejemplo 1.4) se plantean 3 ecuaciones. Se puede observar que c_t es una variable endógena corriente en la ecuación 1, mientras que es una variable exógena corriente en la ecuación 3. La variable y_t es una exógena corriente en la ecuación 1, endógena corriente en la ecuación 3, y exógena retardada en la ecuación 2. Las variables nos observables son las perturbaciones aleatorias: u_t, v_t y g_t.

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

WUOLAH

Descarga la app de Wuolah desde tu store favorita

B) PARÁMETROS Y RELACIONES

Ya se comentó que los **parámetros de posición** del Modelo de Regresión Lineal General (MRLG) son $\beta_1, \beta_2, \dots, \beta_k$, mientras que en el modelo de renta nacional del Ejemplo 1.4 los parámetros son α_1, α_2, y α_3 (en la ecuación 1), y β_1, β_2 y β_3 (en la ecuación 2). Se recuerda también que los parámetros son valores y fijos y desconocidos, y una de las fases del método econométrico es estimar a partir de una muestra el valor de cada uno de los parámetros del modelo. Los parámetros de posición cuantifican el efecto de las variables explicativas sobre las variables explicadas.

En los distintos ejemplos también se ha podido comprobar que las **relaciones** que existen entre los distintos tipos de variables vienen especificadas mediante un sistema de ecuaciones. Estas relaciones pueden ser más simples o complejas en función al problema en estudio. Por ejemplo, el modelo que estudiaremos en esta asignatura es el Modelo de Regresión Lineal General (MRLG), el cual sólo incluye una ecuación, una única variable endógena y una o más variables predeterminadas. Por su parte, el modelo de renta nacional del Ejemplo 1.4 es un sistema de ecuaciones con 3 ecuaciones.

C) FORMAS DE EXPRESAR UN MODELO ECONOMÉTRICO

Los modelos econométricos se pueden clasificar según:

- **Número de ecuaciones.**
  - **Uniecuacionales.** El sistema de ecuaciones está formado por una única ecuación. Por ejemplo:
    - **Uniecuacional simple:**

$$y = \beta_1 + \beta_2 x_2 + u.$$

    - **Uniecuacional múltiple:**

$$y = \beta_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots + \beta_k x_k + u.$$

  - **Multiecuacionales.** El sistema de ecuaciones está formado por más de una ecuación. Un ejemplo es el modelo de renta nacional del Ejemplo 1.4.
- **Forma funcional.**
  - **Modelos lineales.**

$$y = \beta_1 + \beta_2 x_2 + \beta_3 x_3 + u.$$

  - **Modelos no lineales.**

$$y = \beta_1 + \beta_2 e^{x_2} + \beta_3 x_3 + u.$$

$$y = \beta_1 + {x_2}^{\beta_2} + u.$$

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

WUOLAH

Descarga la app de Wuolah desde tu store favorita

- **Características de la perturbación aleatoria.**
  - **Modelos clásicos.** Se considera que la perturbación aleatoria es ruido blanco, es decir, satisface las siguientes propiedades o hipótesis:
    1. $E[u_i] = 0,\ \forall\, i$. (las perturbaciones tienen media igual a 0).
    2. $V[u_i] = \sigma^2,\ \forall\, i$. (las perturbaciones tienen varianza constante, o bien homocedasticidad).
    3. $cov[u_i, u_j] = 0,\ \forall\, i \neq j$. (las perturbaciones son incorreladas).
  - **Modelos generalizados.** En este tipo de modelos las perturbaciones incumplen alguna de las dos últimas hipótesis de los modelos clásicos, es decir:
    2. $V[u_i] = \sigma_i^2, \quad \forall\, i$. (las perturbaciones no tienen varianza constante, o bien existe un problema de heterocedasticidad).
    3. $cov[u_i, u_j] \neq 0,$ para algún $i \neq j$ (existe un problema de autocorrelación).

<u>Nota para el profesor</u>: Ver Tema 5 de la App: Diferencia entre homocedasticidad y heterocedasticidad.

1.3. Naturaleza de la información utilizada en Econometría.

Los datos económicos pueden clasificarse como:

1. **Datos de corte transversal.** Se trata de datos tomados en un mismo instante o momento de tiempo a diferentes individuos, empresas, hogares, centros, etc. Un ejemplo serían los datos procedentes del Ejemplo 1.3 sobre las Facultades en Ciencias Económicas y Empresariales, puesto que en el mismo instante de tiempo, es decir, en el mismo curso académico, se recogen los datos de las 50 facultades de la muestra del estudio.
2. **Datos de series temporales.** Se trata de datos tomados sobre una misma variable pero en distintos instantes de tiempo. Se puede utilizar cualquier unidad de medida relacionada con el tiempo, tal como años, meses semanas, días. Por ejemplo, en un estudio sobre índices de bolsa se podrían recoger datos en cada hora, dado que tales índices pueden sufrir variaciones en periodos inferiores a una hora. En un modelo con datos de series temporales sería deseable que la

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

WUOLAH

Descarga la app de Wuolah desde tu store favorita

Descarga la APP de Wuolah.
Ya disponible para el móvil y la tablet.
Available on the App Store
GET IT ON Google Play

periodicidad con la que se obtienen los datos tengan siempre la misma frecuencia (misma distancia temporal entre dos observaciones consecutivas de la muestra), es decir, si se obtienen datos por años no comenzar a recogerlos de forma mensual.

Ejemplo 1.5. ***Efecto del IVA sobre el consumo.***

Para un determinado país se desea analizar el efecto que puede tener una subida del IVA sobre el consumo, y para ello se consideran datos anuales de ambas variables (IVA y consumo) de los últimos 25 años. Plantear el correspondiente MRLG.

□

Ejemplo 1.6. ***Efecto de la tasa de desempleo sobre el IPC.***

Se desea analizar el efecto de la tasa de desempleo sobre el IPC a partir de un MRLG que relacione ambas variables y considerando los datos globales de todo el país, y los cuales se recogen mensualmente durante 5 años. Plantear el correspondiente MRLG.

□

Ejemplo 1.7. ***Efecto del valor del Euribor sobre el número de hipotecas.***

Se desea analizar si el valor del Euribor tiene efecto sobre el número de hipotecas que se firman. Este índice económico se hace oficial mensualmente desde el año 1999. A modo ilustrativo, la Figura 1 incluye la evolución histórica del Euribor desde el año 1999 hasta el año 2018. Plantear el correspondiente MRLG. ¿Qué otras variables exógenas podrían ser relevantes?

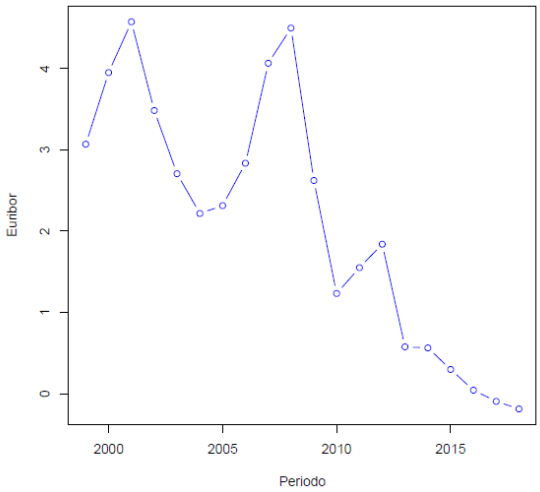


Figura 1.2*. Evolución histórica del Euribor para el periodo 1999-2018.*

No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.
Reservados todos los derechos.
WUOLAH

A su vez, los datos de serie temporales pueden clasificarse en:

2.1. Datos fusionados de sección cruzada. Son los datos tomados a diferentes individuos, empresas, hogares, centros, etc., en dos instantes de tiempo diferentes. Este tipo de datos suelen utilizarse para analizar posibles diferencias del efecto de las variables exógenas sobre la variable endógena del MRLG en los dos instantes de tiempo considerados.

Ejemplo 1.8. (continuación del Ejemplo 1.3) ***Efecto del número de profesores y estudiantes en el rendimiento académico de estudiantes de Ciencias Económicas y Empresariales.***

Un determinado país decide aumentar considerablemente el número de profesores en las Universidades Públicas, y dicha medida se comienza a aplicar en el curso académico 2018/2019. Se desea analizar si esta medida ha sido efectiva, y por tanto afecta de forma positiva al rendimiento académico de los estudiantes de Facultades en Ciencias Económicas y Empresariales, y para ello se toman datos de 50 Facultades de Ciencias Económicas y Empresariales en dos cursos académicos diferentes: 2017/2018 (antes de aplicar la nueva medida política) y 2018/2019 (después de aplicar la medida política). Plantear el correspondiente MRLG.

□

Nota para el profesor: Ver Tema 3 de la App: Cambios en la media de la variable dependiente en dos instantes de tiempo.

2.2. Datos de panel. Se trata de una combinación de datos de corte transversal y datos de series temporales, es decir, son datos tomados para diferentes individuos, empresas, hogares, centros, etc, y para cada uno de estos individuos, empresas, hogares, centros, etc, se toman datos en distintos instantes de tiempo. Por ejemplo, una muestra de distintos empresas, y para cada empresa se considera una serie temporal con datos recogidos de forma mensual.

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

WUOLAH

Descarga la app de Wuolah desde tu store favorita

Ejemplo 1.9. (continuación del Ejemplo 1.7) ***Efecto del valor del Euribor sobre el número de hipotecas, con datos seleccionados por provincias.***

El estudio del Ejemplo 1.7. se puede plantear por provincias, es decir, en lugar de considerar datos globales referidos al país, se recogen datos por provincias. ¿Cuál sería el número total de observaciones suponiendo que el estudio se centra en 10 provincias con datos del Euribor mensuales para el periodo 1999 – 2018?

□

Algunos programas informáticos para Econometría

- **R** (Software con licencia libre - gratuito)

  https://cran.r-project.org/bin/windows/base/

- **RStudio** (Software con licencia libre - gratuito)

  https://www.rstudio.com/products/rstudio/download/

- **Gretl** (Software con licencia libre - gratuito)

  http://gretl.sourceforge.net/es.html

- **Stata** (Software con licencia – no gratuito)
- **SPSS** (Software con licencia – no gratuito)
-

Algunas preguntas de teoría del Tema 1

(Nota: Los exámenes de teoría de la asignatura pueden incluir cualquier pregunta incluida en el programa de la guía docente de la asignatura)

1. Concepto de Econometría. Indique la expresión del Modelo de Regresión Lineal General y explique los distintos componentes de este modelo.
2. Fases del método econométrico.
3. Componentes de un modelo econométrico.
4. Motivos para incluir el término error en el modelo econométrico.
5. Tipos de datos económicos.

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

WUOLAH

Descarga la app de Wuolah desde tu store favorita

2Teoria.pdf

jmgg94

Econometría

3º Doble Grado en Administración y Dirección de Empresas e Informática

Facultad de Ciencias Económicas y Empresariales Universidad de Granada

Descarga la APP de Wuolah.
Ya disponible para el móvil y la tablet.

Available on the App Store
GET IT ON Google Play

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

KEEP CALM AND ESTUDIA UN POQUITO

WUOLAH + #QuédateEnCasa

Descarga la APP de Wuolah.
Ya disponible para el móvil y la tablet.
Available on the App Store
GET IT ON Google Play

TEMA 2. El modelo lineal I

2.0. Introducción

- **Modelo de Regresión Lineal General (MRLG)**

$$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \cdots + \beta_k x_{ki} + u_i, \quad i = 1, \dots, n$$

Ejemplo 2.1. Riqueza (muestra grande)

Considere el siguiente modelo en el que la riqueza por unidad de consumo de individuos (en miles de euros) de un determinado país está explicada en función de la renta por unidad de consumo (variable x_2 y expresada en miles de euros), edad (variable x_3 y medida en años) y número de hijos (x_4):

$$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + u_i, \quad i = 1, \dots, n$$

A partir de una muestra de tamaño 10000 y utilizando el Método de Minimos Cuadrados Ordinarios (MCO), el cual se estudiará más adelante, se obtiene el siguiente modelo estimado:

$$\hat{y}_i = \underset{(0.393)}{20.5} + \underset{(0.257)}{0.5859} x_{2i} + \underset{(0.011)}{0.8953} x_{3i} - \underset{(0.072)}{9.9983} x_{4i}$$

1. Realizar esta estimación usando Gretl y los datos de las prácticas con nombre "000.txt".
2. ¿Cuáles son los valor de *n* y *k*?
3. Explique cómo se denomina cada una de las variables de este modelo.
4. Indique los valores estimados de cada parámetro de posición e interprete los resultados.
5. Justifique si los signos de cada coeficiente estimado son los esperados.
6. ¿Qué indican los valores de los paréntesis ubicados justo debajo de cada coeficiente estimado?

□

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

WUOLAH

- Expresión matricial del MRLG

El Modelo de Regresión Lineal General en expresión matricial es el siguiente:

$$Y = X\beta + u$$

donde

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

$$X = \begin{pmatrix} 1 & x_{21} & \cdots & x_{k1} \\ 1 & x_{22} & \cdots & x_{k2} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & x_{2n} & \cdots & x_{kn} \end{pmatrix}$$

$$\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix}$$

$$u = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}$$

Ejemplo 2.2. Riqueza (muestra pequeña)

Para el MRLG planteado en el Ejemplo 2.1, se obtienen las siguientes observaciones en una de tamaño 3:

Individuo	Riqueza (euros)	Renta (euros)	Edad (años)	Número de hijos
1	50000	1600	60	2
2	30000	1500	30	1
3	10000	2000	50	4

1. Utilizando los datos proporcionados, obtenga la expresión matricial del MRLG.

□

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

WUOLAH

Descarga la app de Wuolah desde tu store favorita

2.1. Hipótesis del modelo / Supuestos básicos requeridos en el modelo

Para que el análisis del MRLG sea válido es necesario que se verifiquen las siguientes hipótesis o supuestos básicos:

S1. Linealidad. Se cumple este supuesto cuando el vector Y puede expresarse como combinación lineal de las columnas de la matriz X, es decir, existe linealidad respecto a los coeficientes del modelo, pero no respecto a las variables.

Ejemplo 2.3. Modelos lineales y no lineales

Algunos ejemplos de modelos lineales son:

$$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + u_i, \quad i = 1, \dots, n$$

$$\ln(y_i) = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + u_i, \quad i = 1, \dots, n$$

$$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 \ln(x_{3i}) + \beta_4 x_{4i} + u_i, \quad i = 1, \dots, n$$

$$\ln(y_i) = \beta_1 + \beta_2 \ln(x_{2i}) + \beta_3 x_{3i} + \beta_4 x_{4i} + u_i, \quad i = 1, \dots, n$$

$$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i}^2 + \beta_4 x_{4i} + u_i, \quad i = 1, \dots, n$$

Un ejemplo de modelo no lineal es:

$$y_i = \beta_1 + \beta_2 x_{2i} + (x_{3i})^{\beta_3} + \beta_4 x_{4i} + u_i, \quad i = 1, \dots, n$$

S2. Supuestos sobre la perturbación aleatoria.

1. Perturbaciones centradas.

2. Perturbaciones homocedásticas.

3. Perturbaciones incorreladas.

Si las perturbaciones son homocedásticas e incorreladas (a estas perturbaciones se les denomina esféricas), la matriz de varianzas-covarianzas del vector u puede expresarse de la siguiente forma:

$$V[u] = \sigma^2 I_n =$$

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

WUOLAH

Descarga la app de Wuolah desde tu store favorita

S3. La matriz X de observaciones de las variables exógenas debe ser de rango completo por columnas.

- Se debe cumplir la siguiente condición: $n > k$.
- Para la muestra con tamaño n, las k columnas de X deben ser linealmente independientes, es decir, no deben existir relaciones lineales entre las variables exógenas. Además, ninguna de las variables exógenas puede ser constante, puesto que en este caso se produciría una relación lineal exacta con la columna de "unos" de la matriz X, e incluida con motivo del término independiente.
- Se dice que se tiene un problema de multicolinealidad exacta en el caso de una combinación lineal exacta entre 2 o más variables explicativas. El problema de la multicolinealidad se estudia en el Tema 4 de la asignatura.
- En la estimación del modelo mediante Mínimos Cuadrados Ordinarios se requiere el cálculo de la inversa de la matriz $X'X$, y el cálculo de esta inversa es posible si se verifica este supuesto.

Ejemplo 2.4. Ejemplos de matrices de observaciones X con relaciones lineales.

□

S4. Exogeneidad

- Las variables exógenas no aportan información que permita predecir la perturbación aleatoria.

- El valor esperado de la perturbación aleatoria es cero para cualquier valor que tomen las variables exógenas, es decir:

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

WUOLAH

Descarga la app de Wuolah desde tu store favorita

Descarga la APP de Wuolah.
Ya disponible para el móvil y la tablet.
Available on the App Store
GET IT ON Google Play

S5. Causalidad

- Las variables consideradas como exógenas son las que explican el comportamiento de la variable dependiente *y*, pero no al contrario.
- La variable *y* tiene un carácter aleatorio, dado que depende de la perturbación aleatoria *u*.
- Las variables explicativas son valores fijos, no son variables aleatorias.

S6. Normalidad

- La perturbación aleatoria tiene distribución Normal, es decir:

  Matricialmente:

$$u \sim N(\mathbf{0}, \sigma^2 I_n)$$

- Este supuesto de Normalidad se requerirá para las fases de verificación y explotación.

2.2. Estimación de los parámetros del modelo por mínimos cuadrados ordinarios. Propiedades.

2.2.1. Estimación de los parámetros del modelo por mínimos cuadrados ordinarios.

Recordemos que el MRLG viene expresado de la siguiente forma.

$$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \cdots + \beta_k x_{ki} + u_i, \quad i = 1, \dots, n$$

Una vez que se disponen de datos muestrales, el siguiente paso consistiría en obtener estimaciones de los parámetros, quedando el modelo estimado de la siguiente forma:

$$\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_{2i} + \hat{\beta}_3 x_{3i} + \cdots + \hat{\beta}_k x_{ki}, \qquad i = 1, \dots, n$$

donde $\hat{y}_i$ denota la predicción de la variable endógena para la observación *i* de la muestra. El ***error o residuo*** de la *i*-ésima observación de la muestra (e_i) se define como la diferencia entre el verdadero valor de la variable endógena (y_i) y su correspondiente predicción, es decir:

$$e_i = y_i - \hat{y}_i$$

WUOLAH

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

El método de Mínimos Cuadrados Ordinarios (MCO) consiste en hacer mínima la suma de los cuadrados de todos los residuos, es decir, los parámetros del modelo se estiman de forma que se minimice la Suma de Cuadrados de los Residuos (***SCR***):

$$SCR = \sum_{i=1}^{n} e_i^2$$

<u>Nota para el profesor</u>: Ver Tema 1 de la App: Nube de puntos.

Para facilitar la obtención de los estimadores por el método MCO y su cálculo correspondiente, se usará la notación matricial. El MRLG en forma matricial se expresó anteriormente como:

$$\boxed{Y = X\beta + u}$$

El modelo estimado en forma matricial se expresa como

$$\boxed{\hat{Y} = X\hat{\beta},}$$

y por tanto el vector de residuos (e) se define como:

$$e = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_i \\ \vdots \\ e_n \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_i \\ \vdots \\ y_n \end{pmatrix} - \begin{pmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_i \\ \vdots \\ \hat{y}_n \end{pmatrix} = Y - \hat{Y} = \boxed{e = Y - X\hat{\beta}}$$

La ***SCR*** en forma matricial viene dada por

$$\boxed{SCR = \sum_{i=1}^{n} e_i^2 = e'e} = \left(Y - X\hat{\beta}\right)'\left(Y - X\hat{\beta}\right) = Y'Y - 2\hat{\beta}'X'Y + \hat{\beta}'X'X\hat{\beta}$$

Para obtener el posible mínimo de la función *SCR* respecto el vector $\hat{\beta}$ se debe realizar su primera derivada respecto $\hat{\beta}$, igualar la función resultante al vector de ceros (**0**) y resolver este sistema de ecuaciones despajando $\hat{\beta}$, es decir:

$$\frac{\partial SCR}{\partial \hat{\beta}} = \mathbf{0}$$

Para confirmar que la solución obtenida es en efecto un mínimo se debe obtener la segunda derivada respecto $\hat{\beta}$ y que verifique:

$$\frac{\partial^2 SCR}{\partial \hat{\beta}} > 0$$

A continuación se obtiene la primera derivada de la *SCR* y se resuelve la ecuación:

WUOLAH

Descarga la app de Wuolah desde tu store favorita

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

$$\frac{\partial SCR}{\partial \hat{\beta}} = -2X'Y + 2X'X\hat{\beta} = \mathbf{0}$$

$$2X'X\hat{\beta} = 2X'Y$$

Por tanto, los estimadores de los parámetros β mediante el método MCO, a los que denominaremos EMCO, se calculan como:

$$\boxed{\hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k \end{pmatrix} = (X'X)^{-1}X'Y}$$

Destacamos que para poder obtener $\hat{\beta}$ según la expresión anterior debe existir la inversa de la matriz $X'X$. Por otra parte se verifica que

$$\frac{\partial^2 SCR}{\partial \hat{\beta}} = 2X'X > 0,$$

puesto que la matriz $X'X$ es definida positiva, y por tanto la solución obtenida a partir de la primera derivada es un mínimo.

Ejemplo 2.5. Cálculo de una inversa.

Ejemplo 2.6. Ejercicio de años de educación.

Ejemplo 2.7. Ejercicio 1 de la relación de ejercicios de los temas 2 y 3.

□

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

WUOLAH

Descarga la app de Wuolah desde tu store favorita

2.2.2. Propiedades del ajuste por MCO.

1. La suma de los residuos es igual a 0:

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

2. La media de la variable endógena coindice con la media de los valores estimados:

$$\bar{y} = \bar{\hat{y}},$$

donde

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \quad ; \quad \bar{\hat{y}}, = \frac{1}{n}\sum_{i=1}^{n} \hat{y}_i$$

3. Los residuos no están correlacionados con las variables explicativas:

$$X'e = 0.$$

A este conjunto de condiciones se le denomina **Ecuaciones Normales** del MRLG.

4. Los residuos tampoco están correlacionados con los valores estimados de la variable dependiente:

$$\hat{Y}'e = (X\hat{\beta})'e = \hat{\beta}'X'e = 0.$$

5. Cuando las variables explicativas toman como valor sus medias muestrales, el correspondiente valor ajustado de la variable dependiente coincide con la media muestral de la variable dependiente, es decir:

$$\bar{y} = \hat{\beta}_1 + \hat{\beta}_2\bar{x}_2 + \hat{\beta}_3\bar{x}_3 + \cdots + \hat{\beta}_k\bar{x}_k$$

2.2.3. Propiedades de los estimadores obtenidos mediante MCO (EMCO).

1. Los EMCO son funciones lineales de las observaciones reales de la variable dependiente.

2. Los EMCO son insesgados:

$$E(\hat{\beta}) = \beta$$

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

WUOLAH

Descarga la app de Wuolah desde tu store favorita

Descarga la APP de Wuolah.
Ya disponible para el móvil y la tablet.
Available on the App Store
GET IT ON Google Play

Demostración:

Teniendo en cuenta que $Y = X\beta + u$, se obtiene:

$$\hat{\beta} = (X'X)^{-1}X'Y = (X'X)^{-1}X'(X\beta + u) = (X'X)^{-1}X'X\beta + (X'X)^{-1}X'u = \beta + (X'X)^{-1}X'u$$

El valor esperado viene dado por:

$$E[\hat{\beta}] = E[\beta + (X'X)^{-1}X'u] = \beta + (X'X)^{-1}X'E[u] = \beta$$

3. La matriz de varianzas-covarianzas de los EMCO viene dada por:

$$V(\hat{\beta}) = \sigma^2(X'X)^{-1}$$

Demostración:

A partir de la demostración anterior se sabe que: $\hat{\beta} - \beta = (X'X)^{-1}X'u$.

$$V(\hat{\beta}) = E\left[(\hat{\beta} - E[\hat{\beta}])(\hat{\beta} - E[\hat{\beta}])'\right] = E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'] =$$

$$E[((X'X)^{-1}X'u)((X'X)^{-1}X'u)'] = E[(X'X)^{-1}X'uu'X(X'X)^{-1}] =$$

$$(X'X)^{-1}X'E[uu']X(X'X)^{-1} = \sigma^2(X'X)^{-1}X'X(X'X)^{-1} =$$

$$\sigma^2(X'X)^{-1}$$

Esta matriz contiene las varianzas y covarianzas de los parámetros del modelo, es decir:

$$V(\hat{\beta}) = \sigma^2(X'X)^{-1} = \begin{pmatrix} V(\hat{\beta}_1) & Cov(\hat{\beta}_1, \hat{\beta}_2) & \cdots & Cov(\hat{\beta}_1, \hat{\beta}_k) \\ Cov(\hat{\beta}_2, \hat{\beta}_1) & V(\hat{\beta}_2) & \cdots & Cov(\hat{\beta}_2, \hat{\beta}_k) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(\hat{\beta}_k, \hat{\beta}_1) & Cov(\hat{\beta}_k, \hat{\beta}_2) & \cdots & V(\hat{\beta}_k) \end{pmatrix}$$

En la diagonal de esta matriz se encuentran las varianzas de los parámetros estimados del modelo. Se define el error estándar de cada parámetro estimado, $SE(\hat{\beta}_j)$, como la raíz cuadrada de su varianza, es decir:

$$SE(\hat{\beta}_j) = \sqrt{V(\hat{\beta}_j)}, \qquad j = 1, \dots, k.$$

4. Teorema de Gauss-Markov: Los EMCO son ELIO (Estimadores Lineales, Insesgados y Óptimos). Se dice que son óptimos en el sentido de tener menor varianza en comparación con cualquier otro vector de estimadores que también sean lineales e insesgados.

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.
WUOLAH

2.2.4. Estimación de la varianza residual.

Mediante el método MCO se ha obtenido el estimador $\hat{\beta}$ de los parámetros β del MRLG. Sin embargo, se comentó con anterioridad que un parámetro adicional del MRLG que también requiere ser estimado es la varianza de las perturbaciones aleatorias, es decir:

$$V(u_i) = \sigma^2, \qquad i = 1, \dots, n$$

Un estimador insesgado de este parámetro viene dado por:

$$\boxed{\hat{\sigma}^2 = \frac{SCR}{n-k}}$$

donde la *SCR* puede calcularse como sigue:

$$\boxed{SCR = Y'Y - \hat{\beta}'X'Y.}$$

Este estimador de la varianza residual puede utilizarse, por ejemplo, para estimar la matriz de varianzas-covarianzas de los EMCO, es decir, estimar:

$$V(\hat{\beta}) = \sigma^2(X'X)^{-1}.$$

Esta matriz de varianzas-covarianzas se estima como sigue:

$$\boxed{\hat{V}(\hat{\beta}) = \hat{\sigma}^2(X'X)^{-1}.}$$

Para el estimador de un determinado parámetro $\hat{\beta}_j$, con $j = 1, \dots, k$, su correspondiente estimador del error estándar se define como:

$$\boxed{\widehat{SE}(\hat{\beta}_j) = \sqrt{\hat{V}(\hat{\beta}_j)}.}$$

<u>Nota para el profesor</u>: Ver Tema 5 de la App: Esperanza y varianza de los estimadores en caso de homocedasticidad y heterocedasticidad.

2.2.5. Distribuciones muestrales de los estimadores obtenidos mediante MCO.

En este apartado se requiere el supuesto S6 sobre la distribución Normal de la perturbación aleatoria. Bajo este supuesto de normalidad se obtienen las siguientes distribuciones:

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

WUOLAH

Descarga la app de Wuolah desde tu store favorita

1. Distribución del vector Y dada la matriz X.

$$Y/_X \sim N(X\beta, \sigma^2 I_n).$$

2. Distribución del vector de valores estimados $\hat{Y}$.

$$\hat{Y} \sim N(X\beta, \sigma^2 X(X'X)^{-1}X').$$

3. Distribución del estimador de la varianza de la perturbación aleatoria.

$$\boxed{\frac{(n-k)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-k}.}$$

4. Distribuciones de los estimadores $\hat{\beta}$.
   a. $\hat{\beta} \sim N\big(\beta, V(\hat{\beta})\big)$

   b. $\frac{\hat{\beta}_j - \beta_j}{\widehat{SE}(\hat{\beta}_j)} \sim t_{n-k}$

Estas distribuciones se usarán en los temas siguientes para la construcción de intervalos de confianza y resolución de contrastes de hipótesis sobre los parámetros del modelo.

<u>Nota para el profesor</u>: Ver Tema 2 de la App: logaritmo en *y* o en *x*.

Ejemplo 2.8. Algunas interpretaciones de parámetros

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

WUOLAH

Descarga la app de Wuolah desde tu store favorita

2.3. Bondad del ajuste: coeficientes de determinación y criterios de Akaike y Schwarz.

A continuación se definen algunas medidas que indican cómo de bien las variables independientes explican los valores de la variable dependiente. Para ello, en primer lugar se introducen algunas definiciones que usaremos en estas medidas:

Suma Cuadrática Total (SCT). Mide la variabilidad de la muestra de la variable dependiente.

$$\boxed{SCT = \sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n} {y_i}^2 - n\bar{y}^2}$$

Suma Cuadrática Explicada (SCE). Mide la variabilidad de los valores estimados ($\hat{y}_i$)

$$\boxed{SCE = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 = \hat{\beta}'X'Y - n\bar{y}^2 = SCT - SCR}$$

Suma Cuadrática Residual (SCR). Mide la variabilidad de los residuos (e_i).

$$\boxed{SCR = \sum_{i=1}^{n} e_i^2 = e'e = Y'Y - \hat{\beta}'X'Y}$$

Estas tres sumas están relacionadas mediante la siguiente expresión en modelos con término independiente:

$$\boxed{SCT = SCE + SCR}$$

- **Coeficiente de determinación**

Destacamos que a medida que la SCE (la variabilidad de los valores estimados) se aproxima a la SCT (variabilidad de los valores observados) mejor será el ajuste del MRLG planteado, y por tanto, mejor explicarán las variables exógenas la variabilidad de la variable endógena. Teniendo en cuenta esta idea se puede definir la siguiente medida de bondad del ajuste, la cual se denomina coeficiente de determinación y denotaremos como R^2:

$$\boxed{R^2 = 1 - \frac{SCR}{SCT} = \frac{SCE}{SCT}}$$

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

WUOLAH

Descarga la app de Wuolah desde tu store favorita

Descarga la APP de Wuolah.
Ya disponible para el móvil y la tablet.
Available on the App Store
GET IT ON Google Play

Propiedades de R^2.

1. El coeficiente de determinación mide la proporción de la variabilidad de la variable endógena que es explicada por las variables exógenas.
2. El coeficiente de determinación siempre está comprendido entre 0 y 1:
   a. $R^2 = 0$ indicaría que $SCE = 0$, y por tanto el modelo no explica nada de la variabilidad de la variable endógena.
   b. $R^2 = 1$ indicaría que $SCE = SCT$, y por tanto el modelo explica perfectamente la variabilidad de la variable endógena.
3. El coeficiente de determinación se usa para comparar modelos con igual número de variables explicativas.
4. No se aconseja su utilización cuando el modelo se formula sin término independiente ($\hat{\beta}_1$).
5. El coeficiente de determinación no depende de las unidades de medida de la variable endógena ni de las variables exógenas.

- Coeficiente de determinación ajustado o corregido.

Uno de los inconvenientes del coeficiente de determinación es que su valor va aumentando a medida que se incluyen variables exógenas en el modelo, incluso cuando estas nuevas variables no sean significativas o no influyan en la variable endógena. Este problema se puede corregir mediante el coeficiente de determinación ajustado o corregido ($\bar{R}^2$), puesto que también mide el porcentaje de variación de la variable endógena que es explicado por las variables exógenas, pero tiene en cuenta el número de tales variables exógenas, de forma que va penalizando la incorporación de nuevas variables. Este coeficiente puede definirse a través de las siguientes expresiones:

$$\bar{R}^2 = 1 - \frac{SCR/(n-k)}{SCT/(n-1)} = 1 - (1 - R^2)\frac{n-1}{n-k}$$

Propiedades de $\bar{R}^2$.

1. Los valores del coeficiente de determinación corregido pueden ser negativos. Por ejemplo, si $R^2 = 0$, entonces $\bar{R}^2 < 0$.
2. Siempre se cumple la siguiente condición: $\bar{R}^2 \leq R^2$.
3. El coeficiente de determinación tampoco depende de las unidades de medida de la variable endógena ni de las variables exógenas.

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.
WUOLAH

- **Criterio de información de Akaike (AIC).**

El *AIC* de un *MRLG* puede obtenerse como:

$$AIC = \ln(\hat{\sigma}^2) + \frac{2k}{n}$$

Propiedades del AIC.

1. En caso de comparación entre varios *MRLG* se seleccionará el que proporcione menor valor del estadístico *AIC*.
2. El *AIC* penaliza la complejidad del modelo con motivo de un mayor número de variables exógenas, y por tanto el AIC sigue el principio de parsimonia: a medida que se incluyen variables exógenas en el modelo, el AIC también aumenta.
3. Para comparar distintos *MRLG*, la variable endógena debe ser la misma.

- **Criterio de información de Schwarz (BIC).**

El *BIC* de un MRLG puede obtenerse como:

$$BIC = \ln(\hat{\sigma}^2) + \frac{k}{n}\ln(n)$$

Propiedades del BIC.

1. En caso de comparación entre varios *MRLG* se seleccionará el que proporcione menor valor del estadístico *BIC*.
2. La principal diferencia entre los criterios *AIC* y *BIC* es que en el *BIC* se penaliza más los modelos con mayor número de variables explicativas, puesto que el número de parámetros $(k+1)$ se multiplica por lnN, en lugar de 2 en el caso del *AIC*.

Ejemplo 2.9.

Se considera el siguiente MRLG del salario mensual (medido en euros) en función de los años de experiencia (x_1) y los años de educación (x_2) para muestras de trabajadores de dos empresas del mismo sector:

$$y_i = \beta_1 + \beta_2 \ln(x_{2i}) + \beta_3 x_{3i} + u_i, \quad i = 1, \dots, n$$

Los modelos estimados para cada empresa son los siguientes:

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

WUOLAH

Descarga la app de Wuolah desde tu store favorita

- Empresa A:

$$\hat{y}_i = \underset{(25)}{-200} + \underset{(10)}{130}\ln(x_{2i}) + \underset{(8)}{70}x_{3i},$$

$$n = 500; \quad SCT = 510000; \quad SCR = 120000$$

$$\widehat{Cov}(\hat{\beta}_1, \hat{\beta}_2) = -5; \quad \widehat{Cov}(\hat{\beta}_1, \hat{\beta}_3) = 10; \quad \widehat{Cov}(\hat{\beta}_2, \hat{\beta}_3) = 5$$

- Empresa B:

$$\hat{y}_i = \underset{(20)}{-300} + \underset{(12)}{150}\ln(x_{2i}) + \underset{(7)}{65}x_{3i},$$

$$n = 400; \quad SCT = 405000; \quad SCR = 90000$$

$$\widehat{Cov}(\hat{\beta}_1, \hat{\beta}_2) = 5; \quad \widehat{Cov}(\hat{\beta}_1, \hat{\beta}_3) = 20; \quad \widehat{Cov}(\hat{\beta}_2, \hat{\beta}_3) = 8$$

Se pide:

a) Interprete los coeficientes estimados de los parámetros de cada modelo.

b) Obtenga la estimación de las varianzas de cada parámetro y para cada empresa.

Sol. Empresa A: $\hat{V}(\hat{\beta}_1) = 625$; $\hat{V}(\hat{\beta}_2) = 100$; $\hat{V}(\hat{\beta}_3) = 64$.

Empresa B: $\hat{V}(\hat{\beta}_1) = 400$; $\hat{V}(\hat{\beta}_2) = 144$; $\hat{V}(\hat{\beta}_3) = 49$.

c) Obtenga la matriz de varianzas-covarianzas del vector de coeficientes estimados del modelo.

d) Obtenga la estimación de la varianza residual para cada empresa.

Sol. Empresa A: $\hat{\sigma}^2 = 241.45$.

Empresa B: $\hat{\sigma}^2 = 226.70$.

e) Obtenga las distintas medidas de ajuste. ¿En qué empresa el modelo planteado se ajusta mejor?

Sol. Empresa A: $R^2 = 0.7647$; $\bar{R}^2 = 0.7638$; $AIC = 5.499$; $BIC = 5.524$.

Empresa B: $R^2 = 0.7778$; $\bar{R}^2 = 0.7767$; $AIC = 5.439$; $BIC = 5.469$.

f) Para cada una de las empresas, ¿cuál sería el salario mensual estimado de un empleado con 10 años de experiencia y 15 años de educación?

g) **Sol.** Empresa A: 1149.336 euros/mes.

Empresa B: 1020.388 euros/mes.

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

WUOLAH

Descarga la app de Wuolah desde tu store favorita

Ejemplo 2.10.

Se considera el siguiente MRLG de la riqueza (medida en decenas de miles de euros) en función de la edad (x_2) y el número de hijos (x_3) para una muestra de 5 personas de un bloque de viviendas:

$$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i, \quad i = 1, \dots, n$$

Los datos recogidos en la mencionada muestra son los siguientes:

y_i	x_{2i}	x_{3i}
4	70	0
-1	30	4
3	65	1
1	36	2
0	30	3

a) Obtenga la estimación de los parámetros del modelo mediante MCO e interprete los resultados.

Sol.

$$X'X = \begin{pmatrix} 5 & 231 & 10 \\ & 12221 & 347 \\ & & 30 \end{pmatrix}$$

$$(X'X)^{-1} = \begin{pmatrix} 21.7605833 & -0.3057888 & -3.7165709 \\ & 0.0044189 & 0.0508175 \\ & & 0.6844012 \end{pmatrix}$$

$$X'Y = \begin{pmatrix} 7 \\ 481 \\ 1 \end{pmatrix}$$

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{pmatrix} = \begin{pmatrix} 1.523099 \\ 0.035787 \\ -0.888378 \end{pmatrix}$$

b) Obtenga la SCR.

Sol.

$$Y'Y = 27$$

$$\hat{\beta}'X'Y = 26.98686$$

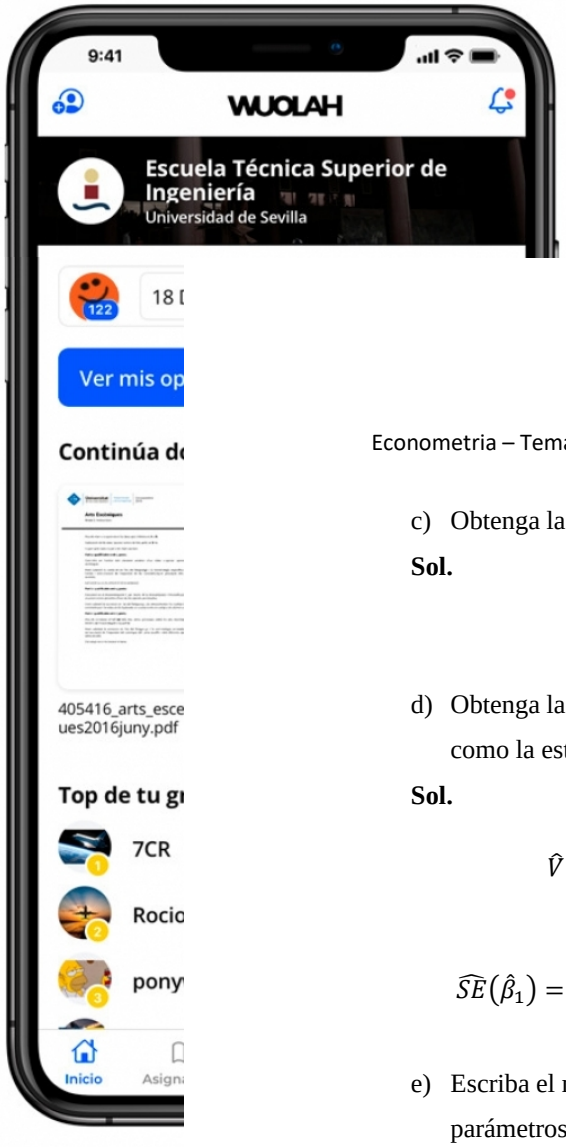
$$SCR = 0.01314$$

WUOLAH

Descarga la app de Wuolah desde tu store favorita

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

Descarga la APP de Wuolah.
Ya disponible para el móvil y la tablet.
Available on the App Store
GET IT ON Google Play



c) Obtenga la estimación de la varianza de la perturbación aleatoria.

Sol.

$$\hat{\sigma}^2 = 0.00657$$

d) Obtenga la matriz de varianzas-covarianzas de los parámetros estimados así como la estimación del error estándar de los parámetros estimados.

Sol.

$$\hat{V}(\hat{\beta}) = \begin{pmatrix} 0.1429670 & -0.0020090 & -0.0244179 \\ & 0.0000290 & 0.0003339 \\ & & 0.0044965 \end{pmatrix}$$

$$\widehat{SE}(\hat{\beta}_1) = 0.37811 \quad ; \quad \widehat{SE}(\hat{\beta}_2) = 0.00539 \quad ; \quad \widehat{SE}(\hat{\beta}_3) = 0.06706$$

e) Escriba el modelo estimado incluyendo la estimación del error estándar de los parámetros estimados.

Sol.

$$\hat{y}_i = \underset{(0.3781)}{1.5231} + \underset{(0.0054)}{0.0358}x_{2i} - \underset{(0.0671)}{0.8884}x_{3i},$$

f) Obtenga una predicción de la riqueza de una persona de 50 años y con 4 hijos.

Sol. -2405 euros.

g) ¿Cómo de fiable es la predicción anterior? Obtenga las distintas medidas de bondad de ajuste.

Sol.

$$SCT = 17.2$$
$$R^2 = 0.99937$$
$$\bar{R}^2 = 0.99874$$
$$AIC = -4.02294$$
$$BIC = -4.25727$$

No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.
Reservados todos los derechos.

WUOLAH

h) Interprete la siguiente salida de resultados del MRLG obtenida en el programa R y relacionada con los datos de este ejercicio. Observe que algunas estimaciones no coinciden con motivo del redondeo.

```
Call:
lm(formula = Y ~ X[, 2] + X[, 3])

Residuals:
       1        2        3        4        5
-0.02863 -0.04339  0.03871 -0.03491  0.06823

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.523111   0.331090   4.600  0.04415 *
X[, 2]       0.035793   0.004718   7.586  0.01694 *
X[, 3]      -0.888378   0.058717 -15.130  0.00434 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.07098 on 2 degrees of freedom
Multiple R-squared:  0.9994,    Adjusted R-squared:  0.9988
F-statistic:  1706 on 2 and 2 DF,  p-value: 0.0005858
```

<u>Algunas</u> preguntas de teoría del Tema 2

(Nota: Los exámenes de teoría de la asignatura pueden incluir cualquier pregunta incluida en el programa de la guía docente de la asignatura)

1. Obtención razonada de los estimadores de los parámetros del MRLG mediante el método MCO.
2. Obtención razonada del valor esperado y de la varianza del estimador MCO de los parámetros del MRLG.
3. Enumerar las hipótesis del MRLG.
4. Enumerar las propiedades del ajuste mediante MCO.
5. Enumerar las propiedades de los estimadores de los parámetros del MRLG obtenidos por el método MCO.
6. Enumerar y definir las medidas de bondad de ajuste del MRLG.

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

WUOLAH

Descarga la app de Wuolah desde tu store favorita

Algunos <u>errores</u> en exámenes

$$(X'X)^{-1} = \frac{1}{719}\begin{bmatrix} 195 & -14 & -113 & -114 \\ & 60 & -132 & -25 \\ & & 578 & 55 \\ & & & 310 \end{bmatrix} \qquad X'y = \begin{bmatrix} 19.5 \\ 38.5 \\ 13.0 \\ 9.5 \end{bmatrix}$$

$$(X' \cdot X) = \frac{1}{719}\begin{pmatrix} 195 & -14 & -113 & -114 \\ -14 & 60 & -132 & -25 \\ 60 & -132 & 578 & 55 \\ -114 & 578 & 55 & 310 \end{pmatrix} \cdot \begin{pmatrix} 19{,}5 \\ 38{,}5 \\ 13{,}0 \\ 9{,}5 \end{pmatrix}$$

WUOLAH

Descarga la app de Wuolah desde tu store favorita

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

Tema-3.pdf

jmgg94

Econometría

3º Doble Grado en Administración y Dirección de Empresas e Informática

Facultad de Ciencias Económicas y Empresariales
Universidad de Granada

Descarga la APP de Wuolah.
Ya disponible para el móvil y la tablet.

Available on the App Store
GET IT ON Google Play

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

KEEP CALM AND ESTUDIA UN POQUITO

WUOLAH + #QuédateEnCasa

Descarga la APP de Wuolah.
Ya disponible para el móvil y la tablet.
Available on the App Store
GET IT ON Google Play

TEMA 3. El modelo lineal II

3.1. Estimación mediante intervalos de confianza de los parámetros del modelo.

3.1.1. Intervalos de confianza para un único coeficiente.

Un intervalo de confianza para un parámetro desconocido es un intervalo de valores cuyos extremos son variables aleatorias, y los cuales contienen al verdadero valor de parámetro con una determinada probabilidad. Para cada muestra que dispongamos se obtendrá un intervalo de confianza, pero a partir de tales datos muestrales los límites del intervalo de confianza ya serán constantes, por lo que dicho intervalo puede contener o no al verdadero valor del parámetro. Por este motivo, una vez que se obtiene un intervalo de confianza basado en datos procedentes de una muestra ya no tiene sentido hablar de la probabilidad de que este intervalo contenga o no al verdadero valor de parámetro. Por los motivos expuestos, la interpretación que puede hacerse de los intervalos de confianza para un parámetro desconocido y con un nivel de confianza de $1-\alpha$ (por ejemplo $1-\alpha=0.95$) es la siguiente:

- *En el caso de disponer de 100 intervalos basados en 100 muestras aleatorias distintas, aproximadamente en el* $(1-\alpha)\%$ *de las veces (por ejemplo el 95%) el verdadero valor del parámetro estará incluido en los intervalos de confianza.*

Suponiendo que las hipótesis básicas establecidas en temas anteriores son ciertas, recordamos una de las distribuciones muestrales del tema anterior:

$$\boxed{\frac{\hat{\beta}_j-\beta_j}{\widehat{SE}(\hat{\beta}_j)}\sim t_{n-k}}$$

Teniendo en cuenta esta distribución se puede deducir fácilmente el intervalo de confianza para el parámetro β_j y con un nivel de confianza $1-\alpha$ como:

$$\boxed{\left[\hat{\beta}_j-t_{n-k;1-\frac{\alpha}{2}}\widehat{SE}(\hat{\beta}_j)\,,\hat{\beta}_j+t_{n-k;1-\frac{\alpha}{2}}\widehat{SE}(\hat{\beta}_j)\right]\quad\equiv\quad\hat{\beta}_j\pm t_{n-k;1-\frac{\alpha}{2}}\widehat{SE}(\hat{\beta}_j)}$$

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.
WUOLAH

Ejemplo 3.1. (Introducción a intervalos de confianza)

Se considera el siguiente MRLG del salario mensual (medido en euros) en función de los años de experiencia (x_2) y los años de educación (x_3) para los trabajadores de una empresa:

$$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i, \quad i = 1, \dots, n$$

El modelo estimado para una muestra de 200 trabajadores es el siguiente:

$$\hat{y}_i = \underset{(12)}{-50} + \underset{(10)}{100} x_{2i} + \underset{(8)}{70} x_{3i},$$

$$SCT = 51000; \quad SCR = 12000$$

$$\widehat{Cov}(\hat{\beta}_1, \hat{\beta}_2) = -5; \quad \widehat{Cov}(\hat{\beta}_1, \hat{\beta}_3) = 10; \quad \widehat{Cov}(\hat{\beta}_2, \hat{\beta}_3) = 5$$

Obtenga un intervalo de confianza para β_3.

Ejemplo 3.2. (Relación entre contrastes de hipótesis e intervalos de confianza)

A partir del intervalo de confianza obtenido en el Ejemplo 3.1., ¿se podría afirmar que el salario se incrementa 80€ cuando los trabajadores tienen un año más de educación y suponiendo que los años de experiencia permanecen constante?

3.1.2. Intervalos de confianza para una combinación lineal de coeficientes.

Un problema con "q" combinaciones lineales de los coeficientes del modelo se expresará matricialmente del siguiente modo:

$$R\beta,$$

donde R es una matriz con q filas (una para cada combinación lineal) y k columnas (una para cada parámetro del modelo) y β es el vector con todos los parámetros del modelo. En esta sección se construirán intervalos de confianza para $R\beta$. La distribución muestral usada en este caso es la siguiente:

$$\frac{R\hat{\beta} - R\beta}{\sqrt{R\,\hat{V}(\hat{\beta})R'}} = \frac{R\hat{\beta} - R\beta}{\widehat{SE}(R\hat{\beta})} \sim t_{n-k}$$

Por tanto, el intervalo de confianza con nivel de confianza $(1-\alpha)$ para la combinación $R\beta$ viene dado por:

$$\boxed{\left(R\hat{\beta} - t_{n-k;1-\frac{\alpha}{2}}\widehat{SE}(R\hat{\beta});\ R\hat{\beta} + t_{n-k;1-\frac{\alpha}{2}}\widehat{SE}(R\hat{\beta})\right) \quad \equiv \quad R\hat{\beta} \pm t_{n-k;1-\frac{\alpha}{2}}\widehat{SE}(R\hat{\beta})}$$

WUOLAH

Descarga la app de Wuolah desde tu store favorita

No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad. Reservados todos los derechos.

Ejemplo 3.3.

Simplificar matricialmente las siguientes combinaciones lineales de los parámetros en un MRLG con 4 variables explicativas.

1. $\beta_1 + \beta_2 + \beta_3$

Sol.

$$R = (1 \quad 1 \quad 1 \quad 0 \quad 0);\ \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{pmatrix};\ R\beta = (1 \quad 1 \quad 1 \quad 0 \quad 0) \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{pmatrix}$$

$$= \quad \beta_1 + \beta_2 + \beta_3$$

2. $\beta_2 + \beta_4 - \beta_5$

Sol.

$$R = (0 \quad 1 \quad 0 \quad 1 \quad -1);\ \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{pmatrix};\ R\beta = (0 \quad 1 \quad 0 \quad 1 \quad -1) \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{pmatrix}$$

$$= \beta_2 + \beta_4 - \beta_5$$

3. $2\beta_3 + \beta_4$

Sol.

$$R = (0 \quad 0 \quad 2 \quad 1 \quad 0);\ \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{pmatrix};\ R\beta = (0 \quad 0 \quad 2 \quad 1 \quad 0) \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{pmatrix}$$

$$= 2\beta_3 + \beta_4$$

4. $\beta_3 - 2\beta_4$
   $3\beta_2$

Sol.

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

WUOLAH

Descarga la app de Wuolah desde tu store favorita

$$R = \begin{pmatrix} 0 & 0 & 1 & -2 & 0 \\ 0 & 3 & 0 & 0 & 0 \end{pmatrix};\ \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{pmatrix};\ R\beta = \begin{pmatrix} 0 & 0 & 1 & -2 & 0 \\ 0 & 3 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{pmatrix}$$

$$= \begin{matrix} \beta_3 - 2\beta_4 \\ 3\beta_2 \end{matrix}$$

5. $\begin{matrix} \beta_3 - \beta_4 \\ 2\beta_2 \\ \beta_1 + \beta_3 - 2\beta_5 \end{matrix}$

Sol.

$$R = \begin{pmatrix} 0 & 0 & 1 & -1 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & -2 \end{pmatrix};\ \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{pmatrix};\ R\beta = \begin{pmatrix} 0 & 0 & 1 & -1 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & -2 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{pmatrix}$$

$$= \begin{matrix} \beta_3 - \beta_4 \\ 2\beta_2 \\ \beta_1 + \beta_3 - 2\beta_5 \end{matrix}$$

□

Ejemplo 3.4. (Intervalo de confianza para una combinación lineal)

Para los datos proporcionados en el Ejemplo 3.1., obtenga un intervalo de confianza que mida el efecto neto en el salario ante un aumento de un año de experiencia y un año de educación.

Sol.: $\hat{V}(\hat{\beta}_2 + \hat{\beta}_3) = 174;\ \ t_{197;0.975} = 1.972;\ \ [143.99;\ \ 196;01]$

Ejemplo 3.5. (Relación entre contraste de hipótesis e intervalos de confianza para una combinación lineal)

A partir del intervalo de confianza del Ejemplo 3.4., ¿se puede afirmar que el salario neto de los trabajadores aumenta 120€ cuando éstos tienen un año más de experiencia y un año más de educación?

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

WUOLAH

Descarga la app de Wuolah desde tu store favorita

Descarga la APP de Wuolah.
Ya disponible para el móvil y la tablet.
Available on the App Store
GET IT ON Google Play

3.2. Contrastes de hipótesis acerca de los parámetros del modelo.

3.2.1. Contrastes de hipótesis sobre un único coeficiente.

Supongamos el MRLG:

$$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \cdots + \beta_k x_{ki} + u_i, \quad i = 1, \dots, n$$

$$t = \frac{\hat{\beta}_j - \beta_j}{\widehat{SE}(\hat{\beta}_j)} \sim t_{n-k}$$

1. Contraste cola a la derecha

$$H_0: \beta_j = \beta_j^0$$
$$H_1: \beta_j > \beta_j^0$$

a) <u>Usando el estadístico experimental.</u>

$$t_{exp} = \frac{\hat{\beta}_j - \beta_j^0}{\widehat{SE}(\hat{\beta}_j)}$$

Se rechaza H_0 si $t_{exp} > t_{n-k;1-\alpha}$.

b) <u>Usando el p-valor.</u>

$$p - valor = P(t > t_{exp})$$

Se rechaza H_0 si $p - valor < \alpha$.

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

WUOLAH

2. Contraste cola a la izquierda

$$H_0: \beta_j = \beta_j^0$$
$$H_1: \beta_j < \beta_j^0$$

a) ***<u>Usando el estadístico experimental.</u>***

$$t_{exp} = \frac{\hat{\beta}_j - \beta_j^0}{\widehat{SE}(\hat{\beta}_j)}$$

Se rechaza H_0 si $t_{exp} < t_{n-k;\alpha} \quad (= -t_{n-k;1-\alpha})$.

b) ***<u>Usando el p-valor.</u>***

$$p - valor = P(t < t_{exp})$$

Se rechaza H_0 si $p - valor < \alpha$.

3. Contraste bilateral

$$H_0: \beta_j = \beta_j^0$$
$$H_1: \beta_j \neq \beta_j^0$$

a) ***<u>Usando el estadístico experimental.</u>***

$$t_{exp} = \frac{\hat{\beta}_j - \beta_j^0}{\widehat{SE}(\hat{\beta}_j)}$$

Se rechaza H_0 si $|t_{exp}| > t_{n-k;1-\frac{\alpha}{2}}$.

b) ***<u>Usando el p-valor.</u>***

$$p - valor = 2 \times P(t > |t_{exp}|)$$

Se rechaza H_0 si $p - valor < \alpha$.

WUOLAH

Descarga la app de Wuolah desde tu store favorita

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

Ejemplo 3.6.

Contraste si un incremento de un año de educación produce un incremento mayor de 60€ en el salario. A continuación se proporcionan capturas de pantalla del software Gretl, y en las cuales puede/n consultarse el/los punto/s critico/s requerido/s en este contraste.

Sol.: $t_{exp} = 1.25; \quad t_{197;0.95} = 1.653.$

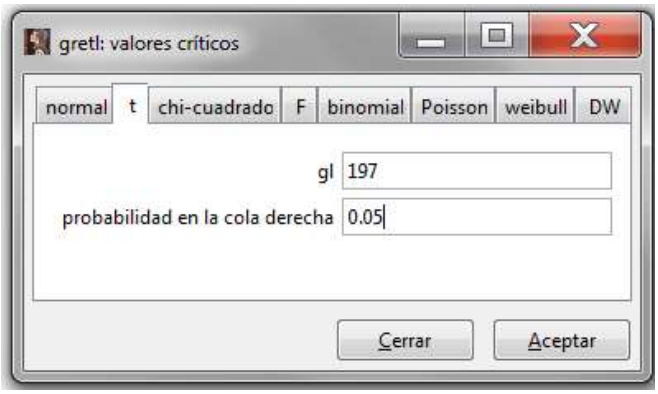


Figura 3.1. *Captura de pantalla del programa estadístico Gretl para el cálculo de un cuantil de la distribución t-Student.*

Ejemplo 3.7.

Contraste si un incremento de un año de educación produce un incremento inferior a 85€ en el salario.

Sol.: $t_{exp} = -1.875; \quad t_{197;0.05} = -t_{197;0.95} = -1.653.$

Ejemplo 3.8.

Contraste si un incremento de un año de educación produce un incremento de 60€ en el salario.

Sol.: $t_{exp} = 1.25; \quad t_{197;0.975} = 1.972.$

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

WUOLAH

Descarga la app de Wuolah desde tu store favorita

Contrastes de significación de parámetros

Un caso particular de un contraste bilateral sobre un parámetro del MRLG son los denominados contrastes de significación, y en los cuales se contrasta si un determinado parámetro del MRLG es igual o distinto de cero. Este contraste es muy utilizado y de gran importancia, puesto que lo que realmente se está contrastando es si la variable explicativa asociada al parámetro del contraste de significación tiene o no algún efecto en la variable explicada, o lo que es lo mismo, si dicha variable explicativa es significativa o no.

$$H_0: \beta_j = 0$$
$$H_1: \beta_j \neq 0$$

a) <u>Usando el estadístico experimental.</u>

$$t_{exp} =$$

Se rechaza H_0 si:

b) <u>Usando el p-valor.</u>

$$p - valor = 2 \times P(t > |t_{exp}|)$$

Se rechaza H_0 si $p - valor < \alpha$.

Ejemplo 3.9.

Contraste si los años de educación tienen efecto sobre los salarios de los trabajadores.

Sol.: $t_{exp} = 8.75$; $t_{197;0.975} = 1.972$.

Ejemplo 3.10. (Ejercicio 10 de la relación de ejercicios)

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

WUOLAH

Descarga la app de Wuolah desde tu store favorita

Descarga la APP de Wuolah.
Ya disponible para el móvil y la tablet.
Available on the App Store
GET IT ON Google Play

3.2.2. Contrastes de hipótesis de una restricción lineal.

En el caso de que las hipótesis básicas del modelo sean ciertas, para resolver contrastes de hipótesis de una restricción lineal se usará el siguiente estadístico muestral *t*:

$$t = \frac{R\hat{\beta} - r}{\sqrt{R\,\hat{V}(\hat{\beta})R'}} = \frac{R\hat{\beta} - r}{\widehat{SE}(R\hat{\beta})} \sim t_{n-k}$$

Para un nivel de significación α, la regla de decisión del contraste de hipótesis dependerá de la hipótesis alternativa:

1. Contraste cola a la derecha

$$H_0: R\beta = r$$
$$H_1: R\beta > r$$

c) <u>Usando el estadístico experimental.</u>

$$t_{exp} = \frac{R\hat{\beta} - r}{\widehat{SE}(R\hat{\beta})}$$

Se rechaza H_0 si $t_{exp} > t_{n-k;1-\alpha}$.

d) <u>Usando el p-valor.</u>

$$p - valor = P(t > t_{exp})$$

Se rechaza H_0 si $p - valor < \alpha$.

No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.
Reservados todos los derechos.

WUOLAH

2. Contraste cola a la izquierda

$$H_0: R\beta = r$$
$$H_1: R\beta < r$$

c) ***<u>Usando el estadístico experimental.</u>***

$$t_{exp} = \frac{R\hat{\beta} - r}{\widehat{SE}(R\hat{\beta})}$$

Se rechaza H_0 si $t_{exp} < t_{n-k;\alpha} \quad (= -t_{n-k;1-\alpha})$.

d) ***<u>Usando el p-valor.</u>***

$$p - valor = P(t < t_{exp})$$

Se rechaza H_0 si $p - valor < \alpha$.

3. Contraste bilateral

$$H_0: R\beta = r$$
$$H_1: R\beta \neq r$$

c) ***<u>Usando el estadístico experimental.</u>***

$$t_{exp} = \frac{R\hat{\beta} - r}{\widehat{SE}(R\hat{\beta})}$$

Se rechaza H_0 si $|t_{exp}| > t_{n-k;1-\frac{\alpha}{2}}$.

d) ***<u>Usando el p-valor.</u>***

$$p - valor = 2 \times P(t > |t_{exp}|)$$

Se rechaza H_0 si $p - valor < \alpha$.

WUOLAH

Descarga la app de Wuolah desde tu store favorita

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

Ejemplo 3.11. (Planteamiento de contrastes de hipótesis para una restricción lineal).

Formular los siguientes contrastes de hipótesis:

a) Contrastar si la experiencia y la educación tienen el mismo efecto en salario.
b) Contrastar si el efecto neto de un año más de experiencia y un año de educación en el salario es mayor 120€.
c) Contrastar si el efecto en el salario de un año más de educación es inferior al de incrementar un año más de experiencia.
d) Contrastar si el efecto de la variable años de experiencia coincide con el doble del efecto de la variable años de educación en el salario.

Ejemplo 3.12. (Contrastes de hipótesis para una restricción lineal)

Contrastar si la experiencia y la educación tienen el mismo efecto en salario.

Sol.: $\hat{V}(\hat{\beta}_2 - \hat{\beta}_3) = 154;\ \ t_{exp} = 2.42$

Ejemplo 3.13. (Contrastes de hipótesis para una restricción lineal)

Contrastar si el efecto neto de un año más de experiencia y un año de educación en el salario es mayor 120€.

Sol.: $\hat{V}(\hat{\beta}_2 + \hat{\beta}_3) = 174;\ \ t_{exp} = 3.79$

Ejemplo 3.14. (Contrastes de hipótesis para una restricción lineal)

Contrastar si el efecto en el salario de un año más de educación es inferior al de incrementar un año más de experiencia.

Sol.: $\hat{V}(\hat{\beta}_2 - \hat{\beta}_3) = 154;\ \ t_{exp} = 2.42$

Ejemplo 3.15. (Contrastes de hipótesis para una restricción lineal)

Contrastar si el efecto de la variable años de experiencia coincide con el doble del efecto de la variable años de educación en el salario.

Sol.: $\hat{V}(\hat{\beta}_2 - 2\hat{\beta}_3) = 336;\ \ t_{exp} = -2.18$

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

WUOLAH

Descarga la app de Wuolah desde tu store favorita

3.2.3. Contrastes de hipótesis de varias restricciones lineales.

El contraste de hipótesis que se plantea es el siguiente:

$$H_0: R\beta = r$$
$$H_1: R\beta \neq r$$

donde R es una matriz con q filas, una fila para cada restricción lineal.

$$F = \frac{(R\hat{\beta} - R\beta)'\left(R\hat{V}(\hat{\beta})R'\right)^{-1}(R\hat{\beta} - R\beta)}{q} \sim F_{q,n-k}$$

a) <u>Usando el estadístico experimental.</u>

$$F_{exp} = \frac{(R\hat{\beta} - r)'\left(R\hat{V}(\hat{\beta})R'\right)^{-1}(R\hat{\beta} - r)}{q}$$

Se rechaza H_0 si $F_{exp} > F_{q,n-k;1-\alpha}$.

b) ***<u>Usando el p-valor.</u>*** $p - valor = P(F > F_{exp})$

Se rechaza H_0 si $p - valor < \alpha$.

Ejemplo 3.16. (Contraste de hipótesis de varias restricciones lineales)

Contrastar si, conjuntamente, si el salario aumenta 90€ y 80€ cuando aumenta un año la experiencia y la educación, respectivamente.

Sol.: $R\hat{V}(\hat{\beta})R' = \begin{pmatrix} 100 & 5 \\ 5 & 64 \end{pmatrix}$; $\det\left[\left(R\hat{V}(\hat{\beta})R'\right)^{-1}\right] = 6375$;

$$\left(R\hat{V}(\hat{\beta})R'\right)^{-1} = \begin{pmatrix} 0.01004 & -0.00078 \\ -0.00078 & 0.01569 \end{pmatrix};\ F_{exp} = 1.36$$

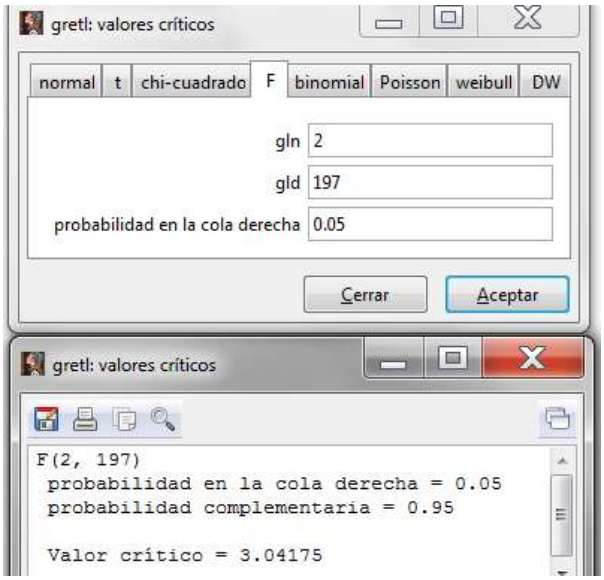


Figura 3.2. *Captura de pantalla del programa estadístico Gretl para el cálculo de un cuantil de la distribución F-Snedecor.*

WUOLAH

Descarga la app de Wuolah desde tu store favorita

No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad. Reservados todos los derechos.

Descarga la APP de Wuolah.
Ya disponible para el móvil y la tablet.
Available on the App Store
GET IT ON Google Play

3.2.4. Contrastes de hipótesis mediante la estimación de modelos restringidos.

Los contrastes de hipótesis con restricciones lineales se pueden resolver mediante la estimación de modelos restringidos. Esta técnica consiste en imponer (asumir que es cierta) la hipótesis nula en el modelo, y volver a estimar mediante MCO el modelo resultante, y el cual se le denomina modelo restringido. El estadístico a utilizar en un contraste de hipótesis basado en modelos restringidos es:

$$F_1 = \frac{\frac{SCR_r - SCR}{q}}{\frac{SCR}{n-k}} \sim F_{q,n-k}$$

Donde:

- q: Número de restricciones lineales.
- SCR: Suma de cuadrado de residuos del MRLG original.
- SCR_r: Suma de cuadrados de residuos del modelo restringido, obtenido una vez que se impone la hipótesis nula en el modelo.

Sólo en el caso de que la variable dependiente del modelo restringido no cambie respecto la variable dependiente del modelo original, se podrá utilizar, de forma adicional, el siguiente estadístico para la resolución de contrastes de hipótesis mediante modelos restringidos:

$$F_2 = \frac{\frac{R^2 - R_r^2}{q}}{\frac{1-R^2}{n-k}} \sim F_{q,n-k}$$

En resumen, el estadístico F_1 se puede utilizar en todas las situaciones, mientras que el estadístico F_2 sólo en aquellas en las que la variable dependiente del modelo restringido no cambia.

Ejemplo 3.17. (Contraste de hipótesis mediante la estimación de modelos restringidos)

Utilizando la estimación de modelos restringidos, contrastar si la experiencia y el salario tienen el mismo efecto en salario. Use la siguiente información:

$$\hat{y}_i = \underset{(12)}{-30} + \underset{(15)}{165}(x_{2i} + x_{3i}), \qquad SCR = 12356.74$$

Sol.: $F_{exp} = 5.86$.

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.
WUOLAH

Ejemplo 3.18.

Dado el modelo

$$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + u_i, \quad i = 1, \dots, n$$

se obtienen las siguientes observaciones:

y_i	11	20	16	17	31	35	45
x_{2i}	1	2	3	4	5	6	7
x_{3i}	6	9	7	11	8	10	12
x_{4i}	4	7	3	2	12	6	7

Mediante los dos procedimientos estudiados, resuelva el siguiente contraste:

$$H_0: \beta_2 = 5;\ \beta_3 + \beta_4 = 0$$
$$H_1: \beta_2 \neq 5 \text{ ó } \beta_3 + \beta_4 \neq 0.$$

a) ***<u>Estadístico experimental obtenido a partir del contraste de hipótesis para varias restricciones lineales.</u>***

En este caso se obtiene la siguiente matriz R, y vectores β y $\boldsymbol{r}$:

$$R = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix};\ \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{pmatrix};\ \boldsymbol{r} = \begin{pmatrix} 5 \\ 0 \end{pmatrix}$$

Bajo la hipótesis nula, el estadístico de este contraste es:

$$F = \frac{\left(R\hat{\beta} - \boldsymbol{r}\right)'\left(R\hat{V}(\hat{\beta})R'\right)^{-1}\left(R\hat{\beta} - \boldsymbol{r}\right)}{q} \sim F_{q,n-k} = F_{2,3}$$

$$F_{exp} = \frac{\left(R\hat{\beta} - r\right)'\left(R\hat{V}(\hat{\beta})R'\right)^{-1}\left(R\hat{\beta} - r\right)}{q}$$

$$= \frac{\begin{pmatrix} -0.965 & 1.8787 \end{pmatrix}\begin{pmatrix} 2.786503 & -2.703102 \\ -2.703102 & 3.834979 \end{pmatrix}^{-1}\begin{pmatrix} -0.965 \\ 1.8787 \end{pmatrix}}{2}$$

$$= \frac{1.066773}{2} = 0.53.$$

Para la obtención del estadístico experimental se requieren estos cálculos:

$$X'Y = \begin{pmatrix} 175 \\ 847 \\ 1683 \\ 1163 \end{pmatrix}$$

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

WUOLAH

Descarga la app de Wuolah desde tu store favorita

$$\hat{\beta} = (X'X)^{-1}X'Y = \begin{pmatrix} -4.6330 \\ 4.0350 \\ 0.7921 \\ 1.0866 \end{pmatrix}$$

$$R\hat{\beta} - \boldsymbol{r} = \begin{pmatrix} \hat{\beta}_2 - 5 \\ \hat{\beta}_3 + \hat{\beta}_4 - 0 \end{pmatrix} = \begin{pmatrix} -0.965 \\ 1.8787 \end{pmatrix}$$

$$Y'Y = 5277$$

$$SCR = Y'Y - \hat{\beta}'X'Y = 73.3099$$

$$\hat{\sigma}^2 = \frac{SCR}{n-k} = 24.43663$$

$$\hat{V}(\hat{\beta}) = \hat{\sigma}^2(X'X)^{-1} = \begin{pmatrix} 130.271621 & 11.1832447 & -16.0786137 & -4.5766562 \\ & 2.7865030 & -2.0671652 & -0.6359363 \\ & & 2.4067284 & 0.4587081 \\ & & & 0.5108341 \end{pmatrix}$$

$$R\hat{V}(\hat{\beta})R' = \begin{pmatrix} 2.786503 & -2.703102 \\ -2.703102 & 3.834979 \end{pmatrix}$$

Usando Gretl o R, se puede obtener que el valor crítico es:

$$F_{q,n-k;1-\alpha} = F_{2,3;0.95} = 9.552$$

Puesto que $F_{exp} < F_{q,n-k;1-\alpha}$ no se rechaza H_0 con un 5% de nivel de significación, y por tanto, no podemos rechazar que $\beta_1 = 5$ y $\beta_2 + \beta_3 = 0$.

b) <u>Estadístico experimental obtenido a partir del modelo restringido.</u>

Se obtiene el modelo restringido asumiendo que $\beta_2 = 5$ y $\beta_4 = -\beta_3$:

$$y_i = \beta_1 + 5x_{2i} + \beta_3 x_{3i} - \beta_3 x_{4i} + u_i, \quad i = 1, \dots, n$$

$$y_i - 5x_{2i} = \beta_1 + \beta_3(x_{3i} - x_{4i}) + u_i, \quad i = 1, \dots, n$$

$$y_{ri} = \beta_1 + \beta_3 x_{ri} + u_i, \quad i = 1, \dots, n$$

donde

$$y_{ri} = y_i - 5x_{2i}$$

$$x_{ri} = x_{3i} - x_{4i}$$

WUOLAH

Descarga la app de Wuolah desde tu store favorita

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

Obtenemos la SCR para el modelo restringido (SCR_r)

$$Y_r = \begin{pmatrix} 11 \\ 20 \\ 16 \\ 17 \\ 31 \\ 35 \\ 45 \end{pmatrix} - 5 \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{pmatrix} = \begin{pmatrix} 6 \\ 10 \\ 1 \\ -3 \\ 6 \\ 5 \\ 10 \end{pmatrix}$$

$$X_r = \begin{pmatrix} 1 & 6-4 \\ 1 & 9-7 \\ 1 & 7-3 \\ 1 & 11-2 \\ 1 & 8-12 \\ 1 & 10-6 \\ 1 & 12-7 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 1 & 2 \\ 1 & 4 \\ 1 & 9 \\ 1 & -4 \\ 1 & 4 \\ 1 & 5 \end{pmatrix}$$

$$Y_r' Y_r = 307$$

$$X_r' Y_r = \begin{pmatrix} 35 \\ 55 \end{pmatrix}$$

$$X_r' X_r = \begin{pmatrix} 7 & 22 \\ 22 & 162 \end{pmatrix}$$

$$\hat{\beta}_r = (X_r' X_r)^{-1} X_r' Y_r = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_3 \end{pmatrix} = \begin{pmatrix} 6.8615 \\ -0.5923 \end{pmatrix}$$

$$SCR_r = Y_r' Y_r - \hat{\beta}_r{}' X_r' Y_r = 99.424$$

Puesto que la variable dependiente del modelo original y la del modelo restringido son diferentes, la expresión del estadístico experimental que debemos utilizar es la siguiente:

$$F_{exp} = \frac{(SCR_r - SCR)/q}{SCR/(n-k)} = \frac{(99.424 - 73.3099)/2}{73.3099/3} = 0.531$$

Puesto que $F_{exp} < F_{q,n-k;1-\alpha}$ no se rechaza H_0 con un 5% de nivel de significación, y por tanto, no podemos rechazar que $\beta_2 = 5$ y $\beta_3 + \beta_4 = 0$.

□

WUOLAH

Descarga la app de Wuolah desde tu store favorita

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

Descarga la APP de Wuolah.
Ya disponible para el móvil y la tablet.
Available on the App Store
GET IT ON Google Play

3.2.5. Contraste de significación global del modelo.

$$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \cdots + \beta_k x_{ki} + u_i, \quad i = 1, \dots, n$$

El contraste de significación global del modelo se expresa de siguiente modo:

$$H_0: \beta_2 = 0, \quad \beta_3 = 0, \cdots \quad \beta_k = 0$$
$$H_1: \beta_2 \neq 0, o\ \beta_3 \neq 0, \cdots o \quad \beta_k \neq 0$$

Ejercicio. Resolver el contraste de significación global mediante modelos restringidos, y teniendo en cuenta que $R_r^2 = 0$.

Es muy común resolver el contraste de significación global con ayuda de la denominada "**Tabla ANOVA**", la cual se incluye a continuación y está basada en la siguiente expresión alternativa del estadístico experimental del contraste de significación global del modelo:

$$F_{exp} = \frac{\frac{R^2}{k-1}}{\frac{1-R^2}{n-k}} = \frac{\frac{SCE}{k-1}}{\frac{SCR}{n-k}}$$

Fuente de variación	**Suma de cuadrados**	**Grados de libertad**	**Medias**	$\boldsymbol{F_{exp}}$
Explicada	SCE	$k-1$	$\frac{SCE}{k-1}$	$\frac{\frac{SCE}{k-1}}{\frac{SCR}{n-k}}$
Residuos	SCR	$n-k$	$\frac{SCR}{n-k}$	
Total	SCT			

Ejemplo 3.19. (Contraste de significación global del modelo y Tabla ANOVA).

Con los datos del Ejemplo 3.1, resuelva el contraste de significación global del modelo y obtenga la tabla ANOVA.

Sol.: $R^2 = 0.7647$; $F_{exp} = 320.125$; $SCE = 39000$.

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.
WUOLAH

3.3. Explotación del modelo.

3.3.1. Predicción en el MRLG.

Dado el modelo

$$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \cdots + \beta_k x_{ki} + u_i, \quad i = 1, \dots, n$$

Para un individuo con las siguientes características para las variables explicativas:

$$x_0 = (1 \quad x_{20} \quad x_{30} \quad \cdots \quad x_{k0})',$$

la correspondiente estimación del valor esperado de la variable dependiente es:

$$\hat{E}(y_0|x_{20}, x_{30}, \dots, x_{k0}) = \hat{\beta}_1 + \hat{\beta}_2 x_{20} + \hat{\beta}_3 x_{30} + \cdots + \hat{\beta}_k x_{k0} = x_0'\hat{\beta} = \hat{y}_0$$

El estimador de la varianza de la predicción $\hat{y}_0$ se define como:

$$\hat{V}(\hat{y}_0) = x_0'\hat{V}(\hat{\beta})x_0$$

Por tanto, el error estándar de $\hat{y}_0$ se define como:

$$\widehat{SE}(\hat{y}_0) = \sqrt{\hat{V}(\hat{y}_0)} = \sqrt{x_0'\hat{V}(\hat{\beta})x_0}$$

3.1.1. Intervalo de predicción para la media de la variable dependiente.

Se dispone de la siguiente distribución muestral:

$$\frac{\hat{y}_0 - x_0'\beta}{\widehat{SE}(\hat{y}_0)} \sim t_{n-k},$$

a partir de la cual puede deducirse el siguiente intervalo de predicción para el valor medio de la variable dependiente para un individuo con características

$$x_0 = (1 \quad x_{20} \quad x_{30} \quad \cdots \quad x_{k0})',$$

y con un nivel de confianza $(1 - \alpha)$:

$$\left[\hat{y}_0 - t_{n-k;1-\frac{\alpha}{2}}\widehat{SE}(\hat{y}_0), \qquad \hat{y}_0 + t_{n-k;1-\frac{\alpha}{2}}\widehat{SE}(\hat{y}_0)\right]$$

3.1.2. Intervalo de predicción para un individuo.

Recordamos que el vector con las características de las variables explicativas de un individuo cualquiera se representa por:

$$x_0 = (1 \quad x_{20} \quad x_{30} \quad \cdots \quad x_{k0})',$$

El valor estimado de la variable dependiente dadas estas características es el mismo que el valor esperado de la variable dependiente:

$$\hat{y}_0 = x_0'\hat{\beta} = \hat{\beta}_0 + \hat{\beta}_1 x_{10} + \hat{\beta}_2 x_{20} + \cdots + \hat{\beta}_k x_{k0}$$

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

WUOLAH

Descarga la app de Wuolah desde tu store favorita

A continuación se define el error de predicción y se proporciona su varianza:

$$e_0 = y_0 - \hat{y}_0$$

$$\hat{V}(e_0) = \hat{\sigma}^2 + \hat{V}(\hat{y}_0) = \hat{\sigma}^2 + x_0'\hat{V}(\hat{\beta})x_0$$

donde

$$y_0 = x_0'\beta = \beta_1 + \beta_2 x_{20} + \beta_3 x_{30} + \cdots + \beta_k x_{k0} + u_0$$

Se define el error estándar del error de predicción como:

$$\widehat{SE}(e_0) = \sqrt{\hat{\sigma}^2 + \hat{V}(\hat{y}_0)} = \sqrt{\hat{\sigma}^2 + x_0'\hat{V}(\hat{\beta})x_0}$$

Se puede deducir la siguiente distribución muestral:

$$\frac{y_0 - \hat{y}_0}{\widehat{SE}(e_0)} \sim t_{n-k}$$

A partir de esta distribución se obtiene el intervalo de predicción para y_0 con un nivel de confianza $(1-\alpha)$ y basado en las características $x_0 = (1 \quad x_{20} \quad x_{30} \quad \cdots \quad x_{k0})'$:

$$\left[\hat{y}_0 - t_{n-k;1-\frac{\alpha}{2}}\widehat{SE}(e_0), \qquad \hat{y}_0 + t_{n-k;1-\frac{\alpha}{2}}\widehat{SE}(e_0)\right]$$

Se puede observar que:

$$\widehat{SE}(e_0) > \widehat{SE}(\hat{y}_0)$$

3.3.2. Contraste para el análisis de la permanencia estructural o validación a posteriori.

En secciones anteriores se han realizado predicciones basadas en los valores x_0, es decir, predicciones basadas en información que no se ha utilizado para la estimación de los parámetros del modelo. En tales predicciones se está asumiendo que dicho modelo estimado sigue siendo válido. Sin embargo, puede ocurrir que el MRLG estimado no sea válido para un individuo con características x_0 para las variables explicativas y un valor y_0^* para la variable dependiente. El contraste de hipótesis que se realiza sobre y_0 se denomina contraste para el análisis de la permanencia estructural del modelo o validación a posteriori. Su formulación es la siguiente:

$$H_0: y_0 = y_0^* \ (Existe\ permanencia\ estructural)$$
$$H_1: y_0 \neq y_0^* (No\ existe\ permanencia\ estrctural)$$

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

WUOLAH

Descarga la app de Wuolah desde tu store favorita

Que exista permanencia estructural para la observación x_0 y y_0^* de un determinado individuo que no pertenezca a la muestra usada en la estimación del modelo significa que el MRLG estimado predice bien las características x_0. Por tanto, este contraste sólo se puede realizar a posteriori, es decir, cuando ya se conoce el dato real y_0^*.

El estadístico para el contraste de permanencia estructural es:

$$\frac{y_0^* - \hat{y}_0}{\widehat{SE}(e_0)} \sim t_{n-k} \; bajo \; H_0.$$

a) <u>Usando el estadístico experimental.</u>

$$t_{exp} = \frac{y_0^* - \hat{y}_0}{\widehat{SE}(e_0)}$$

Se rechaza H_0 si $|t_{exp}| > t_{n-k;1-\frac{\alpha}{2}}$.

b) <u>Usando el p-valor.</u>

$$p - valor = 2 \times P(t > |t_{exp}|)$$

Se rechaza H_0 si $p - valor < \alpha$.

c) <u>Usando el intervalo de predicción para un individuo .</u>

Se rechaza H_0 si y_0^* no está incluido en el intervalo de predicción:

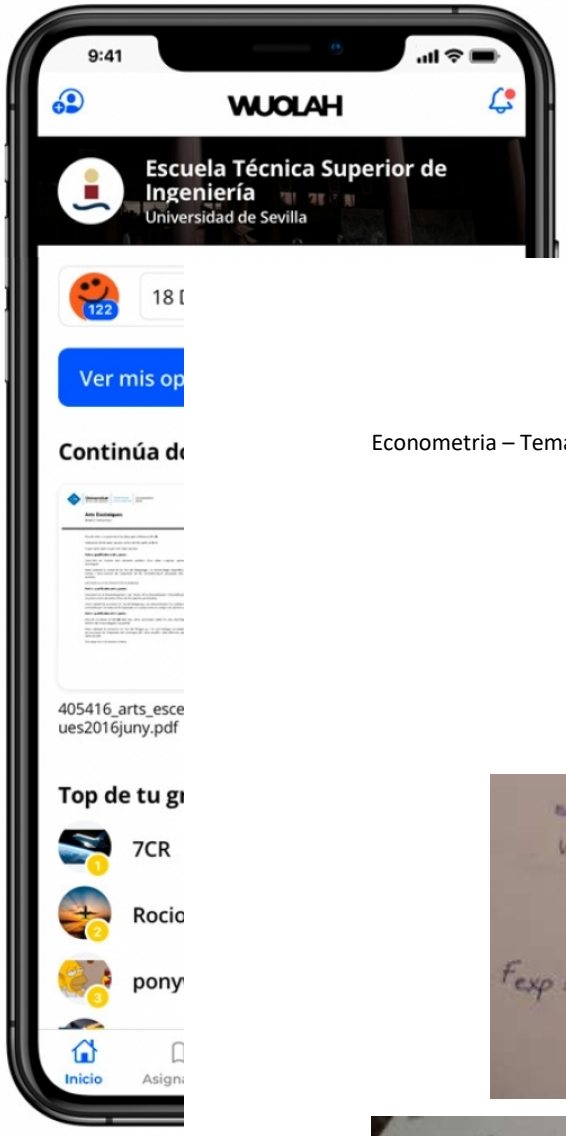
$$\left[\hat{y}_0 - t_{n-k;1-\frac{\alpha}{2}}\widehat{SE}(e_0), \qquad \hat{y}_0 + t_{n-k;1-\frac{\alpha}{2}}\widehat{SE}(e_0)\right]$$

Ejemplo 3.20. (Predicciones y contraste para el análisis de la permanencia estructural).

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

WUOLAH

Descarga la app de Wuolah desde tu store favorita

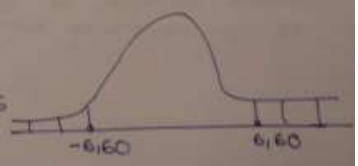


Descarga la APP de Wuolah.
Ya disponible para el móvil y la tablet.
Available on the App Store
GET IT ON Google Play

Algunos errores en exámenes

$$H_0: \hat{\beta}_1 = \hat{\beta}_2 = \beta_3 = \hat{\beta}_4 = 0$$
$$H_1: \beta_1 = \beta_2 = \beta_3 = \beta_4 \neq 0$$

$$SCT = n \cdot V(y_i) = 10 \cdot 0,5725 = 5,725$$
$$V(y_i) = \left(\frac{1}{n} \sum y_i^2\right) - \bar{y}_i^2 = 4,375 - 3,8025 = 0,5725$$
$$y_i^2 = \frac{1}{n} \sum y_i = 1,95$$
$$F_{exp} = \frac{\frac{0,9765}{3}}{\frac{1-0,9765}{6}} = 83,106$$
$$F_{3;6;0,975} = 6,60$$



④ $\hat{y} = -0,2103 + 0,3984 x_2 + 0,2664 x_3$
$(0,5360)$ $(0,1534)$ $(0,4415)$

$n = 10$

a) $H_0: x_2 = x_3 = 0$
$H_1: x_2 \neq 0$
$x_3 \neq 0$

$F_{exp} = \frac{SC}{S}$

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

WUOLAH

4Teoria2.pdf

jmgg94

Econometría

3º Doble Grado en Administración y Dirección de Empresas e Informática

Facultad de Ciencias Económicas y Empresariales Universidad de Granada

Descarga la APP de Wuolah.
Ya disponible para el móvil y la tablet.

Available on the App Store

GET IT ON Google Play

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

KEEP CALM AND ESTUDIA UN POQUITO

WUOLAH + #QuédateEnCasa

Descarga la APP de Wuolah.
Ya disponible para el móvil y la tablet.
Available on the App Store
GET IT ON Google Play

TEMA 4. Multicolinealidad

4.1. Concepto, causas y consecuencias.

- **<u>Concepto de la multicolinealidad</u>**

El problema de multicolinealidad surge por la existencia de relaciones lineales entre dos o más variables explicativas del modelo de regresión lineal general (MRLG). Se distingue entre:

- ***<u>Multicolinealidad perfecta:</u>*** cuando dos variables explicativas están perfectamente correlacionadas.
- ***<u>Multicolinealidad aproximada</u>***: cuando la correlación lineal entre dos o más variables explicativas sea muy elevada.

Ejemplo 4.1.

Para el MRLG:

$$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \cdots + \beta_k x_{ki} + u_i, \quad i = 1, \dots, n$$

se tendrá un problema de ***multicolinealidad perfecta*** cuando la variable x_3, por ejemplo, sea una combinación lineal de la variable x_2. Algunos ejemplos son los siguientes:

1. $x_{3i} = 3x_{2i}, \quad i = 1, \dots, n$
2. $x_{3i} = 2x_{2i} + 5, \quad i = 1, \dots, n$
3. $x_{3i} = \frac{x_{2i}}{2}, \quad i = 1, \dots, n$
4. $x_{3i} = 1 - x_{2i}, \; i = 1, \dots, n$. También se puede expresar como: $x_{3i} + x_{2i} = 1$.

□

Ejemplo 4.2.

Para el MRLG:

$$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i, \quad i = 1, \dots, n$$

se tendrá un problema de ***multicolinealidad perfecta*** cuando, por ejemplo, la matriz de observaciones X sea:

1. $X = \begin{pmatrix} 1 & 2 & 4 \\ 1 & 3 & 6 \\ 1 & 4 & 8 \\ 1 & 5 & 10 \end{pmatrix}$

No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.
Reservados todos los derechos.
WUOLAH

2. $X = \begin{pmatrix} 1 & 2 & 6 \\ 1 & 3 & 6 \\ 1 & 4 & 6 \\ 1 & 5 & 6 \end{pmatrix}$

Ejemplo 4.3.

Para el MRLG:

$$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i, \quad i = 1, \dots, n$$

se tendrá un problema de ***multicolinealidad aproximada*** cuando, por ejemplo, la matriz de observaciones X sea:

1. $X = \begin{pmatrix} 1 & 2 & 3.999 \\ 1 & 3 & 6.001 \\ 1 & 4 & 7.999 \\ 1 & 5 & 10.001 \end{pmatrix}$; $\det(X'X) = 0.000064$

2. $X = \begin{pmatrix} 1 & 2 & 5.999 \\ 1 & 3 & 6.001 \\ 1 & 4 & 6 \\ 1 & 5 & 6.001 \end{pmatrix}$; $\det(X'X) = 0.00003$

□

Ejemplo 4.4 (la trampa de las variables ficticias).

Una ***variable ficticia*** (o variable dummy) es una variable que toma únicamente los valores 0 y 1, y los cuales indican, en general, la presencia o ausencia de una determinada cualidad o atributo, como por ejemplo, sexo, raza, religión, etc. Por ejemplo, si se sospecha que el sexo de los individuos puede tener un efecto sobre la calificación final en la asignatura de Econometría (*notas_econometría*), la forma de incluir el sexo cómo una variable explicativa es mediante una variable ficticia. Para ello, definiremos la variable ficticia hombre de la siguiente forma:

$$hombre_i = \begin{cases} 1 & si\ el\ individuo\ i\ de\ la\ muestra\ es\ hombre. \\ 0 & si\ el\ individuo\ i\ de\ la\ muestra\ es\ mujer. \end{cases}$$

Teniendo en cuenta que el número de horas de estudio a la semana ($núm_horas_i$) es otra variable explicativa, el modelo que se plantea es el siguiente:

$$notas_econometría_i = \beta_1 + \beta_2 núm_horas_i + \beta_3 hombre_i + u_i, \quad i = 1, \dots, n$$

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

WUOLAH

Descarga la app de Wuolah desde tu store favorita

Alternativamente, también se puede plantear el siguiente modelo, el cual aportará las mismas conclusiones:

$$notas_econometría_i = \beta_1 + \beta_2 núm_horas_i + \beta_3 mujer_i + u_i, \quad i = 1, \dots, n$$

donde

$$mujer_i = \begin{cases} 1 & si\ el\ individuo\ i\ de\ la\ muestra\ es\ mujer. \\ 0 & si\ el\ individuo\ i\ de\ la\ muestra\ es\ hombre. \end{cases}$$

Sin embargo, lo que no se puede hacer es incluir ambas variables ficticias en un MRLG con término independiente, es decir, NO se puede plantear el siguiente modelo:

$$notas_econometría_i = \beta_1 + \beta_2 núm_horas_i + \beta_3 hombre_i + \beta_4 mujer_i + u_i, \quad i = 1, \dots, n$$

puesto que en este caso aparecerá un problema de multicolinealidad entre la variable $x_1=1$ asociada al término independiente y las variables ficticias *hombre* y *mujer*. En efecto, se puede comprobar que:

$$hombre_i + mujer_i = 1 = x_{1i}, \quad i = 1, \dots, n$$

A este problema de multicolinealidad generado por la inclusión en un modelo de regresión de dos variables ficticias complementarias se le conoce como la trampa de las variables ficticias.

□

Ejercicio 4.1.

Con el objetivo de contrastar si existen diferencias significativas entre las calificaciones en la asignatura de Econometría de hombres y mujeres, se plantea el siguiente modelo:

$$notas_econometría_i = \beta_1 + \beta_2 núm_horas_i + \beta_3 hombre_i + u_i, \quad i = 1, \dots, n$$

La estimación del modelo a partir de una muestra de 50 estudiantes es la siguiente:

$$\widehat{notas_econometría_i} = \underset{(0.3)}{1} + \underset{(0.1)}{0.5} núm_horas_i - \underset{(0.02)}{0.2}\ hombre_i$$

a) Interprete los coeficientes del modelo.
b) Especifique el modelo en caso de utilizar la variable ficticia $mujer_i$ e indique cuál sería el coeficiente estimado de esta nueva variable ficticia.
c) ¿Existen diferencias significativas entre la calificación final de hombres y mujeres?

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

WUOLAH

Descarga la app de Wuolah desde tu store favorita

- Causas de la multicolinealidad

Algunas de las causas que producen multicolinealidad en el MRLG son:

1. Las observaciones de las variables explicativas tienen muy poca variabilidad.
2. El tamaño de la muestra (n) es muy pequeño.
3. Existe una relación causal entre las variables explicativas.

- Consecuencias de la multicolinealidad

Las principales ***consecuencias o efectos de la multicolinealidad perfecta*** son:

1. La matriz X de observaciones de las variables explicativas no es de rango completo por columnas, es decir, $rango(X) < k$, y por este motivo no es posible obtener la inversa de la matriz $X'X$ y no existe una única solución para la estimación del vector $\hat{\beta}$.

Las principales ***consecuencias o efectos de la multicolinealidad aproximada*** son:

1. En este caso si es posible obtener la inversa de la matriz $X'X$, así como obtener los estimadores del vector $\hat{\beta}$ por el método MCO. Sin embargo, el determinante de la matriz $X'X$ estará próximo a cero, y la inversa de $X'X$ puede tener valores muy elevados. Por este motivo puede aparecer los siguientes problemas en presencia de multicolinealidad aproximada:
   - i) El error estándar de los estimadores de los parámetros del modelo pueden ser muy elevados.
   - ii) Los valores experimentales de los contrastes de significación individual de los parámetros del modelo suelen estar próximos a cero, y en este caso es muy probable que no se rechacen las hipótesis nulas de los contrastes de significación, es decir, los coeficientes no serán significativos.
   - iii) El coeficiente de determinación asociados al MRLG son muy sensibles ante pequeños cambios en los datos.
   - iv) Se tiende a rechazar la hipótesis nula de los contrastes de significación sobre varios parámetros del modelo, a diferencia de lo que ocurre con los ya comentados contrastes de significación individuales.
   - v) Las estimaciones de los coeficientes del modelo son muy sensibles ante pequeños cambios en los datos, y dada la alta variabilidad puede ocurrir que el signo de los coeficientes estimados sean contrarios al valor esperado.

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

WUOLAH

Descarga la app de Wuolah desde tu store favorita

Descarga la APP de Wuolah.
Ya disponible para el móvil y la tablet.
Available on the App Store
GET IT ON Google Play

4.2. Procedimientos de detección de multicolinealidad en la muestra.

4.2.1. Método del número de condición.

Sean λ_{min} y λ_{max} el autovalor más pequeño y más grande, respectivamente, de la matriz $X'X$. El número de condición asociado a la matriz X de observaciones de las variables explicativas se define como:

$$k_\lambda = \sqrt{\frac{\lambda_{max}}{\lambda_{min}}}$$

Según sea el valor del número de condición k_λ nos encontraremos en las siguientes situaciones:

1. $k_\lambda < 20$. Hay evidencias de que no existe un problema de multicolinealidad.
2. $20 \leq k_\lambda \leq 30$. Hay evidencias de un probable problema de multicolinealidad.
3. $k_\lambda > 30$. Hay evidencias claras de un problema de multicolinealidad.

4.2.2. Método del factor de agrandamiento de la varianza.

Sea R_i^2 el coeficiente de determinación obtenido cuando se realiza un modelo de regresión para explicar x_i en función del resto de variables explicativas del modelo original. Para cada coeficiente estimado del MRLG original, $\hat{\beta}_2, \dots, \hat{\beta}_k$, el factor de agrandamiento de la varianza (*FAV*), también denominado factor de inflación de la varianza (*VIF*), se define como:

$$FAV(\hat{\beta}_i) = \frac{1}{1 - R_i^2}$$

El *FAV* es una medida del grado en el que la varianza del coeficiente $\hat{\beta}_i$ se agranda como consecuencia de la relación entre x_i y el resto de variables explicativas. Según sea el valor del *FAV* nos encontraremos en las siguientes situaciones:

1. Si todos los valores del *FAV* son inferiores a 10 se determina que no hay indicios de un problema de multicolinealidad.
2. Si al menos el *FAV* de una de las variables explicativas es superior a 10 se concluye que hay indicios de un problema de multicolinealidad.

De forma equivalente al método *FAV*, se pueden considerar los siguientes criterios para determinar que existen indicios de multicolinealidad grave:

1. Si $R_i^2 \geq 0.9$ para al menos una de las variables explicativas.
2. Si la tolerancia (*TOL*) de una de las variables explicativas es menor de 0.1, donde $TOL_i = 1 - R_i^2$.

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.
WUOLAH

4.2.3. Análisis de contrastes de significación individual y global del modelo.

Siguiendo las comentadas consecuencias de la multicolinealidad, se tendrán indicios de multicolinealidad si se verifica:

1. Un coeficiente de determinación elevado ($R^2 \geq 0.7$) y se rechaza la hipótesis nula del contraste de significación global del modelo.
2. Pocas o ninguna de las variables explicativas del modelo son significativas de forma individual.

Ejemplo 4.5.

Se obtiene la siguiente estimación de un MRLG con tres variables explicativas:

$$\hat{y}_i = 105 - 0.8x_{2i} - 0.05x_{3i} + 0.1x_{4i}$$

Con el fin de determinar la posible existencia de multicolinealidad se plantean los siguientes modelos:

$$\hat{x}_{2i} = -5 + 0.006x_{3i} - 0.2x_{4i}; \quad R^2 = 0.25$$
$$\hat{x}_{3i} = 162 + 87x_{2i} - 0.4x_{4i}; \quad R^2 = 0.70$$
$$\hat{x}_{4i} = 1200 - 524x_{2i} - 176x_{3i}; \quad R^2 = 0.65$$

Además, se obtienen los siguientes autovalores de la matriz $X'X$: 0.7, 0.08, 1.1 y 2. Se pide comprobar la existencia de multicolinealidad.

En primer lugar, disponemos de los coeficientes de determinación R_2^2, R_3^2 y R_4^2, por lo que se puede aplicar el método del factor de agrandamiento de la varianza.

$$FAV(\hat{\beta}_2) = \frac{1}{1 - 0.25} = 1.33.$$
$$FAV(\hat{\beta}_3) = \frac{1}{1 - 0.70} = 3.33.$$
$$FAV(\hat{\beta}_4) = \frac{1}{1 - 0.65} = 2.86.$$

Todos los valores FAV son menores que 10, por lo que se concluye, a partir del método del factor de agrandamiento de la varianza que no existe multicolinealidad en el modelo. Alternativamente, se puede llegar a esta conclusión al verificarse que los coeficientes de determinación R_2^2, R_3^2 y R_4^2 son todos inferiores a 0.9, o bien, todas las tolerancias son mayores de 0.1:

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

WUOLAH

Descarga la app de Wuolah desde tu store favorita

$$TOL_2 = 1 - R_2^2 = 1 - 0.25 = 0.75$$
$$TOL_3 = 1 - R_3^2 = 1 - 0.70 = 0.30$$
$$TOL_4 = 1 - R_4^2 = 1 - 0.65 = 0.35$$

En segundo lugar, se puede aplicar el método del número de condición dado que se disponen de los autovalores de la matriz $X'X$. El número de condición de la matriz X viene dado por:

$$k_\lambda = \sqrt{\frac{\lambda_{max}}{\lambda_{min}}} = \sqrt{\frac{2}{0.08}} = \sqrt{25} = 5$$

Dado que $k_\lambda < 20$, el método del número de condición concluye que no existen evidencias que indiquen un problema de multicolinealidad.

□

4.3. Soluciones al problema de multicolinealidad.

1. **Eliminación o transformación de variables.** Si la variable es irrelevante se puede omitir, y en caso de ser relevante se puede realizar una transformación: primeras diferencias, cocientes, logaritmos, etc.
2. **Aumentar el tamaño de la muestra**. Especialmente en el caso de muestras con pocas observaciones.
3. **Extraer una nueva muestra mediante un mejor diseño muestral.** En el caso de que la muestra inicial no fuese representativa y se pueda extraer mayor información de los individuos, es decir, mayor variabilidad.
4. **Usar información a priori de la relación entre parámetros para estimar el MRLG por mínimos cuadrados restringidos.** Un ejemplo es el apartado b) del Ejercicio 4.4.
5. **Usar el estimador de la regresión alomada.**

$$\hat{\beta}_R = (X'X + cI)^{-1}X'Y$$

Donde $c > 0$ es una constante arbitraria e I es la matriz identidad. Este método presenta dos inconvenientes: es arbitrario y los estimadores de los coeficientes no son interpretables.

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

WUOLAH

Descarga la app de Wuolah desde tu store favorita

Ejercicio 4.2.

Sea el modelo $y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + u_i$ con perturbaciones centradas y esféricas. Con una muestra de 20 observaciones se estimó el modelo obteniendo los siguientes resultados:

$$\hat{y}_i = \underset{(0.8)}{8} + \underset{(0.9)}{0.7x_{2i}} - \underset{(0.03)}{0.4x_{3i}} + 0.1x_{4i}; \qquad R^2 = 0.99$$

Justifique si son coherentes los resultados obtenidos o puede haber algún problema en el modelo, indicando de qué problema se trata, y cómo lo solucionaría.

Sol. $t_{exp.2} = 0.875$; $t_{exp.3} = -0.444$; $t_{exp.4} = 3.333$; $t_{16;0.975} = 2.12$; $F_{exp} = 528$ $F_{3,16;0.95} = 3.24$.

Ejercicio 4.3.

Se obtiene la siguiente estimación de un MRLG con tres variables explicativas:

$$\hat{y}_i = 10 + 1.5x_{2i} - 0.1x_{3i} + 2x_{4i}$$

Con el fin de determinar la posible existencia de multicolinealidad se plantean los siguientes modelos:

$$\hat{x}_{2i} = 5 + 0.06x_{3i} - 2x_{4i}; \quad R^2 = 0.92$$

$$\hat{x}_{3i} = 16 + 8x_{2i} - 1x_{4i}; \quad R^2 = 0.95$$

$$\hat{x}_{4i} = 12 - 5x_{2i} - 2x_{3i}; \quad R^2 = 0.8$$

Además, se obtienen los siguientes autovalores de la matriz $X'X$: 0.5, 1, 0.002, y 0.3. Comprobar la existencia de multicolinealidad mediante distintos procedimientos.

Sol. $k_\lambda = 22.4$; *FAV* = 12.5; 20; 5.

Ejercicio 4.4.

Sea el modelo:

$$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i, \quad i = 1, \dots, n$$

Donde se sabe que

$$x_{2i} - x_{3i} = 1$$

a) Justificar qué parámetros o combinaciones lineales de parámetros son estimables mediante MCO. **Sol.** $(\beta_1 + \beta_2)$ y $(\beta_2 + \beta_3)$
b) Justificar qué parámetros o combinaciones lineales de parámetros son estimables mediante MCO cuando se sabe a priori que $\beta_3 = 2$. **Sol.** β_1 y β_2.

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

WUOLAH

Descarga la app de Wuolah desde tu store favorita

Descarga la APP de Wuolah.
Ya disponible para el móvil y la tablet.
Available on the App Store
GET IT ON Google Play

Ejercicio 4.5.

Para cada una de las siguientes matrices asociadas a un MRLG, analice la posible existencia de multicolinealidad.

$$X = \begin{pmatrix} 1 & 3 & 10 \\ 1 & 1 & 4 \\ 1 & 4 & 13 \end{pmatrix}; \qquad X = \begin{pmatrix} 1 & 3 & 10.001 \\ 1 & 1 & 4 \\ 1 & 4 & 12.999 \end{pmatrix} \qquad X = \begin{pmatrix} 1 & -10 & 3 \\ 1 & 6 & 44 \\ 1 & 0 & 6 \end{pmatrix};$$

Sol. $Det(X'X) = 0$;

$Det(X'X) = 0.000025$;

$Det(X'X) = 131044$.

Ejercicio 4.6.

Para el modelo:

$$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i, \quad i = 1, \dots, n$$

Se cumple que

$$\frac{x_{2i}}{x_{3i}} = \lambda$$

donde λ es una constante.

a) Justificar qué parámetros o combinaciones lineales de parámetros son estimables mediante MCO. **Sol. 1:** β_1 y $(\beta_2\lambda + \beta_3)$. **Sol. 2:** β_1 y $(\beta_2 + \beta_3/\lambda)$.

b) Justificar qué parámetros o combinaciones lineales de parámetros son estimables mediante MCO cuando se sabe a priori que $\beta_2 = 1$. **Sol.** β_1 y β_3.

Ejercicio 4.7.

Sea el modelo:

$$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + u_i, \quad i = 1, \dots, n$$

Donde se sabe que

$$x_{2i} = 3x_{4i}$$

a) Justificar qué parámetros o combinaciones lineales de parámetros son estimables mediante MCO. **Sol. 1:** β_1, β_3 y $(3\beta_2 + \beta_4)$. **Sol. 2:** β_1, β_3 y $(\beta_2 + \beta_4/3)$

b) Justificar qué parámetros o combinaciones lineales de parámetros son estimables mediante MCO cuando se sabe a priori que $\beta_4 = 2$. **Sol.** β_1, β_2 y β_3.

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

WUOLAH

Ejercicio 4.8.

A partir de una muestra de tamaño 30 y después de estimar un MRLG mediante MCO se obtienen los siguientes resultados:

- Modelo estimado y coeficiente de determinación:

$$\hat{y}_i = 40 + 0.4x_{2i} + 0.7x_{3i} + 0.6x_{4i}; \qquad R^2 = 0.998$$

- Estadísticos experimentales de contrastes de significación individual:

$$t_{exp.1} = 4;\ t_{exp.2} = 1.4;\ t_{exp.3} = 1.6;\ t_{exp.4} = 3.1$$

- Autovalores de la matriz $X'X$:

$$\lambda_1 = 0.00842;\ \lambda_2 = 0.00195;\ \lambda_3 = 3.01476$$

- Valores $R^2(x_j)$: Coeficientes de determinación de modelos con variable dependiente x_j y variables explicativas el resto de variables explicativas del modelo original:

$$R^2(x_2) = 0.996$$
$$R^2(x_3) = 0.992$$
$$R^2(x_4) = 0.998$$

Comprobar la existencia de multicolinealidad mediante distintos procedimientos.

Sol. 1: $k_\lambda = 39.3$.

Sol. 2: $FAV = 250; 125; 500$.

Sol. 3: $t_{exp.2} = 1.4$; $t_{exp.3} = 1.6$; $t_{exp.4} = 3.1$; $t_{26;0.975} = 2.06$; $F_{exp} = 4324.667$ $F_{3,26;0.95} = 2.975$.

<u>Algunas</u> preguntas de teoría del Tema 4

(Nota: Los exámenes de teoría de la asignatura pueden incluir cualquier pregunta incluida en el programa de la guía docente de la asignatura)

1. Concepto, causas y consecuencias de la multicolinealidad.
2. Descripción de la trampa de las variables ficticias.
3. Procedimientos de detección de multicolinealidad.
4. Descripción de soluciones al problema de la multicolinealidad.

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

WUOLAH

Descarga la app de Wuolah desde tu store favorita

5Teoria.pdf

jmgg94

Econometría

3º Doble Grado en Administración y Dirección de Empresas e Informática

Facultad de Ciencias Económicas y Empresariales Universidad de Granada

Descarga la APP de Wuolah.
Ya disponible para el móvil y la tablet.

Available on the App Store
GET IT ON Google Play

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

KEEP CALM AND ESTUDIA UN POQUITO

WUOLAH + #QuédateEnCasa

Descarga la APP de Wuolah.
Ya disponible para el móvil y la tablet.
Available on the App Store
GET IT ON Google Play

TEMA 5. Heterocedasticidad

5.1. Concepto, causas y consecuencias.

5.1.1. Concepto de heterocedasticidad.

Teniendo en cuenta el supuesto básico de perturbaciones aleatorias con media 0, $E[u_i] = 0$, el supuesto básico de ***homocedasticidad*** descrito en el Tema 2 establecía que:

1. La varianza del término error es constante para todas las observaciones de la muestra, es decir:

$$V[u_i] = E[u_i^2] = \sigma^2 \,, \quad i = 1, \dots, n$$

2. La ausencia de correlación serial:

$$Cov[u_i, u_j] = E[u_i u_j] = 0 \,, \quad i = 1, \dots, n$$

Estos supuestos implican que la matriz de varianzas-covarianzas del vector u de perturbaciones aleatorias viene dado por:

$$V[u] = \begin{pmatrix} \sigma^2 & 0 & \cdots & \cdots & 0 \\ 0 & \sigma^2 & 0 & \cdots & 0 \\ \vdots & 0 & \ddots & 0 & 0 \\ \vdots & \vdots & \vdots & \sigma^2 & 0 \\ 0 & 0 & 0 & 0 & \sigma^2 \end{pmatrix}$$

En presencia de ***heterocedasticidad*** esta matriz de varianzas-covarianzas viene dada por:

$$V[u] = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & \cdots & 0 \\ 0 & \sigma_2^2 & 0 & \cdots & 0 \\ \vdots & 0 & \ddots & 0 & 0 \\ \vdots & \vdots & \vdots & \sigma_{n-1}^2 & 0 \\ 0 & 0 & 0 & 0 & \sigma_n^2 \end{pmatrix}$$

es decir:

1. La varianza del término error no es constante:

$$V[u_i] = E[u_i^2] = \sigma_i^2 \,, \quad i = 1, \dots, n$$

2. Se sigue asumiendo la ausencia de correlación serial:

$$Cov[u_i, u_j] = E[u_i u_j] = 0 \,, \quad i = 1, \dots, n$$

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

WUOLAH

Ejemplo 5.1.

Sea el siguiente MRLG:

$$y_i = \beta_1 + \beta_2 x_{2i} + u_i, \quad i = 1, \dots, n$$

donde

- y_i: Gasto familiar.
- x_{2i}: Ingreso familiar.

¿Considera que se podría cumplir la hipótesis de homocedasticidad en este ejemplo?

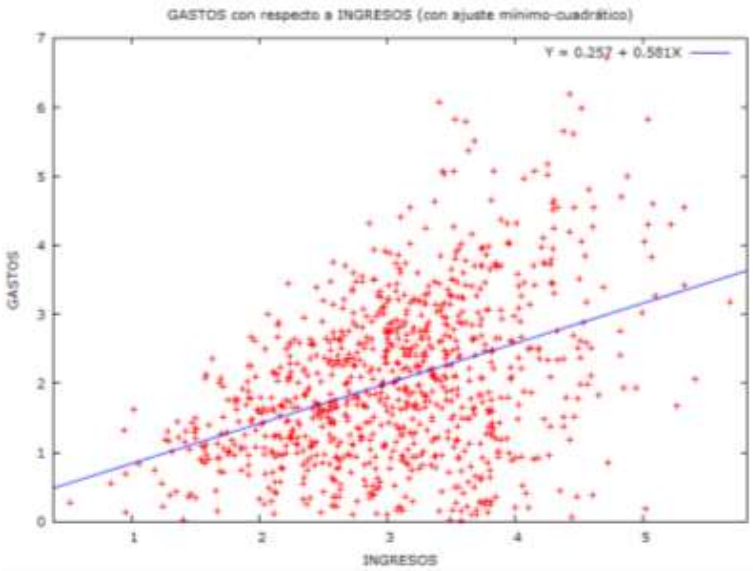


Figura 5.1. *Ejemplo de un MRLG con heterocedasticidad.*

5.1.2. Causas de la heterocedasticidad.

1. Naturaleza del fenómeno. Algunos ejemplos son:
   i) Ahorro (o gasto) familiar e ingreso familiar.
   ii) Consumo de bienes de lujo e ingreso familiar.
2. Uso de datos de corte transversal.
3. Uso de datos agregados.
4. Inclusión en el modelo de variables ficticias.
5. Presencia de factores atípicos.
6. Modelo incorrectamente especificado.

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

WUOLAH

Descarga la app de Wuolah desde tu store favorita

5.1.3. Consecuencias de la heterocedasticidad

1. En presencia de heterocedasticidad, los estimadores obtenidos por el método de Mínimos Cuadrados Ordinarios (EMCO) siguen siendo insesgados.
2. En presencia de heterocedasticidad, los EMCO siguen siendo consistentes.
3. En presencia de heterocedasticidad, los EMCO ya no son eficientes.

5.2. Procedimientos de detección de heterocedasticidad.

5.2.1. Métodos gráficos.

Un método gráfico para detectar heterocedasticidad consiste en representar en un sistema de ejes cartesianos los residuos al cuadrado respecto la variable que puede provocar heterocedasticidad. Existirán indicios de homocedasticidad si se observa una variabilidad constante para las distintas observaciones de la variable que se representa en el gráfico, y en caso de observar una mayor o menor variación para distintos valores de la variable se tendrá un problema de heterocedasticidad. En la Figura 5.2 se representa un MRLG con homocedasticidad, mientras que en la Figura 5.3 se tiene un ejemplo de heterocedasticidad.

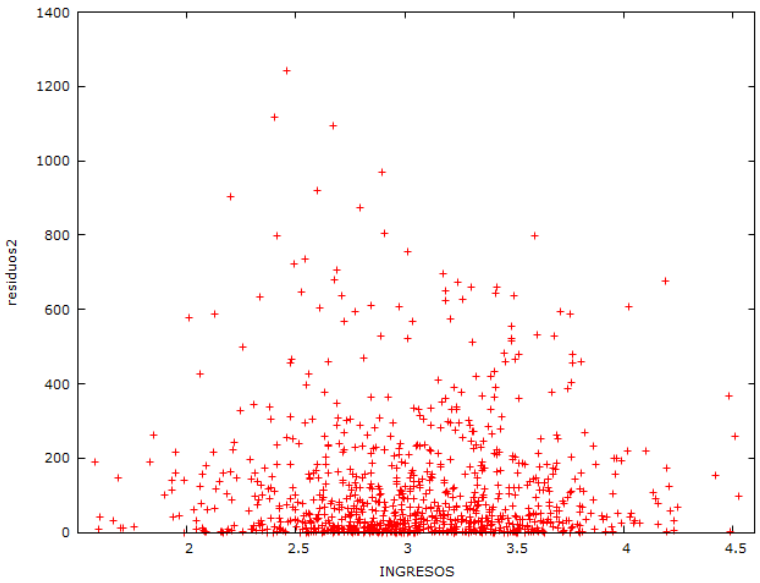


Figura 5.2. *Representación gráfica del cuadrado de residuos en función de una variable explicativa (ingresos) en un MRLG con homocedastidad.*

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

WUOLAH

Descarga la app de Wuolah desde tu store favorita

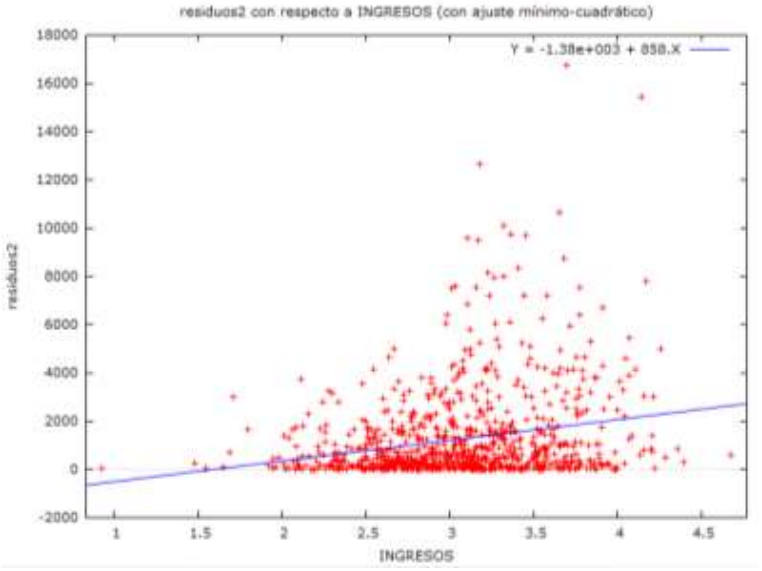


Figura 5.3. *Representación gráfica del cuadrado de residuos en función de una variable explicativa (ingresos) en un MRLG con heterocedasticidad.*

5.2.2. Test de Glesjer.

El test de Glesjer está indicado cuando:

1. El tamaño de la muestra es pequeño.
2. Una variable es la que causa heterocedasticidad.

Supongamos que la variable "*j*" (x_j) del MRLG es la que pensamos que causa heterocedasticidad. Para aplicar el test de Glesjer hay que seguir los siguientes pasos:

Paso 1. Ajustar el modelo mediante MCO y obtener los residuos: $e_i,\ \ i = 1, \dots, n$

Paso 2. Dado que es posible que no se detecte heterocedasticidad directamente a partir de la variable x_j, pero si a partir de una variable transformada basada en x_j, ajustar mediante MCO los siguientes modelos basados en la variable x_j:

$$|e_i| = \alpha_1 + \alpha_2\, x_{ji}^h + v_i,$$

donde *h* suele tomar varios de los siguientes valores: $h = \left\{1, -1, \frac{1}{2}, -\frac{1}{2}, 2, -2\right\}$

Por ejemplo, si $h = \left\{1, -1, \frac{1}{2}\right\}$, los modelos auxiliares que hay que plantear son

$$\begin{aligned} |e_i| &= \alpha_1 + \alpha_2\, x_{ji}^1 + v_i \\ |e_i| &= \alpha_1 + \alpha_2\, x_{ji}^{-1} + v_i \\ |e_i| &= \alpha_1 + \alpha_2\, x_{ji}^{1/2} + v_i \end{aligned}$$

y con tales modelos se está analizando si las variables x_j^1, x_j^{-1} y $x_j^{1/2}$ producen heterocedasticad en el modelo.

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

WUOLAH

Descarga la app de Wuolah desde tu store favorita

Descarga la APP de Wuolah.
Ya disponible para el móvil y la tablet.
Available on the App Store
GET IT ON Google Play

Paso 3. Para cada uno de los modelos adicionales, contrastar mediante el estadístico t-Student la hipótesis: $H_0 : \alpha_2 = 0$.

Paso 4. Para cada valor de h, se concluye que la variable x_j^h produce heterocedasticidad si se rechaza H_0. En el caso de concluir que son varias las variables x_j^h las que producen heterocedasticidad, se considerará la variable que proporciona un mayor coeficiente de determinación en el correspondiente modelo adicional, dado que en este caso el modelo tiene un mejor ajuste, es decir, el modelo explica un mayor porcentaje de la variabilidad del valor absoluto de los residuos.

Ejemplo 5.2.

Considere el siguiente modelo en el que la riqueza por unidad de consumo de un individuo (en miles de euros) de un determinado país está explicada en función de su renta por unidad de consumo (variable x_2 y expresada en miles de euros), edad (variable x_3 y medida en años) y número de hijos (x_4):

$$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + u_i, \quad i = 1, \dots, n$$

Después de aplicar MCO en una muestra de 30 observaciones se obtiene el siguiente modelo estimado:

$$\hat{y}_i = \underset{(1.10)}{19.8} + \underset{(0.71)}{0.83}\, x_{2i} + \underset{(0.03)}{0.91}\, x_{3i} - \underset{(0.20)}{10}\, x_{4i}$$

Se sospecha que la renta por unidad de consumo puede causar heterocedastidad, por lo que se plantean los siguientes modelos basados en los residuos del modelo anterior:

$$\widehat{|e_i|} = \underset{(0.59)}{0.21} + \underset{(0.29)}{9.21}\, x_{2i} \qquad R^2 = 0.536$$

$$\widehat{|e_i|} = \underset{(0.44)}{29.61} - \underset{(0.76)}{20.28}\, x_{2i}^{-1} \qquad R^2 = 0.487$$

$$\widehat{|e_i|} = \underset{(1.11)}{-16.48} + \underset{(0.79)}{25.03}\, x_{2i}^{1/2} \qquad R^2 = 0.512$$

1. ¿Qué test para la detección de heterocedasticidad se puede aplicar en este problema? ¿Está justificado su uso?

El ejercicio proporciona los 3 modelos basados en los residuos del modelo original del paso 2 del test de Glesjer, por lo que se podría aplicar este método para detectar heterocedasticidad. La muestra tiene tamaño pequeño, por lo que su uso estaría justificado.

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.
WUOLAH

2. Con un nivel de significación del 5%, ¿hay evidencias que indiquen que la renta por unidad de consumo provoca heterocedasticidad en el modelo?

En este ejercicio se proporcionan los pasos 1 y 2 del test de Glesjer, por lo que se realiza a continuación los siguientes pasos:

Paso 3. Para cada uno de los tres modelos para el valor absoluto de los residuos se obtiene el estadístico t-Student para el contraste de la hipótesis: $H_0 : \alpha_2 = 0$.

Los estadísticos experimentales t-Student para cada modelo son, respectivamente:

$$t_{exp} = \frac{9.21}{0.29} = 31.76 \quad ; \quad t_{exp} = \frac{-20.28}{0.76} = -26.68 \quad ; t_{exp} = \frac{25.03}{0.79} = 31.68$$

El valor crítico para resolver los contrastes es:

$$t_{n-2;1-\frac{\alpha}{2}} = t_{30-2;1-\frac{0.05}{2}} = t_{28;0.975} = 2.05$$

Paso 4. Para cada uno de los tres modelos se verifica:

$$|t_{exp}| > t_{n-2;1-\frac{\alpha}{2}},$$

por lo que hay evidencias de que las variables x_{2i}, x_{2i}^{-1} y $x_{2i}^{1/2}$ producen heterocedasticidad. Se selecciona la variable x_2 para explicar la varianza de la perturbación aleatoria, dado que con esta variable se obtiene un mayor coeficiente de determinación ($R^2 = 0.536$).

□

5.2.3. Test de Goldfeld-Quandt.

El test de Goldfeld-Quandt está indicado cuando:

1. El tamaño de la muestra es pequeño.
2. Una variable es la que causa heterocedasticidad.

Supongamos que de la variable "j" (x_j) del MRLG es la que pensamos que causa heterocedasticidad. Para aplicar el test de Goldfeld-Quandt hay que seguir los siguientes pasos:

Paso 1. Ordenar de forma creciente las observaciones de la variable x_j, la cual suponemos que puede provocar la heterocedasticidad.

Paso 2. Se omiten c observaciones centrales, y se crean por tanto dos grupos con las observaciones restantes. En general se considera $c \cong n/3$, es decir, un tercio del tamaño muestral. El tamaño del primer grupo se denotará por n_1, mientras que el segundo grupo contendrá n_2 observaciones, de forma que:

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

WUOLAH

Descarga la app de Wuolah desde tu store favorita

$$n = n_1 + c + n_2$$
$$n_1 \cong \frac{n}{3};\ c \cong \frac{n}{3}\ ;\ \ n_2 \cong \frac{n}{3}$$

Paso 3. Ajustar mediante MCO los dos grupos de observaciones restantes, y obtener la estimación de la varianza de la perturbación aleatoria de cada grupo:

$$\hat{\sigma}_1^2 = \frac{SCR_1}{n_1 - k};\quad \hat{\sigma}_2^2 = \frac{SCR_2}{n_2 - k}$$

Paso 4. Bajo la hipótesis nula:

$$H_0: Existe\ homocedasticidad$$

el estadístico para dicho contraste viene dado por

$$F = \frac{\hat{\sigma}_2^2}{\hat{\sigma}_1^2} \sim F_{n_2-k;n_1-k}$$

Paso 5. Con un nivel de significación α, Se rechaza H_0 si $F_{exp} > F_{n_2-k;n_1-k;1-\alpha}$, es decir, se concluye que la variable x_j proporciona heterocedasticidad en el modelo.

Ejemplo 5.3.

Considere el siguiente modelo en el que la riqueza por unidad de consumo de un individuo (en miles de euros) de un determinado país está explicada en función de su renta por unidad de consumo (variable x_2 y expresada en miles de euros), edad (variable x_3 y medida en años) y número de hijos (x_4):

$$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + u_i, \quad i = 1, \dots, n$$

Después de aplicar MCO en una muestra de 30 observaciones se obtiene el siguiente modelo estimado:

$$\hat{y}_i = 19.8 + 0.83\, x_{2i} + 0.91\, x_{3i} - 10\, x_{4i}$$
$$(1.10)\ \ (0.71) \qquad (0.03) \qquad (0.20)$$

Se sospecha que la renta por unidad de consumo puede causar heterocedastidad, por lo que se ordenan todas las observaciones según el orden creciente de la variable x_2, se omiten las c observaciones centrales y se obtienen las siguientes sumas de cuadrados de residuos para los dos grupos restantes de observaciones (con 10 observaciones en cada grupo): $SCR_1 = 240;\ \ SCR_2 = 1200$

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

WUOLAH

Descarga la app de Wuolah desde tu store favorita

1. ¿Qué test para la detección de heterocedasticidad se puede aplicar en este problema? ¿Está justificado su uso?

Se omiten *c* observaciones centrales, y ajusta modelos mediante MCO en los dos grupos restantes, es decir, aplica los primeros pasos del test de Goldfeld-Quandt. Además, el tamaño de la muestra es pequeño, por lo que el uso de este test estaría justificado.

2. ¿Cuál es el valor asignado a la constante *c*?

Se ha aplicado la regla general que indican que se omiten las

$$c = \frac{n}{3} = \frac{30}{3} = 10$$

observaciones centrales.

3. Con un nivel de significación del 5%, ¿hay evidencias que indiquen que la renta por unidad de consumo provoca heterocedasticidad en el modelo?

En este ejercicio se proporcionan los pasos 1 y 2 del test de Goldfeld-Quandt, y ajusta los modelos requeridos en el paso 3, por lo que se finaliza a continuación este paso 3, y se resuelven además los pasos siguientes del test de Goldfeld-Quandt:

Paso 3. Se obtiene la estimación de la varianza de la perturbación aleatoria de cada grupo:

$$\hat{\sigma}_1^2 = \frac{SCR_1}{n_1 - k} = \frac{240}{10-4} = 40$$

$$\hat{\sigma}_2^2 = \frac{SCR_2}{n_2 - k} = \frac{1200}{10-4} = 200$$

Paso 4. Bajo la hipótesis nula:

$$H_0: Existe\ homocedasticidad$$

el estadístico para este contraste viene dado por

$$F = \frac{\hat{\sigma}_2^2}{\hat{\sigma}_1^2} \sim F_{n_2-k;\, n_1-k} = F_{10-4;10-4} = F_{6;6}.$$

Paso 5. El estadístico experimental viene dado por:

$$F_{exp} = \frac{\hat{\sigma}_2^2}{\hat{\sigma}_1^2} = \frac{200}{40} = 5$$

Con un nivel de significación del 5%, el valor crítico del contraste es:

$$F_{n_2-k;n_1-k;1-\alpha} = F_{6;6;0.95} = 4.28$$

Se rechaza H_0 puesto que $F_{exp} > F_{n_2-k;n_1-k;1-\alpha}$, es decir, se concluye que la variable x_2 (renta por unidad de consumo) produce heterocedasticidad en el modelo.

□

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

WUOLAH

Descarga la app de Wuolah desde tu store favorita

Descarga la APP de Wuolah.
Ya disponible para el móvil y la tablet.
Available on the App Store
GET IT ON Google Play

5.2.4. Test de White.

El test de White está indicado cuando:

1. El tamaño de la muestra es elevado.
2. No se requiere especificar la naturaleza de la heterocedasticidad, al tratarse de un test asintótico.
3. Un conjunto de variables explicativas son las que causan heterocedasticidad.

Para aplicar el test de White se deben seguir los siguientes pasos:

Paso 1. Estimar el MRLG original mediante MCO y obtener los residuos.

Paso 2. Estimar un segundo modelo mediante MCO donde su variable endógena serán los residuos al cuadrado del modelo original, y sus variables explicativas serán todas las variables regresoras del modelo original, sus cuadrados, y los productos cruzados entre tales variables, y después de omitir posibles elementos redundantes.

Paso 3. Bajo la hipótesis nula:

$$H_0: Existe\ homocedasticidad$$

el estadístico para este contraste viene dado por

$$\chi^2 = n\, R_2^2 \sim \chi_q^2,$$

donde:

- n: es el tamaño de la muestra.
- R_2^2: es el coeficiente de determinación del segundo modelo planteado donde se explican los cuadrados de los residuos.
- q: es el número de variables explicativas del segundo modelo planteado donde se explican los cuadrados de los residuos.

Paso 4. Con un nivel de significación α, Se rechaza H_0 si $\chi_{exp}^2 > \chi_{q;1-\alpha}^2$, es decir, se concluye que el modelo tiene un problema de heterocedasticidad.

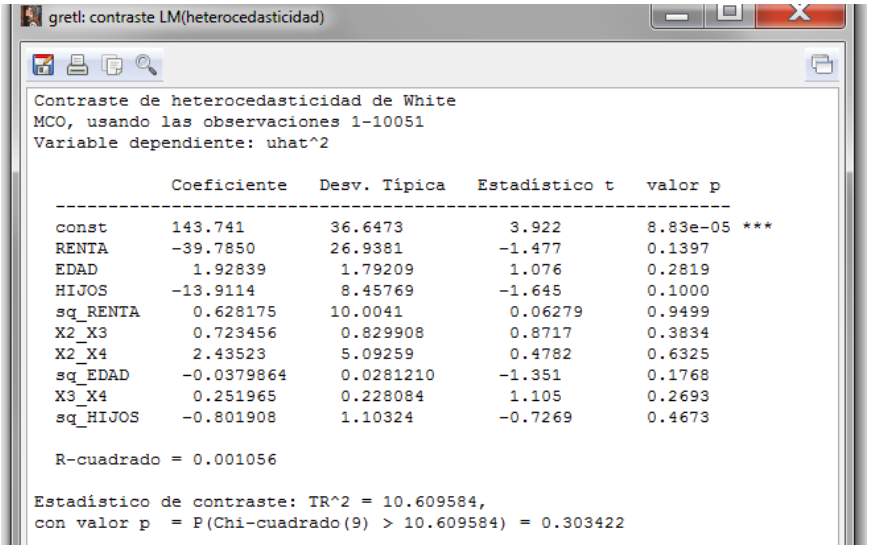
Ejemplo 5.4.

Considere el siguiente modelo en el que la riqueza por unidad de consumo de un individuo (en miles de euros) de un determinado país está explicada en función de su renta por unidad de consumo (variable x_2 y expresada en miles de euros), edad (variable x_3 y medida en años) y número de hijos (x_4):

$$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + u_i, \quad i = 1, \dots, n$$

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.
WUOLAH

Se estima dicho modelo mediante MCO y se aplica el test de White para detectar un posible problema de heterocedasticidad. Los resultados de este contraste pueden verse en la Figura 5.4.



```
gretl: contraste LM(heterocedasticidad)

Contraste de heterocedasticidad de White
MCO, usando las observaciones 1-10051
Variable dependiente: uhat^2

             Coeficiente   Desv. Típica   Estadístico t   valor p
  ---------------------------------------------------------------
  const      143.741       36.6473          3.922         8.83e-05 ***
  RENTA      -39.7850      26.9381         -1.477         0.1397
  EDAD         1.92839      1.79209         1.076         0.2819
  HIJOS      -13.9114       8.45769        -1.645         0.1000
  sq_RENTA     0.628175    10.0041          0.06279       0.9499
  X2_X3        0.723456     0.829908        0.8717        0.3834
  X2_X4        2.43523      5.09259         0.4782        0.6325
  sq_EDAD     -0.0379864    0.0281210      -1.351         0.1768
  X3_X4        0.251965     0.228084        1.105         0.2693
  sq_HIJOS    -0.801908     1.10324        -0.7269        0.4673

  R-cuadrado = 0.001056

Estadístico de contraste: TR^2 = 10.609584,
con valor p  = P(Chi-cuadrado(9) > 10.609584) = 0.303422
```

Figura 5.4. *Salida proporcionada por Gretl para el contraste de heterocedasticidad de White en el modelo lineal que explica la riqueza por unidad de consumo.*

1. Según los resultados de la Figura 5.4, ¿cuál es el valor de R_2^2?

El coeficiente de determinación del segundo modelo planteado donde se explican los cuadrados de los residuos tiene un valor de $R_2^2 = 0.001056$.

2. ¿Cuál es el valor de q?

El numero de variables explicativas del segundo modelo planteado donde se explican los cuadrados de los residuos es 9, es decir, $q = 9$.

3. ¿Cuál es el tamaño de la muestra?

$$n = 10051.$$

4. Con un nivel de significación del 5%, ¿A qué conclusión se llega con este contraste?

Bajo la hipótesis nula:

$$H_0: Existe\ homocedasticidad$$

el estadístico experimental viene dado por $\chi^2_{exp} = n\, R_2^2 = 10.6$, y el cual tiene asociado un p-valor igual a 0.3. No se rechaza H_0 puesto que p-valor > 0.05, y por tanto no hay indicios de un problema de heterocedasticidad en este modelo.

□

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

WUOLAH

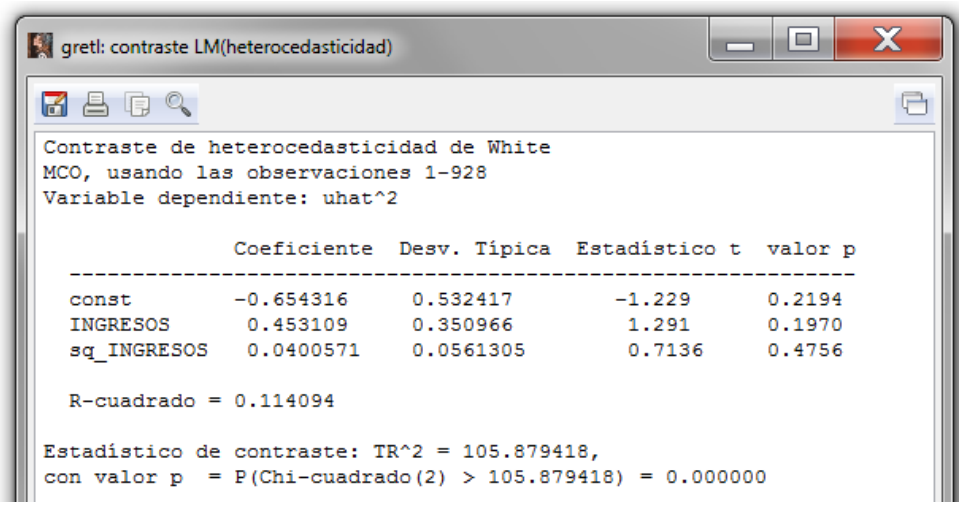
Descarga la app de Wuolah desde tu store favorita

Ejemplo 5.5.

Considere el siguiente modelo en el que el gasto familiar mensual (en miles de euros) está explicado en función del ingreso familiar mensual (variable x_2 y expresada en miles de euros):

$$y_i = \beta_1 + \beta_2 x_{2i} + u_i, \quad i = 1, \dots, n$$

Se estima dicho modelo mediante MCO y se aplica el test de White para detectar un posible problema de heterocedasticidad. Los resultados de este contraste pueden verse en la Figura 5.5.



```
gretl: contraste LM(heterocedasticidad)

Contraste de heterocedasticidad de White
MCO, usando las observaciones 1-928
Variable dependiente: uhat^2

                Coeficiente  Desv. Típica  Estadístico t  valor p
  ---------------------------------------------------------------
  const         -0.654316     0.532417        -1.229       0.2194
  INGRESOS       0.453109     0.350966         1.291       0.1970
  sq_INGRESOS    0.0400571    0.0561305        0.7136      0.4756

  R-cuadrado = 0.114094

Estadístico de contraste: TR^2 = 105.879418,
con valor p  = P(Chi-cuadrado(2) > 105.879418) = 0.000000
```

Figura 5.5. *Salida proporcionada por Gretl para el contraste de heterocedasticidad de White en el modelo lineal que explica la riqueza por unidad de consumo.*

1. Con un nivel de significación del 5%, ¿A qué conclusión se llega con este contraste?

Bajo la hipótesis nula:

$$H_0: Existe\ homocedasticidad$$

el estadístico experimental viene dado por $\chi^2_{exp} = nR_2^2 = 105.88$, y el cual tiene asociado un p-valor igual a 0. Se rechaza H_0 puesto que el p-valor < 0.05, y por tanto hay indicios de un problema de heterocedasticidad en este modelo.

□

WUOLAH

Descarga la app de Wuolah desde tu store favorita

No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.
Reservados todos los derechos.

5.2.5. Test de Breusch-Pagan.

El test de Breusch-Pagan está indicado cuando:

1. El tamaño de la muestra es elevado.
2. Un conjunto de variables explicativas son las que causan heterocedasticidad.

Para aplicar el test de Breusch-Pagan se deben seguir los siguientes pasos. Por simplicidad asumiremos que todas variables explicativas del modelo original son la causa de la heterocedasticidad.

Paso 1. Estimar el MRLG original mediante MCO y obtener los residuos.

Paso 2. Estimar un segundo modelo mediante MCO donde su variable endógena serán los residuos al cuadrado del modelo original, y sus variables explicativas serán las variables explicativas del modelo original, es decir:

$$e_i^2 = \alpha_1 + \alpha_2 x_{2i} + \cdots + \alpha_k x_{ki} + \varepsilon_i, \quad i = 2, \dots, n$$

Paso 3. Bajo la hipótesis nula:

$$H_0: Existe\ homocedasticidad$$

el estadístico para este contraste viene dado por

$$\chi^2 = n\, R_2^2 \sim \chi_{k-1}^2,$$

donde:

- *n*: es el tamaño de la muestra.
- R_2^2: es el coeficiente de determinación del segundo modelo planteado donde se explican los cuadrados de los residuos en función de las *k-1* variables explicativas.

Paso 4. Con un nivel de significación α, Se rechaza H_0 si $\chi_{exp}^2 > \chi_{k-1;1-\alpha}^2$, es decir, se concluye que el modelo tiene un problema de heterocedasticidad.

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

WUOLAH

Descarga la app de Wuolah desde tu store favorita

Descarga la APP de Wuolah.
Ya disponible para el móvil y la tablet.
Available on the App Store
GET IT ON Google Play

5.3. Estimación de modelos con heterocedasticidad.

5.3.1. Matriz de varianzas-covarianzas conocida de los errores salvo una constante multiplicativa (Estimador Mínimos Cuadrados Generalizados – MCG).

Asumimos que en el MRLG:

$$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \cdots + \beta_k x_{ki} + u_i, \quad i = 1, \dots, n$$

se satisfacen todos los supuestos básicos establecidos en el tema 2, excepto el supuesto básico de homocedasticidad. En este apartado consideramos que conocemos parte de la matriz de varianzas-covarianzas de los errores, es decir, esta matriz viene dada por:

$$V[u] = \sigma^2 \Omega$$

donde:

- σ^2 es una constante desconocida.
- Ω es una matriz diagonal y conocida, y cuyos elementos son:

$$\Omega = \begin{pmatrix} \omega_1 & 0 & 0 & \cdots & 0 \\ 0 & \omega_2 & 0 & \cdots & 0 \\ 0 & 0 & \ddots & \cdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & 0 & \omega_n \end{pmatrix}$$

Puede observarse que la varianza del error para cada individuo de la muestra viene dado por:

$$V[u_i] = \sigma^2 \omega_i, , \qquad i = 1, \dots, n$$

En esta sección asumiremos que los valores ω_i, con $i = 1, \dots, n$, son conocidos. Por ejemplo, la varianza del error para cada individuo de la muestra puede depender de la variable explicativa x_j mediante la siguiente expresión:

$$V[u_i] = \sigma^2 x_{ji}, , \qquad i = 1, \dots, n$$

es decir:

$$\omega_i = x_{ji}, \qquad i = 1, \dots, n$$

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.
WUOLAH

Bajo estas condiciones, la idea que se sigue es:

1. Realizar una transformación en el MRLG de forma que el modelo transformado dependa de errores con varianza constante.
2. A continuación estimar los parámetros del modelo transformado mediante MCO. A los estimadores obtenidos de esta forma se le denominan estimadores obtenidos por Mínimos Cuadrados Generalizados (MCG).

La transformación que permite obtener homocedasticidad en el modelo transformado es dividir el MRLG original por $\sqrt{\omega_i}$. Si aplicamos esta transformación en los errores:

$$u_i^* = \frac{u_i}{\sqrt{\omega_i}},$$

se puede comprobar que la varianza del término u_i^* es constante:

$$V[u_i^*] = V\left[\frac{u_i}{\sqrt{\omega_i}}\right] = \frac{1}{\omega_i}V[u_i] = \frac{\sigma^2\omega_i}{\omega_i} = \sigma^2 \ (constante = homocedasticidad)$$

Por tanto, los pasos para obtener el estimador por MCG son:

Paso 1. Dividir el MRLG original por $\sqrt{\omega_i}$:

$$\frac{y_i}{\sqrt{\omega_i}} = \beta_1\frac{1}{\sqrt{\omega_i}} + \beta_2\frac{x_{2i}}{\sqrt{\omega_i}} + \beta_3\frac{x_{3i}}{\sqrt{\omega_i}} + \cdots + \beta_k\frac{x_{ki}}{\sqrt{\omega_i}} + \frac{u_i}{\sqrt{\omega_i}}, \quad i = 1, \dots, n$$

Por motivos de simplicidad denotaremos:

$$y_i^* = \frac{y_i}{\sqrt{\omega_i}}; \quad x_{1i}^* = \frac{1}{\sqrt{\omega_i}}; \quad x_{2i}^* = \frac{x_{2i}}{\sqrt{\omega_i}}; \quad x_{3i}^* = \frac{x_{3i}}{\sqrt{\omega_i}}; \quad \dots; \quad x_{ki}^* = \frac{x_{ki}}{\sqrt{\omega_i}}; \quad u_i^* = \frac{u_i}{\sqrt{\omega_i}}$$

Y por tanto el modelo transformado es:

$$y_i^* = \beta_1 x_{1i}^* + \beta_2 x_{2i}^* + \beta_3 x_{3i}^* + \cdots + \beta_k x_{ki}^* + u_i^*, \quad i = 1, \dots, n$$

Paso 2. Estimar el modelo transformado por MCO. Al estimador obtenido por *MCG* se le denotará como $\hat{\beta}_{MCG}$, y por tanto su expresión será:

$$\hat{\beta}_{MCG} = (X^{*\prime}X^*)^{-1}X^{*\prime}y^*$$

$$y^* = \begin{pmatrix} y_1^* \\ y_2^* \\ \vdots \\ y_n^* \end{pmatrix}; \quad X^* = \begin{pmatrix} x_{11}^* & x_{21}^* & \cdots & x_{k1}^* \\ x_{12}^* & x_{22}^* & \cdots & x_{k2}^* \\ \vdots & \vdots & \ddots & \vdots \\ x_{1n}^* & x_{2n}^* & \cdots & x_{kn}^* \end{pmatrix}$$

WUOLAH

Descarga la app de Wuolah desde tu store favorita

No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad. Reservados todos los derechos.

5.3.2. Matriz de varianzas-covarianzas desconocida (Estimador Mínimos Cuadrados Generalizados Factible – MCGF).

En la sección anterior se asumía que la varianza del error viene dado por:

$$V[u_i] = \sigma^2 \omega_i, , \qquad i = 1, \dots, n$$

y además, se consideraba que los valores ω_i eran conocidos, con $i = 1, \dots, n$. Por ejemplo, si la variable explicativa x_j es la que provoca heterocedasticidad, podría ocurrir que $\omega_i = x_{ji}$. En esta sección se analiza el caso en el que se desconoce la variable que produce un efecto en la varianza del término error. En el caso de no conocer qué variable es la que afecta a la varianza del término error, una solución es asumir que:

$$\omega_i = exp(\delta_1 + \delta_2 x_{2i} + \delta_3 x_{3i} + \cdots + \delta_k x_{ki}).$$

Destacamos que su usa la función exponencial para garantizar que la función lineal de las variables explicativas sea positiva, al igual que la varianza del término error. En caso de no considerar la función exponencial se podría estimar una varianza negativa, lo cual no es factible por definición. Teniendo en cuenta esta expresión para ω_i, los pasos para obtener el estimador por *MCGF* son:

Paso A. Estimar por MCO el siguiente modelo:

$$\log(e_i^2) = \delta_1 + \delta_2 x_{2i} + \delta_3 x_{3i} + \cdots + \delta_k x_{ki} + \epsilon_i.$$

Los parámetros estimados de este modelo se denotan como: $\hat{\delta}_1,\ \hat{\delta}_2, \hat{\delta}_3, \cdots, \hat{\delta}_k$.

Paso B. Estimar los valores ω_i de la siguiente forma:

$$\widehat{\omega}_i = exp\left(\hat{\delta}_1 + \hat{\delta}_2 x_{2i} + \hat{\delta}_3 x_{3i} + \cdots + \hat{\delta}_k x_{ki}\right).$$

Paso C. Realizar los Pasos 1 y 2 del Estimador MCG, pero usando $\widehat{\omega}_i$ en lugar de ω_i.

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

WUOLAH

Descarga la app de Wuolah desde tu store favorita

Ejercicio 5.1.

Considere el siguiente modelo en el que la riqueza por unidad de consumo de un individuo (en miles de euros) de un determinado país está explicada en función de su renta por unidad de consumo (variable x_2 y expresada en miles de euros), edad (variable x_3 y medida en años) y número de hijos (x_4):

$$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + u_i, \quad i = 1, \dots, n$$

Después de aplicar MCO en una muestra de 57 observaciones se obtiene el siguiente modelo estimado:

$$\widehat{y_i} = \underset{(1.10)}{10} + \underset{(0.71)}{0.8}\, x_{2i} + \underset{(0.03)}{0.9}\, x_{3i} - \underset{(0.20)}{11}\, x_{4i}$$

Se sospecha que la renta por unidad de consumo puede causar heterocedastidad, por lo que se plantean los siguientes modelos basados en los residuos del modelo anterior:

$$\widehat{|e_i|} = \underset{(0.6)}{0.2} + \underset{(2)}{9}.\, x_{2i} \qquad R^2 = 0.6$$

$$\widehat{|e_i|} = \underset{(0.44)}{20} - \underset{(3)}{15}\, x_{2i}^{-1} \qquad R^2 = 0.5$$

$$\widehat{|e_i|} = \underset{(2)}{-10} + \underset{(10)}{15}\, x_{2i}^{1/2} \qquad R^2 = 0.4$$

Además, se ordenan todas las observaciones según el orden creciente de la variable x_2, se omiten las 19 observaciones centrales y se obtienen las siguientes sumas de cuadrados de residuos para los dos grupos restantes de observaciones: $SCR_1 = 900$; $SCR_2 = 1500$, y los cuales tienen el mismo número de observaciones.

Con un nivel de significación del 5%, ¿hay evidencias que indiquen que la renta por unidad de consumo provoca heterocedasticidad en el modelo? Utilice todos los procedimientos que se puedan realizar en este ejercicio.

Sol. t_{exp}= 4.5; -5; 1.5; F_{exp}= 1.67 , $F_{15,15,0.95}$=2.4.

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

WUOLAH

Descarga la app de Wuolah desde tu store favorita

Descarga la APP de Wuolah.
Ya disponible para el móvil y la tablet.
Available on the App Store
GET IT ON Google Play

Ejercicio 5.2.

Considere el siguiente modelo:

$$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 x_{5i} + u_i, \quad i = 1, \dots, n$$

a) Si se desea aplicar el Test de White, ¿cuántas variables explicativas tendrá el modelo adicional que se debe estimar y donde la variable endógena son los residuos al cuadrado del modelo original? Es decir, ¿cuál es el valor del parámetro q?

b) Si a partir de una muestra con 1000 observaciones se obtiene en el modelo adicional comentado en el apartado anterior un coeficiente de determinación de 0.03. ¿Hay evidencias de heterocedasticidad en el modelo con un 5% de nivel de significación?

c) Explique cómo se resolvería el problema de heterocedásticidad si se comprueba que el cuadrado de la variable x_4 es la que produce heterocedásticidad en el modelo.

Sol. a) q= 14. b) Valor crítico = 23.7; Valor Experimental = 30.

Ejercicio 5.3.

Dada una muestra de tamaño 1000 se estima el modelo $y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + u_t$ por MCO, se obtienen los residuos, y se realiza una regresión auxiliar donde es explican el cuadrado de los residuos en función de las variables x_2, x_3, los cuadrados de estas dos variables, el producto de ambas variables, y se incluye el término independiente en esta regresión auxiliar.

a) ¿Qué test se puede aplicar para detectar heterocedasticad? Formule el correspondiente contraste de hipótesis.

b) Si el coeficiente de determinación de la regresión auxiliar fue de 0.0095, ¿qué conclusión se obtiene a partir del correspondiente test? Sol: $\chi^2_{exp} = n\, R_2^2 = 9.5$; $\chi^2_{5;0.95} = 11.07$.

No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.
Reservados todos los derechos.

WUOLAH

Ejercicio 5.4.

A partir de una muestra con 40 observaciones se estima un MRLG con 3 variables explicativas, se obtiene los residuos de dicho modelo, y con éstos se estiman los siguientes modelos adicionales mediante MCO:

$$|e_i| = \alpha_1 + \alpha_2\, x_{3i}^h + v_i,$$

Los valores de h, los estadísticos experimentales del contraste de significación individual de la variable x_3^h, así como los correspondientes coeficientes de determinación de cada modelo se describen en la siguiente tabla:

h	t_{exp}	R^2
1	3.25	0.54
-1	2.94	0.61
1/2	1.45	0.42
-1/2	3.56	0.59
2	2.83	0.65
-2	1.63	0.39

a) ¿Qué variables de las sugeridas en los modelos adicionales producen heterocedásticidad en el modelo? Justifique la respuesta.
b) Justifique qué variable se debe seleccionar para realizar la estimación en presencia de heterocedasticidad, y explique este procedimiento de estimación.

<u>Algunas</u> preguntas de teoría del Tema 5

(Nota: Los exámenes de teoría de la asignatura pueden incluir cualquier pregunta incluida en el programa de la guía docente de la asignatura)

1. Concepto, causas y consecuencias de la heterocedasticidad.
2. Descripción de un método gráfico para detectar la heterocedastidad.
3. Explique el test de Glesjer incluyendo cuando está indicado su uso y los pasos para su realización.
4. Explique el test de Goldfeld-Quandt incluyendo cuando está indicado su uso y los pasos para su realización.
5. Explique el test de White incluyendo cuando está indicado su uso y los pasos para su realización.
6. Explique el test de Breush-Pagan incluyendo cuando está indicado su uso y los pasos para su realización.

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

WUOLAH

Descarga la app de Wuolah desde tu store favorita

6Teoria.pdf

jmgg94

Econometría

3º Doble Grado en Administración y Dirección de Empresas e Informática

Facultad de Ciencias Económicas y Empresariales Universidad de Granada

Descarga la APP de Wuolah.
Ya disponible para el móvil y la tablet.

Available on the App Store
GET IT ON Google Play

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

KEEP CALM AND ESTUDIA UN POQUITO

WUOLAH + #QuédateEnCasa

Descarga la APP de Wuolah.
Ya disponible para el móvil y la tablet.
Available on the App Store
GET IT ON Google Play

TEMA 6. Autocorrelación

6.1. Concepto, causas y consecuencias.

6.1.1. Concepto.

Teniendo en cuenta el supuesto de perturbaciones aleatorias con media 0, $E[u_i] = 0$, el supuesto básico de homocedasticidad descrito en el Tema 2 establecía que:

1. La varianza del término error es constante para todas las observaciones de la muestra, es decir:

$$V[u_i] = E[u_i^2] = \sigma^2 , \quad i = 1, \dots, n$$

2. La ***ausencia de correlación serial***:

$$Cov[u_i, u_j] = E[u_i u_j] = 0 , \quad i = 1, \dots, n$$

Estos supuestos implican que la matriz de varianzas-covarianzas del vector u de perturbaciones aleatorias viene dado por:

$$V[u] = \begin{pmatrix} \sigma^2 & 0 & \cdots & \cdots & 0 \\ 0 & \sigma^2 & 0 & \cdots & 0 \\ \vdots & 0 & \ddots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \sigma^2 & 0 \\ 0 & 0 & \cdots & 0 & \sigma^2 \end{pmatrix}$$

En presencia de ***autocorrelación*** asumiremos que las perturbaciones son homocedásticas pero autorreladas, es decir, la matriz de varianzas-covarianzas viene dada por:

$$V[u] = \begin{pmatrix} \sigma^2 & \sigma_{12} & \cdots & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma^2 & \cdots & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \sigma^2 & \sigma_{n-1\,n} \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{n\,n-1} & \sigma^2 \end{pmatrix}$$

es decir:

1. La varianza del término error es constante:

$$V[u_i] = E[u_i^2] = \sigma^2 , \quad i = 1, \dots, n.$$

2. No se cumple el supuesto de ausencia de correlación serial:

$$Cov[u_i, u_j] = E[u_i u_j] = \sigma_{ij} \neq 0 , \quad \forall\, i \neq j.$$

No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.
Reservados todos los derechos.
WUOLAH

En este tema consideraremos la siguiente notación:

$$y_t = \beta_1 + \beta_2 x_{2t} + \cdots + \beta_k x_{kt} + u_t, \quad t = 1, \dots, T$$

donde la etiqueta *t* hace referencia a un instante de tiempo, y los datos se recogen en un total de *T* periodos de tiempo. Destacamos que en el MRLG con perturbaciones homocedásticas, si existe autocorrelación, aumenta de modo considerable el número de parámetros a estimar, y en esta situación no se dispone de suficiente información para resolver este problema. Por este motivo, es decir, para poder realizar la estimación de parámetros, es común asumir una estructura de autocorrelación para el término error. Por ejemplo, un proceso autorregresivo[1] de orden 1 para la perturbación aleatoria, AR(1), consiste en la siguiente relación:

$$u_t = \rho u_{t-1} + \varepsilon_t \quad con\ -1 < \rho < 1.$$

Se asume que el término ε_t es un "ruido blanco": ε_t son independientes e idénticamente distribuidos a lo largo del tiempo con media cero e igual varianza:

$$\varepsilon_t \sim N(0, \sigma_\varepsilon^2); \quad cov(\varepsilon_t, \varepsilon_s) = 0 \ \forall\, t \neq s.$$

Si $\rho > 0$ estaríamos indicando que si en un año *t* (o el periodo de tiempo correspondiente: días, meses, trimestres, etc) el término error es positivo ($u_t > 0$), entonces para el año siguiente también se esperaría un error positivo ($u_{t+1} > 0$). Se puede demostrar que ρ es el coeficiente de correlación lineal entre las perturbaciones y las perturbaciones con un retardo.

□

6.1.2. Causas.

1. ***Datos de series temporales.*** Es muy común que en datos de series temporales el término error en un instante de tiempo esté relacionado con una o varias perturbaciones anteriores, puesto que la serie de datos puede presentar una tendencia en el tiempo.

2. ***Manipulación de datos.*** Algunas transformaciones en los datos (datos agregados, promedios, interpolaciones, etc) puede provocar la presencia de autocorrelación en las perturbaciones aleatorias.

[1] Un modelo autorregresivo es una representación de un proceso aleatorio en el que la variable de interés depende de sus observaciones pasadas.

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

WUOLAH

Descarga la app de Wuolah desde tu store favorita

3. ***Uso de variables retardadas en el MRLG.*** Cuando se plantea un modelo basado en datos de series temporales es muy común trabajar con modelos dinámicos[2], es decir, modelos que incluyen variables retardadas, ya sea la variable endógena o las exógenas. Por ejemplo:

$$y_t = \beta_1 + \beta_2 y_{t-1} + \beta_3 x_{3t} + u_t, \quad t = 1, \dots, T$$

4. ***Sesgo de especificación.***
   a) Utilización de una forma funcional incorrecta. Por ejemplo, si se tiene el siguiente modelo:

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{2t}^2 + u_t, \quad t = 1, \dots, T$$

   pero en su lugar se usa el modelo:

$$y_t = \beta_1 + \beta_2 x_{2t} + u_t, \quad t = 1, \dots, T$$

   b) Exclusión de variables explicativas autocorrelacionadas.

6.1.3. Consecuencias (si se aplica MCO a un MRLG con perturbaciones autocorrelacionadas).

1. El estimador del vector de parámetros β obtenido por MCO, $\hat{\beta}$, sigue siendo insesgado, es decir: $E[\hat{\beta}] = \beta$.
2. No se cumple el Teorema de Gauss-Markov, es decir, $\hat{\beta}$ no es el estimador con menor varianza entre la clase de estimadores lineales e insesgados.
3. Con motivo de la propiedad anterior, se ve afectada la estimación por intervalos de confianza y los contrastes de hipótesis.
4. El predictor óptimo no es el dado en MCO, sino el MCG.

6.2. Procedimientos de detección de autocorrelación

6.2.1. Métodos gráficos.

- **Gráfico del término error en el tiempo**

Un método gráfico para detectar autocorrelación consiste en representar en un sistema de ejes cartesianos los residuos (Eje *Y*) respecto el tiempo (Eje *X*). Existirán indicios de autocorrelación si se observa que, en general, el valor del residuo en un determinado instante de tiempo depende del valor del residuo del periodo anterior. En las Figuras 6.1 y 6.2 se representan varios MRLG: sin autocorrelación; con autocorrelación positiva; con autocorrelación negativa.

[2] Un modelo econométrico dinámico es un modelo econométrico en el que las variables explicativas presentan retardos, por lo que tiene sentido en presencia de una serie temporal.

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

WUOLAH

Descarga la app de Wuolah desde tu store favorita

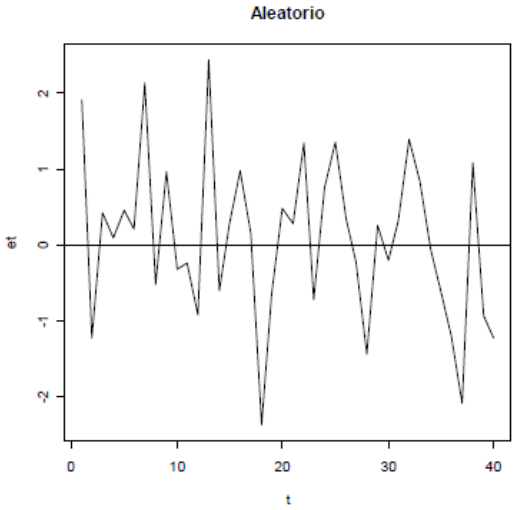


Figura 6.1. *Representación gráfica de residuos en función del tiempo en un MRLG sin autocorrelación.*

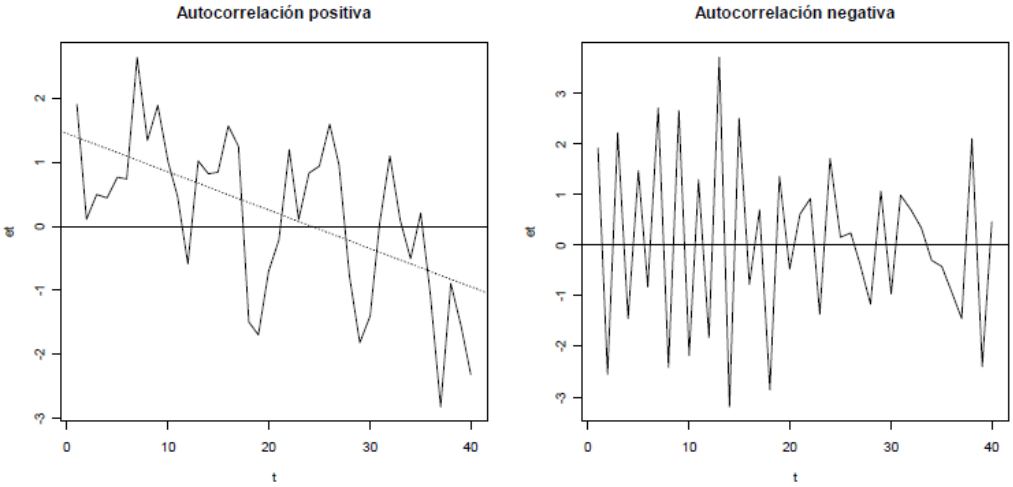


Figura 6.2. *Representación gráfica de residuos en función del tiempo en dos MRLG: con autocorrelación positiva y autocorrelación negativa.*

- **Gráfico de dispersión entre el término error en un instante de tiempo y el término error en el periodo anterior.**

En las Figuras 6.3 y 6.4 se representan estos gráficos de dispersión para varios MRLG: sin autocorrelación; con autocorrelación positiva; con autocorrelación negativa.

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

WUOLAH

Descarga la app de Wuolah desde tu store favorita

Descarga la APP de Wuolah.
Ya disponible para el móvil y la tablet.
Available on the App Store
GET IT ON Google Play

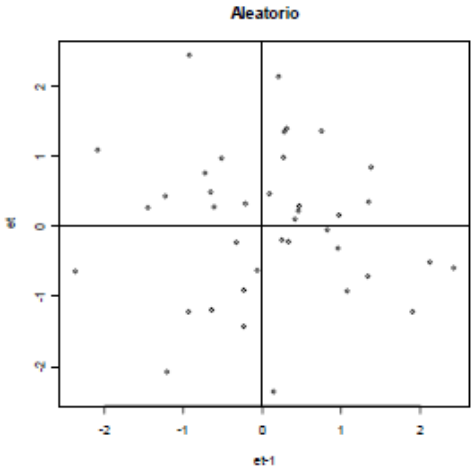


Figura 6.3. *Gráfico de dispersión entre los residuos en un instante de tiempo (e_t) y los residuos en el periodo anterior (e_{t-1}) en un MRLG sin autocorrelación.*

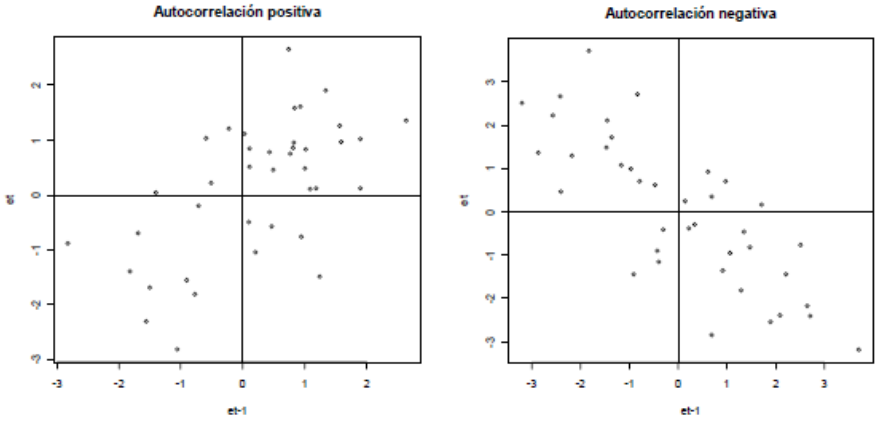


Figura 6.4. *Gráfico de dispersión entre los residuos en un instante de tiempo (e_t) y los residuos en el periodo anterior (e_{t-1}) en dos MRLG: con autocorrelación positiva y autocorrelación negativa.*

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

WUOLAH

6.2.2. Contraste de Durbin-Watson.

El test de Durbin-Watson está indicado cuando:

1. El modelo es lineal y con término independiente.
2. Las variables explicativas también son no estocásticas[3].
3. Las perturbaciones aleatorias se expresan mediante un modelo AR(1).

El test de Durbin-Watson no está indicado cuando:

1. Los modelos son autorregresivos. Por ejemplo, no está indicado cuando una de las variables explicativas se la variable endógena retardada, es decir:

$$y_t = \beta_1 + \beta_2 y_{t-1} + \beta_3 x_{3t} + u_t, \quad t = 1, \dots, T$$

2. Los datos son de corte transversal.

Supongamos un MRLG donde las perturbaciones aleatorias siguen un modelo AR(1), es decir,

$$u_t = \rho u_{t-1} + \varepsilon_t \quad con\ -1 < \rho < 1.$$

y donde el término ε_t es un "ruido blanco". El contraste de hipótesis de Durbin-Watson es el siguiente:

$$H_0: \rho = 0$$
$$H_1: \rho \neq 0$$

Para aplicar el test de Durbin-Watson hay que seguir los siguientes pasos:

Paso 1. Ajustar el modelo original mediante MCO y obtener los residuos:

$$e_t = y_t - \hat{y}_t, \qquad t = 1, \dots, T$$

Paso 2. Obtener el estadístico de Durbin-Watson:

$$DW = \frac{\sum_{t=2}^{T}(e_t - e_{t-1})^2}{\sum_{t=1}^{T} {e_t}^2} \cong 2(1-\hat{\rho}); \qquad \hat{\rho} \cong \frac{\sum_{t=2}^{T} e_t e_{t-1}}{\sum_{t=1}^{T} {e_t}^2}$$

Se observa que:

1. $\hat{\rho} = 0$ implicaría que no existe autocorrelación, y: $DW \cong 2$.
2. $\hat{\rho} = 1$ implicaría autocorrelación positiva, y: $DW \cong 0$.
3. $\hat{\rho} = -1$ implicaría autocorrelación negativa, y: $DW \cong 4$.

[3] Estocástico: Teoría estadística de los procesos cuya evolución en el tiempo es aleatoria.

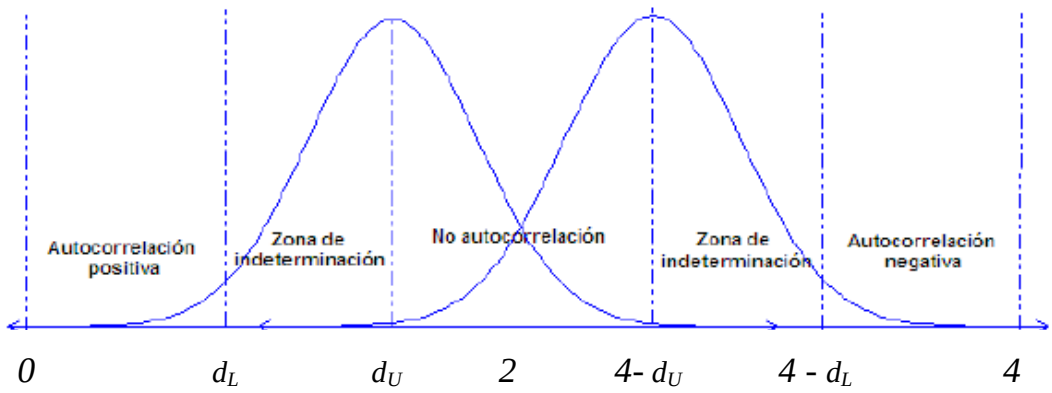
Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

WUOLAH

Descarga la app de Wuolah desde tu store favorita

Paso 3. Obtener los valores críticos del contraste de Durbin-Watson: d_L y d_U. Para ello se pueden usar tablas estadísticas del contraste de Durbin-Watson, o bien programas estadísticos como Gretl.

Paso 4. Concluir la presencia o no de autocorrelación según se sitúe el estadístico DW en el siguiente esquema:



Como se puede observar, este método tiene el inconveniente de que posee zonas de indeterminación, es decir, en estos casos el contraste de Durbin-Watson no es concluyente.

Ejemplo 6.1.

Consideremos el siguiente modelo:

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \beta_4 x_{4t} + u_t, \quad t = 1, \dots, T$$

A partir de una muestra con 17 observaciones se obtiene el siguiente modelo estimado:

$$\hat{y}_t = 25 + 0.9x_{2t} - 1.1x_{3t} + 0.1x_{4t}$$
$$(15) \quad (2.3) \qquad (0.3) \qquad (0.05) \quad DW = 1.55$$

1. Con un nivel de significación del 5%, ¿existe autocorrelación de orden 1 a partir del contraste de Durbin-Watson?

Paso 3. Obtenemos los valores críticos del contraste de Durbin-Watson, teniendo en cuenta que $T = 17$ y $k^* = k - 1 = 3$:

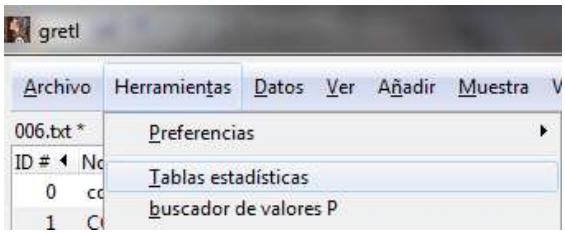
- $d_L = 0.897$.
- $d_U = 1.710$.

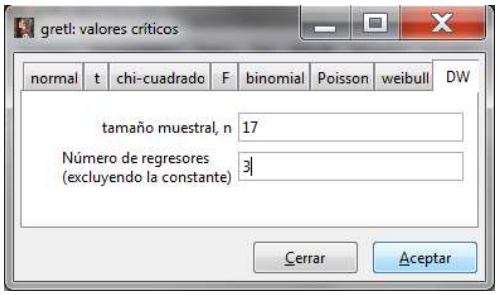
Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

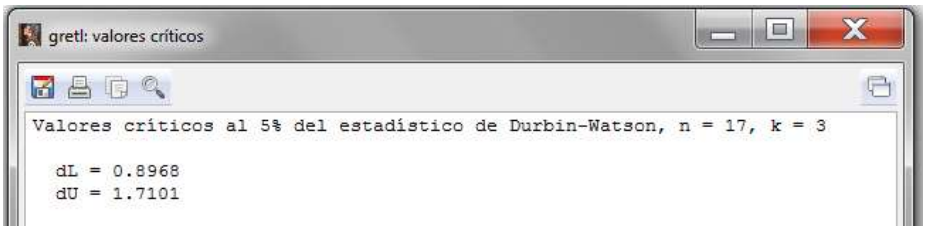
WUOLAH

Descarga la app de Wuolah desde tu store favorita

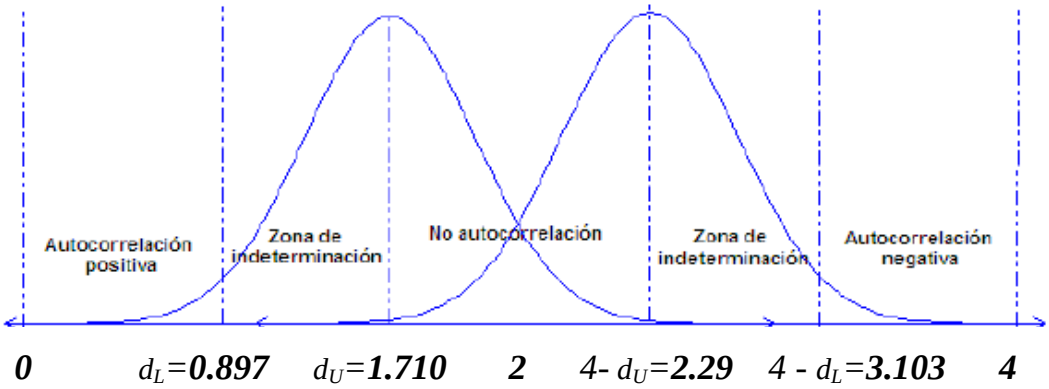
Estos valores críticos se pueden obtener usando Gretl y siguiendo las siguientes instrucciones:







Paso 4. Elaboramos el esquema de posibles zonas del contraste de Durwin-Watson a partir de los valores críticos:



El estadístico DW = 1.55 se sitúa en una zona de indeterminación, por lo que el contraste de Durwin-Watson no es concluyente para determinar la presencia de autocorrelación.

□

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

WUOLAH

Descarga la app de Wuolah desde tu store favorita

Descarga la APP de Wuolah.
Ya disponible para el móvil y la tablet.
Available on the App Store
GET IT ON Google Play

6.2.3. Constraste *h* de Durbin.

El test *h* de Durbin está indicado para el siguiente modelo autorregresivo:

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \cdots + \beta_j \boldsymbol{y_{t-1}} + \cdots + \beta_k x_{kt} + u_t, \quad t = 1, \dots, T$$

donde

$$u_t = \rho u_{t-1} + \varepsilon_t \quad con \ -1 < \rho < 1.$$

La variable retardada y_{t-1} se ha ubicado junto al parámetro β_j, pero sería posible ubicarla junto a cualquier otro parámetro. Para aplicar el contraste h de Durbin habría que sustituir β_j por el parámetro que acompaña a y_{t-1}.

El contraste de hipótesis del test h de Durbin es el siguiente:

$$H_0: \rho = 0$$
$$H_1: \rho \neq 0$$

Paso 1. Ajustar el modelo autorregresivo mediante MCO y obtener los residuos y los errores estándar de los coeficientes estimados.

Paso 2. Obtener el estadístico h de Durbin, el cual tiene distribución Normal estándar bajo la hipótesis nula:

$$h = \hat{\rho}\sqrt{\frac{T}{1 - T \times \hat{V}(\hat{\beta}_j)}} \to N(0,1); \qquad \hat{\rho} \cong \frac{\sum_{t=2}^{T} e_t e_{t-1}}{\sum_{t=1}^{T} {e_t}^2}$$

Al final del tema se describen otros métodos para estimar ρ.

Paso 3. Se rechaza H_0 si $|h| > Z_{1-\frac{\alpha}{2}}$.

El inconveniente del test h de Durbin es que se obtiene la raíz cuadrad de un valor negativo cuando $T \times \hat{V}(\hat{\beta}_j) > 1$.

Ejemplo 6.2.

Para el periodo 2010-2017 se estima el siguiente modelo por MCO:

$$\hat{y}_t = \underset{(1.1)}{3} - \underset{(0.15)}{0.55} x_{2t} + \underset{(0.20)}{0.9} x_{3t} - \underset{(0.05)}{1.1} y_{t-1} + \underset{(0.20)}{0.8} x_{5t},$$

$$R^2 = 0.95; \quad DW = 3.6$$

WUOLAH

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

1. Suponiendo que las perturbaciones aleatorias siguen un esquema *AR(1)*, Contrastar, con un 5% de nivel de significación, la existencia de autorrelación en el modelo.

Una de las variables explicativas del modelo es la propia variable endógena pero retardada, por lo que el contraste de Durbin-Watson no es apropiado. Por este motivo se usa la prueba h de Durbin. El contraste de hipótesis del test h de Durbin es el siguiente:

$$H_0: \rho = 0$$
$$H_1: \rho \neq 0$$

donde el modelo de las perturbaciones aleatorias es:

$$u_t = \rho u_{t-1} + \varepsilon_t$$

Paso 2. Obtenemos el estadístico h de Durbin:

El valor de $\hat{\rho}$ se puede despejar de la siguiente expresión, dado que el estadístico DW lo proporciona el ejercicio:

$$DW = 3.6 \cong 2(1 - \hat{\rho})$$
$$1.8 \cong 1 - \hat{\rho}$$
$$\hat{\rho} = -0.8$$

$$h = -0.8\sqrt{\frac{8}{1 - 8 \times 0.05^2}} = -2.29$$

Paso 3. Se rechaza H_0 puesto que $|h| = 2.29 > Z_{1-\frac{\alpha}{2}} = 1.96$, es decir, hay indicios de autocorrelación en el modelo.

2. ¿Puede afirmarse que la correlación es negativa según el contraste h de Durbin?

El contraste de hipótesis es el siguiente:

$$H_0: \rho = 0$$
$$H_1: \rho < 0$$

El estadístico h de Durbin $h = -2.29$ es menor que $Z_\alpha = -1.645$, y por tanto se rechaza H_0, es decir, existen evidencias de una autocorrelación negativa en el modelo.

□

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

WUOLAH

Descarga la app de Wuolah desde tu store favorita

6.2.4. Contraste de Ljung-Box.

El contraste de Ljung-Box es más general que los contrastes anteriores, de forma que si los residuos son independientes, sus primeras *m* autocorrelaciones son cero. El contraste de hipótesis asociados es:

$$H_0: \rho_1 = \rho_2 = \cdots = \rho_m = 0$$
$$H_1: \rho_j \neq 0 \; para \; algún \; j \in \{1,2,\dots,m\}$$

El valor de *m* se puede fijar de forma arbitraria con la restricción de que no puede ser elevado. Para aplicar el contraste de Ljung-Box se deben seguir los siguientes pasos:

Paso 1. Estimar el MRLG original mediante MCO y obtener los residuos.

Paso 2. Obtener el estadístico *Q* de Ljung-Box, el cual tiene distribución Chi-cuadrado con *m* grados de libertad bajo la hipótesis nula:

$$Q = T(T+2)\sum_{j=1}^{m}\frac{\hat{\rho}_j^2}{T-j} \rightarrow \chi_m^2;$$

donde el coeficiente de autocorrelación muestral de orden *j* viene dado por:

$$\hat{\rho}_j = \frac{\sum_{t=j+1}^{T} e_t e_{t-j}}{\sum_{t=1}^{T} {e_t}^2}.$$

Paso 3. Se rechaza H_0 si $Q > \chi_{m;1-\alpha}^2$.

Ejemplo 6.3.

Para el modelo estimado del Ejemplo 6.2 se obtienen los siguientes coeficientes de autocorrelación muestral: $\hat{\rho}_1 = -0.35$; $\hat{\rho}_2 = -0.3$; $\hat{\rho}_3 = -0.4$; $\hat{\rho}_4 = -0.5$.
Con un nivel de significación del 5%, contraste mediante el test de Ljung-Box la presencia de autocorrelación en el modelo.

Paso 2. Obtenemos el estadístico *Q* de Ljung-Box, teniendo en cuenta que $m = 4$:

$$Q = T(T+2)\sum_{j=1}^{m}\frac{\hat{\rho}_j^2}{T-j} = 8 \times 10 \times \left(\frac{(-0.35)^2}{8-1} + \frac{(-0.3)^2}{8-2} + \frac{(-0.4)^2}{8-3} + \frac{(-0.5)^2}{8-4}\right)$$
$$= 10.16$$

Paso 3. Se rechaza H_0 puesto que $Q = 10.16 > \chi_{m;1-\alpha}^2 = 9.49$, es decir, hay indicios de autocorrelación en el modelo.

□

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

WUOLAH

Descarga la app de Wuolah desde tu store favorita

6.3. Estimación de modelos con perturbaciones autocorreladas.

Asumimos el MRLG:

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \cdots + \beta_k x_{kt} + u_t, \quad t = 1, \dots, T,$$

con

$$u_t = \rho u_{t-1} + \varepsilon_t \,.$$

y donde suponemos que ε_t es ruido blanco, es decir,

$$\varepsilon_t \sim N(0, \sigma_\varepsilon^2); \quad cov(\varepsilon_t, \varepsilon_s) = 0 \;\; \forall\, t \neq s.$$

y cumple las hipótesis básicas establecidas para la estimación mediante MCO:

- Perturbaciones aleatorias con media cero.
- Perturbaciones aleatorias homocedásticas.
- Perturbaciones aleatorias incorreladas.

Por tanto, la idea es transformar el modelo de forma que dicho modelo transformado tenga como perturbaciones aleatorias los valores ε_t. Bajo estas condiciones la estimación de modelo se realiza mediante los siguientes pasos:

Paso 1. Retardar el modelo un periodo y multiplicar por ρ:

$$\rho y_{t-1} = \rho\beta_1 + \rho\beta_2 x_{2t-1} + \rho\beta_3 x_{3t-1} + \cdots + \rho\beta_k x_{kt-1} + \rho u_{t-1}, \quad t = 2, \dots, T$$

Paso 2. Al modelo original se le resta el modelo del Paso 1, obteniendo la denominada ***<u>ecuación en primeras diferencias generalizada</u>***.

$$\begin{aligned} y_t &= \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \cdots + \beta_k x_{kt} + u_t \\ \rho y_{t-1} &= \rho\beta_1 + \rho\beta_2 x_{2t-1} + \rho\beta_3 x_{3t-1} + \cdots + \rho\beta_k x_{kt-1} + \rho u_{t-1} \end{aligned}$$

$$y_t - \rho y_{t-1} = \beta_1(1-\rho) + \beta_2(x_{2t} - \rho x_{2t-1}) + \cdots + \beta_k(x_{kt} - \rho x_{kt-1}) + (u_t - \rho u_{t-1})$$

$$y_t^* = \beta_1^* + \beta_2 x_{2t}^* + \beta_3 x_{3t}^* + \cdots + \beta_k x_{kt}^* + \varepsilon_t, \quad t = 2, \dots, T.$$

Paso 3. Estimar por MCO el modelo resultante en el Paso 2. Se observa que el coeficiente del término independiente cambia de β_1 a β_1^*. La relación entre ambos es la siguiente:

$$\beta_1^* = \beta_1(1-\rho)\,; \qquad \beta_1 = \frac{\beta_1^*}{1-\rho}$$

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

WUOLAH

Descarga la app de Wuolah desde tu store favorita

Descarga la APP de Wuolah.
Ya disponible para el móvil y la tablet.
Available on the App Store
GET IT ON Google Play

En la práctica, el valor de ρ no suele ser conocido, por lo que se requiere una estimación para poder obtener la estimación del modelo con perturbaciones autocorreladas. En las siguientes secciones se presentan varias alternativas para estimar ρ, usando métodos directos y métodos por etapas.

6.3.1. Estimación directa de ρ.

1. **Usando la ecuación en primeras diferencias generalizada**. Por ejemplo:
   a) Estimar el modelo original por MCO y obtener $\hat{\beta}_1$.
   b) Estimar el modelo en primeras diferencias y obtener $\hat{\beta}_1^*$.
   c) A partir de la ecuación:
   $$\beta_1 = \frac{\beta_1^*}{1-\rho}$$
   se obtiene:
   $$1-\rho = \frac{\beta_1^*}{\beta_1}; \quad \rho = 1 - \frac{\beta_1^*}{\beta_1}; \quad \hat{\rho} = 1 - \frac{\hat{\beta}_1^*}{\hat{\beta}_1}.$$

2. **Usando MCO en el modelo $e_t = \rho e_{t-1} + \varepsilon_t$:**
   $$\hat{\rho} = \frac{\sum_{t=2}^{T} e_t e_{t-1}}{\sum_{t=2}^{T} e_{t-1}^2}$$

3. **Coeficiente de correlación lineal entre el modelo e_t y e_{t-1}:**
   $$\hat{\rho} = \frac{\sum_{t=2}^{T} e_t e_{t-1}}{\sqrt{\sum_{t=2}^{T} e_{t-1}^2}\sqrt{\sum_{t=1}^{T} e_t^2}}$$

4. **Despejando ρ del estadístico de Durbin-Watson: $DW \cong 2(1-\hat{\rho})$.**
   $$\hat{\rho} = 1 - \frac{DW}{2}$$

6.3.2. Estimación por etapas de ρ.

1. **Procedimiento en 2 etapas de Cochrane – Orcutt.**
   a) **Primera etapa.** Se estima ρ mediante un determinado método de estimación directa, obteniendo así el valor inicial $\hat{\rho}$.

No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.
Reservados todos los derechos.

WUOLAH

b) Segunda etapa. Usando $\hat{\rho}$ se obtiene la ecuación en primeras diferencias generalizada, es decir:

$$y_t - \hat{\rho} y_{t-1} = \beta_1(1-\hat{\rho}) + \beta_2(x_{2t} - \hat{\rho} x_{2t-1}) + \cdots + \beta_k(x_{kt} - \hat{\rho} x_{kt-1}) + (u_t - \hat{\rho} u_{t-1})$$

$$y_t^* = \beta_1^* + \beta_2 x_{2t}^* + \beta_3 x_{3t}^* + \cdots + \beta_k x_{kt}^* + \varepsilon_t, \quad t = 2, \dots, T.$$

Y este modelo se estima mediante MCO y se estima de nuevo ρ a partir de sus residuos.

2. Método iterativo de Cochrane – Orcutt.

Consiste en repetir el procedimiento en 2 etapas de Cochrane – Orcutt hasta que la diferencia (en valor absoluto) de dos estimaciones consecutivas de ρ es menor que un valor *d* prefijado. En general, se considera $d = 0.001$.

3. Método de Hildreth – Lu.

a) Primera etapa. Se consideran todos los posibles valores de ρ con 1 décima de diferencia, es decir, la parrilla inicial de valores son:

$$\hat{\rho}_j = \{-0.9, -0.8, -0.7, \dots, 0.7, 0.8, 0.9\}$$

b) Segunda etapa. Usando cada $\hat{\rho}_j$ se obtiene la ecuación en primeras diferencias generalizada, es decir:

$$y_t - \hat{\rho}_j y_{t-1} = \beta_1(1-\hat{\rho}_j) + \beta_2(x_{2t} - \hat{\rho}_j x_{2t-1}) + \cdots + \beta_k(x_{kt} - \hat{\rho}_j x_{kt-1}) + (u_t - \hat{\rho}_j u_{t-1})$$

$$y_{jt}^* = \beta_1^* + \beta_2 x_{2jt}^* + \beta_3 x_{3jt}^* + \cdots + \beta_k x_{kjt}^* + \varepsilon_{jt}, \quad t = 2, \dots, T.$$

Y estos modelos se estiman mediante MCO y calcula la correspondiente SCR:

$$SCR_j$$

c) Tercera etapa. Se elige el valor de $\hat{\rho}_j$ correspondiente al modelo con menor SCR_j. Se vuelve a la primera etapa pero con 1 centésima de diferencia y con valores entre $\hat{\rho}_j - 0.09$ y $\hat{\rho}_j + 0.09$. Por ejemplo, si $\hat{\rho}_j = 0.7$, la nueva parrilla de valores son:

$$\hat{\rho}_j = \{0.61, 0.62, 0.63, \dots, 0.77, 0.78, 0.79\}$$

Se repite el proceso hasta que la SCR de una etapa y la siguiente son prácticamente iguales.

WUOLAH

Descarga la app de Wuolah desde tu store favorita

No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad. Reservados todos los derechos.

Ejercicio 6.1.

Se han tomado datos anuales, durante el periodo 2003-2017 de las variables Renta bruta anual (variable x_2, medida en miles de euros), Consumo anual en alimentación (variable y, medida en miles de euros), e Impuestos anuales (variable x_3, medida en miles de euros) de una misma familia. Una vez estimado el modelo, los residuos obtenidos son:

Año	e_t
2003	-1.67
2004	0.39
2005	1.29
2006	-2.33
2007	-0.70
2008	0.17
2009	0.38
2010	-0.56
2011	0.08
2012	0.24
2013	0.01
2014	-0.89
2015	1.72
2016	1.15
2017	0.72

a) ¿Hay indicios de autocorrelación a partir del contraste de Durbin-Watson?

Sol.: DW= 1.867; DW (aproximación) = 2.066; dl = 0.976; du = 1.543.

b) ¿Hay indicios de autocorrelación a partir del contraste de Ljung-Box? Considere m = 4.

Sol.: $\hat{\rho}_1 = -0.0328$; $\hat{\rho}_2 = -0.2885$; $\hat{\rho}_3 = 0.1893$; $\hat{\rho}_4 = 0.2363$; $Q = 3.71$; $\chi^2_{4;0.95} = 9.49$.

Ejercicio 6.2.

Se han tomado datos anuales, durante el periodo 2009-2018 de las variables Renta bruta anual (variable x_2, medida en miles de euros), Consumo anual en alimentación (variable y medida en miles de euros), e Impuestos anuales (variable x_3, medida en miles de euros) de una misma familia. Una vez estimado el modelo, los residuos obtenidos son:

2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
-2.0	0.5	-1.0	1.0	-0.5	1.0	0.0	1.0	-2.0	2

¿Hay indicios de autocorrelación a partir del contraste de Durbin-Watson?

Sol.: DW = 2.6667; DW (aproximación) = 3.1515; dl = 0.697; du = 1.641.

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

WUOLAH

Descarga la app de Wuolah desde tu store favorita

Ejercicio 6.3.

Se han tomado datos anuales, durante el periodo 2009-2018 de las variables Renta bruta anual (variable x_2, medida en miles de euros), Consumo anual en alimentación (variable y medida en miles de euros), e Impuestos anuales (variable x_3, medida en miles de euros) de una misma familia. Una vez estimado el modelo, los residuos obtenidos son:

2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
-2.0	-1.0	-2.0	-2.0	-1.0	-2.0	1.0	-1.0	-2.0	-1.0

a) ¿Hay indicios de autocorrelación a partir del contraste de Durbin-Watson?

Sol.: DW = 0.76; DW (aproximación) = 0.96; dl = 0.697; du = 1.641.

b) ¿Hay indicios de autocorrelación a partir del contraste de Ljung-Box? Use *m = 4.*

Sol.: $\hat{\rho}_1 = 0.52$; $\hat{\rho}_2 = 0.48$; $\hat{\rho}_3 = 0.44$; $\hat{\rho}_4 = 0.32$; $Q = 12.43$; $\chi^2_{4;0.95} = 9.49$.

Ejercicio 6.4.

Para el periodo 2010-2018 se estima el siguiente modelo por MCO:

$$\hat{y}_t = \underset{(1.5)}{4} - \underset{(0.15)}{0.6}x_{2t} + \underset{(0.30)}{0.9}x_{3t} - \underset{(0.1)}{1.1}y_{t-1},$$

$$R^2 = 0.9$$

Además, se obtienen los siguientes residuos:

2010	2011	2012	2013	2014	2015	2016	2017	2018
-1.0	0.0	2.0	-1.0	2.0	-2.0	1.0	-1.0	1.0

Suponiendo que las perturbaciones aleatorias siguen un esquema *AR(1).*

a) Contrastar, con un 5% de nivel de significación, la existencia de autorrelación en el modelo.

Sol.: $\hat{\rho} = -0.7$; $h = -2.2$; $Z_{0.975} = 1.96$

b) A partir de los resultados obtenidos, ¿es razonable contrastar si existe autocorrelación negativa? Formule y resuelva el correspondiente contraste de hipótesis.

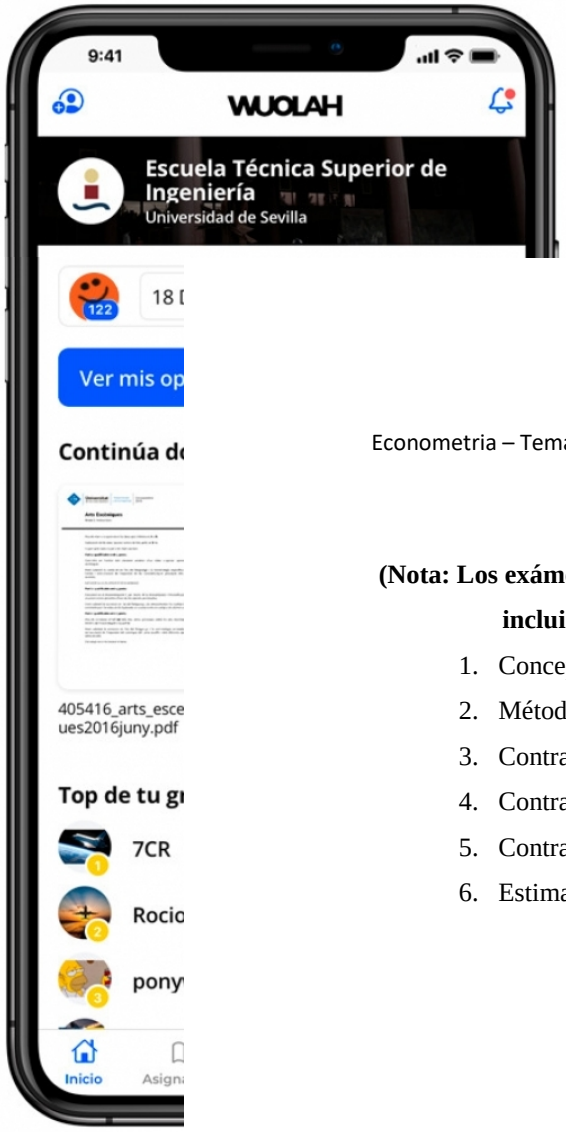
Sol.: $\hat{\rho} = -0.7$; $h = -2.2$; $-Z_{0.95} = -1.65$

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

WUOLAH

Descarga la app de Wuolah desde tu store favorita

Descarga la APP de Wuolah.
Ya disponible para el móvil y la tablet.
Available on the App Store
GET IT ON Google Play



Algunas preguntas de teoría del Tema 6

(Nota: Los exámenes de teoría de la asignatura pueden incluir cualquier pregunta incluida en el programa de la guía docente de la asignatura)

1. Concepto, causas y consecuencias de la autocorrelación.
2. Métodos gráficos para detectar la autocorrelación.
3. Contraste de Durbin-Watson.
4. Contraste *h* de Durbin.
5. Contraste de Ljung-Box.
6. Estimación de modelos con perturbaciones autocorreladas.

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

WUOLAH