Sarriko-On

Estadística Actuarial: Regresión Lineal

ISBN: 978-84-691-9178-1

Mª Victoria Esteban González

03-08

Facultad de Ciencias Económicas y Empresariales
Ekonomia eta Enpresal-Zientzien Fakultatea

eman ta zabal zazu
Universidad del País Vasco
Euskal Herriko Unibertsitatea

Estadística Actuarial: Regresión Lineal

Mª Victoria Esteban González

Departamento de Economía Aplicada III. Econometría y Estadística
Facultad de Ciencias Económicas y Empresariales
Universidad del País Vasco/Euskal Herriko Unibertsitatea

Preámbulo

Las notas que se desarrollan a continuación no tienen más ambición que servir como apoyo al proceso de aprendizaje de los estudiantes de la asignatura *Estadística Actuarial: Regresión* de la Licenciatura en Ciencias Actuariales y Financieras.

El contenido de la asignatura, de una dificultad media-alta esta estructurado en cuatro capítulos. El primero de ellos introduce el concepto de Econometría y define algunos de los términos más habituales. En los capítulos dos y tres se desarrolla el grueso del contenido. En el capítulo dos se especifica y estima el Modelo de Regresión Lineal General. Se desarrolla el estimador Mínimo Cuadrático Ordinario, sus propiedades y se muestra como hacer inferencia con él. Se revisa su comportamiento bajo mala especificación del modelo y las consecuencias de disponer de una muestra de variables altamente correlacionadas. En el capítulo tres se muestra como utilizar variables ficticias. En el cuarto y último capítulo se introduce al alumno en las técnicas de validación del modelo. Al final de las notas aparece la bibliografía completa.

Como decía anteriormente, estas notas sirven de apoyo al estudio y sobre todo permiten agilidad en las clases presenciales, en ningún caso deben utilizarse como sustituto de los libros incluidos en la bibliografía. De igual manera recomiendo la realización de ejercicios tanto los recomendados en clase como los que aparecen en la bibliografía. La unión del estudio de los conceptos y la utilización de los mismos en los ejercicios permite adquirir la agilidad necesaria para el dominio de la asignatura.

Contenido

Tema 1

Introducción

1.1. Modelo económico y modelo econométrico

Definición: Econometría en sentido estricto significa medida de la economía. La Econometría se ocupa de formular, cuantificar y valorar las relaciones entre variables económicas, para ello necesita de otras materias como son la Teoría Económica, la Estadística y las Matemáticas.

Definición: La Econometría se ocupa del estudio de estructuras que permitan analizar características o propiedades de una variable económica utilizando como causas explicativas otras variables económicas. (Novales, 1993)

Como es sabido la Teoría Económica se ocupa del análisis de la economía, como consecuencia del mismo formula las relaciones existentes entre las variables económicas objeto de estudio. Sin embargo la teoría Económica no se ocupa de cuantificarlas, éste es un cometido específico de la Econometría, que sí tiene como objetivo cuantificar las relaciones entre variables. Unido a este objetivo aparece un pilar clave para la Econometría que es la disponibilidad de información cuantificada sobre las variables que son objeto de estudio, en definitiva lo que llamamos **datos**. Las Matemáticas nos servirán para escribir en términos de ecuaciones las teorías económicas objeto de estudio y la Estadística nos proporciona instrumentos para el tratamiento de datos que nos permiten cuantificar las relaciones y valorar los resultados de acuerdo a criterios establecidos. En ocasiones nos encontraremos con problemas específicos para los que la estadística no tiene solución y por ello necesitaremos desarrollar los instrumentos y métodos apropiados para llevar a cabo los objetivos.

Resumiendo podríamos decir que los **objetivos de la Econometría** son: verificación de una teoría, estudio del pasado, descripción del presente, predicción del futuro y orientación de la acción política.

1.2. Etapas en la elaboración de un modelo econométrico

En la construcción de un modelo econométrico podemos distinguir según el ejemplo anterior, las siguientes etapas:

a) **Especificación:** En esta fase hay que dar forma al problema inicial en términos de un modelo. Determinar la variable a explicar y las variables explicativas, la forma funcional del modelo y la distribución probabilística de la perturbación aleatoria.

b) **Estimación:** Utilizando los datos disponibles y mediante métodos estadísticos adecuados se obtendrán valores para los parámetros desconocidos del modelo.

c) **Valoración del modelo o contraste de hipótesis:** En esta fase se debe valorar si el modelo propuesto y los resultados obtenidos en la fase de estimación son adecuados a solucionar el problema objetivo de partida. Para ello se utilizan métodos de inferencia estadística que permitirán rechazar o aceptar hipótesis de comportamiento sobre el modelo.

   El resultado de la validación del modelo puede ser que el modelo es adecuado a nuestros propósitos y nos sirve para tomar decisiones. En caso contrario, es decir que el modelo no es útil, habrá que repasar todas las fases y corregir los errores que pudiéramos haber llevado a cabo, lo que puede incluir una reformulación inicial. Es decir, este proceso de en principio tres fases puede ser en realidad un proceso iterativo.

d) **Predicción:**
   Finalmente un modelo correctamente especificado y estimado ha de ser utilizado para predecir. Este concepto implica tanto determinar los valores futuros de la variable endógena como contestar a preguntas del tipo ¨qué pasaría sí...?, en definitiva debe servirnos para dar consejos de política económica.

1.3. Tipología de datos y variables en Econometría

El modelo econométrico genérico completamente especificado tiene la siguiente forma:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \ldots + \beta_K X_{Kt} + u_t \qquad t = 1, 2, \ldots, T \tag{1.1}$$

Donde Y es la variable a explicar o variable endógena, X_2, X_3, …, X_K son las variables explicativas, o regresores, del modelo, el subíndice que las acompaña indica el número de variables explicativas del modelo, el modelo anterior tiene K-variables explicativas. Los coeficientes $\beta_k \quad k = 1, 2, \ldots, K$ son los parámetros a estimar, que se suponen constantes. Además es de interés notar que el parámetro β_1 acompaña a la variable explicativa X_1 constante e igual a la unidad en todo momento del tiempo. El subíndice t hace referencia al tiempo y por tanto T indica el tamaño de la muestra de observaciones disponible.

La diferencia entre un modelo económico y un modelo econométrico es la perturbación aleatoria que incluimos en el modelo econométrico. A partir de este elemento en el modelo econométrico podemos distinguir dos partes **la parte sistemática** del modelo y **la parte aleatoria**. La primera corresponde al comportamiento medio o estable de la relación y la segunda se corresponde con la perturbación aleatoria, u_t.

El objetivo sobre el modelo genérico representado por la ecuación (1.1) es conocer los valores de los parámetros desconocidos $\beta_k \quad k = 1, 2, \ldots, K$. Para llevar a cabo este objetivo utilizaremos métodos estadísticos. Para ello al modelo especificado deberemos de añadir hipótesis sobre el comportamiento probabilístico de la perturbación aleatoria que caractericen su distribución. En general, supondremos que dicha perturbación tiene una distribución centrada en cero, o sea media cero, lo que implica que el comportamiento medio de la variable a explicar está recogido en su totalidad por la parte sistemática del modelo:

$$E(Y_t) = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \ldots + \beta_K X_{Kt} \qquad t = 1, 2, \ldots, T \tag{1.2}$$

Además de la media debemos caracterizar también la varianza, covarianzas y distribución de la perturbación.

1.3.1. Conceptos básicos

En los puntos anteriores han surgido algunos conceptos que deberían quedar claros para poder referirnos a ellos con propiedad. Revisaremos algunos de ellos.

- **Población y muestra:**
  Población son todos los posibles valores que toma la variable objeto de estudio. Muestra sería la parte de la población que vamos a utilizar en el estudio para extraer conclusiones. Por tanto la muestra está contenida en la población y nosotros la utilizaremos para establecer conclusiones que puedan extrapolarse a la población.

- **Datos:**
  Los datos son los valores numéricos que toman tanto la variable a explicar como las variables explicativas. Generalmente los obtenemos de series estadísticas cuyas fuentes pueden ser oficiales o privadas. La importancia de los datos está determinada por la unidad de medida. Los podemos clasificar en:

  a) Datos de serie temporal: Reflejan la evolución de una variable a lo largo del tiempo, según esto la variable estará ordenada cronológicamente con un orden lógico. Las variables medidas en series temporales se denotan con el subíndice t y este puede referirse a observaciones temporales mensuales, trimestrales, diarias cuatrimestrales, anuales, etc. Ejemplo: el Producto Nacional Bruto (PNB) de 1965-2000. En este caso la población serían todos los posibles valores del PNB a lo largo del tiempo y la muestra el período que vamos a estudiar, de 1965 al 2000.

  b) Datos de sección cruzada o corte transversal: Son datos atemporales dado que miden el comportamiento de una variable en diferentes unidades y en el mismo momento del tiempo. Ejemplo: ventas de las empresas metalúrgicas en el País Vasco en el año 1999. Esta sería la muestra a utilizar y la población estaría constituida por todas las unidades.

  c) Datos de panel: es la unión de datos de serie temporal y datos de sección cruzada. Están fuera del objetivo del curso de Introducción a la Econometría y también del de Econometría.

- **Variables:**
  Una variable es un ente económico que toma diferentes valores. Podemos distinguir entre variables exógenas, aquellas que inciden en el modelo desde el exterior y variables endógenas, aquellas que queremos explicar con el modelo. A las variables exógenas también se las denomina variables explicativas o independientes y a la variable endógena también se le puede denominar como variable a explicar o dependiente. Además debemos tener en cuenta que podemos encontrarnos con relaciones simultáneas como:

  $$Y_t = \alpha + \beta Y_{t-1} + u_t$$

  o como

  $$C_t = a + bY_t + u_t \qquad\qquad Y_t = C_t + I_t$$

  donde las variables cambian su papel según miremos a una ecuación u otra. Podemos distinguir los siguientes tipos de variables:

  a) - Fijas: aquellas que toman valores que el investigador puede controlar.

     - Estocásticas: aquellas cuyo valor cambia según una ley de probabilidad.

  b) - Cuantitativas: aquellas que podemos valorar numéricamente.

- Cualitativas : aquellas que miden cualidades y que por lo tanto no se miden con un valor numérico y será el investigador el que se lo asigne según un criterio.

- **Los parámetros:**
  Los parámetros son los valores que permanecen desconocidos del modelo. En un modelo econométrico podemos distinguir dos tipos de parámetros:

  a) Los parámetros de la relación económica: Son las ponderaciones que aplicadas a las variables exógenas nos permiten calcular la endógena. En el modelo siguiente son a y b:

  $$C_t = a + bY_t + u_t \tag{1.3}$$

  b) Los parámetros de la estructura probabilística: son los parámetros que determinan la estructura de la parte aleatoria del modelo, media y varianza de la perturbación aleatoria y de la variable endógena.

- **Modelo:**
  Hemos visto que un modelo no es más que un conjunto de relaciones entre variables económicas y que representamos mediante relaciones matemáticas. Clasificación de los modelos:

  a) - Modelos exactos: aquellos que determinan exactamente el valor de una variable conocido el valor de otra-s.:

  $$Y = a + bX$$

  - Modelos estocásticos: aquellos que incluyen alguna variable aleatoria:

  $$Y_t = a + bX_t + u_t \qquad u \sim (m(u), var(u))$$

  b) - Modelos uniecuacionales: aquellos que se componen de una única ecuación:

  $$C_t = a + bY_t + u_t$$

  - Modelos multiecuacionales: aquellos que se componen de más de una ecuación. Por ejemplo cuando una variable influye en otra-s y a la vez es influida por éstas:

  $$C_t = a + bY_t + u_t \qquad\qquad Y_t = C_t + I_t$$

  c) - Modelos estáticos: Cuando el tiempo no aparece de forma explícita en la ecuación y todas las variables se miden en el mismo momento.

  - Modelos dinámicos: Aquellos que tienen variables definidas en diferentes momentos del tiempo o el tiempo aparece como variable explícita en la ecuación. Un ejemplo de los primeros sería:

  $$C_t = a + bY_t + cC_{t-1} + u_t$$

  mientras que un ejemplo de los segundos sería el siguiente modelo no explícitamente dinámico, generalmente llamado estático histórico

  $$C_t = a + bY_t + ct + u_t$$

  donde el parámetros c recoge la tendencia de la variable endógena a lo largo del tiempo.

  d) - Modelos basados en series temporales: pueden ser dinámicos u estáticos.

- Modelos basado en datos de corte transversal: son siempre estáticos.

- **Parámetro, estimador y estimación:**
En el modelo:
$$Y_t = \alpha + \beta X_t + u_t \qquad t = 1, 2, \ldots, T$$
tenemos diferentes parámetros desconocidos. En la parte aleatoria aparecerían los que caracterizan a la distribución probabilística de la perturbación aleatoria y en la parte sistemática aparecen α y β. Todos son parámetros desconocidos. Los llamaremos parámetros poblacionales ya que lo que nosotros hemos especificado es un modelo general que debería recoger el comportamiento medio de las variables en la población. Para obtener resultados del modelo anterior nosotros lo aplicamos a la muestra, de tamaño T. Nuestro objetivo es determinar el valor de estos parámetros poblacionales desconocidos de la muestra. Para aproximarnos a ese valor utilizamos técnicas estadísticas, en concreto estimadores. Un estimador no es más que una fórmula que nos dice como debemos obtener los valores numéricos de α y β mediante la muestra. Al valor finalmente obtenido en la muestra le llamamos estimación. En concreto la notación matemática para estos conceptos, aplicada al parámetro β sería:

β	parámetro poblacional
$\hat{\beta}$	estimador
0,5	estimación

donde por ejemplo:
$$\hat{\beta} = \frac{\sum_{t=1}^{T}(Y_t - \bar{Y})(X_t - \bar{X})}{\sum_{t=1}^{T}(X_t - \bar{X})^2} = 0,5$$
Los estimadores van a ser variables aleatorias con distribución a determinar ya los que exigiremos ciertas propiedades que van a determinar esta distribución.

- **Estructura:**
Cuando estudiamos la relación entre las variables económicas especificamos un modelo econométrico. En la especificación elegimos la forma funcional del modelo y las variables explicativas a incluir así como las propiedades de la perturbación. Una vez que el modelo está totalmente especificado le estimaremos y tendremos unos valores para los parámetros. A la relación resultante le llamamos estructura. Un modelo especificado sería:
$$Y_t = \alpha + \beta X_t + u_t \qquad t = 1, 2, \ldots, T$$
mientras que una estructura para ese modelo dada una muestra de tamaño T podría ser:
$$\hat{Y}_t = 20 + 5X_t$$
Notar que un modelo puede tener diferentes estructuras según los valores que las variables exógena y endógena tomen en la muestra.

1.4. Tratamiento de la información con Gretl: inclusión de datos en Gretl y análisis descriptivo básico

Gretl es un programa que permite obtener de manera sencilla mediante ventana resultados estadísticos y econométricos. Empezaremos viendo como leer datos y obtener sus principales estadísticos. Una vez ejecutado el programa Gretl en la ventana principal aparece un menú de ventanas que nos permite diferentes posibilidades. En primer lugar necesitaremos leer los datos con los que trabajar. Dependiendo del origen de éstos si están en una archivo de muestra incluido en Gretl, si están disponibles en papel, en la web o en un archivo propio procederemos de una manera u otra.

• Para leer **datos incluidos en la base del programa Gretl**:
Pinchar *Archivo* → *Abrir datos* → *Archivo de muestra* → Aquí seleccionamos la base de datos que necesitemos, por ejemplo *ETM* → y ahora seleccionamos el archivo, por ejemplo *monthly-crsp.gdt*

Aparecerán las variables de la muestra y en la barra superior diferentes etiquetas, por ejemplo en Datos podremos ver las observaciones y sus características. Algunas de las opciones que contiene la etiqueta Datos son las siguientes:

Mostrar valores
Editar valores
Leer información
Ver descripción
Estructura del conjunto de datos

Para obtener lo que necesitamos sólo tenemos que pinchar la etiqueta correspondiente y la variable o variables a estudiar. Por ejemplo para ver la estructura del conjunto de datos pinchamos en esta etiqueta y obtendremos una pantalla en la que aparecerá seleccionado el tipo de datos con el que estamos trabajando, en este caso *Serie temporal.* Pinchamos aceptar y veremos la frecuencia, *mensual*, y el inicio y final de la muestra *1968:1 a 1998:12.* La etiqueta estructura del conjunto de datos es muy útil cuando necesitamos cambiar alguno de ellos por ejemplo si añadimos nuevas observaciones.

La misma información contenida en la estructura del conjunto de datos podemos encontrarla en la etiqueta: Ver descripción, que describe el conjunto de datos junto con cada una de las variables que lo componen.

En el menú inicial aparece también la etiqueta Ver en la cual podemos obtener gráficos de las variables y sus estadísticos principales entre otros.

• Para hacer **Gráficos**:
Por ejemplo para hacer el gráfico de serie temporal de la serie CRSP.

Pinchar *Ver* → *Gráficos* → *Gráficos de series temporales.*

Para guardar el gráfico: situar el ratón sobre el gráfico y pinchar con el botón derecho. Elegir opción. Podemos guardarlos en postcript (.eps) o .png, etc. En la ventana que aparece para guardarlo escribir *a*:\ y poner un nombre por ejemplo CRSPVW.

• Para obtener los **Estadísticos principales** de las variables de la muestra:
Pinchar en *Ver* → *Estadísticos principales.*

La ventana de output mostrará la media, moda, valor máximo y mínimo de la serie, desviación típica, coeficiente de variación, curtosis y asimetría. Podemos obtener los estadísticos para una única serie o para el conjunto de ellas seleccionándolo previamente.

Si queremos guardar el output pinchamos en el icono del diskette arriba a la izquierda y seleccionamos cómo queremos que lo guarde, texto plano, Word o Latex y en la ventana damos el nombre que deseemos al fichero de resultados, por ejemplo *estadVW* para la serie CRSP o *estadmuestra* para el conjunto.

En el menú inicial también aparece la etiqueta Variable para trabajar con una única serie de la muestra. Algunas de las opciones que incluye esta etiqueta son:

Buscar
Mostrar valores
Estadísticos principales
Distribución de frecuencias
Gráfico de frecuencias (simple, contra la normal, contra la gamma)
Gráfico de series temporales
Editar atributos
etc

• Obtener **datos que están en el servidor**:
Queremos estudiar una serie que se encuentra en el servidor, *Crédito más de 5 años a hogares.* Esta serie aparece publicada en la base de datos del Banco de España con el código BE182704.

Pinchar *Archivo* → *Abrir datos* → *Bases de datos* → *sobre servidor*

En el listado de bases de datos que aparece vamos a
bde18 Banco de España (Tipo de interés)
y pinchamos en *Obtener listado de series* comprobando que contienen la serie que queremos y representarla gráficamente:

Series → *Mostrar*

Series → *Representar*

Para importar los datos a Gretl situamos el cursor sobre la serie de interés, BE182704, y vamos a

Series → *Importar*

Además tenemos opción de hacer lo siguiente:

- Añadir o cambiar información sobre la variable: en menú *Variable* → *Editar atributos.* En esta ventana podremos cambiar también el nombre de la serie utilizado en los gráficos.
- Añadir notas explicativas: en menú *Datos* → *Editar información*
- Consultar las notas informativas: en menú *Datos* → *Leer información* o en *Datos* → *Descripción*

• Para **crear un conjunto de datos:**

Pinchar *Archivo* → *Nuevo conjunto de datos* y completar la información que pide sobre: *número de observaciones*
estructura del conjunto de datos (serie temporal o sección cruzada
frecuencia
Observación inicial

A la pregunta *¿Desea empezar a introducir los valores de los datos usando la hoja de cálculo de Gretl?* contestar *Sí*

- Introducir el nombre de la variable. El máximo de caracteres que acepta es 15, no usar acentos ni la letra *ñ*. Pinchar *Aceptar*

- En la hoja de cálculo situarnos en la primera celda y teclear la observación correspondiente, a continuación pintar *intro*. Si nos saltamos alguna observación podemos insertar una fila en el lugar correspondiente con solo situarnos en la celda posterior e ir a *observación* $\rightarrow$ *insertar obs.* Una vez introducidas todas las variables pinchar *Aplicar*.

- Para guardar los datos: en menú *Archivo* $\rightarrow$ *Guardar datos*. Dar nombre al conjunto de datos, por ejemplo *Azar* y se grabará automáticamente con la extensión *gdt*.

  Si en otro momento queremos usar este conjunto de datos solo habrá que clickear en él dos veces para que se active.

- Si queremos añadir variables en menú: *Añadir* $\rightarrow$ tenemos las siguientes posibilidades:

  - Logaritmos de las variables seleccionadas
  - Cuadrados de las variables seleccionadas
  - Retardos de las variables seleccionadas
  - Primeras diferencias de las variables seleccionadas
  - Diferencias del logaritmo las variables seleccionadas
  - Diferencias estacionales de las variables seleccionadas
  - Variable índice
  - Tendencia temporal
  - Variable aleatoria (uniforme, normal, chi cuadrado y t-Student) Por ejemplo para crear una variable normal de media 0 y desviación 1 haremos *nombre de la variable 0 1*
  - Variables ficticias, etc.
  - Definir una nueva variable. Esta opción podemos utilizarla para crear combinaciones de variables por ejemplo $Z_t = 4 + \epsilon_t \qquad \epsilon_t \sim N(0,1)$. Permite los operadores,

    `+, -, *, /, ^`

    (suma, resta, producto, potencia) entre otros.

Tema 2

Modelo de Regresión Lineal General

2.1. Especificación del Modelo de Regresión Lineal General (MRLG): supuestos básicos

Vamos a estudiar la relación existente entre una variable Y, y un conjunto de K variables, $X_1, X_2, \ldots, X_K$, todas ellas cuantitativas, mediante la especificación de un modelo lineal.

Supondremos, por tanto, el siguiente modelo lineal

$$Y_t = \beta_1 X_{1t} + \beta_2 X_{2t} + \ldots + \beta_K X_{Kt} + u_t \quad t = 1, 2, \ldots, T$$

donde estamos considerando K-variables explicativas pero habitualmente $X_{1t} = 1 \; \forall t$, de forma que β_1 es un término independiente y el Modelo de Regresión Lineal General (MRLG) queda especificado como,

$$Y_t = \beta_1 + \beta_2 X_{2t} + \ldots + \beta_K X_{Kt} + u_t \qquad t = 1, 2, \ldots, T.$$

Nuestro objetivo será estimar lo mejor posible esta función de regresión poblacional, para ello estimaremos los parámetros desconocidos $\beta_1, \beta_2, \ldots, \beta_K$ a partir de una muestra dada. La muestra disponible se compone de T-datos u observaciones de cada una de las variables $Y, X_1, X_2, \ldots, X_K$. Por tanto, la muestra disponible es de tamaño T, y T es el número de observaciones disponibles sobre cada una de las variables. Los elementos del modelo y su notación son:

$$\begin{aligned} Y_t &= \text{observación } t\text{-ésima de } Y \\ X_{kt} &= \text{observación } t\text{-ésima de } X_k \quad \forall k = 1, \ldots, K \end{aligned}$$

donde X_{kt} es una observación de las disponibles en la muestra $t = 1, 2, \ldots, T$.

Elementos del MRLG

- Y es la variable a explicar o variable endógena o regresando.
- $X_k \;\; k = 1, \ldots, K$ son las K variables explicativas o variables exógenas o regresores.
- $\beta_k \;\; k = 1, \ldots, K$ son los coeficientes o parámetros (desconocidos).

- u es la perturbación aleatoria.

La perturbación aleatoria u_t es una variable aleatoria no observable que pretende recoger:

- Variables no incluidas en el modelo.
- Comportamiento aleatorio de los agentes económicos.
- Errores de medida.

Notar que dado que u_t es una variable aleatoria también lo es Y_t.

Representación del MRLG en forma matricial

En el modelo

$$Y_t = \beta_1 + \beta_2 X_{2t} + \ldots + \beta_K X_{Kt} + u_t \qquad t = 1, 2, \ldots, T \tag{2.1}$$

el tiempo es un elemento importante, dado que el tiempo varía de $t = 1, 2, \ldots, T$ el modelo puede escribirse para todas las observaciones disponibles como el siguiente sistema de T ecuaciones:

$$\left\{\begin{array}{lll}
Y_1 = \beta_1 + \beta_2 X_{21} + \beta_3 X_{31} + \ldots + \beta_K X_{K1} + u_1 & & t = 1 \\
Y_2 = \beta_1 + \beta_2 X_{22} + \beta_3 X_{32} + \ldots + \beta_K X_{K2} + u_2 & & t = 2 \\
\vdots \quad \vdots \qquad \vdots \qquad \vdots \qquad \vdots & & \\
Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \ldots + \beta_K X_{Kt} + u_t & & t = t \\
\vdots \quad \vdots \qquad \vdots \qquad \vdots & & \\
Y_T = \beta_1 + \beta_2 X_{2T} + \beta_3 X_{3T} + \ldots + \beta_K X_{KT} + u_T & & t = T
\end{array}\right.$$

En forma matricial se escribe:

$$\underset{(T\times 1)}{Y} = \underset{(T\times K)}{X} \; \underset{(K\times 1)}{\beta} + \underset{(T\times 1)}{u}$$

donde:
$Y_{(T\times 1)}$: Contiene las observaciones de la variable endógena.
$X_{(T\times K)}$: Contiene las observaciones de las variables exógenas.
$\beta_{(K\times 1)}$: Contiene los parámetros desconocidos o coeficientes del modelo.
$u_{(T\times 1)}$: contiene las perturbaciones aleatorias.

$$\underset{(T\times 1)}{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_t \\ \vdots \\ Y_T \end{bmatrix} \qquad \underset{(T\times K)}{X} = \begin{bmatrix} 1 & X_{21} & X_{31} & \cdots & X_{K1} \\ 1 & X_{22} & X_{32} & \cdots & X_{K2} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_{2t} & X_{3t} & \cdots & X_{Kt} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_{2T} & X_{3T} & \cdots & X_{KT} \end{bmatrix}$$

$$\underset{(K\times 1)}{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \vdots \\ \beta_K \end{bmatrix} \qquad \underset{(T\times 1)}{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_t \\ \vdots \\ u_T \end{bmatrix}$$

2.1.1. Hipótesis básicas.

a) Hipótesis sobre la perturbación aleatoria

- La perturbación u_t tiene media cero para todo t, $E(u_t) = 0 \quad \forall t$. La perturbación mide las diferencias con respecto a un promedio, $u_t = Y_t - E(Y_t)$ y a priori no tenemos razones para suponer que todas las desviaciones están por encima o por debajo de ese promedio, por ello parece lógico pensar que en media las desviaciones son cero. Para la perturbación en t lo escribimos como $E(u_t) = 0 \quad \forall t$, cuando miramos al modelo en forma matricial escribimos esta hipótesis como $E(u) = \vec{0}$:

$$E(u) = E \begin{bmatrix} u_1 \\ u_2 \\ \dots \\ u_T \end{bmatrix} = \begin{bmatrix} E(u_1) \\ E(u_2) \\ \dots \\ E(u_T) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \dots \\ 0 \end{bmatrix} = \vec{0}$$

- $E(u_t^2) = \sigma_u^2 \quad \forall t$ es decir la varianza de la perturbación es desconocida e igual a σ^2 en todo momento del tiempo. Estamos suponiendo igual dispersión o variabilidad. A esta hipótesis se le conoce con el nombre de Homocedasticidad:

$$Var(u_t) = E(u_t - E(u_t))^2 = E(u_t^2) = \sigma^2$$

$$E(u_1^2) = E(u_2^2) = E(u_3^2) = \ldots = E(u_T^2) = \sigma^2$$

El caso contrario, cuando la dispersión varía a lo largo de la muestra se denomina heterocedasticidad ($E(u_t^2) = \sigma_t^2$). El Gráfico 2.1 ilustra ambas situaciones:

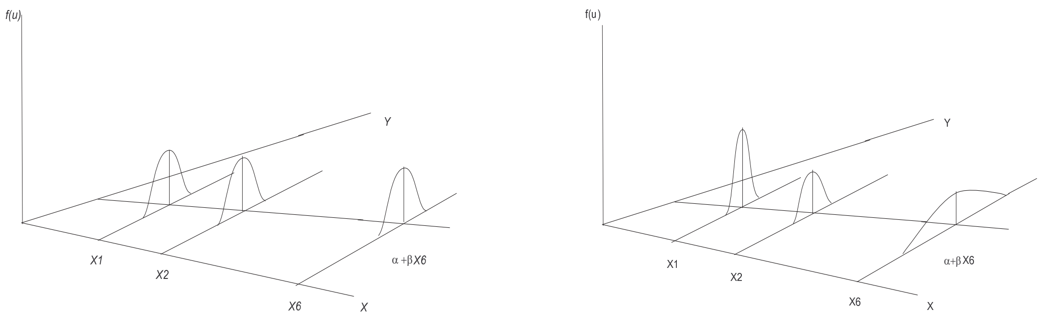


Gráfico 2.1: Perturbaciones homocedásticas versus heterocedásticas

Hay que notar que generalmente σ^2 será desconocida y por tanto en el modelo tendremos que estimar $(k + 1)$ incógnitas, los k-coeficientes poblacionales desconocidos más la varianza poblacional de la perturbación σ^2.

- $E(u_t u_s) = 0 \quad \forall t, s \quad t \neq s$. La covarianza entre perturbaciones de distinto momento del tiempo es cero.

$$Cov(u_t, u_s) = E(u_t - E(u_t))(u_s - E(u_s)) = E(u_t u_s) = 0 \quad \forall t, s \quad t \neq s$$

$$E(u_1 u_2) = E(u_1 u_3) = E(u_4 u_{20}) = 0$$

A esta hipótesis también se la llama hipótesis de No Autocorrelación.

Definimos la matriz de varianzas y covarianzas de la perturbación como:

$$E(uu') = \sigma^2 I_T$$

$$E(uu') = E \begin{bmatrix} u_1 \\ u_2 \\ \dots \\ u_T \end{bmatrix} \begin{bmatrix} u_1' & u_2' & \dots & u_T' \end{bmatrix} = E \begin{bmatrix} u_1u_1' & u_1u_2' & \dots & u_1u_T' \\ u_2u_1' & u_2u_2' & \dots & u_2u_T' \\ \dots & \dots & \dots & \dots \\ u_tu_1' & u_tu_2' & \dots & u_Tu_T' \end{bmatrix} =$$

$$= E \begin{bmatrix} u_1^2 & u_1u_2' & \dots & u_1u_T' \\ u_2u_1' & u_2^2 & \dots & u_2u_T' \\ \dots & \dots & \dots & \dots \\ u_tu_1' & u_tu_2' & \dots & u_T^2 \end{bmatrix} = \begin{bmatrix} E(u_1^2) & E(u_1u_2') & \dots & E(u_1u_T') \\ E(u_2u_1') & E(u_2^2) & \dots & E(u_2u_T') \\ \dots & \dots & \dots & \dots \\ E(u_tu_1') & E(u_tu_2') & \dots & E(u_T^2) \end{bmatrix} =$$

$$= \begin{bmatrix} \sigma^2 & 0 & 0 & \dots & 0 \\ 0 & \sigma^2 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \sigma^2 \end{bmatrix} = \sigma^2 \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix} = \sigma^2 I_T$$

En la matriz de varianzas y covarianzas de la perturbación así definida se incluyen las dos hipótesis, homocedasticidad y la hipótesis de no autocorrelación.

A la hipótesis que reconoce que las varianzas de la perturbación no son constantes en el tiempo (o los individuos) se le conoce como hipótesis de Heterocedasticidad. A la hipótesis que reconoce que las covarianzas entre perturbaciones de distinto momento del tiempo, o entre distintos individuos, son distintas de cero se le conoce con el nombre de Autocorrelación.

- Las perturbaciones siguen una distribución normal. Si definimos la perturbación aleatoria como la suma de errores independientes entre sí, a través del Teorema Central del Límite podremos suponer una distribución normal y escribir esta hipótesis junto con las anteriores como:

$$u \sim NID(0, \sigma^2 I_T)$$

donde decimos que las perturbaciones siguen una distribución normal, idéntica e independientemente distribuidas, de media cero y varianza constante igual a σ^2. Son independientes dado que su covarianza es cero y dado que todas tienen igual varianza y covarianza su distribución es idéntica, por ello para una perturbación en t escribimos su distribución como $u_t \sim N(0, \sigma^2)$.

Estas propiedades pueden también escribirse conjuntamente como

$$u_t \sim NID(0, \sigma_u^2) \quad \forall t = 1, \dots, T$$

ó en forma matricial,

$$\underset{(T\times 1)}{u} \sim N(\underset{(T\times 1)}{0_T}, \underset{(T\times T)}{\sigma_u^2 I_T})$$

b) Hipótesis sobre las variables exógenas X.

- Las variables explicativas son variables no aleatorias (no estocásticas o fijas). Esto quiere decir que cuando digamos que cambiamos la muestra los valores de las variables exógenas no cambian y sólo cambian los valores de la variable endógena Y_t.
  Como consecuencia de que las variables exógenas sean fijas tendremos que son incorrelacionadas con las perturbaciones:

$$E(X'u) = X'E(u) = 0$$

- La matriz X es de rango completo e igual a K con $K < T, \quad rg(X) = K$, es decir no hay ninguna combinación lineal exacta entre las columnas de X, son todas linealmente independientes con lo que el rango de la matriz es igual al número de coeficientes desconocido ya que en X tenemos una columna por parámetro. A esta hipótesis se le conoce con el nombre de No Multicolinealidad. El que además exijamos que $K < T$ es porque necesitamos tener más observaciones que coeficientes a estimar en el modelo.

c) Hipótesis de carácter general:

- Los coeficientes permanecen constantes a lo largo de toda la muestra.
- Sobre la forma funcional suponemos:
  - Linealidad en los coeficientes.
  - El modelo está correctamente especificado. Esto quiere decir que todas las variables $X_1, X_2, \ldots, X_K$ explican Y y no hay ninguna otra de fuera del modelo que explique a Y. Es decir no falta ninguna ni sobra ninguna.

Observación:

$$E(uu') = \sigma^2 I_T \qquad\qquad E(u'u) = E\left[[u_1 u_2 \ldots u_T] \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_T \end{bmatrix} \right]$$

2.2. Forma funcional. Interpretación de los coeficientes.

Dados los supuestos básicos del MRLG el valor esperado de la variable endógena lo encontramos como,

$$\begin{aligned} E(Y_t) &= E(\beta_1 + \beta_2 X_{2t} + \ldots + \beta_K X_{Kt} + u_t) \\ &= \beta_1 + \beta_2 X_{2t} + \ldots + \beta_K X_{Kt} + \underbrace{E(u_t)}_{=0} = \\ &= \beta_1 + \beta_2 X_{2t} + \ldots + \beta_K X_{Kt}. \end{aligned}$$

Llamamos **Función de Regresión Poblacional** (FRP) a $E(Y_t)$. Los coeficientes, $\beta_1, \beta_2, \ldots, \beta_K$ se interpretan del siguiente modo:

- $\beta_1 = E(Y_t|X_{2t} = \ldots = X_{Kt} = 0)$. Valor esperado de Y_t cuando las variables explicativas son todas cero.

- $\beta_k = \frac{\partial E(Y_t)}{\partial X_{kt}} = \frac{\Delta E(Y_t)}{\Delta X_{kt}} \quad \forall k = 2, \ldots, K$. Incremento (o decremento) en el valor esperado de Y_t cuando la variable explicativa X_k se incrementa en una unidad, manteniéndose constantes el resto de las variables.

Cuando decimos que el MRLG es un modelo lineal queremos decir que Y o alguna transformación de Y es lineal en las X o en alguna transformación lineal de las X. Hay dos tipos de linealidad, linealidad en variables y linealidad en parámetros. Nosotros estamos interesados en la linealidad en parámetros: es decir las derivadas con respecto a los coeficientes desconocidos son una función lineal sólo de las X.

$$\beta_i = \frac{\partial E(Y_t)}{\partial X_{it}} \qquad i = 2, 1, \ldots, K$$

El modelo lineal más sencillo es el Modelo de Regresión Lineal Simple donde la variable endógena Y queda explicada por una única variable exógena X

$$Y_t = \beta_1 + \beta_2 X_t + u_t \quad t = 1, 2, \ldots, T$$

De igual forma es lineal el Modelo de Regresión Lineal General donde la variable endógena Y se explica con un conjunto de k-variables explicativas $(X_{1t}, X_{2t}, \ldots, X_{Kt})$

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \ldots + \beta_k X_{kt} + u_t \quad t = 1, 2, \ldots, T$$

dado que estamos interesados sólo en la linealidad en parámetros también serán considerados lineales los siguientes modelos:

$$Y_t = \alpha + \beta \frac{1}{X_t} + u_t \qquad \text{ó} \qquad Y_t = \alpha + \beta X_t^2 + u_t$$

que son lineales en parámetros según lo dicho anteriormente aunque no lo sean en variables. Ahora bien, existen otras relaciones que aunque en principio no son lineales pueden transformarse en lineales y por tanto son perfectamente estimables en nuestros términos. Por ejemplo:

a) Sea el siguiente modelo:

$$X_t = AB^{Y_t} u_t$$

podemos transformar el modelo en lineal en parámetros tomado logaritmos y obtener:

$$Y_t = \alpha + \beta LnX_t + u_t \tag{2.2}$$

donde $\beta = (LnB)^{-1}$ y $\alpha = (\frac{LnA}{LnB})$ a esta transformación se le llama semilogarítmica.

b) Sea el modelo:

$$Y_t = AX_t^B u_t \longrightarrow LnY_t = \alpha + \beta LnX_t + u_t \tag{2.3}$$

donde $\alpha = LnA$, a esta transformación se le llama doblemente logarítmica.

c) Otro ejemplo es la función Cobb-Douglas:

$$Q_t = AL_t^{\beta_2} K_t^{\beta_3} u_t \longrightarrow LnQ_t = \beta_1 + \beta_2 LnL_t + \beta_3 LnK_t + u_t \tag{2.4}$$

siendo $\beta_1 = LnA$. Una ventaja de este tipo de modelos como el recogido en la ecuación (2.4), en los que **todas** las variables están medidas en logaritmos, es que los parámetros de pendiente además de recibir la interpretación habitual pueden interpretarse en términos de elasticidades:

$$\beta_2 = \frac{\partial E(LnQ_t)}{\partial LnL_t} = \frac{\partial E(Q_t)}{\partial L_t}\frac{L_t}{Q_t}$$

$$\beta_3 = \frac{\partial E(LnQ_t)}{\partial LnK_t} = \frac{\partial E(Q_t)}{\partial K_t}\frac{K_t}{Q_t}$$

Es decir $\beta_i \quad i = 2,3$, miden el cambio porcentual (o elasticidad) generado en la variable endógena como consecuencia de un cambio porcentual (un 1 por ciento) en la variable exógena correspondiente, ceteris paribus. En el ejemplo anterior β_2 y β_3 representan las elasticidades de la función de producción con respecto a los factores de producción trabajo y capital respectivamente.

Es importante notar que para la ecuación (2.2) está interpretación no es posible ya que:

$$\beta = \frac{\partial E(Y_t)}{\partial LnX_t} = \frac{\partial E(Y_t)}{\partial X_t}X_t$$

2.3. Estimación por Mínimos Cuadrados Ordinarios

• Nuestro objetivo es estimar los parámetros desconocidos $\beta_1, \beta_2, \ldots, \beta_K$ del modelo

$$Y_t = \beta_1 + \beta_2 X_{2t} + \ldots + \beta_K X_{Kt} + u_t \qquad t = 1,2,\ldots,T \Leftrightarrow Y = X\beta + u$$

Denotamos los parámetros estimados como $\hat{\beta}_k$ y el modelo estimado o **Función de Regresión Muestral** (FRM) se escribe:

$$\hat{Y}_t = \hat{\beta}_1 + \hat{\beta}_2 X_{2t} + \ldots + \hat{\beta}_K X_{Kt} \qquad t = 1,2,\ldots,T$$

$$\hat{Y} = X\hat{\beta} \qquad \text{en forma matricial,}$$

• Dada la Función de Regresión Poblacional las perturbaciones pueden expresarse así:

$$u_t = Y_t - E(Y_t) \qquad t = 1,2,\ldots,T$$

$$u = Y - X\beta \qquad \text{en forma matricial.}$$

• Dada la Función de Regresión Muestral definimos los residuos como:

$$\hat{u}_t = Y_t - \hat{Y}_t = Y_t - \hat{\beta}_1 - \hat{\beta}_2 X_{2t} - \ldots - \hat{\beta}_K X_{Kt}$$

$$t = 1,2,\ldots,T$$

$$\hat{u} = Y - \hat{Y} = Y - X\hat{\beta} \qquad \text{en forma matricial.}$$

Los residuos son a la FRM lo que las perturbaciones a la FRP pero no tienen las mismas propiedades.

• **Representación gráfica**: Cuando $K = 2$,

$$Y_t = \beta_1 + \beta_2 X_{2t} + u_t$$

el modelo suele escribirse

$$Y_t = \alpha + \beta X_t + u_t$$

y se denomina Modelo de Regresión Lineal Simple (MRLS). Entonces, la relación entre la FRM, FRP, residuos y perturbaciones puede visualizarse en el Gráfico 2.2

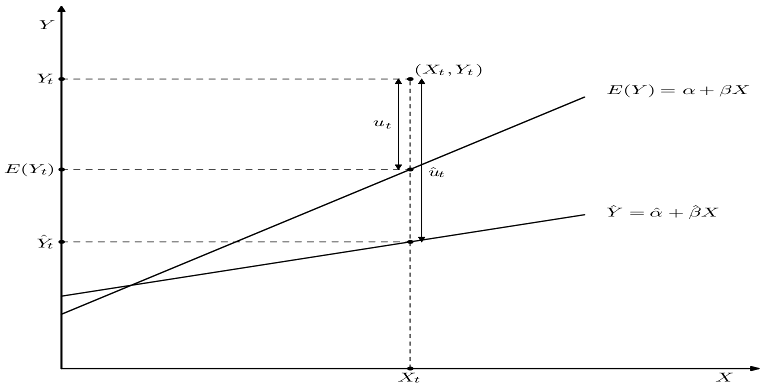


Gráfico 2.2: FRM, FRP, residuos y perturbaciones en un MRLS.

2.3.1. Método de estimación de Mínimos Cuadrados Ordinarios (MCO)

El objetivo es estimar los parámetros ó coeficientes $\beta_1, \ldots, \beta_K$, para el cual buscamos K estimadores respectivamente. A los parámetros estimados los denotamos como $\hat{\beta}_1, \ldots, \hat{\beta}_K$. En forma matricial el vector de parámetros estimados es $\hat{\beta}_{(K \times 1)}$ A ambos los denotamos $\hat{\beta}$.

El criterio de estimación que vamos a estudiar este curso es el criterio mínimo cuadrático ordinario (MCO) que consiste en minimizar el sumatorio de los residuos al cuadrado:

$$min_{\hat{\beta}} \textstyle\sum_{t=1}^{T} \hat{u}_t^2 = min_{\hat{\beta}} \sum_{t=1}^{T} (Y_t - \hat{Y}_t)^2$$

El resultado de aplicar el criterio MCO al modelo de regresión lineal general proporciona los K-estimadores, la aplicación de los estimadores a una muestra nos devuelve K-estimaciones, una para cada parámetro desconocido.

• **Estimador MCO del MRLG**

El problema es:

$$min_{\hat{\beta}_1,\ldots,\hat{\beta}_K} \textstyle\sum_{t=1}^{T} \hat{u}_t^2 = min_{\hat{\beta}_1,\ldots,\hat{\beta}_K} \sum_{t=1}^{T} (Y_t - \hat{Y}_t)^2 =$$

$$min_{\hat{\beta}_1,\ldots,\hat{\beta}_K} \textstyle\sum_{t=1}^{T} (Y_t - \hat{\beta}_1 - \hat{\beta}_2 X_{2t} - \ldots - \hat{\beta}_K X_{Kt})^2 \qquad (2.5)$$

Para encontrar los K estimadores tomamos derivadas con respecto a los elementos desconocidos o lo que es lo mismo buscamos las Condiciones de Primer Orden (C.P.O.) de mínimo:

$$\frac{\partial \sum_{t=1}^{T} \hat{u}_t^2}{\partial \hat{\beta}_1} = 0$$

$$\begin{aligned}
\frac{\partial \sum_{t=1}^{T} \hat{u}_t^2}{\partial \hat{\beta}_2} &= 0 \\
\frac{\partial \sum_{t=1}^{T} \hat{u}_t^2}{\partial \hat{\beta}_3} &= 0 \\
\ldots & \quad \ldots \\
\frac{\partial \sum_{t=1}^{T} \hat{u}_t^2}{\partial \hat{\beta}_K} &= 0
\end{aligned}$$

de las cuales obtenemos las siguientes K-ecuaciones normales:

$$\begin{aligned}
-2\sum_{t=1}^{T}(Y_t - \hat{\beta}_1 - \hat{\beta}_2 X_{2t} - \ldots - \hat{\beta}_K X_{Kt}) &= 0 \\
-2\sum_{t=1}^{T}(Y_t - \hat{\beta}_1 - \hat{\beta}_2 X_{2t} - \ldots - \hat{\beta}_K X_{Kt})X_{2t} &= 0 \\
\ldots \quad \ldots \quad \ldots & \quad \ldots \\
-2\sum_{t=1}^{T}(Y_t - \hat{\beta}_1 - \hat{\beta}_2 X_{2t} - \ldots - \hat{\beta}_K X_{Kt})X_{Kt} &= 0
\end{aligned}$$

que podemos escribir como

$$\begin{aligned}
\sum Y_t &= T\hat{\beta}_1 + \hat{\beta}_2 \sum X_{2t} + \ldots + \hat{\beta}_K \sum X_{Kt} \\
\sum X_{2t} Y_t &= \hat{\beta}_1 \sum X_{2t} + \hat{\beta}_2 \sum X_{2t}^2 + \ldots + \hat{\beta}_K \sum X_{2t} X_{Kt} \\
\vdots \quad & \quad \vdots \qquad \vdots \qquad\qquad \vdots \\
\sum X_{Kt} Y_t &= \hat{\beta}_1 \sum X_{Kt} + \hat{\beta}_2 \sum X_{Kt} X_{2t} + \ldots + \hat{\beta}_K \sum X_{Kt}^2
\end{aligned}$$

Matricialmente,

$$\sum_{t=1}^{T} \hat{u}_t^2 = \underset{(1 \times 1)}{\hat{u}'\hat{u}}$$

y el criterio puede escribirse

$$min_{\hat{\beta}} \hat{u}'\hat{u} = min_{\hat{\beta}} (Y - X\hat{\beta})'(Y - X\hat{\beta}) = min_{\hat{\beta}} (Y'Y - 2\hat{\beta}'X'Y + \hat{\beta}'X'X\hat{\beta}).$$

Las K Condiciones de Primer Orden (C.P.O.) de mínimo son

$$\frac{\partial \hat{u}'\hat{u}}{\partial \hat{\beta}} = 0 \Rightarrow -2X'(Y - X\hat{\beta}) = 0.$$

Despejando, obtenemos las ecuaciones normales en forma matricial

$$X'Y = X'X\hat{\beta}_{MCO}. \tag{2.6}$$

y la expresión matricial del estimador MCO es:

$$\hat{\beta}_{MCO} = (X'X)^{-1}X'Y \tag{2.7}$$

donde $(X'X)$ es una matriz de orden $K \times K$, $X'Y$ un vector de orden $K \times 1$ y $\hat{\beta}$ un vector de orden $K \times 1$, tales que

$$\underset{(K \times K)}{X'X} = \begin{bmatrix} T & \sum X_{2t} & \sum X_{3t} & \cdots & \sum X_{Kt} \\ \sum X_{2t} & \sum X_{2t}^2 & \sum X_{2t}X_{3t} & \cdots & \sum X_{2t}X_{Kt} \\ \sum X_{3t} & \sum X_{3t}X_{2t} & \sum X_{3t}^2 & \cdots & \sum X_{3t}X_{Kt} \\ \vdots & \vdots & \vdots & & \vdots \\ \sum X_{Kt} & \sum X_{Kt}X_{2t} & \sum X_{Kt}X_{3t} & \cdots & \sum X_{Kt}^2 \end{bmatrix}$$

$$\underset{(K \times 1)}{X'Y} = \begin{bmatrix} \sum Y_t \\ \sum X_{2t}Y_t \\ \sum X_{3t}Y_t \\ \vdots \\ \sum X_{Kt}Y_t \end{bmatrix} \qquad \underset{(K \times 1)}{\hat{\beta}} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \vdots \\ \hat{\beta}_K \end{bmatrix}.$$

Además se cumplen también las condiciones de segundo orden de mínimo.

• Estimador MCO con datos centrados (o desviaciones a la media)

Existe una manera alternativa de calcular el estimador MCO (2.7): *centrando* previamente las variables. Al estimador calculado de esta manera lo llamamos estimador de MCO con datos centrados o desviaciones a la media.

A partir de las ecuaciones normales, si se despeja $\hat{\beta}_1$ de la primera ecuación normal

$$\begin{aligned} \sum Y_t &= T\hat{\beta}_1 + \hat{\beta}_2 \sum X_{2t} + \ldots + \hat{\beta}_K \sum X_{Kt} \\ \Longrightarrow \hat{\beta}_1 &= \bar{Y} - \hat{\beta}_2 \bar{X}_2 - \ldots - \hat{\beta}_K \bar{X}_K \end{aligned}$$

y es sustituido en la segunda,

$$\sum X_{2t}Y_t = (\bar{Y} - \hat{\beta}_2 \bar{X}_2 - \ldots - \hat{\beta}_K \bar{X}_K) \sum X_{2t} + \hat{\beta}_2 \sum X_{2t}^2 + \ldots + \hat{\beta}_K \sum X_{2t}X_{Kt}$$

y se agrupan términos en torno a los coeficientes $\hat{\beta}_k$,

$$\sum X_{2t}Y_t - \bar{Y} \sum X_{2t} = \hat{\beta}_2 (\sum X_{2t}^2 - \bar{X}_2 \sum X_{2t}) + \ldots + \hat{\beta}_K (\sum X_{2t}X_{Kt} - \bar{X}_K \sum X_{2t}).$$

Multiplicando y dividiendo por T a $\sum X_{2t}$,

$$\begin{aligned} \sum X_{2t}Y_t - T\bar{Y}\bar{X}_2 = \hat{\beta}_2 (\sum X_{2t}^2 - T\bar{X}_2^2) + \\ \ldots + \hat{\beta}_K (\sum X_{2t}X_{Kt} - T\bar{X}_K \bar{X}_2). \end{aligned}$$

Se denota:

$x_{kt} = X_{kt} - \bar{X}_k$, $\forall k$ a la variable K-ésima centrada o en desviaciones a la media.

$y_t = Y_t - \bar{Y}$ a la variable endógena centrada o en desviaciones a la media.

Además se cumple que la suma del producto cruzado entre dos variables centradas, por ejemplo x_{2t} y x_{3t}, es

$$\sum x_{2t}x_{3t} = \sum (X_{2t} - \bar{X}_2)(X_{3t} - \bar{X}_3) = \sum X_{2t}X_{3t} - T\bar{X}_2\bar{X}_3$$

y, por tanto, la segunda ecuación normal anterior puede escribirse como

$$\sum x_{2t}y_t = \hat{\beta}_2 \sum x_{2t}^2 + \ldots + \hat{\beta}_K \sum x_{2t}x_{Kt}.$$

Si se sustituye la primera ecuación normal, en las restantes $K-1$ ecuaciones se obtiene el siguiente sistema de ecuaciones normales con datos centrados

$$\begin{array}{rcl}
\sum x_{2t}y_t & = & \hat{\beta}_2 \sum x_{2t}^2 + \hat{\beta}_3 \sum x_{2t}x_{3t} + \ldots + \hat{\beta}_K \sum x_{2t}x_{Kt} \\
\sum x_{3t}y_t & = & \hat{\beta}_2 \sum x_{3t}x_{2t} + \hat{\beta}_3 \sum x_{3t}^2 + \ldots + \hat{\beta}_K \sum x_{3t}x_{Kt} \\
\vdots & & \vdots \qquad\qquad \vdots \qquad\qquad \vdots \\
\sum x_{Kt}y_t & = & \hat{\beta}_2 \sum x_{Kt}x_{2t} + \hat{\beta}_3 \sum x_{Kt}x_{3t} + \ldots + \hat{\beta}_K \sum x_{Kt}^2,
\end{array}$$

que puede escribirse, en forma matricial,

$$x'y = x'x\hat{\beta}^*$$

Aquí, $\hat{\beta}^*$ es un vector $(K-1)\times 1$ que incluye sólo a los coeficientes $\hat{\beta}_2, \ldots, \hat{\beta}_K$.

Por lo tanto, una forma alternativa de obtener el estimador MCO (2.7) de los coeficientes del modelo es

$$\hat{\beta}^* = (x'x)^{-1}x'y$$

donde

$$\underset{(K-1)\times(K-1)}{x'x} = \begin{bmatrix} \sum x_{2t}^2 & \sum x_{2t}x_{3t} & \cdots & \sum x_{2t}x_{Kt} \\ \sum x_{3t}x_{2t} & \sum x_{3t}^2 & \cdots & \sum x_{3t}x_{Kt} \\ \vdots & \vdots & & \vdots \\ \sum x_{Kt}x_{2t} & \sum x_{Kt}x_{3t} & \cdots & \sum x_{Kt}^2 \end{bmatrix}$$

$$\underset{((K-1)\times 1)}{x'y} = \begin{bmatrix} \sum x_{2t}y_t \\ \sum x_{3t}y_t \\ \vdots \\ \sum x_{Kt}y_t \end{bmatrix} \qquad \underset{((K-1)\times 1)}{\hat{\beta}^*_{MCO}} = \begin{bmatrix} \hat{\beta}_2 \\ \hat{\beta}_3 \\ \vdots \\ \hat{\beta}_K \end{bmatrix}$$

y finalmente, se estima el término independiente (a partir de la primera ecuación normal) como

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2\bar{X}_2 - \ldots - \hat{\beta}_K\bar{X}_K.$$

Interpretación de los coeficientes estimados por MCO

- $\hat{\beta}_1 = \widehat{E}(Y_t|X_{kt} = 0, \forall k = 2, \ldots, K)$.

  Valor esperado *estimado* de Y_t cuando las variables explicativas son todas cero.

- $\hat{\beta}_k = \frac{\partial \widehat{E(Y_t)}}{\partial X_{kt}} = \frac{\Delta \widehat{E(Y_t)}}{\Delta X_{kt}} \quad \forall k = 2, \ldots, K.$

  Incremento esperado *estimado* (ó *decremento esperado estimado*) en Y_t cuando la variable X_k se incrementa en una unidad, manteniéndose constantes el resto de las variables explicativas.

Algunas equivalencias de notación

$$\begin{aligned} Y_t = \beta_1 + \beta_2 X_{2t} + \ldots + \beta_K X_{Kt} + u_t \quad & t = 1,2,\ldots,T \\ \Leftrightarrow Y = X\beta + u & \end{aligned}$$

$$\begin{aligned} E(Y_t) = \beta_1 + \beta_2 X_{2t} + \ldots + \beta_K X_{Kt} \quad & t = 1,2,\ldots,T \\ \Leftrightarrow E(Y) = X\beta & \end{aligned}$$

$$\begin{aligned} \hat{Y}_t = \hat{\beta}_1 + \hat{\beta}_2 X_{2t} + \ldots + \hat{\beta}_K X_{Kt} \quad & t = 1,2,\ldots,T \\ \Leftrightarrow \hat{Y} = X\hat{\beta} & \end{aligned}$$

$$\begin{aligned} \hat{u}_t = Y_t - \hat{Y}_t \quad & t = 1,2,\ldots,T \\ \Leftrightarrow \hat{u} = Y - \hat{Y} & \end{aligned}$$

2.3.2. Propiedades de la Función de Regresión Muestral, FRM

a) La FRM pasa por el vector de medias:
$\bar{Y} = \hat{\beta}_1 + \hat{\beta}_2 \bar{X}_2 + \ldots + \hat{\beta}_K \bar{X}_K$. Por la primera ecuación normal tenemos:

$$\begin{aligned} \sum Y_t &= T\hat{\beta}_1 + \hat{\beta}_2 \sum X_{2t} + \ldots + \hat{\beta}_K \sum X_{Kt} \\ \frac{1}{T}\sum Y_t &= \hat{\beta}_1 + \hat{\beta}_2 \frac{1}{T}\sum X_{2t} + \ldots + \hat{\beta}_K \frac{1}{T}\sum X_{Kt} \\ \bar{Y} &= \hat{\beta}_1 + \hat{\beta}_2 \bar{X}_2 + \ldots + \hat{\beta}_K \bar{X}_K \end{aligned}$$

Por tanto esta propiedad se cumple sólo si el modelo tiene término independiente ya que si no no existe la primera ecuación normal.

b) La suma de los residuos es cero:
$\sum_{t=1}^{T} \hat{u}_t = 0$.

$$\sum_{t=1}^{T} \hat{u}_t = \sum_{t=1}^{T} (Y_t - \hat{Y}_t) = \sum_{t=1}^{T} (Y_t - \hat{\beta}_1 + \hat{\beta}_2 X_{2t} + \ldots + \hat{\beta}_K X_{Kt}) = 0$$

por la primera ecuación normal. Por tanto de nuevo sólo es cierto si el modelo tiene término independiente ya que si no no existe la primera ecuación normal.

c) La media muestral de Y es igual a la media muestral de las estimaciones de Y: $\bar{Y} = \bar{\hat{Y}}$.

$$\begin{aligned} \hat{u}_t = Y_t - \hat{Y}_t \Longleftrightarrow Y_t &= \hat{Y}_t + \hat{u}_t \\ \sum Y_t &= \sum \hat{Y}_t + \underbrace{\sum \hat{u}_t}_{=0} \\ \frac{1}{T}\sum Y_t &= \frac{1}{T}\sum \hat{Y}_t \Longrightarrow \bar{Y} = \bar{\hat{Y}} \end{aligned}$$

d) Los residuos son ortogonales a las variables explicativas: $X'\hat{u} = 0$ ($\hat{u}'X = 0$).

$$X'\hat{u} = X'(Y - \hat{Y}) = X'(Y - X\hat{\beta}) = 0$$

por las ecuaciones normales. Notar que:

$$X'\hat{u} = 0 \Leftrightarrow \begin{bmatrix} \sum_1^T \hat{u}_t \\ \sum_1^T X_{2t}\hat{u}_t \\ \sum_1^T X_{3t}\hat{u}_t \\ \vdots \\ \sum_1^T X_{Kt}\hat{u}_t \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \Rightarrow \sum_{t=1}^{T} \hat{u}_t = 0$$

c) Los residuos son ortogonales a las estimaciones de la variable endógena: $\hat{Y}'\hat{u} = 0$ ($\hat{u}'\hat{Y} = 0$).

$$\hat{Y}'\hat{u} = (X\hat{\beta})'\hat{u} = \hat{\beta}' \underbrace{X'\hat{u}}_{=0} = 0$$

Las propiedades 1, 2, y 3 se cumplen **sólo** si el modelo tiene un término independiente mientras que las propiedades 4 y 5 se cumplen siempre.

2.3.3. Medidas de bondad del ajuste:

Consideramos ahora la bondad del ajuste de la recta de regresión a la muestra. Definimos la variación de la variable Y como la distancia de los valores observados de la variable a su media muestral. La suma de esas variaciones al cuadrado es la variación de la variable endógena que se quiere explicar con la variación de las variables explicativas. Se le denota como:

$$\sum(Y_t - \bar{Y})^2 = SCT \rightarrow \text{Suma de Cuadrados Total}$$

Cuando el modelo tenga término independiente podremos dividir la variación total en dos partes, variación explicada y variación sin explicar.

$$\underbrace{\sum(Y_t - \bar{Y})^2}_{\text{Variación Total}} = \underbrace{\sum(\hat{Y}_t - \bar{\hat{Y}})^2}_{\text{Variación Explicada}} + \underbrace{\sum \hat{u}_t^2}_{\text{Variación sin Explicar}}$$

Dado que $Y = \hat{Y} + \hat{u}$, tenemos:

$$\begin{aligned} Y'Y &= (\hat{Y} + \hat{u})'(\hat{Y} + \hat{u}) = \\ &= \hat{Y}'\hat{Y} + \underbrace{\hat{Y}'\hat{u}}_{=0} + \underbrace{\hat{u}'\hat{Y}}_{=0} + \hat{u}'\hat{u} = \\ &= \hat{Y}'\hat{Y} + \hat{u}'\hat{u} \end{aligned}$$

Restando en ambos lados $T\bar{Y}^2$,

$$Y'Y - T\bar{Y}^2 = \hat{Y}'\hat{Y} - T\bar{Y}^2 + \hat{u}'\hat{u}$$

Si el modelo tiene término independiente, $\bar{Y} = \bar{\hat{Y}}$ de donde,

$$\begin{aligned} Y'Y - T\bar{Y}^2 &= \hat{Y}'\hat{Y} - T\bar{\hat{Y}}^2 + \hat{u}'\hat{u} \\ \sum Y_t^2 - T\bar{Y}^2 &= \sum \hat{Y}_t^2 - T\bar{\hat{Y}}^2 + \sum \hat{u}_t^2 \\ \sum(Y_t - \bar{Y})^2 &= \sum(\hat{Y}_t - \bar{\hat{Y}})^2 + \sum \hat{u}_t^2 \\ \\ SCT &= SCE + SCR \end{aligned}$$

donde[1]:

SCT: Suma de Cuadrados Total, mide la variación total.

SCE: Suma de Cuadrados Explicada, mide la variación explicada.

SCR: Suma de Cuadrados Residual, mide la variación sin explicar.

$$SCT = \sum (Y_t - \bar{Y})^2 = \sum y_t^2 = Y'Y - T\bar{Y}^2$$

$$SCE = \sum (\hat{Y}_t - \bar{\hat{Y}})^2 = \sum \hat{y}_t^2 = \hat{Y}'\hat{Y} - T\bar{Y}^2$$

$$SCR = \sum \hat{u}_t^2 = Y'Y - \hat{Y}'\hat{Y} = Y'Y - \hat{\beta}'X'Y$$

Coeficiente de determinación, R^2

El coeficiente de determinación o R^2 mide la proporción de la variación de Y que explicamos con la regresión:

$$\begin{aligned} R^2 &= \frac{SCE}{SCT} = \frac{\sum (\hat{Y}_t - \bar{\hat{Y}})^2}{\sum (Y_t - \bar{Y})^2} \\ &= 1 - \frac{SCR}{SCT} = 1 - \frac{\sum \hat{u}_t^2}{\sum (Y_t - \bar{Y})^2} \end{aligned}$$

• Características:

- $0 \leq R^2 \leq 1$ si existe término independiente en el modelo.
- Si no existe término independiente el R^2 no tiene sentido.
- El R^2 depende del número de variables explicativas que introduzcamos en el modelo de regresión porque la inclusión de nuevas variables exógenas hará que el porcentaje de variación explicada sea por lo menos igual o mayor, es decir el R^2 nunca será menor. Siempre que incluyamos variables explicativas en un modelo el R^2 va a aumentar aunque estas no sean significativas. Añadir variables explicativas aunque mejora el ajuste añade otro problema que es que tenemos que estimar más parámetros, con lo que perdemos grados de libertad. Por ello buscamos otro coeficiente que mida la bondad del ajuste y tenga en cuenta el efecto que hay al incluir nuevas variables en el modelo. Este coeficiente es el coeficiente de determinación corregido, $\bar{R}^2$ que no es más que una ponderación del R^2 por sus grados de libertad.

Coeficiente de determinación corregido, $\bar{R}^2$.

$$\begin{aligned} \bar{R}^2 = 1 - \frac{\frac{SCR}{(T-K)}}{\frac{SCT}{(T-1)}} &= 1 - \frac{(T-1)}{(T-K)} \frac{SCR}{SCT} \\ &= 1 - \frac{(T-1)}{(T-K)} (1 - R^2) \end{aligned}$$

- Cualquiera que sea el número de variables incluidas en un modelo la SCT será constante y por tanto si incluimos una nueva variable la SCR será menor y la SCE será mayor.

[1] En el anexo 1 aparecen distintas expresiones de la SCT, SCE y SCR que pueden resultar útiles

- Este coeficiente, penaliza la inclusión de nuevas variables explicativas. Si la nueva variable incluida explica a la variable endógena compensando la pérdida de grados de libertad el $\bar{R}^2$ aumenta. Sin embargo si la nueva variable incluida no explica a la variable endógena compensando la pérdida de grados de libertad el $\bar{R}^2$ disminuye.
- Si $K = 1$, $R^2 = \bar{R}^2$.
- Si $K > 1$, $\bar{R}^2 \leq R^2$.

Coeficientes de correlación

El coeficiente de correlación lineal simple mide el grado de asociación lineal entre dos variables, para X e Y se define

$$r_{\mathrm{xy}} = \frac{\frac{\sum(X_t-\bar{X})(Y_t-\bar{Y})}{T}}{\sqrt{\frac{\sum(X_t-\bar{X})^2}{T}}\sqrt{\frac{\sum(Y_t-\bar{Y})^2}{T}}} = \frac{\frac{\sum x_t y_t}{T}}{\sqrt{\frac{\sum x_t^2}{T}}\sqrt{\frac{\sum y_t^2}{T}}} \qquad \text{además} \qquad -1 \leq r_{XY} \leq 1$$

En el MRLG tendremos una matriz de coeficientes de correlación habitualmente denotada por R:

$$R = \begin{bmatrix} r_{11} & r_{12} & \dots & r_{1K} \\ r_{21} & r_{22} & \dots & r_{2K} \\ \dots & \dots & \dots & \dots \\ r_{K1} & r_{K2} & \dots & r_{KK} \end{bmatrix}$$

La matriz de correlación R se define como aquella matriz cuyos elementos son el coeficiente de correlación simple entre dos variables i y j, siendo:

- r_{1k} representa la correlación entre Y y X_k $\qquad k = 1, 2, \dots K$
- $r_{kk} = 1$, los elementos de la diagonal principal son todos unos.
- Además es una matriz simétrica.

En el modelo lineal general la correlación entre Y y X_2 no está adecuadamente recogida por el coeficiente de correlación simple ya que parte de la variación de Y será debida al resto de variables exógenas. Será necesario descontar este efecto tanto de Y como de X_2. Por ejemplo, en el modelo

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t$$

para estudiar la influencia de X_2 en Y utilizaremos el coeficiente de correlación parcial entre Y y X_2 que mide la correlación que queda entre estas dos variables después de eliminar el efecto de X_3 sobre Y y sobre X_2.

$$r_{12\cdot3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{1 - r_{13}^2}\sqrt{1 - r_{23}^2}}$$

2.4. Propiedades de los estimadores MCO

El método de MCO es sólo uno de los posibles métodos de estimación, la pregunta es ¿Cómo podemos elegir entre estimadores? obviamente en base a sus propiedades sobre su comportamiento en muestras repetidas. Estas propiedades son insesgadez, varianza pequeña y error cuadrático medio.

• **Insesgadez**
Un estimador es insesgado si su valor esperado coincide con el verdadero valor del parámetro. Sea $\hat{\theta}$ un estimador del parámetro θ, será insesgado si $E(\hat{\theta}) = \theta$.

• **Varianza mínima**
Desearemos que la varianza de un estimador sea lo más pequeña posible ya que cuanto menor sea la varianza muestral mayor es la precisión del estimador.

Si estamos comparando dos estimadores insesgados elegiremos aquel que tenga la menor varianza. Pero si estamos comparando dos estimadores sesgados o un estimador sesgado y uno insesgado este criterio no nos sirve y debemos introducir uno nuevo, el concepto de error cuadrático medio.

• **Error cuadrático Medio (ECM)**

$$ECM(\hat{\theta}) = E(\hat{\theta} - \theta)^2 = V(\hat{\theta}) + sesgo(\hat{\theta})^2$$

donde $Sesgo(\hat{\theta}) = E(\hat{\theta}) - \theta$. En base a este criterio elegimos el estimador con menor ECM.

2.4.1. Propiedades de los estimadores MCO

Sea el modelo de regresión lineal general

$$Y = X\beta + u$$

donde se cumplen todas las hipótesis básicas. El estimador MCO de los coeficientes

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

tiene las siguientes propiedades:

- Es lineal en las perturbaciones.
- Es insesgado.
- Tiene varianza mínima entre todos los estimadores lineales e insesgados

Demostración:

- **Linealidad.** Como las variables explicativas son no aleatorias, el estimador MCO es una combinación lineal de las perturbaciones.

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

- **Insesgadez.** Dado que: $E(u) = 0$ y la matriz X es no aleatoria, $\hat{\beta}_{MCO}$ es insesgado, es decir, su valor esperado es igual al vector de coeficientes del modelo.

$$\begin{aligned} E(\hat{\beta}) &= E(\beta + (X'X)^{-1}X'u) = \\ &= E(\beta) + (X'X)^{-1}X' \underbrace{E(u)}_{=0} = \\ &= \beta \end{aligned}$$

- **Varianza mínima.** Dado que: $E(u) = 0$ $E(uu') = \sigma^2 I_T$ y la matriz X es no aleatoria,

$$\begin{aligned} Var(\hat{\beta}) &= E[(\hat{\beta} - E(\hat{\beta}))(\hat{\beta} - E(\hat{\beta}))'] = \\ &= E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'] = \\ &= E\left[\left[(X'X)^{-1}X'u\right]\left[(X'X)^{-1}X'u\right]'\right] = \\ &= E[(X'X)^{-1}X'uu'X(X'X)^{-1}] = \\ &= (X'X)^{-1}X' \underbrace{E[uu']}_{=\sigma^2 I_T} X(X'X)^{-1} = \\ &= (X'X)^{-1}X'\sigma^2 I_T X(X'X)^{-1} = \\ &= \sigma^2(X'X)^{-1}X'X(X'X)^{-1} = \\ &= \sigma^2(X'X)^{-1} \end{aligned}$$

Esta matriz de varianzas y covarianzas es mínima y nos lo garantiza el Teorema de Gauss-Markov.

$$\underset{(K \times K)}{V(\hat{\beta})} =$$

$$\begin{bmatrix} Var(\hat{\beta}_1) & Cov(\hat{\beta}_1, \hat{\beta}_2) & Cov(\hat{\beta}_1, \hat{\beta}_3) & \cdots & Cov(\hat{\beta}_1, \hat{\beta}_K) \\ Cov(\hat{\beta}_2, \hat{\beta}_1) & Var(\hat{\beta}_2) & Cov(\hat{\beta}_2, \hat{\beta}_3) & \cdots & Cov(\hat{\beta}_2, \hat{\beta}_K) \\ Cov(\hat{\beta}_3, \hat{\beta}_1) & Cov(\hat{\beta}_3, \hat{\beta}_2) & Var(\hat{\beta}_3) & \cdots & Cov(\hat{\beta}_3, \hat{\beta}_K) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ Cov(\hat{\beta}_K, \hat{\beta}_1) & Cov(\hat{\beta}_K, \hat{\beta}_2) & Cov(\hat{\beta}_K, \hat{\beta}_3) & \cdots & Var(\hat{\beta}_K) \end{bmatrix} =$$

$$= \sigma^2 \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1K} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2K} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3K} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{K1} & a_{K2} & a_{K3} & \cdots & a_{KK} \end{bmatrix} = \sigma^2(X'X)^{-1}$$

Como toda matriz de varianzas y covarianzas, es simétrica.

Teorema de Gauss-Markov: Dados los supuestos básicos del modelo de regresión lineal general, "dentro de la clase de estimadores lineales e insesgados, $\hat{\beta}$ es el estimador eficiente, es decir, $\hat{\beta}$ tiene mínima varianza".

2.4.2. Consecuencias del incumplimiento de algunos supuestos: colinealidad

A la hora de estimar un modelo económico, los datos disponibles sobre las variables explicativas o regresores pueden presentar un alto grado de correlación, especialmente en un contexto de series temporales y con series macroeconómicas.

Cuando dos o más variables explicativas en un modelo están altamente correlacionadas en la muestra, es muy difícil separar el efecto parcial de cada una de estas variables sobre la variable dependiente. La información muestral que incorpora una de estas variables es casi la misma que el resto de las correlacionadas con ella. En este tema analizaremos las implicaciones que tiene en la estimación por el método de Mínimos Cuadrados Ordinarios este fenómeno muestral.

- El problema de multicolinealidad es un problema relacionado con la matriz de variables exógenas X .
- Se refiere no tanto a si existe o no relación lineal entre las variables exógenas del modelo de regresión, que existirá, como al grado de correlación lineal entre las variables explicativas del modelo de regresión lineal.
- En todo momento nosotros vamos a suponer que tenemos un modelo correctamente especificado y que al estimarlo detectamos los problemas en la matriz de datos X. Así, estamos enfocando el problema como un problema muestral.
- Podemos distinguir dos casos:
  - Multicolinealidad exacta: se produce cuando existe una relación lineal exacta.
  - Alta colinealidad: cuando la correlación entre las variables exógenas es muy alta pero no exacta.

Para verlo más claramente vamos a seguir un ejemplo. Para el modelo:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t \qquad t = 1, \ldots, T \tag{2.8}$$

Sea la siguiente secuencia de posibles correlaciones entre las variables X_2 y X_3.

$(x'x)$	$(x'x)^{-1}$	$\|x'x\|$
$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	1
$\begin{pmatrix} 1 & 0,9 \\ 0,9 & 1 \end{pmatrix}$	$\begin{pmatrix} 5,26 & -4,74 \\ -4,74 & 5,26 \end{pmatrix}$	0,19
$\begin{pmatrix} 1 & 0,99 \\ 0,99 & 1 \end{pmatrix}$	$\begin{pmatrix} 50,25 & -49,74 \\ -49,74 & 50,25 \end{pmatrix}$	0,0199
$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$	$\nexists$	0

Multicolinealidad exacta

- Los efectos directos de la correlación exacta entre regresores es que el valor del determinante $|X'X| = 0$, por tanto no podemos encontrar $(X'X)^{-1}$ y por tanto, no podemos estimar el modelo por MCO ya que el estimador se define como $\hat{\beta}_{MCO} = (X'X)^{-1}X'Y$.
- En este caso lo que ocurre es que tenemos combinaciones lineales en las columnas de la matriz X con lo que $rg(X) \neq k$ por lo que $(X'X)$ es una matriz singular.
- Relajamos la hipótesis básica:

$$rg(X) \neq k$$

- Cuando la correlación entre regresores es perfecta el problema de multicolinealidad exacta se convierte en un problema de especificación ya que **no podemos estimar todos los parámetros del modelo de forma individual**. Vamos a probar este resultado con el siguiente ejemplo:

Sea el modelo :

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t \qquad t = 1, \ldots, T \tag{2.9}$$

donde $X_{3t} = \alpha X_{2t}$.

- Si nosotros estimamos el modelo por MCO aplicamos el estimador $\hat{\beta}_{MCO} = (x'x)^{-1}(x'y)$ donde:
$$(x'x) = \begin{pmatrix} \sum x_{2t}^2 & \alpha \sum x_{2t}^2 \\ \alpha \sum x_{2t}^2 & \alpha^2 \sum x_{2t}^2 \end{pmatrix} = \sum x_{2t}^2 \begin{pmatrix} 1 & \alpha \\ \alpha & \alpha^2 \end{pmatrix}$$
donde $|x'x| = 0$ y no podemos encontrar la expresión $\hat{\beta}_{MCO}$.

- Si intentamos resolver el problema utilizando las ecuaciones normales tendremos que estas se definen como: $(x'x)\hat{\beta} = (x'y)$ y son:
$$\begin{pmatrix} \sum x_{2t}^2 & \alpha \sum x_{2t}^2 \\ \alpha \sum x_{2t}^2 & \alpha^2 \sum x_{2t}^2 \end{pmatrix} \begin{pmatrix} \hat{\beta}_2 \\ \hat{\beta}_3 \end{pmatrix} = \begin{pmatrix} \sum x_{2t} y_t \\ \alpha \sum x_{2t} y_t \end{pmatrix}$$
es decir las dos ecuaciones normales del sistema serían:
$$\sum x_{2t}^2 (\hat{\beta}_2 + \alpha \hat{\beta}_3) = \sum x_{2t} y_t$$
$$\alpha \sum x_{2t}^2 (\hat{\beta}_2 + \alpha \hat{\beta}_3) = \alpha \sum x_{2t} y_t$$
siendo la segunda ecuación normal redundante y lo único que seríamos capaces de estimar en el modelo sería la combinación lineal:
$$\beta_2 \widehat{+ \alpha} \beta_3 = \frac{\sum x_{2t} y_t}{\sum x_{2t}^2}$$

pero no cada uno de sus parámetros de forma individual. Además **no** importa la solución arbitraria de las ecuaciones normales, esta combinación lineal tiene siempre un único valor y siempre el mismo. Por ejemplo si $x_{3t} = 2x_{2t}$ y disponemos de la siguiente información muestral:

$$(x'x) = \begin{pmatrix} 10 & 20 \\ 20 & 40 \end{pmatrix} \qquad (x'y) = \begin{pmatrix} 5 \\ 10 \end{pmatrix}$$

las ecuaciones normales serían:

$$10\hat{\beta}_2 + 20\hat{\beta}_3 = 5$$

$$20\hat{\beta}_2 + 40\hat{\beta}_3 = 10$$

de donde al ser redundante la segunda ecuación normal tendríamos:

$$2\hat{\beta}_2 + 4\hat{\beta}_3 = 1$$

o lo que es igual:

$$2(\hat{\beta}_2 + 2\hat{\beta}_3) = 1$$

$$\widehat{\beta_2 + 2\beta_3} = 0,5$$

Conclusiones:

- Si existe multicolinealidad exacta:

$$rg(X) \neq k \Rightarrow |X'X| = 0 \Rightarrow \nexists (X'X)^{-1}$$

Por tanto, no podemos estimar de forma individual todos los parámetros del modelo.

- Podremos estimar:
  - individualmente: aquellos parámetros cuyas variables exógenas no est,n afectadas de correlación exacta con otras variables exógenas del modelo y
  - combinaciones lineales de los parámetros cuyas variables exógenas est,n implicadas en las relaciones lineales exactas.

- **Detección**: basta con ver que $|X'X| = 0$.

- **Predicción**: Desde el punto de vista de la predicción la multicolinealidad exacta no plantea problemas ya que en el periodo de predicción el modelo se definirá como:

$$Y_{T+1} = \beta_1 + \beta_2 X_{2,T+1} + \beta_3 X_{3,T+1} + u_{T+1} \tag{2.10}$$

y se mantendrá la relación $X_{3,T+1} = \alpha X_{2,T+1}$ de donde:

$$E(Y_{T+1}) = \beta_1 + (\beta_2 + \alpha\beta_3) X_{2,T+1} \tag{2.11}$$

y el estimador por punto de la variable endógena sería:

$$\hat{Y}_{T+1} = \hat{\beta}_1 + (\widehat{\beta_2 + \alpha\beta_3}) X_{2,T+1} \tag{2.12}$$

Alta colinealidad

- En este caso el valor del $|X'X|$ está muy próximo a cero, pero será distinto de cero, por tanto $\exists (X'X)^{-1}$ y podremos calcular los estimadores MCO. Además estos estimadores serán **lineales, insesgados y de varianza mínima**.

  Sin embargo la existencia de alta colinealidad entre variables produce efectos importantes que deben ser tenidos en cuenta y que son los siguientes:

- Varianzas y covarianzas cuantitativamente muy grandes:
  Dado que $(X'X)$ es casi singular, el valor de $|X'X|$ será muy pequeño, por lo que, $(X'X)^{-1}$ tendrá elementos muy grandes. Así, encontraremos varianzas y covarianzas muy grandes, pero estos valores serán los más pequeños que podemos encontrar en estas circunstancias. Cualquier otro estimador tendrá varianza mayor y por tanto el estimador MCO seguirá siendo de varianza mínima. Aunque como consecuencia del tamaño de $(X'X)^{-1}$, las estimaciones sean muy imprecisas.

- El mayor tamaño de las varianzas hará que aumente la probabilidad de no rechazar la hipótesis nula de significatividad individual, cuando en realidad la variable sea significativa, sólo que los datos no permiten detectar esta significatividad.
  Esto se debe a que cuando contrastamos: $H_o : \beta_i = 0$ utilizamos como estadístico de contraste:

$$\frac{\hat{\beta}_i}{\hat{des}(\hat{\beta}_i)} \overset{H_o}{\sim} t_{(T-k)}$$

  a mayor varianza, menor valor del estadístico t calculado y por tanto, mayor probabilidad de no rechazar la hipótesis nula, ya que son los valores pequeños del estadístico los que nos llevan a no rechazar y los grandes a rechazar.

- Como consecuencia de lo anterior, podremos encontrar R^2 grandes, que indican que las variables exógenas son conjuntamente significativas, unidos a variables explicativas no significativas a nivel individual.

- Pequeños cambios en los datos producen cambios importantes en las estimaciones de los parámetros. Volvamos al ejemplo inicial, suponiendo que $\sum x_{2t}y_t = 2{,}8$ y $\sum x_{3t}y_t = 2,9$ tendremos los siguientes estimadores en los casos 1 a 3.

  En los casos dos y tres vemos que un cambio en la covarianza entre X_2 y X_3 de 0,9 a 0,99 ha producido fuertes cambios en las estimaciones de los parámetros.

  De todas formas, si nos fijamos en la combinación lineal $\hat{\beta}_2 + \hat{\beta}_3$ en el segundo caso $\hat{\beta}_2 + \hat{\beta}_3 = 3$ mientras que en el segundo caso $\hat{\beta}_2 + \hat{\beta}_3 = 2,9$, por lo que parece que la suma de los coeficientes se ha estimado de forma precisa, pero este resultado depende de la covarianza entre $\hat{\beta}_2$ y $\hat{\beta}_3$. Teníamos varianzas muy grandes, si la colinealidad es (+) la covarianza será (−) con lo que habrá una compensación, pero si la colinealidad es (−), la covarianza será (+) y la $Var(\hat{\beta}_2 + \hat{\beta}_3$ explotaría.

$(x'x)$	$(x'x)^{-1}$	$\|x'x\|$	$\hat{\beta}_{MCO}$
$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	1	$\hat{\beta}_2 = 2,8$ $\hat{\beta}_3 = 2,9$ $\hat{\beta}_2 + \hat{\beta}_3 = 5,7$
$\begin{pmatrix} 1 & 0,9 \\ 0,9 & 1 \end{pmatrix}$	$\begin{pmatrix} 5,26 & -4,74 \\ -4,74 & 5,26 \end{pmatrix}$	0,19	$\hat{\beta}_2 = 1$ $\hat{\beta}_3 = 2$ $\hat{\beta}_2 + \hat{\beta}_3 = 3$
$\begin{pmatrix} 1 & 0,99 \\ 0,99 & 1 \end{pmatrix}$	$\begin{pmatrix} 50,25 & -49,74 \\ -49,74 & 50,25 \end{pmatrix}$	0,0199	$\hat{\beta}_2 = -3,567$ $\hat{\beta}_3 = 6,432$ $\hat{\beta}_2 + \hat{\beta}_3 = 2,9$
$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$	$\nexists$	0	

¿Cómo podemos analizar si existe un problema de alta colinealidad?

- Una primera aproximación consiste en obtener los coeficientes de correlación muestral simples para cada par de variables explicativas y ver si el grado de correlación entre estas variables es alto.

- El valor del determinante decrece cuando aumenta la colinealidad, tendiendo a cero cuando esta se hace exacta. Este hecho podemos interpretarlo como un aviso pero no tenemos una medida que nos permita afirmar cuando es grave o muy grave.

- Valores altos del R^2 unidos a coeficientes individualmente no significativos estarían indicando la posible existencia de colinealidad alta.

- Otra forma de **detectar la multicolinealidad** consiste en realizar la regresión de cada una de las variables explicativas sobre el resto[2] y analizar los coeficientes de determinación de cada regresión. Si alguno o algunos de estos coeficientes de determinación (R_j^2) son altos, estaría señalando la posible existencia de un problema de multicolinealidad.

- Belsley, Kuh y Welsch (1980) consideran una serie de indicadores para analizar el grado de multicolinealidad entre los regresores de un modelo, como por ejemplo los llamados **Tolerancia** (TOL) y **Factor de Inflación de la Varianza** (VIF) que se definen:

$$VIF_j = \frac{1}{\left(1 - R_j^2\right)} \qquad TOL_j = \frac{1}{VIF_j}$$

  siendo R_j^2 el coeficiente de determinación de la regresión auxiliar de la variable X_j sobre el resto de las variables explicativas y $1 \leq VIF_j \leq \infty$.

[2]En cada regresión se incluye el término constante como regresor pero no como variable dependiente.

La varianza de cada uno de los coeficientes de la regresión MCO ($\hat{\beta}_j$) de un modelo de regresión lineal general se puede expresar como:

$$var(\hat{\beta}_j) = \frac{\sigma^2}{\sum \left(X_{ji} - \bar{X}_j\right)^2} \frac{1}{\left(1 - R_j^2\right)} = \frac{\sigma^2}{\sum \left(X_{ji} - \bar{X}_j\right)^2} VIF_j$$

donde β_j, es el coeficiente que acompaña a la variable X_j y R_j^2 es el coeficiente de determinación de la regresión auxiliar de la variable X_j en función del resto de las variables explicativas. Como vemos existe una relación inmediata entre el valor VIF_j y la varianza del coeficiente estimado. Cuanto más se acerque R_j^2 a la unidad, es decir, cuanto mayor sea la colinealidad de la variable X_j con el resto, mayor es el valor de VIF_j y mayor es la varianza del coeficiente estimado, porque tal y como hemos dicho, la multicolinealidad "infla" la varianza. Según estos autores, si $VIF_j > 10$, entonces concluiremos que la colinealidad de X_j con las demás variables es alta.

La utilización de los coeficientes TOL y VIF para detectar la presencia de la multicolinealidad ha recibido múltiples críticas, porque la conclusión obtenida con estos valores no siempre recoge adecuadamente la información y problema de los datos. Tal y como hemos visto anteriormente, las varianzas de los estimadores depende del VIF_j, σ^2 y $\sum \left(X_{ji} - \bar{X}_j\right)^2$, por lo que un alto VIF_j no es condición suficiente ni necesaria para que dichas varianzas sean elevadas ya que es posible que σ^2 sea pequeño o $\sum \left(X_{ji} - \bar{X}_j\right)^2$ grande y se compensen.

En la literatura se han propuesto muchas soluciones al posible problema de alta colinealidad y ninguna de ellas es totalmente satisfactoria por ello parece sensato aprender a convivir con el problema y tener cuidado de no omitir aquellas variables que esconden su significatividad bajo un problema de colinealidad y no incurrir así en un problema de mala especificación. Aunque no es fácil, se pueden considerar las siguientes "soluciones" para intentar resolver el problema:

- Si realmente es un problema muestral, una posibilidad es cambiar de muestra porque puede ser que con nuevos datos el problema se resuelva, aunque esto no siempre ocurre. La idea consiste en conseguir datos menos correlacionados que los anteriores, bien cambiando toda la muestra o simplemente incorporando más datos en la muestra inicial. De todas formas, no siempre resulta fácil obtener mejores datos por lo que muy probablemente debamos convivir con el problema teniendo cuidado con la inferencia realizada y las conclusiones de la misma.

- En ocasiones, si se incorpora información a priori sobre los coeficientes del modelo desaparece el problema. Aún así, sería conveniente tener en cuenta dicha información antes de la detección del problema de multicolinealidad y no posteriormente, ya que así estimaremos el modelo más eficientemente.

2.4.3. Consecuencias del incumplimiento de algunos supuestos: omisión de variables relevantes e inclusión de variables irrelevantes

Dentro de las hipótesis básicas hemos supuesto que el modelo estaba correctamente especificado, esto en ocasiones no es así bien porque faltan variables (omisión de variables relevantes) o porque hay más de las necesarias (inclusión de variables irrelevantes). Estas situaciones influyen en las propiedades del estimador MCO y es necesario tenerlo en cuenta.

Omisión de variables relevantes

Suponemos que el correctamente especificado es:

$$Y = X\beta + u = [\ X_1 \quad X_2\ ] \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + u = X_1\beta_1 + X_2\beta_2 + u \tag{2.13}$$

donde X_1 es una submatriz de orden $(T \times K_1)$ y X_2 es una submatriz de orden $(T \times K_2)$ y por tanto β_1 es un subvector de orden $(K_1 \times 1)$ y β_2 es un subvector de orden $(K_2 \times 1)$. Pero nosotros estimamos el siguiente modelo incorrectamente especificado:

$$Y = X_1\beta_1 + v \qquad \text{donde} \qquad v = X_2\beta_2 + u \tag{2.14}$$

El modelo (2.14) incurre en un error de especificación ya que se omiten las variables relevantes recogidas en X_2. Esto es lo mismo que imponer la restricción vectorial $\beta_2 = 0$ cuando no es cierta.

El estimador MCO de β_1 es $\hat{\beta}_1 = (X_1'X_1)^{-1}X_1'Y$, y $\hat{v} = Y - X_1\hat{\beta}_1$. Consecuencias:

- En general los estimadores son sesgados:

$$E(\hat{\beta}_1) = E((X_1'X_1)^{-1}X_1'Y) = \beta_1 + (X_1'X_1)^{-1}X_1'X_2\beta_2$$

  $Sesgo[\hat{\beta}_1] = (X_1'X_1)^{-1}X_1'X_2\beta_2$ y se anulara si $X_1'X_2 = 0$, es decir, si las variables omitidas son ortogonales a las no omitidas. Notar que el sesgo se anula también para $\beta_2 = 0$ pero esta es una solución trivial dado que al ser X_2 regresores relevantes necesariamente $\beta_2 \neq 0$

- $V(\hat{\beta}_1) = \sigma^2(X_1'X_1)^{-1}$

- El estimador de la varianza de la perturbación es sesgado, y lo es **siempre** incluso cuando los regresores son ortogonales:

$$\hat{\sigma}^2 = \frac{\hat{v}'\hat{v}}{T - k_1} \longrightarrow E(\hat{\sigma}^2) = \frac{E(\hat{v}'\hat{v})}{T - k_1} \neq \sigma^2$$

- Inferencia no válida ya que al ser $\hat{\sigma}^2$ sesgada los estadísticos de contraste habituales t y F no siguen las distribuciones t-Student y F-Snedecor habituales.

Inclusión de variables irrelevantes

Este caso formalmente es justo el inverso del anterior. El modelo correctamente especificado es:

$$Y = X_1\beta_1 + u \qquad u \sim N(0, \sigma^2 I) \tag{2.15}$$

y el modelo estimado es:

$$Y = X_1\beta_1 + X_2\beta_2 + v \tag{2.16}$$

donde aparecen las variables irrelevantes en la matriz X_2 de orden $T \times K_2$ con unos coeficientes, β_2, de orden $K_2 \times 1$, que son cero, poblacionalmente. Es decir, no aplicamos la restricción $\beta_2 = 0$ cuando es cierta. Consecuencias:

- Los estimadores de los coeficientes son **insesgados**. Podemos escribir el modelo a estimar como:

$$Y = X_1\beta_1 + X_2\beta_2 + u \longrightarrow Y = [X_1 \; X_2] \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + u \longrightarrow Y = X^*\beta^* + u$$

$$\begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{bmatrix}^{-1} \begin{bmatrix} X_1'Y \\ X_2'Y \end{bmatrix}$$

$$\hat{\beta}^* = (X^{*'}X^*)^{-1}X^{*'}Y = (X^{*'}X^*)^{-1}X^{*'}(X\beta + u)$$
$$E(\hat{\beta}^*) = (X^{*'}X^*)^{-1}X^{*'}X\beta = \begin{bmatrix} I_{K_1} \\ 0 \end{bmatrix}\beta$$

$$E(\hat{\beta}^*) = E\begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} I_{K_1} \\ 0 \end{bmatrix}\beta = \begin{bmatrix} I_{K_1}\beta \\ 0 \end{bmatrix} = \begin{bmatrix} \beta \\ 0 \end{bmatrix}$$

Por lo tanto, el estimador de (2.16) sigue siendo insesgado.

- Las matriz de varianzas y covarianzas:

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

- El estimador de la varianza de las perturbaciones del modelo (2.16),

$$\hat{\sigma}^2 = \frac{\hat{v}'\hat{v}}{T - (K_1 + K_2)}$$

es un estimador **insesgado** de σ^2.

- La inferencia es aún válida para los estadísticos de contraste de hipótesis basados en $\hat{\beta}_1$, $\hat{\beta}_2$ y $\hat{\sigma}^2$.

2.5. Distribución del estimador MCO. Estimación por intervalo

2.5.1. Distribución del estimador de MCO bajo Normalidad

Si $Y = X\beta + u$, donde $u \sim N(0, \sigma^2 I_T)$, el estimador MCO, dado que es lineal en las perturbaciones, también seguirá una distribución Normal Multivariante, con vector de medias

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

y matriz de varianzas y covarianzas

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

Es decir,

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

Para el k-ésimo coeficiente,

$$\hat{\beta}_k \sim N(\beta_k, \sigma^2 a_{kk})$$

donde a_{kk} es el elemento (k,k) de la matriz $(X'X)^{-1}$.

2.5.2. Estimación de la varianza de las perturbaciones

En la matriz de varianzas y covarianzas del estimador MCO aparece la varianza de las perturbaciones, que si es desconocida ha de ser estimada. Habitualmente se utiliza el siguiente estimador insesgado[3] de σ^2:

$$\hat{\sigma}^2 = \frac{\hat{u}'\hat{u}}{T-K} = \frac{SCR}{T-K} = \frac{\sum \hat{u}_t^2}{T-K}$$

Y por tanto podremos utilizarlo como el estimador apropiado de la varianza de la perturbación. Para trabajar con él es útil escribirlo en términos de las variables observables Y, X, así:

$$\hat{\sigma}^2 \quad = \quad \frac{\hat{u}'\hat{u}}{T-K} = \frac{Y'Y - \hat{\beta}'X'Y}{T-K} = \frac{Y'Y - \hat{\beta}X'X\hat{\beta}}{T-K} = \frac{y'y - \hat{\beta}^{*'}x'y}{T-K}$$

Bajo las hipótesis básicas, un estimador insesgado de la matriz de varianzas y covarianzas, de $\hat{\beta}_{MCO}$ es

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

2.5.3. Estimación por intervalo

Hemos visto que bajo las hipótesis básicas:

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

Para el k-ésimo coeficiente,

$$\hat{\beta}_k \sim N(\beta_k, \sigma^2 a_{kk})$$

donde a_{kk} es el elemento (k,k) de la matriz $(X'X)^{-1}$. Una vez estimada la varianza de la perturbación con el estimador insesgado $\hat{\sigma}^2$ se puede demostrar que:

$$\frac{\hat{\beta}_k}{\hat{\sigma}\sqrt{a_{kk}}} \sim t_{(T-K)}$$

donde $t_{(T-K)}$ denota la distribución t-Student con $(T-K)$ grados de libertad, y $\hat{\sigma}\sqrt{a_{kk}}$ es la desviación estimada del coeficiente estimado. (Notación $\hat{\sigma}\sqrt{a_{kk}} = \hat{\sigma}_{\hat{\beta}_k}$).

El intervalo de confianza asociado es:

$$Pr\left[\hat{\beta}_k - t_{\frac{\alpha}{2}(T-K)}\hat{\sigma}_{\hat{\beta}_k} < \beta_k < \hat{\beta}_k + t_{\frac{\alpha}{2}(T-K)}\hat{\sigma}_{\hat{\beta}_k}\right] = 1-\alpha$$

Con lo que podemos escribir el intervalo de confianza del $(1-\alpha)$ por ciento para un coeficiente cualquiera β_k como:

$$IC(\beta_k)_{1-\alpha} = \left(\hat{\beta}_k \pm t_{\frac{\alpha}{2}(T-K)}\,\hat{\sigma}_{\hat{\beta}_k}\right)$$

Este es un *estimador por intervalo* porque en los extremos inferior y superior del intervalo aparecen $\hat{\beta}_k$ y $\hat{\sigma}_{\hat{\beta}_k}$, que son estimadores. Este intervalo es aleatorio, porque para cada muestra se obtiene un valor numérico distinto de $\hat{\beta}_k$ y $\hat{\sigma}_{\hat{\beta}_k}$. Cuando usamos una muestra para obtener las estimaciones, tendremos [un número $\leq \beta_k \leq$ otro número] y se denomina *estimación por*

[3]En el Anexo 2 aparece la demostración.

intervalo de β_k ó intervalo de confianza $(1-\alpha)$ para β_k. Un intervalo de confianza nos dice que, con probabilidad $(1-\alpha)$ se estima que el parámetro β_k estará dentro de ese rango de valores.

Las propiedades de la variable aleatoria $IC(\beta_k)$ se basan en la noción del muestreo repetido: si obtuviéramos infinitas muestras de tamaño T de una misma población, y para cada una de ellas construyésemos el intervalo, entonces $(1-\alpha) \times 100\,\%$ de todos los intervalos construidos contendrían el verdadero valor (desconocido) de β_k.

¿Para qué sirven las estimaciones por intervalo? La respuesta es que nos dan una información muy valiosa sobre la precisión de las estimaciones por punto, esto es, nos dicen hasta qué punto nos podemos fiar de ellas. Si un intervalo de confianza es ancho (debido a una $\widehat{Var}(\hat{\beta}_k)$ grande) nos está diciendo que no hay mucha información en la muestra sobre β_k. Además, como veremos más adelante, los intervalos sirven para realizar contraste de hipótesis.

2.6. Contraste de hipótesis sobre los coeficientes de la regresión

Un problema fundamental de la econometría es aportar un conocimiento descriptivo de una economía real, los economistas desarrollan teorías sobre el comportamiento económico y las evalúan. Los contrastes de hipótesis son los procedimientos que se usan para evaluar estas teorías[4]. Para ello vamos a utilizar el modelo $Y = X\beta + u$ donde consideramos que se cumplen las hipótesis básicas y en especial las siguientes:

- Linealidad en u.
- X no estocástica.
- $\beta_{(T\times 1)}$ constante en el periodo muestral.
- Normalidad de la perturbación.

Las tres primeras hipótesis las hemos utilizado para determinar las propiedades del estimador de Mínimos Cuadrados Ordinarios. La normalidad no es necesaria para estimar por MCO ni para determinar las propiedades del estimador pero si lo es para realizar inferencia dado que al ser $\hat{\beta}_{MCO}$ lineal en u tendrá su misma distribución y podremos derivar estadísticos de contraste basándonos en ella. Por ejemplo, dado que

$$u_t \sim N(0,\sigma^2) \longrightarrow \hat{\beta}_i \sim N(\beta_i, \sigma^2 a_{ii})$$

si conocemos todos los elementos incluido σ^2 podríamos contrastar hipótesis de la forma $H_0 : \beta_i = c$ con el siguiente estadístico:

$$\frac{\hat{\beta}_i - c}{\sigma\sqrt{a_{ii}}} \overset{H_0}{\sim} N(0,1)$$

Si σ^2 es desconocida habría de ser estimada y derivar el correspondiente estadístico de contraste, que sería:

$$\frac{\frac{\hat{\beta}_i - c}{\sigma\sqrt{a_{ii}}}}{\sqrt{\frac{\sum \hat{u}_t^2/\sigma^2}{T-K}}} \overset{H_0}{\sim} t_{(T-K)} \qquad \text{si simplificamos} \qquad \frac{\hat{\beta}_i - c}{\hat{\sigma}\sqrt{a_{ii}}} \overset{H_0}{\sim} t_{(T-K)}$$

En general nosotros lo que queremos es contrastar conjuntos lineales de hipótesis por ello nos resulta cómodo derivar una única expresión general para hacer inferencias que sólo sea necesario

[4] En el anexo 3 aparece un recordatorio estadístico sobre el contraste de hipótesis

adaptarla al caso particular. Antes de derivar esta expresión general vamos a recordar o derivar las distribuciones de los estadísticos básicos más importantes que hemos visto[5].

• **Distribuciones que nos interesan**

- Distribución de la variable endógena o regresando:
  $Y = X\beta + u$
  $E(Y) = X\beta$
  $Var(Y) = E[(Y - E(Y))(Y - E(Y))'] = E(uu') = \sigma^2 I_T$

  Dado
  $$u \sim N(0, \sigma^2 I_T) \longrightarrow Y \sim N(X\beta, \sigma^2 I_T)$$

- Distribución del vector de estimadores MCO:
  $\hat{\beta} = (X'X)^{-1}X'Y = \beta + (X'X)^{-1}X'u$
  $E(\hat{\beta}) = \beta$
  $Var(\hat{\beta}) = \sigma^2(X'X)^{-1}$

  Dado
  $$\begin{aligned} u \sim N(0, \sigma^2 I_T) \quad &\longrightarrow \quad \hat{\beta} \sim N(\beta, \sigma^2(X'X)^{-1}) \\ &\longrightarrow \quad \hat{\beta}_i \sim N(\beta_i, \sigma^2 a_{ii}) \end{aligned}$$

- Distribución de los residuos:
  $\hat{u} = Mu$
  $E(\hat{u}) = E(Mu) = ME(u) = 0$
  $Var(\hat{u}) = E[\hat{u}\hat{u}'] = E(Muu'M'] = ME(uu')M = \sigma^2 M$

  Dado
  $$u \sim N(0, \sigma^2 I_T) \longrightarrow \hat{u} \sim N(0, \sigma^2 M)$$

- Distribución de la SCR:
  $SCR = \hat{u}'\hat{u} = u'Mu$
  Utilizando el resultado 3 del Anexo 4 dado
  $$u \sim N(0, \sigma^2 I_T) \quad \longrightarrow \quad \frac{u'Mu}{\sigma^2} \sim \chi^2_{(T-K)} \longrightarrow \frac{\hat{u}'\hat{u}}{\sigma^2} \sim \chi^2_{(T-K)}$$

  ya que: $\hat{u} \sim N(0, \sigma^2 M)$ tenemos
  $$\frac{\hat{u}'M^{-1}\hat{u}}{\sigma^2} = \frac{u'MM^{-1}Mu}{\sigma`2} = \frac{u'Mu}{\sigma^2} \sim \chi^2_{(T-k)}$$

2.6.1. Expresión general para contrastar restricciones lineales

Nosotros queremos contrastar conjuntos lineales de hipótesis, para ello vamos a expresar la hipótesis nula como:
$$H_0 : \underset{(q \times K)}{R} \cdot \underset{(K \times 1)}{\beta} = \underset{(q \times 1)}{r}$$

Esto significa que estamos contrastando q hipótesis y:

[5]En el Anexo 4 se recoge un recordatorio de distribuciones asociadas a la normal que puede ser de utilidad.

- q es el número de restricciones que impone la hipótesis nula
- R es una matriz de constantes conocidas.
- r es un vector de constantes conocidas.
- β es el vector de orden $(K \times 1)$ de parámetros desconocidos.

Las constantes de R y de r dependen de cuál sea la hipótesis nula en concreto que se desee contrastar. Vamos a repasar diferentes ejemplos:

a) **Contrastes individuales**:

- Ejemplo 1: **Contraste de significatividad individual**:

$$H_0 : \beta_i = 0 \qquad \text{en este caso } q = 1$$

Definimos:

$$R = \begin{bmatrix} 0 & \dots & 0 & \quad 1 & 0 & \dots & 0 \end{bmatrix}$$

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_i \\ \vdots \\ \beta_K \end{bmatrix} \qquad \text{y} \qquad r = \begin{bmatrix} 0 \end{bmatrix}$$

por tanto expresarlo como $H_0 : R\beta = r$

$$\begin{bmatrix} 0 & \dots & 0 & \quad 1 & 0 & \dots & 0 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_i \\ \vdots \\ \beta_K \end{bmatrix} = \begin{bmatrix} 0 \end{bmatrix}$$

- Ejemplo 2: Si la hipótesis nula es $H_0 : \beta_i = c$, cambia el valor de la constante en la matriz r

$$r = \begin{bmatrix} c \end{bmatrix}$$

b) **Contrastes de combinaciones lineales**:

- Ejemplo 1:

$$H_0 : \beta_2 + \beta_3 = 1$$

En este caso $q = 1$, la forma $H_0 : R\beta = r$ corresponde a

$$\begin{bmatrix} 0 & 1 & 1 & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \vdots \\ \beta_K \end{bmatrix} = [1]$$

- Ejemplo 2:

$$H_0 : \beta_2 = \beta_3$$

En este caso $q = 1$, la forma $H_0 : R\beta = r$ corresponde a

$$\begin{bmatrix} 0 & 1 & -1 & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \vdots \\ \beta_K \end{bmatrix} = [0]$$

donde lo que hemos hecho es escribir la hipótesis nula como:
$H_0 : \beta_2 - \beta_3 = 0$

- Ejemplo 3:

$$H_0 : \begin{cases} 2\beta_2 + 3\beta_3 = 5 \\ \beta_1 - 2\beta_4 = 3 \end{cases}$$

En este caso $q = 2$

$$\begin{bmatrix} 0 & 2 & 3 & 0 & \cdots & 0 \\ 1 & 0 & 0 & -2 & \cdots & 0 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \vdots \\ \beta_K \end{bmatrix} = \begin{bmatrix} 5 \\ 3 \end{bmatrix}$$

c) Contraste de significatividad de la regresión o contraste de significatividad conjunta:

$$\begin{cases} H_0 : & \beta_2 = \beta_3 = \cdots = \beta_K = 0 \quad \rightarrow q = K - 1 \\ H_A : & \text{alguna igualdad no se da} \end{cases}$$

la hipótesis nula en forma vectorial es:

$$\begin{bmatrix} \beta_2 \\ \beta_3 \\ \vdots \\ \beta_K \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

así que la matriz R de orden $(K-1) \times K$ es:

$$R = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 \\ & & & & & \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{bmatrix} = [0 \quad I_{K-1}]$$

y la matriz r es un vector de ceros de orden $(K-1) \times 1$

Ahora ya sabemos cómo formular cualquier hipótesis nula de una o varias restricciones lineales.

El estadístico general que usaremos para realizar los contrastes para el caso en que σ^2 sea desconocida es el siguiente:

$$\frac{\left(R\hat{\beta}-r\right)'\left(R\,(\,X'X\,)^{-1}\,R'\right)^{-1}\left(R\hat{\beta}-r\right)/q}{\hat{\sigma}^2} \overset{H_0}{\sim} \mathcal{F}_{(q,T-K)} \tag{2.17}$$

Rechazaremos la hipótesis nula para valores grandes del estadístico, ya que si H_0 no es cierta, la diferencia $R\hat{\beta}-r$ será grande y el estadístico calculado (al que llamaremos F) tomará valores altos. El criterio de contraste es por tanto el de rechazar H_0 con un nivel de significación α cuando

$$\frac{\left(R\hat{\beta}-r\right)'\left(R\,(\,X'X\,)^{-1}\,R'\right)^{-1}\left(R\hat{\beta}-r\right)/q}{\hat{\sigma}^2} > \mathcal{F}_{\alpha(q,T-K)}$$

y no rechazarla en caso contrario[6].

Cuando tengamos los datos en desviaciones a la media podemos escribir:

$$\frac{\left(R\hat{\beta}^*-r\right)'\left[R\,(x'x)^{-1}\,R'\right]^{-1}\left(R\hat{\beta}^*-r\right)/q}{\hat{u}'\hat{u}/T-K} \overset{H_0}{\sim} \mathcal{F}_{(q,T-K)}$$

Hay que darse cuenta de que ahora el vector $\hat{\beta}^*$ y la matriz R tienen un elemento menos y una columna menos respectivamente, porque con variables centradas no aparece de manera explícita el término independiente.

2.6.2. Aplicación del procedimiento general

Vamos a aplicar el estadístico general a los ejemplos anteriores.

- **Contraste de significatividad individual**

$$H_0 : \beta_i = 0$$
$$H_a : \beta_i \neq 0$$

En este caso $q = 1$. La distribución del parámetro estimado $\hat{\beta}_i$ es la siguiente:

$$\hat{\beta}_i \sim N(\beta_i, \sigma^2 a_{ii})$$

La expresión $R\hat{\beta} = r$ es:

$$\begin{bmatrix} 0 & \dots & 0 & 1 & 0 & \dots & 0 \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_{i-1} \\ \hat{\beta}_i \\ \hat{\beta}_{i+1} \\ \vdots \\ \hat{\beta}_K \end{bmatrix} = [0]$$

[6]En el Anexo 5 se muestra cómo obtener la distribución del estadístico general cuando σ^2 es conocida y cuando es desconocida.

Así, $R\hat{\beta} - r = \hat{\beta}_i$. La expresión $R(X'X)^{-1}R'$ corresponde a

$$\begin{bmatrix} 0 & \dots & 1 & \dots & 0 \end{bmatrix} \begin{bmatrix} \ddots & & \\ & a_{ii} & \\ & & \ddots \end{bmatrix} \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} = a_{ii}$$

Por tanto,

$$\frac{\left(R\hat{\beta} - r\right)' \left(R\,(\,X'X\,)^{-1}\,R'\right)^{-1} \left(R\hat{\beta} - r\right)/q}{\hat{\sigma}^2} =$$

$$\frac{\hat{\beta}_i'(a_{ii})^{-1}\hat{\beta}_i}{\hat{\sigma}^2} \overset{H_0}{\sim} \mathcal{F}_{(q,T-K)}$$

$$\frac{\hat{\beta}_i^2}{\hat{\sigma}^2 a_{ii}} \overset{H_0}{\sim} \mathcal{F}_{(q,T-K)}$$

Utilizando el resultado 8 del Anexo 4 podemos derivar el siguiente estadístico de contraste:

$$\frac{\hat{\beta}_i}{\hat{\sigma}\sqrt{a_{ii}}} \overset{H_0}{\sim} t_{(T-K)}$$

O lo que es lo mismo:

$$\frac{\hat{\beta}_i}{\widehat{desv}(\hat{\beta}_i)} \overset{H_0}{\sim} t_{(T-K)}$$

Si el estadístico calculado para la muestra es mayor que el estadístico en tablas, para un α dado, se rechaza la hipótesis nula. En este caso $\beta_i \neq 0$ y la variable explicativa asociada X_i es significativa para explicar el comportamiento de la variable endógena. Por tanto este contraste sirve para decidir si la variable X_i debe mantenerse en el modelo.

En secciones anteriores hablamos de la estimación por intervalo y se mencionó que también podíamos realizar inferencia con intervalos de confianza. Pues bien si recordamos el intervalo de confianza asociado a β_i:

$$Pr\left[\hat{\beta}_i - t_{\frac{\alpha}{2}(T-K)}\hat{\sigma}_{\hat{\beta}_i} < \beta_i < \hat{\beta}_i + t_{\frac{\alpha}{2}(T-K)}\hat{\sigma}_{\hat{\beta}_i}\right] = 1 - \alpha$$

$$IC(\beta_i)_{1-\alpha}: \quad \left(\hat{\beta}_i \pm t_{\frac{\alpha}{2}(T-K)}\,\hat{\sigma}_{\hat{\beta}_i}\right)$$

y la regla de decisión es que si la constante c (en este caso $c = 0$) pertenece al intervalo, no rechazamos H_0 con un nivel de significación α y si no pertenece al intervalo, rechazamos H_0 con un nivel de significación α. Claramente, con esta manera de hacer los contrastes individuales se obtienen exactamente los mismos resultados que con el enfoque del contraste de significación.

- **Contraste de significatividad conjunto**

$$\begin{cases} H_0: & \beta_2 = \beta_3 = \cdots = \beta_K = 0 \quad \rightarrow q = K - 1 \\ H_A: & \text{alguna igualdad no se da} \end{cases}$$

Donde $q = K - 1$. En este caso

$$R = \begin{bmatrix} 0 & I_{K-1} \end{bmatrix} \qquad r = [0] \qquad \hat{\beta} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_K \end{bmatrix}$$

LLamamos:

$$\hat{\beta}_2^\star = \begin{bmatrix} \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_K \end{bmatrix}$$

Así

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2^\star \end{bmatrix}$$

de donde

$$R\hat{\beta} - r = \hat{\beta}_2^\star \longrightarrow \hat{\beta} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2^\star \end{bmatrix}$$

y

$$[R(X'X)^{-1}R']^{-1} = (x'x)$$

El estadístico de contraste es:

$$\frac{\hat{\beta}_2^{\star\prime}(x'x)\hat{\beta}_2^\star / K - 1}{\hat{\sigma}^2} \overset{H_0}{\sim} \mathcal{F}_{(q,T-K)}$$

dado que $\hat{\beta}_2^{\star\prime}(x'x)\hat{\beta}_2^\star = \hat{\beta}_2^{\star\prime} x'y$ podemos escribir el estadístico anterior como:

$$\frac{R^2/K-1}{1-R^2/T-K} \overset{H_0}{\sim} \mathcal{F}_{(q,T-K)}$$

Si el estadístico calculado para la muestra es mayor que el estadístico en tablas, para un α dado, se rechaza la hipótesis nula. Las variables son conjuntamente significativas para explicar el comportamiento de la variable endógena.

- **Contraste de combinaciones lineales**
  Parael ejemplo 1, para la hipótesis

$$H_0 : \beta_2 + \beta_3 = 1$$

$$H_0 : \beta_2 + \beta_3 \neq 1$$

$q = 1$ por lo que podemos aplicar el resultado 8 del Anexo 4 $\mathcal{F}_{(1,q)} = t^2_{(T-K)}$. Renombrando $R\hat{\beta} = \hat{w}$ y $r = c$ se puede expresar el estadístico de contraste como:

$$t = \frac{\hat{w} - c}{\hat{\sigma}_{\hat{w}}} \overset{H_0}{\sim} t_{(T-K)} \quad \text{si H}_0 \text{ es cierta}$$

Aplicación:

$$H_0 : \beta_2 + \beta_3 = 1 \qquad H_0 : w = 1$$

$$H_a : \beta_2 + \beta_3 \neq 1 \qquad H_a : w \neq 1$$

$$\hat{w} - c = R\hat{\beta} - r = \begin{bmatrix} 0 & 1 & 1 & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \vdots \\ \hat{\beta}_K \end{bmatrix} - [1] = \hat{\beta}_2 + \hat{\beta}_3 - 1$$

Buscamos la distribución del estadístico $\hat{w}$:

$$\hat{w} = \hat{\beta}_2 + \hat{\beta}_3$$

dado que $\hat{\beta}_2 \sim N(\beta_2, \sigma^2 a_{22})$ $\quad \hat{\beta}_3 \sim N(\beta_3, \sigma^2 a_{33})$

$$\begin{aligned} E(\hat{w}) &= E(\hat{\beta}_2 + \hat{\beta}_3) = \beta_2 + \beta_3 \\ Var(\hat{w}) &= E[\hat{w} - E(\hat{w})]^2 \\ &= E[(\hat{\beta}_2 + \hat{\beta}_3) - (\beta_2 + \beta_3)]^2 \\ &= Var(\hat{\beta}_2) + Var(\hat{\beta}_3) + 2Cov(\hat{\beta}_2, \hat{\beta}_3) \\ &= \sigma^2(a_{22} + a_{33} + 2a_{23}) \end{aligned}$$

Por tanto

$$\hat{\beta}_2 + \hat{\beta}_3 \sim N(\beta_2 + \beta_3, \sigma^2(a_{22} + a_{33} + 2a_{23}))$$

y el estadístico de contraste es:

$$\frac{\hat{\beta}_2 + \hat{\beta}_3 - 1}{\hat{\sigma}\sqrt{a_{22} + a_{33} + 2a_{23}}} \overset{H_0}{\sim} t_{(T-K)}$$

Con la regla de decisión habitual. En el resto de ejemplo de combinaciones lineales se procederá igual.

2.6.3. Estimación mínimo cuadrática sujeta a restricciones

Un aspecto básico de la inferencia estadística que se lleva a cabo en Economía es que el investigador sólo contrasta hipótesis en cuya validez está dispuesto a creer a priori, de modo que si su contraste no las rechaza, entonces pasa a imponerlas en la representación estructural que está considerando. Si la hipótesis nula no se rechaza, entonces sería muy interesante disponer de un procedimiento para estimar de nuevo el modelo, pero esta vez imponiendo ese conjunto de hipótesis que hemos contrastado y no rechazado. La idea de eficiencia está ligada a la utilización óptima de toda la información disponible. Si se cree que los coeficientes del modelo satisfacen ciertas restricciones, entonces se ganaría eficiencia introduciendo dichas restricciones en el proceso de información.

En este caso vamos a encontrar el estimador que minimice la suma de cuadrados de los residuos, pero esta vez imponiendo las restricciones $\mathbf{R}\beta = \mathbf{r}$, es decir, se trata esta vez de resolver un problema de optimización sujeto a restricciones lineales. Nuestro problema a resolver es:

$$\begin{aligned} &Min(Y - X\hat{\beta})'(Y - X\hat{\beta}) \\ &\quad s.a. \qquad R\hat{\beta} = r \end{aligned}$$

El lagrangiano de tal problema será:

$$L = (Y - X\hat{\beta})'(Y - X\hat{\beta}) - 2\lambda'(R\hat{\beta} - r)$$

donde λ es un vector de dimensión $(q \times 1)$ de multiplicadores de Lagrange (tantos como restricciones). Tomando derivadas parciales de este lagrangiano con respecto a $\hat{\beta}'$, así como respecto a λ, se tiene:

$$\frac{\partial L}{\partial \hat{\beta}} = -2X'Y + 2X'X\hat{\beta} - 2R'\lambda \quad (K \quad derivadas)$$

$$\frac{\partial L}{\partial \lambda} = -2(R\hat{\beta} - r) \quad (q \quad derivadas)$$

Igualando a cero estas derivadas parciales y resolviendo el sistema de $K+q$ ecuaciones que así se obtienen resulta:

$$(X'X)\hat{\beta}_r - X'Y - R'\lambda = 0_k \tag{2.18}$$

$$R\hat{\beta}_r - r = 0_q \tag{2.19}$$

y las soluciones a este sistema de ecuaciones, $\hat{\beta}_r$, son el **estimador de Mínimos Cuadrados Restringidos (MCR)** y el vector de precios sombra (multiplicadores de Lagrange) de las q restricciones. La expresión del estimador restringido es la siguiente:

$$\hat{\beta}_r = \hat{\beta}_{MCO} + (X'X)^{-1}R'\left(R(X'X)^{-1}R'\right)^{-1}(r - R\hat{\beta})$$

donde $\hat{\beta}_{MCO}$ es el estimador Mínimo Cuadrático Ordinario sin restringir.
La matriz de varianzas y covarianzas de este estimador es:

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

Resultados:

a) $\hat{\beta}_r$ es lineal en u.

b) **Si las restricciones que hemos impuesto son ciertas el estimador $\hat{\beta}_r$ es insesgado**.

c) Comparando las matrices de varianzas y covarianzas de los estimadores de mínimos cuadrados ordinarios y mínimos cuadrados restringidos se puede demostrar que

$$V(\hat{\beta}) - V(\hat{\beta}_r)$$

es una matriz semidefinida positiva aunque la restricción no se cumpla.

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

Si la restricción no se cumple el estimador restringido será sesgado, por lo tanto para comparar los estimadores MCR y MCO habrá, en general, que utilizar el criterio del error cuadrático medio.

Análisis de residuos:

Los residuos del modelo estimado por MCR se definen:

$$\hat{u}_r = Y - X\hat{\beta}_r \tag{2.20}$$

tal que:

$$\hat{u}_r = Y - X\hat{\beta} - X\hat{\beta}_r + X\hat{\beta} = \hat{u} - X(\hat{\beta}_r - \hat{\beta})$$

de donde podemos definir la suma residual de cuadrados del modelo restringido como:

$$\begin{aligned} \hat{u}_r'\hat{u}_r &= \left(\hat{u} - X(\hat{\beta}_r - \hat{\beta})\right)' \left(\hat{u} - X(\hat{\beta}_r - \hat{\beta})\right) = \\ &= \hat{u}'\hat{u} - (\hat{\beta}_r - \hat{\beta})'X'\hat{u} - \hat{u}'X(\hat{\beta}_r - \hat{\beta}) + \\ &\quad (\hat{\beta}_r - \hat{\beta})'(X'X)(\hat{\beta}_r - \hat{\beta}) \end{aligned}$$

dado que $X'\hat{u} = 0$ tenemos que:

$$\hat{u}_r'\hat{u}_r - \hat{u}'\hat{u} = (\hat{\beta}_r - \hat{\beta})'(X'X)(\hat{\beta}_r - \hat{\beta})$$

de donde:

$$\hat{u}_r'\hat{u}_r - \hat{u}'\hat{u} = (R\hat{\beta} - r)' \left(R(X'X)^{-1}R'\right)^{-1} (R\hat{\beta} - r)$$

con lo que el estadístico general para el contraste de hipótesis de la forma $R\beta = r$:

$$\frac{(R\hat{\beta}_{MCO} - r)' \left(R(X'X)^{-1}R'\right)^{-1} (R\hat{\beta}_{MCO} - r)/q}{\hat{\sigma}^2} \overset{H_0}{\sim} F(q, T-K)$$

también puede escribirse de la forma:

$$\frac{\hat{u}_r'\hat{u}_r - \hat{u}'\hat{u}/q}{\hat{u}'\hat{u}/(T-k)} \overset{H_0}{\sim} F(q, T-k) \tag{2.21}$$

siendo:

- $\hat{u}_r'\hat{u}_r$ la suma residual de cuadrados del modelo restringido.
- $\hat{u}'\hat{u}$ la suma residual del modelo no restringido.

y por tanto la ecuación (2.21) define un estadístico válido para contrastar cualquier hipótesis lineal con tal de definir y estimar correctamente el modelo restringido. Generalmente se le conoce con el nombre de **estadístico de diferencias en las sumas residuales de cuadrados**.

2.6.4. Contrastes basados en sumas de cuadrados de residuos

Hemos visto cómo se puede contrastar cualquier hipótesis nula escrita de la forma $H_0 : R\beta = r$ mediante el estadístico F. Una expresión equivalente de este estadístico es el estadístico de diferencias en las sumas residuales de cuadrados visto en la sección anterior.

$$F = \frac{(\hat{u}_R'\hat{u}_R - \hat{u}'\hat{u})/q}{\hat{u}'\hat{u}/T-K} \overset{H_0}{\sim} \mathcal{F}_{(q,T-K)}$$

Para su aplicación sólo es necesario obtener la SCR del modelo restringido y no restringido. Hay dos formas equivalentes de obtener la SCR del modelo restringido:

a) a partir del vector completo $\underset{(K\times 1)}{\hat{\beta}_R}$ como $\hat{u}'_R\hat{u}_R = (Y - X\hat{\beta}_R)'(Y - X\hat{\beta}_R)$.

Precaución: $\hat{u}'_R\hat{u}_R \neq Y'Y - \hat{\beta}'_R X'Y$

b) Estimando el modelo restringido y a continuación calculando su suma residual de cuadrados. Esta forma es la más habitual de trabajar y vamos a estudiarla en detalle.

Estimación del modelo restringido y cálculo de la SCR_R
El modelo restringido es aquel que cumple la hipótesis nula. Por ejemplo sea el MRLG

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + u_t$$

donde queremos contrastar la hipótesis nula $H_0 : \beta_2 + \beta_3 = 1$ sustituyendo la restricción en el modelo encontramos el modelo restringido:

$$\text{MR:} \quad Y_t = \beta_1 + \beta_2 X_{2t} + (1 - \beta_2) X_{3t} + u_{Rt}$$

$$\underbrace{Y_t - X_{3t}}_{=Y_t^\star} = \beta_1 + \beta_2(\underbrace{X_{2t} - X_{3t}}_{=X_t^\star}) + u_{Rt}$$

La aplicación de MCO en el modelo resultante son los estimadores de MCR. Los demás $\hat{\beta}_R$ se obtienen con las restricciones. En el ejemplo en el modelo restringido se calculan $\hat{\beta}_1^R$ y $\hat{\beta}_2^R$ y finalmente se calcula $\hat{\beta}_3^R = 1 - \hat{\beta}_2^R$.
En este modelo restringido estimado por MCO se calcula la $SCR = \hat{u}'_R\hat{u}_R$. Si escribimos el MR en términos matriciales

$$Y^\star = X^\star\beta^\star + u^\star$$

entonces

$$\hat{u}'_R\hat{u}_R = Y^{\star\prime}Y^\star - \hat{\beta}_R^{\star\prime} X^{\star\prime} Y^\star$$

donde $Y^\star$ y $X^\star$ son las variables que quedan en el modelo restringido.

- Ejemplo: Contraste de significatividad conjunta: $H_0 : \beta_2 = \beta_3 = \ldots = \beta_K = 0$. Para esta hipótesis el modelo restringido es

$$Y_t = \beta_1 + u_t$$

si estimamos el MR por MCO obtenemos:

$$min_{\hat{\beta}_1} \sum \hat{u}_t^2 = min_{\hat{\beta}_1} \sum (Y_t - \hat{\beta}_1)^2$$

$$\frac{\partial \sum \hat{u}_t^2}{\partial \hat{\beta}_1} = -2\sum(Y_t - \hat{\beta}_1) = 0 \longrightarrow \hat{\beta}_1^R = \bar{Y}$$

de donde

$$\begin{aligned} \hat{u}'_R\hat{u}_R &= \sum(Y_t - \hat{Y}_t)^2 = \\ &= \sum(Y_t - \hat{\beta}_1^R)^2 = \sum(Y_t - \bar{Y})^2 = SCT \end{aligned}$$

Así

$$\frac{\hat{u}'_r\hat{u}_r - \hat{u}'\hat{u}/q}{\hat{u}'\hat{u}/(T-k)} = \frac{SCT - SCR}{SCR/T - K}$$

dividiendo el numerador y el denominador de entre $SCT = y'y$ obtenemos.

$$F = \frac{(\hat{u}'_R\hat{u}_R - \hat{u}'\hat{u})/q}{\hat{u}'\hat{u}/T - K} = \frac{R^2/K - 1}{(1 - R^2/T - K} \overset{H_0}{\sim} F(K-1, T-K)$$

estadístico que coincide con el obtenido para el contraste de significatividad conjunta.

2.7. Predicción

Aunque pueda considerarse que la obtención de un buen conjunto de estimaciones es el objetivo principal de la Econometría, a menudo también tiene gran importancia el logro de unas predicciones precisas.

Supongamos que con T observaciones se ha estimado el modelo:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \ldots + \beta_K X_{Kt} + u_t.$$

Dada una nueva observación de las variables explicativas,

$$X'_p = \left[\begin{array}{cccc} 1 & X_{2p} & \cdots & X_{Kp} \end{array} \right] \qquad p \notin \{1, 2, \ldots, T\}$$

se puede utilizar el modelo estimado por MCO para predecir el valor que tendrá la variable endógena (desconocido en ese momento).

Dado el modelo de regresión, la ecuación para Y_p es:

$$Y_p = \beta_1 + \beta_2 X_{2p} + \ldots + \beta_K X_{Kp} + u_p$$

Para abreviar, utilizaremos la expresión vectorial:

$$Y_p = X'_p \beta + u_p$$

Dada la información muestral disponible (no conocemos β ni u_p) la **predicción por punto de $\boldsymbol{Y_p}$** es:

$$\hat{Y}_p = X'_p \hat{\beta}_{MCO}$$

O lo que es lo mismo:

$$\hat{Y}_p = \hat{\beta}_1 + \hat{\beta}_2 X_{2p} + \ldots + \hat{\beta}_K X_{Kp}.$$

Hay cuatro fuentes potenciales de error al realizar una predicción:

a) El error de especificación. El modelo de regresión en que nos basamos puede ser incorrecto: pueden faltar variables explicativas que afectan de manera clave a Y, puede que la forma funcional propuesta no sea correcta, puede que se incumpla alguna hipótesis básica, etc.

b) Error en los valores de X_p. La predicción se hace para unos valores dados de X_p, pero estos pueden ser desconocidos en el momento en que se hace la predicción.

c) El error muestral. No hay más remedio que usar $\hat{\beta}$ en vez de los valores verdaderos β para hacer la predicción.

d) El error aleatorio. Y_p dependerá de u_p, la perturbación aleatoria (desconocida) correspondiente a esa observación. Cuanto más diferente sea de cero, mayor será este error.

Dadas todas estas fuentes de incertidumbre a la hora de predecir Y, es muy recomendable que la predicción puntual de Y se acompañe con una medida de lo precisa que esperamos que sea esa predicción. En esto consiste la predicción por intervalo.

• **Predicción por intervalo del valor de la variable endógena**

Es muy difícil que el valor predicho para Y_p, $\hat{Y}_p$ coincida con el valor real. Si la predicción por punto se hace para el mes siguiente, o para el año siguiente, llegará un momento en que conoceremos el error cometido. Este error se denomina **error de predicción** y es igual a

$$e_p = Y_p - \hat{Y}_p$$

En el momento en que hacemos la predicción, tenemos cierta información sobre e_p, ya que es una variable aleatoria con una distribución conocida. En concreto,

$$e_p \sim N(0, \sigma^2(1 + X_p'\,(X'X)^{-1}\,X_p))$$

<u>Demostración:</u>

$$\begin{aligned} e_p &= Y_p - \hat{Y}_p = X_p'\,\beta + u_p - X_p'\,\hat{\beta} = \\ &= u_p - X_p'\,(\hat{\beta} - \beta) \end{aligned} \tag{2.22}$$

Buscamos su distribución. Si u_p es normal el estimador MCO dado que es lineal en la perturbación también lo será y por tanto el error de predicción también lo es. En cuanto a su media y varianza:

$$E(e_p) = E\left[u_p - X_p'\,(\hat{\beta} - \beta)\right] = 0 - X_p'\,(\beta - \beta) = 0$$

$$\begin{aligned} Var(e_p) &= E\,[e_p - E(e_p)]\,[e_p - E(e_p)]' = \\ &= E\left(e_p\,e_p'\right) = \\ &= E\left[\left(u_p - X_p'\,(\hat{\beta} - \beta)\right)\left(u_p - X_p'\,(\hat{\beta} - \beta)\right)'\right] = \\ &= E\left[u_p\,u_p'\right] + E\left[X_p'\,(\hat{\beta} - \beta)\,(\hat{\beta} - \beta)'\,X_p\right] - \\ &\quad -2X_p'\,E\left[\,(\hat{\beta} - \beta)\,u_p'\right] = \\ &= E\left(u_p^2\right) + X_p'\,E\left[\,(\hat{\beta} - \beta)\,(\hat{\beta} - \beta)'\right]X_p - \\ &\quad -2X_p'\,E\left[(X'X)^{-1}\,X'\,u\,u_p\right] = \\ &= \sigma^2 + \sigma^2 X_p'\,(X'X)^{-1}\,X_p - 0 = \\ &= \sigma^2\left(1 + X_p'\,(X'X)^{-1}\,X_p\right) \end{aligned}$$

Por tanto:

$$e_p \sim N(0, \sigma^2\left(1 + X_p'\,(X'X)^{-1}\,X_p\right))$$

Tipificando el error de predicción queda:

$$\frac{e_p - 0}{\sigma\sqrt{1 + X_p'\,(X'X)^{-1}\,X_p}} \sim N(0,1)$$

El problema es que σ^2 es desconocida. Utilizando que e_p y $\hat{\sigma}^2$ son variables aleatorias independientes, y el Resultado 6 del Anexo 4 obtenemos:

$$\frac{\dfrac{e_p}{\sigma\sqrt{1 + X_p'\,(X'X)^{-1}\,X_p}}}{\sqrt{\dfrac{\hat{\sigma}^2(T-K)}{\sigma^2}\dfrac{1}{T-K}}} \sim t_{(T-K)}$$

Simplificando:

$$\frac{e_p}{\hat{\sigma}\sqrt{1 + X_p' \, (\, X'X \,)^{-1} \, X_p}} \sim t_{(T-K)}$$

De hecho el denominador final es $\hat{\sigma}_{e_p}$ (la desviación estimada del error de predicción). Tras sustituir $e_p = Y_p - \hat{Y}_p$, se puede utilizar dicha distribución para obtener el siguiente intervalo de predicción para la variable endógena:

$$Pr\left[-\, t_{\frac{\alpha}{2}(T-K)} \leq \frac{Y_p - \hat{Y}_p}{\hat{\sigma}_{e_p}} \leq t_{\frac{\alpha}{2}(T-K)}\right] = 1 - \alpha$$

$$Pr\left[\hat{Y}_p - \, t_{\frac{\alpha}{2}(T-K)} \cdot \hat{\sigma}_{e_p} \leq Y_p \leq \hat{Y}_p + \, t_{\frac{\alpha}{2}(T-K)} \cdot \hat{\sigma}_{e_p}\right] = 1 - \alpha$$

$$IC_{1-\alpha}(Y_p) = \left(\hat{Y}_p - \, t_{\frac{\alpha}{2}(T-K)} \, \hat{\sigma}_{e_p} \; , \; \hat{Y}_p + \, t_{\frac{\alpha}{2}(T-K)} \, \hat{\sigma}_{e_p} \right)$$

2.8. Estimación del MRLG con GRETL: principales resultados, forma funcional, predicción.

• Instrucciones para realizar una estimación **MCO**:
Una vez abierto el fichero de datos con el que vamos a trabajar, vamos a

Modelo → Mínimos Cuadrados Ordinarios
Aparecerá una ventana para especificar la parte sistemática del modelo donde debemos:

Seleccionar la variable dependiente pinchando a la izquierda sobre ella y a continuación pinchar en la derecha *Elegir*
Seleccionar las variables independientes pinchando a la izquierda sobre ella-s y a continuación pinchar en la derecha *Añadir*
Para obtener los resultados de la estimación MCO pinchar en *Aceptar*

En esta ventana aparecerán los resultados básicos de la estimación del modelo. Los resultados que Gretl nos devuelve muestran entre otros estadísticos la estimación de los parámetros de la recta de ajuste, sus desviaciones típicas y estadísticos de significatividad individual. Además podemos hacer contraste de restricciones lineales en la pestaña correspondiente, guardar variables como los residuos, las estimaciones de la variable endógena etc y realizar gráficos.

Los resultados los podemos guardar como texto plano de la manera habitual o como icono con *Archivo → Guardar como icono*

En la pantalla de resultados de la estimación aparecen en la barra de menú otros estadísticos o resultados que pueden ser de interés, por ejemplo:

- Podemos hacer gráficos de interés: En la opción *Gráficos* podemos hacer gráficos que nos ayudan a interpretar los resultados de la estimación, por ejemplo

  Gráficos → Gráfico de la variable estimada y observada

  Gráficos → Gráfico de residuos → contra alguna de las variables explicativas del modelo

- Guardar variables asociadas a la regresión: Podemos ver y guardar los valores de $\hat{Y}$, $\hat{u}$ y otros resultados de utilidad para ello:

  - Ver los valores: Pinchar en *Análisis* → *Mostrar variable* y seleccionar *observada, estimada o residuos* según nuestro interés.
  - Guardar los valores: Pinchar en *Guardar* → *seleccionar la variable de interés.*

  Gretl utiliza por defecto la denominación *yhat*, *uhat* para designar a la variable endógena estimada y a los residuos, respectivamente y en la descripción de la variable indicará por ejemplo para *uhat: residuos del modelo 1*, lo cual resulta muy útil pues en general trabajaremos con varios modelos a la vez y hay que distinguir claramente las variables de cada uno.

- en *Análisis* encontramos la *Matriz de varianzas y covarianzas de los coeficientes estimados*, podemos hacer *Predicciones* u obtener *Intervalos de confianza para los coeficientes*

- en *Contrastes* podemos *Omitir u añadir variables, contrastar combinaciones lineales o restricciones lineales* además podremos realizar contrastes sobre los residuos, de los cuales nos ocuparemos en el último tema del curso.

2.9. Anexos

ANEXO 1.- Distintas expresiones de SCT, SCE y SCR

$$\begin{aligned}
SCT &= \sum y_t^2 = y'y = \\
&= \sum (Y_t - \bar{Y})^2 = \\
&= \sum Y_t^2 - T\bar{Y}^2 = Y'Y - T\bar{Y}^2
\end{aligned}$$

$$\begin{aligned}
SCE &= \sum \hat{y}_t^2 = \sum (\hat{Y}_t - \bar{\hat{Y}})^2 = \\
&= \sum \hat{Y}_t^2 - T\bar{\hat{Y}}^2 = \\
&= \hat{Y}'\hat{Y} - T\bar{\hat{Y}}^2 = \\
&= (X\hat{\beta})'(X\hat{\beta}) - T\bar{\hat{Y}}^2 = \\
&= \hat{\beta}' \underbrace{X'X\hat{\beta}}_{=X'Y} - T\bar{Y}^2 = \\
&= \hat{\beta}'X'Y - T\bar{\hat{Y}}^2 = \\
&= \hat{y}'\hat{y} = \\
&= \hat{\beta}^{*'}x'x\hat{\beta}^* = \\
&= \hat{\beta}^{*'}x'y
\end{aligned}$$

$$\begin{aligned}
SCR &= \sum \hat{u}_t^2 = \hat{u}'\hat{u} = \\
&= (Y - \hat{Y})'(Y - \hat{Y}) = (Y - X\hat{\beta})'(Y - X\hat{\beta}) = \\
&= Y'Y - \hat{\beta}'X'Y - Y'X\hat{\beta} + \hat{\beta}'X'X\hat{\beta} = \\
&= Y'Y - Y'X(X'X)^{-1}X'Y - \\
&\quad Y'X(X'X)^{-1}X'Y + \\
&\quad Y'X(X'X)^{-1}X'X(X'X)^{-1}X'Y = \\
&= Y'Y - Y'X(X'X)^{-1}X'Y = \\
&= Y'Y - \hat{\beta}'X'X\hat{\beta} = \\
&= Y'Y - \hat{\beta}'X'Y = \\
&= y'y - \hat{\beta}^{*'}x'y
\end{aligned}$$

ANEXO 2. Demostración de la insesgadez de $\hat{\sigma}^2$

1. Propiedades de los residuos MCO

Los residuos MCO se pueden escribir en función de las perturbaciones:

$$\begin{aligned}
\hat{u} &= Y - \hat{Y} = \\
&= Y - X\hat{\beta} = \\
&= Y - X(X'X)^{-1}X'Y = \\
&= [I_T - X(X'X)^{-1}X']Y = \\
&= [I_T - X(X'X)^{-1}X'](X\beta + u) = \\
&= X\beta - X(X'X)^{-1}X'X\beta + [I_T - X(X'X)^{-1}X']u = \\
&= \underbrace{[I_T - X(X'X)^{-1}X']}_{=M} u = Mu
\end{aligned}$$

Las características de la matriz M son:

- Es simétrica: $M = M'$

$$\begin{aligned}
M' &= [I_T - X(X'X)^{-1}X']' \\
&= I_T' - (X(X'X)^{-1}X')' \\
&= I_T - (X')'[(X'X)^{-1}]'X' \\
&= I_T - X(X'X)^{-1}X' \\
&= M
\end{aligned}$$

- Es idempotente: $MM = M$

$$\begin{aligned}
MM &= [I_T - X(X'X)^{-1}X'][I_T - X(X'X)^{-1}X'] = \\
&= I_T - X(X'X)^{-1}X' - X(X'X)^{-1}X' + \\
&\quad X\underbrace{(X'X)^{-1}X'X}_{=I_K}(X'X)^{-1}X' = \\
&= I_T - X(X'X)^{-1}X' = M
\end{aligned}$$

- rg(M)= tr(M)= $T - K$

$$\begin{aligned}
tr(M) &= tr[I_T - X(X'X)^{-1}X'] \\
&= tr(I_T) - tr[X(X'X)^{-1}X'] \\
&= tr(I_T) - tr[(X'X)^{-1}X'X] \\
&= tr(I_T) - tr(I_K) \\
&= T - K
\end{aligned}$$

- M es ortogonal a X: $MX = 0$

$$\begin{aligned}
MX &= [I_T - X(X'X)^{-1}X']X \\
&= X - X(X'X)^{-1}X'X = 0
\end{aligned}$$

2. Demostración: $E(\hat{\sigma}^2) = \sigma^2$

Como

$$\begin{aligned} E(\hat{u}'\hat{u}) &= E(u'Mu) = \\ &= E(tr(u'Mu)) = E(tr(Muu')) = \\ &= tr(E(Muu')) = tr(ME(uu')) = \\ &= tr(M\sigma^2 I_T) = \\ &= \sigma^2 tr(M) = \sigma^2(T-K) \end{aligned}$$

se puede demostrar fácilmente que $\hat{\sigma}^2$ es un estimador insesgado:

$$E(\hat{\sigma}^2) = E\left(\frac{\hat{u}'\hat{u}}{T-K}\right) = \frac{E(\hat{u}'\hat{u})}{T-K} = \frac{\sigma^2(T-K)}{T-K} = \sigma^2$$

Y por tanto podremos utilizarlo como el estimador apropiado de la varianza de la perturbación.

ANEXO 3. Recordatorio estadístico: contrastes de hipótesis

Una **hipótesis** es una determinada afirmación o conjetura sobre algunos parámetros de la función de densidad de una variable aleatoria. Por ejemplo, en el Tema 2 se ha visto que en el modelo de regresión lineal simple, la distribución del estimador MCO de la pendiente es:

$$\hat{\beta} \sim N\left(\beta, \frac{\sigma^2}{\sum(X_t - \bar{X})^2}\right)$$

Un ejemplo de hipótesis sería decir que β –que es desconocido– es igual a 0,8.

La realización de los contrastes de hipótesis tiene tres fases:

a) Formular dos hipótesis excluyentes.

b) Obtener un estadístico de contraste e identificar su distribución muestral.

c) Fijar una regla de decisión y elegir una de las dos hipótesis.

Siguiendo nuestro ejemplo, las dos hipótesis excluyentes serían la **hipótesis nula**, $H_0 : \beta = 0{,}8$, que es una afirmación concreta sobre la función de densidad, y la **hipótesis alternativa**, $H_A : \beta \neq 0{,}8$, que en Econometría será casi siempre la complementaria de la hipótesis nula.

Si suponemos que σ^2 es conocida, la fase dos se consigue simplemente tipificando la distribución de $\hat{\beta}$:

$$z = \frac{\hat{\beta} - \beta}{\sqrt{\frac{\sigma^2}{\sum(X_t - \bar{X})^2}}} \sim N(0,1)$$

A la izquierda está el estadístico de contraste, z, y a la derecha su distribución muestral.

La regla de decisión divide todos los valores que puede tomar el estadístico de contraste en dos regiones:

- **Región crítica**: es el conjunto de valores del estadístico para los cuales el mecanismo de contraste recomienda que se rechace la H_0.
- **Región de aceptación**: es el conjunto de valores del estadístico para los cuales el mecanismo de contraste recomienda que no se rechace la H_0.

En el ejemplo, si $H_0 : \beta = 0{,}8$ es cierta, el estadístico z queda:

$$z = \frac{\hat{\beta} - 0,8}{\sqrt{\frac{\sigma^2}{\sum(X_t - \bar{X})^2}}} \sim N(0,1)$$

y los valores cercanos a cero son los que con mayor probabilidad tomará z. En cambio, si H_0 es falsa, la media de z no será cero, y los valores de z cercanos a cero son menos probables. Consecuentemente, consideraremos que:

- los valores cercanos a cero son evidencia en favor de la H_0, así que estarán en la región de aceptación.
- los valores alejados de cero son evidencia en favor de la H_A, así que estarán en la región crítica.

Dependiendo del valor que tome z, nos decidiremos por la H_0 o por la H_A.

Errores de contraste
El contraste puede llevarnos a tomar una decisión correcta, o a caer en un error, que puede ser de dos tipos:

- **error de Tipo I**: rechazar H_0 cuando es verdadera.
- **error de Tipo II**: no rechazar H_0 cuando es falsa.

El **nivel de significación** o **tamaño del contraste** es la probabilidad máxima de cometer error de tipo I.

Ejemplo: en un juicio, las dos hipótesis excluyentes pueden ser:
$H_0 \longrightarrow$ el defensor: el acusado es inocente.
$H_A \longrightarrow$ el fiscal: el acusado es culpable.

Error de Tipo I: probabilidad de condenar a un inocente; viene dada por el nivel de significación.

Error de Tipo II: probabilidad de absolver a un culpable. En la contrastación estadística se desconoce la probabilidad exacta de este tipo de error. Depende de cuál sea en realidad el valor de β.

La función de **potencia del contraste** es la probabilidad de rechazar la H_0 cuando sea falsa. O sea, es $1 - Pr$(Error de tipo II).

ANEXO 4: Recordatorio de Distribuciones asociadas a la Normal

- Resultado 1:
  Sea X una variable aleatoria tal que $X \sim N(0,1)$ entonces $X^2 \sim \chi^2_{(1)}$
- Resultado 2:
  Sea X v.a. n-multivariante tal que $X \sim N(0,I)$ entonces $X'X \sim \chi^2_{(n)}$ ya que es la suma de n-variables independientes tipificadas al cuadrado.
- Resultado 3:
  Sea $X \sim N(0,\sigma^2 I)$, las variables son independientes, con igual varianza σ^2, entonces
  $$\frac{X_1^2}{\sigma^2} + \ldots + \frac{X_n^2}{\sigma^2} \sim \chi^2_{(n)} \longrightarrow \frac{X'X}{\sigma^2} \sim \chi^2_{(n)}$$
- Resultado 4:
  Sea $X \sim N(0,\sum)$ donde las variables no son independientes, entonces, $X'(\sum)^{-1}X \sim \chi^2_{(n)}$
- Resultado 5:
  Sean Z y W dos v.a. independientes tal que $Z \sim \chi^2_{(m)}$ y $W \sim \chi^2_{(n)}$ entonces $Z+W \sim \chi^2_{(m+n)}$
- Resultado 6: Definición de la **t de Student**
  Sean X y Z dos v.a. independientes al que $X \sim N(0,1)$ y $Z \sim \chi^2_{(m)}$, entonces, $\frac{X}{\sqrt{\frac{Z}{m}}} \sim t_{(m)}$
- Resultado 7: Definición de la **F de Snedecor**
  Sean Z y W dos v.a. independientes tal que $Z \sim \chi^2_{(m)}$ y $W \sim \chi^2_{(n)}$, entonces $\frac{Z/m}{W/n} \sim F_{(m,n)}$
- Resultado 8:
  Si $F \sim F_{(1,n)}$ entonces $t^2 \sim F_{(1,n)}$ o lo que es igual $\sqrt{F} \sim t_{(n)}$
- Lema 1:
  Sea A una matriz idempotente, entonces $rg(A) = tr(A)$
- Lema 2:
  Sea $X \sim NI(0,I)$ y A una matriz idempotente con $rg(A) = m$, entonces $X'AX \sim \chi^2_{(m)}$
- Lema 3:
  Sea $Z \sim N(0,\sigma^2 I_T)$, y sea A una matriz idempotente y simétrica de orden $(T \times T)$ tal que $rg(A) = m$, entonces $\frac{X'AX}{\sigma^2} \sim \chi^2_{(m)}$
- Lema 4:
  Sean $X \sim NI(0,I)$, A una matriz simétrica e idempotente de orden $(T \times T)$ con $rg(A) = m$, B una matriz simétrica e idempotente de orden $(T \times T)$ con $rg(B) = n$, entonces
  $$X'AX \sim \chi^2_{(m)}$$
  $$X'BX \sim \chi^2_{(n)}$$
  Y si $AB = 0$, $X'AX$ y $X'BX$ son independientes.
- Lema 5:
  Sean.
  $X \sim NI(0,I)$, A una matriz simétrica e idempotente de rango m, B una matriz simétrica e idempotente de rango n, $AB = 0$ entonces se verifica que:
  $$\frac{X'AX/m}{X'BX/n} \sim F_{(m,n)}$$

ANEXO 5. Estadístico general para hacer inferencia. Obtención de la distribución
El siguiente paso será hallar un estadístico que nos permita contrastar cualquiera de estas hipótesis frente a

$$H_a : R\beta \neq r$$

Queremos utilizar la expresión $R\beta = r$ para hacer inferencia, por lo tanto necesitamos conocer la distribución de:

$$R\hat{\beta} = r$$

que seguirá una distribución normal dado que:

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

Buscamos la media y matriz de varianzas y covarianzas de la forma $R\hat{\beta}$

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

$$\begin{aligned} Var(R\hat{\beta}) &= E\left[R\hat{\beta} - E(R\hat{\beta})\right]\left[R\hat{\beta} - E(R\hat{\beta})\right]' \\ &= E\left[R\hat{\beta} - R\beta\right]\left[R\hat{\beta} - R\beta\right]' \\ &= E\left[R(\hat{\beta} - \beta)\right]\left[R(\hat{\beta} - \beta)\right]' \\ &= E\left[R(\hat{\beta} - \beta)(\hat{\beta} - \beta)'R'\right] \\ &= R\underbrace{E(\hat{\beta} - \beta)(\hat{\beta} - \beta)'}_{=Var(\hat{\beta})=\sigma^2(X'X)^{-1}}R' \\ &= \sigma^2 R\,(X'X)^{-1}\,R' \end{aligned}$$

Puesto que $\hat{\beta}$ es normal multivariante y R es una matriz de constantes conocidas,

$$R\hat{\beta} \sim N\left(R\beta, \sigma^2 R\,(X'X)^{-1}\,R'\right)$$

- **Estadístico de contraste general ($\boldsymbol{\sigma^2}$ conocida)**

Restando su media a la variable aleatoria $R\hat{\beta}$ se obtiene:

$$R\hat{\beta} - R\beta \sim N\left(0, \sigma^2 R\,(X'X)^{-1}\,R'\right)$$

Si la hipótesis nula es cierta, $R\beta$ es igual a r, y haciendo esta sustitución en la expresión anterior:

$$\underset{(q\times 1)}{\left(R\hat{\beta} - r\right)} \overset{H_0}{\sim} N\left(0, \sigma^2 R\,(X'X)^{-1}\,R'\right)$$

La matriz de varianzas y covarianzas es definida positiva, así que aplicando la generalización del resultado 4 del Anexo 4 obtenemos que:

$$\left(R\hat{\beta} - r\right)'\left[\sigma^2 R\,(X'X)^{-1}\,R'\right]^{-1}\left(R\hat{\beta} - r\right) \overset{H_0}{\sim} \chi^2_{(q)} \qquad (2.23)$$

En el caso (nada realista) de que σ^2 sea conocida, se puede utilizar este estadístico para contrastar cualquier hipótesis nula planteada de la forma general.

La idea principal es que, dado que la hipótesis nula afirma que $R\beta$ y r son iguales, esperaremos que la diferencia $R\hat{\beta} - r$ sea pequeña si la hipótesis nula es cierta y al contrario, esperaremos que $R\hat{\beta} - r$ sea grande (en valor absoluto) si la hipótesis nula es falsa. Por esta razón, valores del estadístico próximos a 0 formarán la región de aceptación, y valores grandes del estadístico formarán la región crítica.

Además cuanto mayor sea la $Var(\hat{\beta})$, mayor será la $Var(R\hat{\beta})$ y más fácil será que la diferencia $R\hat{\beta} - r$ sea grande aunque H_0 sea cierta; por ello se pondera el estadístico con la matriz de varianzas y covarianzas.

El criterio de contraste es rechazar H_0 con un nivel de significación α cuando

$$\left(R\hat{\beta} - r\right)' \left[\sigma^2 R\left(X'X\right)^{-1} R'\right]^{-1} \left(R\hat{\beta} - r\right) > \chi^2_{\alpha(q)}$$

y no rechazarla en caso contrario.

• Estadístico de contraste general($\boldsymbol{\sigma^2}$ desconocida)

El problema para la aplicación del contraste descrito en el apartado anterior es que habitualmente no conocemos σ^2. En este caso, se debe sustituir esta varianza en el estadístico anterior por un estimador suyo, por ejemplo $\hat{\sigma}^2 = \frac{\hat{u}'\hat{u}}{T-K}$. Pero esta sustitución cambia la distribución que sigue el estadístico, que pasará a ser una F-Snedecor. Vamos a derivar este estadístico a continuación. Recordamos la definición de F-Snedecor:

$$\frac{\chi_{(n)}/n}{\chi_{(m)}/m} \sim F_{(n,m)}$$

donde necesitamos independencia entre numerador y denominador.

Disponemos de:

$$\left(R\hat{\beta} - r\right)' \left[\sigma^2 R\left(X'X\right)^{-1} R'\right]^{-1} \left(R\hat{\beta} - r\right) \sim \chi^2_{(q)}$$

y de:

$$\frac{\hat{u}'\hat{u}}{\sigma^2} \sim \chi^2_{(T_K)}$$

o lo que es igual

$$\frac{\hat{\sigma}^2(T-k)}{\sigma^2} \sim \chi^2_{(T-K)}$$

Ambas expresiones son independientes por lo que podemos construir el siguiente estadístico: (Demostración en el Anexo 4).

$$\frac{\left(R\hat{\beta} - r\right)' \left[\sigma^2 R\left(X'X\right)^{-1} R'\right]^{-1} \left(R\hat{\beta} - r\right)/q}{\frac{\hat{\sigma}^2(T-K)}{\sigma^2}/(T-K)} \overset{H_0}{\sim} \mathcal{F}_{(q,T-K)}$$

Simplificando obtenemos el siguiente estadístico, que usaremos en los contrastes para el caso en que σ^2 sea desconocida, y que llamaremos F:

$$\frac{\left(R\hat{\beta} - r\right)' \left(R\left(X'X\right)^{-1} R'\right)^{-1} \left(R\hat{\beta} - r\right)/q}{\hat{\sigma}^2} \overset{H_0}{\sim} \mathcal{F}_{(q,T-K)} \tag{2.24}$$

Este será el estadístico general que utilizaremos para contrastar hipótesis lineales cuando σ^2 sea desconocida. La idea del contraste es la misma que la anterior; rechazaremos la hipótesis nula para valores grandes del estadístico, ya que si H_0 no es cierta, la diferencia $R\hat{\beta} - r$ será grande y F tomará valores altos. El criterio de contraste es por tanto el de rechazar H_0 con un nivel de significación α cuando

$$\frac{\left(R\hat{\beta} - r\right)' \left(R\,(X'X)^{-1}R'\right)^{-1}\left(R\hat{\beta} - r\right)/q}{\hat{\sigma}^2} > \mathcal{F}_{\alpha(q,T-K)}$$

y no rechazarla en caso contrario.

• **Demostración de que $\hat{\beta}$ y $\hat{\sigma}^2$ son independientes**

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

Aplicando el Resultado 2 del Anexo 4 se llega a una χ^2:

$$(\hat{\beta}-\beta)'\left(\sigma^2\,(X'X)^{-1}\right)^{-1}(\hat{\beta}-\beta) \sim \chi^2_{(K)}$$

$$\frac{(\hat{\beta}-\beta)'X'X\,(\hat{\beta}-\beta)}{\sigma^2} \sim \chi^2_{(K)} \tag{2.25}$$

El resultado de independencia entre dos estadísticos que siguen una χ^2 es el Resultado 4. Para poder aplicarlo,

$$\frac{\sigma^2(T-K)}{\sigma^2} \sim \chi^2_{(T-K)} \tag{2.26}$$

y (2.25) han de expresarse como formas cuadráticas de un mismo vector, u. Para (2.26), la forma cuadrática era $\frac{u'Mu}{\sigma^2} \sim \chi^2_{(T-K)}$. Lo mismo se consigue para (2.25), sin más que sustituir $(\hat{\beta}-\beta) = (X'X)^{-1}X'u$ en el estadístico anterior:

$$\frac{(\hat{\beta}-\beta)'X'X\,(\hat{\beta}-\beta)}{\sigma^2} =$$

$$\frac{u'X\,(X'X)^{-1}X'X\,(X'X)^{-1}X'u}{\sigma^2} =$$

$$= \frac{u'\left[X\,(X'X)^{-1}X'\right]u}{\sigma^2} = \frac{u'\,(I-M)\,u}{\sigma^2} \sim \chi^2_{(K)}$$

$I - M$ es simétrica porque M es simétrica. $I - M$ es idempotente porque M es idempotente y entonces $(I-M)(I-M) = I - M - M + M = I - M$.

Además, $rg\,(I-M) = tr\,(I-M) = tr(I_T) - tr(M) = T - (T-K) = K$.

Con todo esto, ya estamos en condiciones de usar el Resultado 4: como $M(I-M) = 0$, las formas cuadráticas (2.26) y (2.25) son independientes.

Las únicas variables aleatorias en (2.26) y (2.25) son $\hat{\sigma}^2$ y $\widehat{\beta}$ respectivamente, así que son independientes. Por lo tanto, cualquier componente del vector $\widehat{\beta}$ es independiente de $\hat{\sigma}^2$.

Tema 3

Variables explicativas cualitativas

3.1. Variables Ficticias: definición y utilización en el MRLG

A lo largo del curso únicamente se han especificado modelos con variables de naturaleza cuantitativa, es decir, aquéllas que toman valores numéricos. Sin embargo, las variables también pueden ser cualitativas, es decir, pueden tomar valores no numéricos como categorías, clases o atributos. Por ejemplo, son variables cualitativas el género de las personas, el estado civil, la raza, el pertenecer a diferentes zonas geográficas, momentos históricos, estaciones del año, etc. De esta forma, el salario de los trabajadores puede depender del género de los mismos; la tasa de criminalidad puede venir determinada por la zona geográfica de residencia de los individuos; el PIB de los países puede estar influenciado por determinados acontecimientos históricos como las guerras; las ventas de un determinado producto pueden ser significativamente distintas en función de la época del año, etc.

En este tema, aunque seguimos manteniendo que la variable dependiente es cuantitativa, vamos a considerar que ésta puede venir explicada por variables cualitativas y/o cuantitativas.

Dado que las categorías de las variables no son directamente cuantificables, las vamos a cuantificar construyendo unas variables artificiales llamadas ficticias, binarias o dummies, que son numéricas. Estas variables toman arbitrariamente el valor 1 si la categoría está presente en el individuo y 0 en caso contrario[1].

$$D_i = \begin{cases} 1 & \text{si la categoría está presente} \\ 0 & \text{en otro caso} \end{cases}$$

Por ejemplo si queremos estudiar la dependencia del salario (W_i) con respecto al sexo del individuo definiremos dos variables ficticias:

$$S_{1i} = \begin{cases} 1 & \text{si el individuo i es hombre} \\ 0 & \text{en otro caso} \end{cases}$$

$$S_{2i} = \begin{cases} 1 & \text{si el individuo i es mujer} \\ 0 & \text{en otro caso} \end{cases}$$

[1] Las variables ficticias pueden tomar dos valores cualesquiera, sin embargo, la interpretación de los coeficientes es más sencilla si se consideran los valores 0 y 1.

la variable sexo tiene dos categorías o estados de la naturaleza: hombre y mujer, para recogerlos utilizamos dos variables ficticias que dividen la muestra en dos clases hombres y mujeres, y asignamos un valor arbitrario a cada clase. Elegir los valores (0,1) es muy cómodo pero podríamos elegir otros por ejemplo

$$S_{1i} = \begin{cases} 1 & \text{si el individuo i es hombre} \\ 0 & \text{en otro caso} \end{cases}$$

$$S_{2i} = \begin{cases} 2 & \text{si el individuo i es mujer} \\ 0 & \text{en caso contrario} \end{cases}$$

Especificaremos la función de salario e interpretaremos la función y sus parámetros con un único conjunto de variables ficticias: Sexo. Después incluiremos otro nuevo conjunto: Nivel de educación y repetiremos el estudio.

3.1.1. Modelo que recoge sólo efectos cualitativos: comparando medias. Sólo un conjunto de variables ficticias.

Supongamos que tenemos datos de salarios entre hombres y mujeres y creemos que, en media, existen diferencias salariales entre estos dos grupos. Para contrastar que esto es cierto podemos recoger el efecto cualitativo sexo sobre el salario utilizando las variables ficticias:

$$S_{1i} = \begin{cases} 1 & \text{si el individuo i es hombre} \\ 0 & \text{en otro caso} \end{cases}$$

$$S_{2i} = \begin{cases} 1 & \text{si el individuo i es mujer} \\ 0 & \text{en otro caso} \end{cases}$$

y podemos especificar el siguiente modelo:

$$W_i = \alpha_1 S_{1i} + \alpha_2 S_{2i} + u_i \qquad i = 1, \ldots, N_H + N_M \tag{3.1}$$

siendo N_H el número de individuos varones y N_M el número de mujeres.

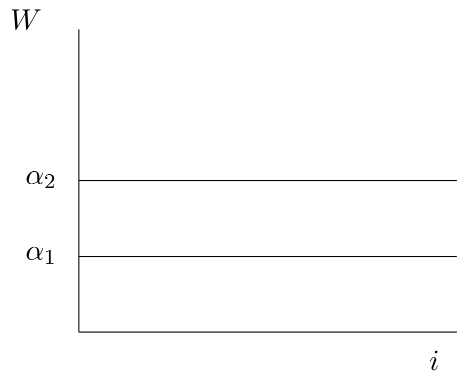
El modelo o función de salarios para cada grupo sería

$$\begin{array}{lll} W_i = \alpha_1 + u_i & i = 1, \ldots, N_H & \text{para los hombres} \\ W_i = \alpha_2 + u_i & i = 1, \ldots, N_M & \text{para las mujeres} \end{array}$$

de donde suponiendo $u_i \sim NID(0, \sigma^2)$

$$\alpha_1 = E(W_i/\text{hombre})$$
$$\alpha_2 = E(W_i/\text{mujer})$$

por tanto estos coeficientes recogen el salario medio dentro del grupo. Gráficamente:



La hipótesis de que no existe discriminación salarial por sexo es:

$$H_0 : \alpha_1 = \alpha_2 \qquad \Rightarrow H_0 : \alpha_1 - \alpha_2 = 0$$

que podemos contrastar con el estadístico F habitual, siendo $R\beta = r$:

$$\begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} = 0 \qquad \text{y} \qquad q = 1$$

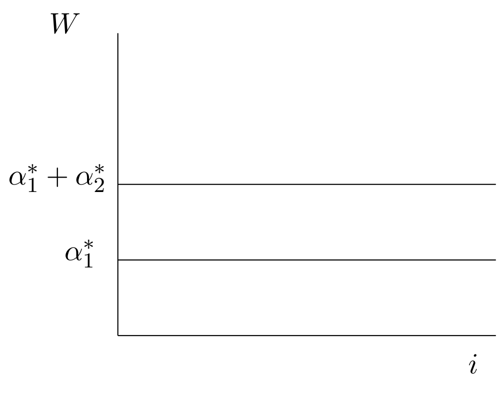
alternativamente podemos especificar el modelo (3.1) como.

$$W_i = \alpha_1^* + \alpha_2^* S_{2i} + u_i \qquad i = 1, \ldots, N_H + N_M \tag{3.2}$$

en este caso

$$\begin{aligned} \alpha_1^* &= E(W_i/\text{hombre}) \\ \alpha_1^* + \alpha_2^* &= E(W_i/\text{mujer}) \end{aligned}$$

y por tanto α_2^* recoge el diferencial en el salario medio entre hombres y mujeres. Gráficamente:



Si en este modelo queremos contrastar la existencia de discriminación salarial por sexo contrastaríamos una única hipótesis $H_0 : \alpha_2^* = 0$ con el estadístico t-habitual:

$$\frac{\hat{\alpha}_2^*}{\widehat{des}(\hat{\alpha}_2^*)} \overset{H_0}{\sim} t_{(N_H+N_M)-2}$$

Este contraste y el anterior son equivalentes, ya que $\alpha_1 = \alpha_1^*$ y $\alpha_2 = \alpha_1^* + \alpha_2^*$. No cabe duda que la estimación del modelo (3.2) facilita el contraste de discriminación salarial por sexo lo cual debería ser tenido en cuenta a la hora de proponer una especificación adecuada.

Lo que no debemos hacer es especificar el modelo:

$$W_i = \alpha_0 + \alpha_1 S_{1i} + \alpha_2 S_{2i} + u_i \qquad i = 1, \ldots, N_H + N_M \tag{3.3}$$

ya que en este caso existe multicolinealidad exacta y no podríamos estimar separadamente α_0, α_1 y α_2, notar que en (3.3):

$$\begin{aligned} E(W_i/\text{hombre}) &= \alpha_0 + \alpha_1 \\ E(W_i/\text{mujer}) &= \alpha_0 + \alpha_2 \end{aligned}$$

• <u>Estimación del modelo (3.1):</u>

$$W_i = \alpha_1 S_{1i} + \alpha_2 S_{2i} + u_i \qquad i = 1, \ldots, N_H + N_M$$

$$\begin{bmatrix} W_H \\ W_M \end{bmatrix} = \begin{bmatrix} i_H & 0 \\ 0 & i_M \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} + \begin{bmatrix} u_H \\ u_M \end{bmatrix} \Rightarrow Y = X\beta + u$$

$$\hat{\beta}_{MCO} = (X'X)^{-1}X'Y$$

$$\begin{bmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \end{bmatrix} = \left[\begin{bmatrix} i'_H & 0 \\ 0 & i'_M \end{bmatrix} \begin{bmatrix} i_H & 0 \\ 0 & i_M \end{bmatrix} \right]^{-1} \begin{bmatrix} i'_H & 0 \\ 0 & i'_M \end{bmatrix} \begin{bmatrix} W_H \\ W_M \end{bmatrix} =$$

$$= \begin{bmatrix} N_H & 0 \\ 0 & N_M \end{bmatrix}^{-1} \begin{bmatrix} \sum W_H \\ \sum W_M \end{bmatrix} = \begin{bmatrix} \sum W_H/N_H \\ \sum W_M/N_M \end{bmatrix} = \begin{bmatrix} \bar{W}_H \\ \bar{W}_M \end{bmatrix}$$

$$\hat{W}_i = \hat{\alpha}_1 S_{1i} + \hat{\alpha}_2 S_{2i} = \bar{W}_H S_{1i} + \bar{W}_M S_{2i}$$

Notación utilizada: W_H, W_M son vectores columna que recogen los salarios de hombres y mujeres, por tanto de orden $N_H \times 1$ y $N_M \times 1$, respectivamente. i_H, i_M son vectores de unos de tamaño $N_H \times 1$ y $N_M \times 1$ respectivamente.

Los mismos resultados se obtendrían si hubiésemos estimados las ecuaciones por separado en las dos ecuaciones a que da lugar el modelo (3.1):

$$W_i = \alpha_1 + u_i \quad i = 1, \ldots, N_H \qquad \text{y} \qquad W_i = \alpha_2 + u_i \quad i = 1, \ldots, N_H$$

ya que $E(u_i^2) = \sigma^2 \quad \forall i$ y la muestra de hombres es independiente de la muestra de mujeres.

• <u>Estimación del modelo (3.2):</u>

$$W_i = \alpha_1^* + \alpha_2^* S_{2i} + u_i \qquad i = 1, \ldots, N_H + N_M$$

$$\begin{bmatrix} W_H \\ W_M \end{bmatrix} = \begin{bmatrix} i_H & 0 \\ i_M & i_M \end{bmatrix} \begin{bmatrix} \alpha_1^* \\ \alpha_2^* \end{bmatrix} + \begin{bmatrix} u_H \\ u_M \end{bmatrix} \Rightarrow Y = X\beta + u$$

$$\hat{\beta}_{MCO} = (X'X)^{-1}X'Y$$

$$\begin{bmatrix} \hat{\alpha}_1^* \\ \hat{\alpha}_2^* \end{bmatrix} = \left[\begin{bmatrix} i'_H & i'_M \\ 0 & i'_M \end{bmatrix} \begin{bmatrix} i_H & 0 \\ i_M & i_M \end{bmatrix} \right]^{-1} \begin{bmatrix} i'_H & i'_M \\ 0 & i'_M \end{bmatrix} \begin{bmatrix} W_H \\ W_M \end{bmatrix} =$$

$$= \begin{bmatrix} N_H + N_M & N_M \\ N_M & N_M \end{bmatrix}^{-1} \begin{bmatrix} \sum W_H + \sum W_M \\ \sum W_M \end{bmatrix} = \begin{bmatrix} \bar{W}_H \\ \bar{W}_M - \bar{W}_H \end{bmatrix}$$

que sería el equivalente a estimar cada ecuación por separado, en las dos ecuaciones a las que da lugar el modelo (3.2):

$$W_i = \alpha_1^* + u_i \quad i = 1, \ldots, N_H \qquad \text{y} \qquad W_i = \alpha_1^* + \alpha_2^* + u_i \quad i = 1, \ldots, N_M$$

Una forma alternativa de contrastar la existencia de discriminación salarial por sexo en los modelos (3.1) y (3.2) sería utilizar el estadístico:

$$\frac{\hat{u}'_r \hat{u}_r - \hat{u}'\hat{u}/q}{\hat{u}'\hat{u}/N_H + N_M - K} \overset{H_0}{\sim} F_{(q, N_H + N_M - K)}$$

donde el modelo restringido en cualquiera de las dos especificaciones es $W_i = \mu + u_i$ del cual obtendríamos $SCR_r = \hat{u}'_r \hat{u}_r$ y $\hat{u}'\hat{u}$ la obtendríamos de realizar dos regresiones separadas, para hombres y para mujeres: $\hat{u}'\hat{u} = \hat{u}'_H \hat{u}_H + \hat{u}'_M \hat{u}_M$ que sería la que obtendríamos de estimar (3.1).

3.1.2. Dos o más conjuntos de variables ficticias

Supongamos que pensamos que en el nivel de salarios influye además del sexo el nivel de educación. Para recoger estos efectos podemos definir dos conjuntos de variables ficticias, sexo y educación, la primera con dos categorías o estados de la naturaleza y la segunda con tres, y recoger cada categoría o estado de la naturaleza con un variable ficticia. Así, definimos:

$$S_{1i} = \begin{cases} 1 & \text{si el individuo i es hombre} \\ 0 & \text{en otro caso} \end{cases} \qquad E_{1i} = \begin{cases} 1 & \text{si i tiene hasta estudios primarios} \\ 0 & \text{en otro caso} \end{cases}$$

$$S_{2i} = \begin{cases} 1 & \text{si el individuo i es mujer} \\ 0 & \text{en otro caso} \end{cases} \qquad E_{2i} = \begin{cases} 1 & \text{si i tiene hasta estudios secundarios} \\ 0 & \text{en otro caso} \end{cases}$$

$$E_{3i} = \begin{cases} 1 & \text{si i tiene hasta estudios universitarios} \\ 0 & \text{en otro caso} \end{cases}$$

siendo E_{ij} sucesos excluyentes. La especificación correspondiente es:

$$W_i = \mu + \alpha_2 S_{2i} + \beta_2 E_{2i} + \beta_3 E_{3i} + u_i \qquad i = 1, \ldots, N_H + N_M \tag{3.4}$$

donde para evitar problemas de multicolinealidad exacta hemos excluido una categoría de cada factor cualitativo. Podemos interpretar los parámetros de la forma siguiente:

$E(W_i/\text{hombre hasta estudios primarios}) = \mu$
$E(W_i/\text{hombre hasta estudios secundarios}) = \mu + \beta_2$
$E(W_i/\text{hombre hasta estudios universitarios}) = \mu + \beta_3$
$E(W_i/\text{mujer hasta estudios primarios}) = \mu + \alpha_2$
$E(W_i/\text{mujer hasta estudios secundarios}) = \mu + \alpha_2 + \beta_2$
$E(W_i/\text{mujer hasta estudios universitarios}) = \mu + \alpha_2 + \beta_3$

Esta información podemos resumirla en la siguiente tabla:

$E(W_i)$	E_{1i}	E_{2i}	E_{3i}
S_{1i}	μ	$\mu+\beta_2$	$\mu+\beta_3$
S_{2i}	$\mu+\alpha_2$	$\mu+\alpha_2+\beta_2$	$\mu+\alpha_2+\beta_3$

y podemos interpretar los parámetros como sigue:

μ	base de comparación.
α_2	efecto diferencial en el salario medio debido al factor sexo. Por tanto es el diferencial en el salario medio entre hombres y mujeres independientemente de su nivel de educación.
β_2	Efecto diferencial en el salario medio debido a tener un nivel de estudios secundarios. Por tanto es el diferencial en el salario medio, para hombres y mujeres, entre tener un nivel de estudios primarios y tener secundaria.
β_3	Efecto diferencial en el salario medio debido a tener un nivel de estudios universitarios. Por tanto es el diferencial en el salario medio, para hombres y mujeres, entre tener un nivel de estudios primarios y tener universidad.

En el modelo anterior se pueden realizar los siguientes contrastes:

a) En media, no hay discriminación salarial por sexo, $H_0 : \alpha_2 = 0$, mediante un contraste de significatividad individual.

b) En media, no hay discriminación salarial por razones de estudios: $H_0 : \beta_2 = \beta_3 = 0$.

c) En media, no hay diferencias salariales entre el grupo de individuos con estudios primarios y secundarios: $H_0 : \beta_2 = 0$

d) En media, no hay diferencias salariales entre el grupo de individuos con estudios primarios y universitarios: $H_0 : \beta_3 = 0$

e) En media, no hay diferencias salariales entre el grupo de individuos con estudios secundarios y universitarios: $H_0 : \beta_2 = \beta_3 \Leftrightarrow H_0 : \beta_2 - \beta_3 = 0$

En los casos 1, 3 y 4 $q = 1$ y podemos realizar el contraste con el estadístico t-habitual. En el caso 5 ocurre lo mismo $q = 1$ y construimos el estadístico t para la combinación lineal $(\beta_2 - \beta_3)$. En el caso 2 $q = 2$ y utilizamos la forma $R\beta = r$ con el estadístico F-habitual.

La matriz de regresores del modelo sería:

$$X = \begin{bmatrix} i_{N_1} & 0 & 0 & 0 \\ i_{N_2} & 0 & i_{N_2} & 0 \\ i_{N_3} & 0 & 0 & i_{N_3} \\ i_{N_4} & i_{N_4} & 0 & 0 \\ i_{N_5} & i_{N_5} & i_{N_5} & 0 \\ i_{N_6} & i_{N_6} & i_{N_6} & i_{N_6} \end{bmatrix}$$

donde i_{N_j} es un vector de unos de tamaño el número de individuos que cumplen las condiciones, por ejemplo i_{N_6} es un vector de unos de tamaño el número de mujeres con estudios universitarios.

Cuando existen dos o más conjuntos de variables ficticias lo que no debemos hacer es incluir todas las variables ficticias y un término independiente. En el caso anterior tenemos dos conjuntos con dos y tres estados de la naturaleza respectivamente, si proponemos la especificación:

$$W_i = \mu^* + \alpha_1^* S_{1i} + \alpha_2^* S_{2i} + \beta_1^* E_{1i} + \beta_2^* E_{2i} + \beta_3^* E_{3i} + u_i \qquad i = 1, \ldots, N_H + N_M$$

existiría multicolinealidad exacta en la matriz de regresores y no podríamos estimar separadamente ninguno de los coeficientes. La matriz de regresores es:

La matriz de regresores del modelo sería:

$$X = \begin{bmatrix} i_{N_1} & i_{N_1} & 0 & i_{N_1} & 0 & 0 \\ i_{N_2} & i_{N_2} & 0 & 0 & i_{N_2} & 0 \\ i_{N_3} & i_{N_3} & 0 & 0 & 0 & i_{N_3} \\ i_{N_4} & 0 & i_{N_4} & i_{N_4} & 0 & 0 \\ i_{N_5} & 0 & i_{N_5} & 0 & i_{N_5} & 0 \\ i_{N_6} & 0 & i_{N_6} & 0 & 0 & i_{N_6} \end{bmatrix} \Rightarrow rg(X) < K$$

3.1.3. Inclusión de variables cuantitativas

En cualquiera de los modelos anteriores puede incluirse una-s variable-s cuantitativas, por ejemplo si creemos que el salario depende no solo de sexo sino también del número de horas trabajadas, variable que denotamos como X_i propondremos:

$$W_i = \alpha_1 S_{1i} + \alpha_2 S_{2i} + \beta X_i + u_i \qquad i = 1, \ldots, N_H + N_M \tag{3.5}$$

Donde el coeficiente β se interpreta de la forma habitual, $\beta = \frac{E(W_i)}{\partial X_i}$. En forma matricial el modelo sería:

$$\begin{bmatrix} W_H \\ W_M \end{bmatrix} = \begin{bmatrix} i_H & 0 & X_H \\ 0 & i_M & X_M \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \beta \end{bmatrix} + \begin{bmatrix} u_H \\ u_M \end{bmatrix} \Rightarrow Y = X\beta + u$$

que da lugar a las siguientes ecuaciones:

$$W_H = \alpha_1 + \beta X_H + u_H \quad i = 1, \ldots, N_H \qquad \text{y} \qquad W_M = \alpha_2 + \beta X_M + u_M \quad i = 1, \ldots, N_M$$

La especificación alternativa correspondiente sería:

$$W_i = \alpha_1^* + \alpha_2^* S_{2i} + \beta X_i + u_i \qquad i = 1, \ldots, N_H + N_M \tag{3.6}$$

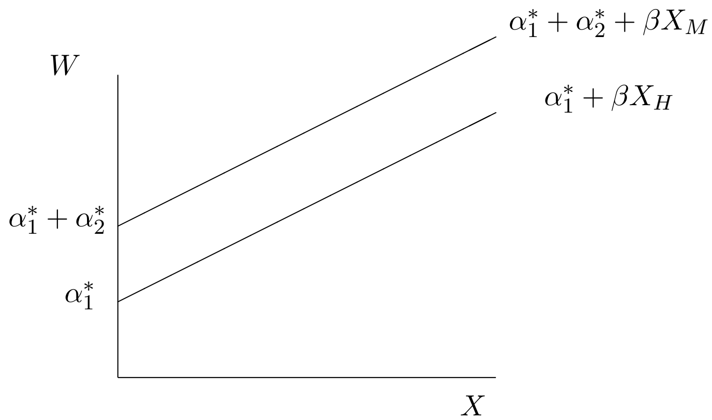
Donde el coeficiente β se interpreta de la forma habitual. En forma matricial el modelo sería:

$$\begin{bmatrix} W_H \\ W_M \end{bmatrix} = \begin{bmatrix} i_H & 0 & X_H \\ i_M & i_M & X_M \end{bmatrix} \begin{bmatrix} \alpha_1^* \\ \alpha_2^* \\ \beta \end{bmatrix} + \begin{bmatrix} u_H \\ u_M \end{bmatrix} \Rightarrow Y = X\beta + u$$

que da lugar a las siguientes ecuaciones:

$$W_H = \alpha_1^* + \beta X_H + u_H \quad i = 1, \ldots, N_H \qquad \text{y} \qquad W_M = \alpha_1^* + \alpha_2^* + \beta X_M + u_M \quad i = 1, \ldots, N_M$$

Gráficamente:



3.2. Comportamiento estacional

Las variables ficticias permiten recoger fácilmente comportamientos estacionales. Por ejemplo que las ventas de una empresa sean sistemáticamente superiores en alguno de los trimestres del año y que ese comportamiento se repita sistemáticamente año tras año es un clásico patrón de comportamiento sistemático estacional. Este comportamiento se produce en datos de series temporales de periodo inferior al anual y puede ser estudiado fácilmente mediante variables ficticias.

Por ejemplo para recoger el comportamiento estacional de una variable Y_t muestreada trimestralmente podemos proponer el modelo:

$$Y_t = \gamma_1 W_{1t} + \gamma_2 W_{2t} + \gamma_3 W_{3t} + \gamma_4 W_{4t} + \epsilon_t$$

donde t es el tiempo y W_i son las variables ficticias estacionales que toman valor uno en el trimestre i-ésimo y cero en el resto. La especificación alternativa sería:

$$Y_t = \alpha + \gamma_2^* W_{2t} + \gamma_3^* W_{3t} + \gamma_4^* W_{4t} + \epsilon_t$$

3.3. Efectos de interacción

3.3.1. Entre factores cualitativos y cuantitativos

En las ecuaciones (3.5) y (3.6) se recogen cambios en ordenada pero no en pendiente, sin embargo podemos pensar que el número de horas trabajadas cambia según el sexo del individuo con lo cual debemos recoger cambios en pendiente. Este efecto podemos analizarlo asociando las variables ficticias a la variable cuantitativa. Así proponemos el siguiente modelo:

$$W_i = \alpha_1 S_{1i} + \alpha_2 S_{2i} + \beta_1 (S_{1i} \times X_i) + \beta_2 (S_{2i} \times X_i) + u_i \qquad i = 1, \ldots, N_H + N_M \tag{3.7}$$

$$E(W_i/\text{hombre}) = \alpha_1 + \beta_1 X_i$$

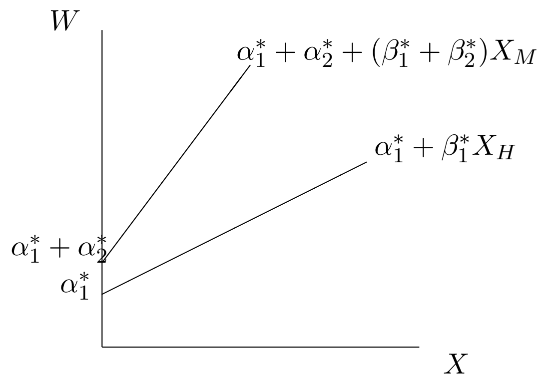
$$E(W_i/\text{mujere}) = \alpha_2 + \beta_2 X_i$$

donde β_1 y β_2 recogen el incremento en el salario medio ante un aumento unitario en las horas trabajadas, para los hombres y para las mujeres respectivamente.

La especificación alternativa sería:

$$W_i = \alpha_1^* + \alpha_2^* S_{2i} + \beta_1^* X_i + \beta_2^* (S_{2i} \times X_i) + u_i \qquad i = 1, \ldots, N_H + N_M \tag{3.8}$$

siendo β_2^* el incremento en el salario medio de una mujer con respecto a un hombre ante un aumento de una hora en el número de horas trabajado. Gráficamente:



3.3.2. Entre factores cualitativos

En el modelo (3.4) se supone que el efecto de cada factor es constante para todos los niveles de los demás factores. Sin embargo si suponemos que el efecto diferencial del sexo variase con el nivel de educación existiría un efecto interacción entre las variables ficticias sexo y educación, que podemos recoger así:

$$W_i = \mu + \alpha_2 S_{2i} + \beta_2 E_{2i} + \beta_3 E_{3i} + \gamma_2 (S_{2i} \times E_{2i}) + \gamma_3 (S_{2i} \times E_{3i}) + u_i \qquad i = 1, \ldots, N_H + N_M \quad (3.9)$$

donde la tabla que resume el comportamiento de la recta de regresión poblacional sería:

$E(W_i)$	E_{1i}	E_{2i}	E_{3i}
S_{1i}	μ	$\mu + \beta_2$	$\mu + \beta_3$
S_{2i}	$\mu + \alpha_2$	$\mu + \alpha_2 + \beta_2 + \gamma_2$	$\mu + \alpha_2 + \beta_3 + \gamma_3$

y podemos interpretar los parámetros como sigue:

μ	base de comparación.
β_2	Efecto diferencial en el salario medio debido a tener un nivel de estudios secundarios, con respecto a tener estudios primarios, para los hombres.
β_3	Efecto diferencial en el salario medio debido a tener un nivel de estudios universitarios, con respecto a tener estudios primarios, para los hombres.
α_2	Efecto diferencial en el salario medio entre los hombres y las mujeres para un nivel de educación primaria.
$\alpha_2 + \gamma_2$	Efecto diferencial en el salario medio, entre hombres y mujeres, para un nivel de educación secundaria.
$\alpha_2 + \gamma_3$	Efecto diferencial en el salario medio, entre hombres y mujeres, para un nivel de educación universitaria.
$\beta_2 + \gamma_2$	Efecto diferencial en el salario medio debido a tener un nivel de estudios secundarios, con respecto a tener estudios primarios, para las mujeres.
$\beta_3 + \gamma_3$	Efecto diferencial en el salario medio debido a tener un nivel de estudios universitarios, con respecto a tener estudios primarios, para las mujeres.

3.4. Tratamiento de las variables ficticias en Gretl.

Gretl permite trabajar tanto con variables ficticias cuantitativas como cualitativas y su tratamiento no difiere, solo debemos de ocuparnos de especificar correctamente el modelo. En el caso de que la variable ficticia no esté construida Gretl permite hacerlo. En la pantalla inicial en *Añadir* podemos añadir *Variables ficticias periódicas* que se ajustarán lógicamente a la periodicidad muestral del conjunto de datos, *Variables ficticias para las variables discretas seleccionadas* donde por ejemplo si tenemos una variable que toma valores 1, 2 y 3 podremos construir tres variables ficticias tal como

$$D_1 = \begin{cases} 1 & \text{si la variable toma valor 1} \\ 0 & \text{en otro caso} \end{cases}$$
$$D_2 = \begin{cases} 1 & \text{si la variable toma valor 2} \\ 0 & \text{en otro caso} \end{cases}$$
$$D_3 = \begin{cases} 1 & \text{si la variable toma valor 3} \\ 0 & \text{en otro caso} \end{cases}$$

Por supuesto también podremos introducirlas con el editor tal y como se aprendió en el Tema 1.

Veamos un ejemplo aplicado. Abrimos el fichero de datos data7-3 de Ramanathan, que contiene datos para 14 viviendas sobre el precio de venta de la vivienda (PRICE), pies cuadrados habitables (SQFT), número de habitaciones (BEDRMS) y número de baños (BATHS), y una variable ficticia que toma el valor 1 si la vivienda tiene piscina y 0 en caso contrario (POOL), una variable ficticia que toma el valor 1 si la vivienda tiene sala de estar y 0 en caso contrario (FAMROOM) y una variable ficticia que toma el valor 1 si la vivienda tiene chimenea y 0 en caso contrario (FIREPL). Seleccionamos las variables PRICE y POOL y observamos los valores de estas dos variables:

Obs	price	pool
1	199,9	1
2	228,0	0
3	235,0	1
4	285,0	0
5	239,0	0
6	293,0	0
7	285,0	0
8	365,0	1
9	295,0	0
10	290,0	0
11	385,0	1
12	505,0	1
13	425,0	0
14	415,0	0

Por ejemplo, la primera vivienda de la muestra tiene un precio de 199.900 dólares y tiene piscina (ya que la variable POOL toma el valor 1), mientras que la segunda no tiene piscina (la variable POOL toma el valor 0) y su precio de venta es de 228.000 dólares, etc.

Con los datos anteriores podemos obtener fácilmente que el precio medio de la vivienda es 317.493 dólares:

Estadísticos principales, usando las observaciones 1 - 14
para la variable price (14 observaciones válidas)

Media	Mediana	Mínimo	Máximo
317,49	291,50	199,90	505,00
Desv. Típ.	**C.V.**	**Asimetría**	**Exc. de curtosis**
88,498	0,27874	0,65346	−0,52983

Sin embargo, también es posible obtener el precio medio para las viviendas que tienen piscina, por un lado, y para las que no la tienen, por otro. Para ello, en primer, lugar se selecciona el precio para aquellas viviendas con piscina. Para ello, seleccionamos la variable PRICE, pinchamos en *Muestra → Definir a partir de v. ficticia...*, seleccionamos la variable POOL y aceptamos. De esta forma hemos seleccionado el precio para aquellas viviendas que tienen piscina[2]. A continuación, se obtienen los estadísticos principales:

Estadísticos principales, usando las observaciones 1 - 5
para la variable price (5 observaciones válidas)

Media	Mediana	Mínimo	Máximo
337,98	365,00	199,90	505,00
Desv. Típ.	**C.V.**	**Asimetría**	**Exc. de curtosis**
122,99	0,36390	0,15896	−1,2798

Para seleccionar el precio de las viviendas que no tienen piscina, pinchamos en *Muestra* → *Restringir a partir de criterio*, introducimos la condición $POOL = 0$ y aceptamos. Los estadísticos principales son los siguientes:

Estadísticos principales, usando las observaciones 1 - 9
para la variable price (9 observaciones válidas)

Media	Mediana	Mínimo	Máximo
306,11	290,00	228,00	425,00
Desv. Típ.	**C.V.**	**Asimetría**	**Exc. de curtosis**
68,959	0,225275	0,87575	−0,52255

Por tanto, el precio medio de las viviendas con piscina es de 337.980 dólares frente a los 306.111 de las viviendas sin piscina. Dado el modelo una vivienda con piscina es en promedio 31.869 dólares más cara que la que no tiene piscina. Notar que no se están teniendo en cuenta otros factores que pueden afectar al precio de la vivienda (número de pies cuadrados habitables, número de habitaciones, etc.).

El sencillo análisis anterior podemos realizarlo mediante un análisis de regresión. Podemos especificar un modelo econométrico utilizando la variable ficticia POOL como regresor, estimarlo, hacer inferencia e ir incorporando otras características que pueden afectar a los precios de las viviendas. Para comenzar, consideramos el siguiente modelo:

$$PRICE_i = \alpha_1 + \alpha_2 POOL_i + u_i \qquad i = 1, \ldots, 14 \tag{3.10}$$

donde

- α_1: precio medio de una vivienda sin piscina.
- $\alpha_1 + \alpha_2$: precio medio de una vivienda con piscina.
- α_2: diferencia en el precio medio de una vivienda con piscina con respecto a una que no la tiene.

Los resultados de estimar el modelo por Mínimos Cuadrados Ordinarios utilizando Gretl son los siguientes:

[2]Para restablecer el tamaño muestral inicial pinchar en *Muestra* → *Recuperar el rango completo*.

Modelo 1: estimaciones MCO utilizando las 14 observaciones 1–14
Variable dependiente: price

Variable	Coeficiente	Desv. típica	Estadístico t	valor p
const	306,111	30,2077	10,1335	0,0000
pool	31,8689	50,5471	0,6305	0,5402

Media de la var. dependiente	317,493
D.T. de la variable dependiente	88,4982
Suma de cuadrados de los residuos	98550,5
Desviación típica de los residuos ($\hat{\sigma}$)	90,6231
R^2	0,0320632
$\bar{R}^2$ corregido	−0,0485982
Grados de libertad	12
Log-verosimilitud	−81,880
Criterio de información de Akaike	167,760
Criterio de información Bayesiano de Schwarz	169,038

$$\widehat{PRICE}_i = \underset{(10,13)}{306,111} + \underset{(0,63)}{31,869}\, POOL_i \qquad i = 1, \ldots, 14$$

Para contrastar en el modelo (3.10) si hay diferencias significativas en el precio medio de la vivienda entre aquéllas que tienen piscina y las que no, la hipótesis de contraste es $H_0 : \beta = 0$. Este contraste se puede realizar utilizando el estadístico t habitual cuyo p-valor es 0,5405, por lo que no se rechaza la hipótesis nula para un nivel de significación del 5 %, es decir, el precio medio de la vivienda no es significativamente diferente por el hecho de tener piscina. Alternativamente, se puede realizar el contraste utilizando el estadístico F basado en las sumas de cuadrados de los residuos siendo en este caso el modelo (3.10) el modelo no restringido mientras que el modelo restringido es $PRICE_i = \alpha_1 + u_i \quad i = 1, \ldots, 14$.

Supongamos que ampliamos el modelo (3.10) incorporando regresores que podrían explicar el precio de la vivienda como: el hecho de que la vivienda tenga sala de estar o no, el hecho que tenga chimenea o no, el número de habitaciones y el número de baños. Las dos primeras son variables ficticias que pueden definirse así:

$$FIREPL_i = \begin{cases} 1 & \text{si la vivienda i-ésima tiene chimenea} \\ 0 & \text{en otro caso} \end{cases}$$

$$FAMROOM_i = \begin{cases} 1 & \text{si la vivienda i-ésima tiene sala de estar} \\ 0 & \text{en otro caso} \end{cases}$$

Mientras que el número de baños y el número de habitaciones se definen como en los temas anteriores:

$BEDRMS$ número de habitaciones de la vivienda i-ésima
$BATHS$ número de cuartos de baño de la vivienda i-ésima

Con todas ellas podemos definir el siguiente modelo para explicar el precio de la vivienda:

$$\begin{aligned} PRICE_i = {} & \gamma_1 + \gamma_2 POOL_i + \gamma_3 FAMROOM_i + \gamma_4 FIREPL_i \\ & + \beta_1 SQFT_i + \beta_2 BEDRMS_i + \beta_3 BATHS_i + u_i \qquad i = 1, \ldots, 14 \end{aligned} \tag{3.11}$$

Donde lo primero a notar es que en el modelo (3.11), afectando a la ordenada, conviven tres conjuntos de variables ficticias con dos categorías cada una, el hecho de tener o no piscina, el hecho de tener o no chimenea y el hecho de tener o no sala de estar, de las cuales sólo se incluye una de cada conjunto y se mantiene el término independiente. Esta forma de definir el modelo es muy cómoda ya que sigue manteniendo los resultados de los modelos con término independiente y permite una fácil interpretación de los coeficientes que acompañan a las variables ficticias. Así, γ_i $i = 2, 3, 4$ recogen el diferencial en el valor esperado de una vivienda por el hecho de poseer la característica correspondiente manteniéndose constante el resto de variables. El resultado de la estimación es:

Modelo 1: estimaciones MCO utilizando las 14 observaciones 1–14
Variable dependiente: price

Variable	Coeficiente	Desv. típica	Estadístico t	valor p
const	39,0571	89,5397	0,4362	0,6758
pool	53,1958	22,0635	2,4110	0,0467
famroom	−21,344	42,8734	−0,4979	0,6338
firepl	26,1880	53,8454	0,4864	0,6416
sqft	0,146551	0,0301014	4,8686	0,0018
bedrms	−7,0455	28,7363	−0,2452	0,8134
baths	−0,263691	41,4547	−0,0064	0,9951

Media de la var. dependiente	317,493
D.T. de la variable dependiente	88,4982
Suma de cuadrados de los residuos	9010,24
Desviación típica de los residuos ($\hat{\sigma}$)	35,8773
R^2	0,911504
$\bar{R}^2$ corregido	0,835650
$F(6, 7)$	12,0166
valor p para $F()$	0,00221290
Log-verosimilitud	−65,134
Criterio de información de Akaike	144,269
Criterio de información Bayesiano de Schwarz	148,743

La interpretación de los coeficientes estimados es la siguiente:

- $\hat{\gamma}_1 = 39,057$: el precio medio estimado de las viviendas sin piscina, baños, habitaciones, sala de estar ni chimenea y con 0 pies cuadrados habitables es de 39.057 dólares.

- $\hat{\gamma}_2 = 53,1958$: la diferencia estimada en el precio medio de las viviendas con piscina con respecto a las que no la tienen, siendo iguales en el resto de características (pies cuadrados habitables, habitaciones, baños, sala de estar y chimenea) es de 53.196 dólares.

- $\hat{\gamma}_3 = -21,34$: el precio medio estimado de una vivienda con sala de estar es 21.340 dólares inferior al de una sin sala de estar, siendo idénticas en el resto de características. Esto se debe a que, al mantener constante el número de pies cuadrados de la vivienda y el número de habitaciones y baños, incluir una sala de estar hará que el resto de habitaciones o baños sean de menor tamaño.

- $\hat{\gamma}_4 = 26,188$: el precio medio estimado de una vivienda con chimenea es 26.188 dólares más caro que el de una sin chimenea, siendo idénticas en el resto de características.

- $\hat{\beta}_1 = 0,147$: el precio medio estimado de una vivienda se incrementa en 147.000 dólares al aumentar en 1 pie cuadrado habitable su superficie, permaneciendo constantes el número de baños y habitaciones.
- $\hat{\beta}_2 = -7,046$: el precio medio estimado de una vivienda disminuye en 7.046 dólares al aumentar en 1 el número de habitaciones, permaneciendo constantes el número de baños y los pies cuadrados habitaciones. Esto se debe a que las habitaciones serán de menor tamaño.
- $\hat{\beta}_3 = -0,264$: el precio medio estimado de una vivienda disminuye en 264 dólares al aumentar en 1 el número de baños, permaneciendo constantes el número de habitaciones y los pies cuadrados habitables. De nuevo, las habitaciones serán de menor tamaño.

Contraste de hipótesis

Para contrastar, por ejemplo, que no existen diferencias significativas en el precio medio de la vivienda por el hecho de tener chimenea, se realiza un contraste de significatividad individual de la variable FIREPL. En este caso, observando el p-valor correspondiente, 0,6416, se puede concluir que a un nivel de significación del 5 %, no existen diferencias significativas en el precio medio de una vivienda por el hecho de tener chimenea.

Si comparamos los modelos (3.10) y (3.11), ninguna de las variables añadidas en el último es significativa individualmente[3]. Además, el $\bar{R}^2$ es inferior. El contraste de significatividad conjunta para las variables añadidas se puede realizar con el estadístico F basado en las sumas de cuadrados residuales de los modelos restringido (modelo (3.10)) y no restringido (modelo (3.11)). En este caso, el resultado es:

Contraste de omisión de variables –
  Hipótesis nula: los parámetros son cero para las variables
    bedrms
    baths
    famroom
    firepl
  Estadístico de contraste: $F(4, 7) = 0{,}0864517$
  con valor p = $P(F(4, 7) > 0{,}0864517) = 0{,}983881$

por lo que no se rechaza la hipótesis nula de que las variables añadidas al modelo (**??**) son conjuntamente no significativas. Al omitir dichas variables el modelo mejora en cuanto a la significación de sus coeficientes y el $\bar{R}^2$. Por tanto, manteniendo las variables POOL y SQFT, la inclusión del resto (FIREPL, FAMROOM, BATHS, BEDRMS) no añade capacidad explicativa al modelo.

[3]Un problema añadido es que tenemos un bajo tamaño muestral, T=14, y hemos aumentado significativamente el número de parámetros a estimar, K=7, por lo que tenemos muy pocos grados de libertad.

Tema 4

Validación del Modelo de Regresión

4.1. Sobre constancia de los coeficientes: contraste de cambio estructural

En ocasiones puede ocurrir que la relación entre la variable dependiente y los regresores cambie a lo largo del periodo muestral, es decir, puede que exista un cambio estructural. Por ejemplo, si estamos analizando el consumo de tabaco y durante el período muestral se ha producido una campaña de salud pública informando sobre los peligros que conlleva el consumo de tabaco, podemos pensar que tras dicha campaña el comportamiento de la demanda de tabaco haya cambiado, reduciéndose significativamente. Si esto ocurre no podemos especificar una única función de demanda para todo el período muestral si no que deberíamos especificar dos funciones, una hasta la campaña antitabaco y otra para el período siguiente. Por tanto ante sospechas de que exista un cambio estructural debemos de contrastar la estabilidad de los parámetros de nuestra relación.

El contraste de cambio estructural, llamado habitualmente contraste de Chow puede realizarse de manera sencilla mediante el estadístico de sumas de cuadrados de los residuos sin más que especificar adecuadamente el modelo restringido y el no restringido. También podemos llevarlo a cabo utilizando variables ficticias. Veamos un ejemplo para repasar las posibilidades. Hay cambio estructural cuando el modelo

$$Y_t = \alpha_1 + \beta_1 X_t + u_t \qquad t = 1, 2, \ldots, t_1 < T \tag{4.1}$$

y el modelo

$$Y_t = \alpha_2 + \beta_2 X_t + u_t \qquad t = t_1, t_2, \ldots, T \tag{4.2}$$

se verifica sólo desde el momento t_1 hasta T.

En este caso no podemos escribir una única ecuación del tipo:

$$Y_t = \alpha + \beta X_t + u_t \qquad t = 1, 2, \ldots, T \tag{4.3}$$

ya que no se verifica durante todo el periodo muestral y nuestro modelo estaría formado por las ecuaciones (4.1) y (4.2).

Si existe cambio estructural rechazaríamos $H_0 : \alpha_1 = \alpha_2, \beta_1 = \beta_2$ donde $q = 2$.

Este contraste podemos llevarlo a cabo utilizando el estadístico F basado en las sumas de cuadrados de los residuos siendo en este caso el modelo restringido el recogido en la ecuación

(4.3) mientras que el modelo no restringido está constituido por las ecuaciones (4.1) y (4.2). Matricialmente:

$$\text{Modelo no restringido} \Leftrightarrow \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} i_1 & X_1 & 0 & 0 \\ 0 & 0 & i_2 & X_2 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \beta_1 \\ \alpha_2 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \Rightarrow \hat{u}'\hat{u} = \hat{u}_1'\hat{u}_1 + \hat{u}_2'\hat{u}_2$$

$$\text{Modelo restringido} \Leftrightarrow \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} i_1 & X_1 \\ i_2 & X_2 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \Rightarrow \hat{u}_r'\hat{u}_r$$

bajo el supuesto $u_1 \sim N(0, \sigma_1^2), u_2 \sim N(0, \sigma_2^2)$ el estadístico de contraste y distribución es:

$$\frac{\hat{u}_r'\hat{u}_r - \hat{u}'\hat{u}/2}{\hat{u}'\hat{u}/T_1 + T_2 - 4} \overset{H_0}{\sim} F_{(2, T_1+T_2-4)}$$

Notar que para la validez del contraste además de suponer normalidad en las perturbaciones se supone también que $\sigma_1^2 = \sigma_2^2$.

4.2. Sobre las perturbaciones: contrastes de homocedasticidad y ausencia de correlación temporal

4.2.1. Contraste de homocedasticidad

Hasta el momento uno de los supuestos básicos del modelo de regresión lineal es que la varianza de cada término de perturbación u_t condicionada a los valores de las variables explicativas, es constante e igual a σ^2. Llamábamos a este supuesto *homocedasticidad* y lo denotábamos: $Var(u_t) = \sigma^2$ ó lo que es igual $E(u_t^2) = \sigma^2 \quad \forall t$. La varianza σ^2 es una medida de dispersión de u_t alrededor de su media , $E(u_t) = 0$, o equivalentemente, una medida de dispersión de la variable dependiente Y_t alrededor de su media $\beta_1 + \beta_2 X_{2t} + \ldots + \beta_k X_{kt}$. Así, homocedasticidad significa que la dispersión es la misma a través de todas las observaciones.

Supongamos que disponemos de observaciones sobre consumo y renta para un conjunto de familias, en un año determinado. Las familias con rentas bajas no tienen mucha flexibilidad en sus gastos, en general el grueso de la misma se gastará en cosas básicas, por ello la forma de consumo entre familias de renta baja no variará demasiado. Sin embargo, las familias de rentas altas tienen más posibilidades de consumo, ser grandes consumidores o ahorradores o llevar un gasto equilibrado. En cualquier caso su consumo puede ser muy distinto entre sí por lo que pueden tener una gran dispersión alrededor de su consumo medio mientras que las familias con rentas bajas no. En esta situación suponer que existe homocedasticidad no es sensato, deberíamos suponer que existe heterocedasticidad.

Llamamos heterocedasticidad al caso en que la varianza del término de error varía a través del tiempo si miramos a series temporales, $Var(u_t) = \sigma_t^2$, o cambia de un individuo a otro si miramos datos de sección cruzada, (familias, países, etc.), $Var(u_i) = \sigma_i^2$. Seguimos suponiendo que no existe autocorrelación entre perturbaciones de distinto momento del tiempo, es decir, $E(u_t u_s) = 0 \quad \forall t, s \quad t \neq s$ por lo que sólo consideramos la existencia de heterocedasticidad.

En el Gráfico 4.2.1 se puede apreciar la diferencia entre el comportamiento de las perturbaciones homocedásticas, a la izquierda) y heterocedásticas, a la derecha. En la figura de la izquierda se

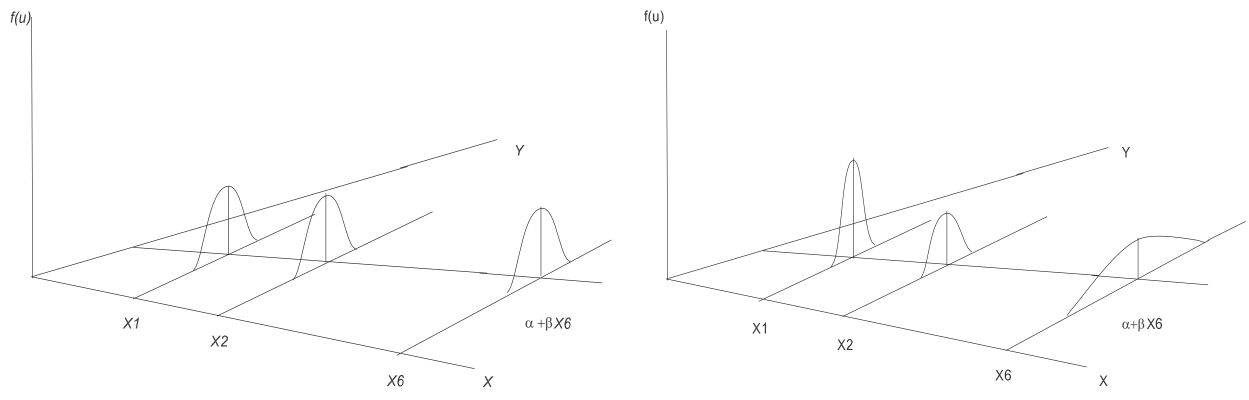


Gráfico 4.1: Perturbaciones homocedásticas versus heterocedásticas

puede observar que la varianza condicional de Y_t a las X_t permanece igual sin importar los valores que tome la variable X. Recordar que la varianza condicional de Y_t es la misma que la de u_t, por tanto, en el gráfico estamos observando cómo la varianza de la perturbación permanece constante independientemente del valor que tome el regresor. En la figura de la derecha se puede observar que la varianza de Y_t aumenta a medida que X_t aumenta y por tanto hay heterocedasticidad:

$$E(u_t^2) = \sigma_t^2$$

Hay diversas razones por las cuales las varianzas de u_t pueden no ser constantes:

- Modelos que tengan en cuenta expectativas: una expectativa no es más que una medida de lo que un agente espera que ocurra, la formación de esa medida conlleva un proceso de aprendizaje. Es de esperar que los agentes aprendan de sus errores y según avance el tiempo se confundan menos, en este caso σ_i^2 se reducirá.

- Si estamos analizando la relación entre consumo y renta podemos esperar que a medida que aumente la renta aumente σ_i^2. Una familia con mayor renta tiene mayores posibilidades de consumo, no sólo consumir más variedad de productos, sino que aumentará el valor del consumo real. Si la renta es suficientemente grande, podrá diferir consumo entre periodos y podrá ahorrar.

- Por razonamientos parecidos a los anteriores las empresas con mayores beneficios podrán presentar mayor variabilidad en sus políticas de dividendos. Si las ganancias son muy bajas simplemente no podrán repartir dividendos.

- En ocasiones es el investigador el que causa la heterocedasticidad. Por ejemplo, si el investigador trabaja con medias muestrales de las observaciones dentro de un grupo el modelo a estimar sería $\bar{Y}_t = \alpha + \beta\bar{X}_t + \bar{u}_t$ donde $Var(\bar{u}_t) = \sigma^2/n_t$ siendo n_t el número de observaciones dentro del grupo. Si el número de observaciones de cada grupo varía, la varianza de la perturbación del modelo sería heterocedástica.

- Otra causa de heterocedasticidad puede encontrarse en la mala especificación de un modelo. Si en un modelo se ha omitido una variable relevante su exclusión puede llevar a pensar que existe heterocedasticidad en las perturbaciones del modelo. Por ejemplo, si consideramos la función de demanda de un producto y excluimos los precios de los bienes complementarios a él o de sus competidores, los estimadores MCO serán sesgados y el estudio de los residuos mínimo cuadráticos del modelo puede dar la impresión de que la varianza de la perturbación no es constante. Si incluimos la variable o variables omitidas la impresión puede desaparecer. En este caso la solución al problema pasa por especificar correctamente el modelo.

Consecuencias de ignorar la heterocedasticidad

Vamos a analizar las consecuencias de utilizar el estimador MCO en presencia de heterocedasticidad:

- **En las propiedades del estimador MCO:** El estimador MCO bajo heterocedasticidad sigue siendo una combinación lineal de las perturbaciones. También sigue siendo insesgado ya que la media de la perturbación es cero. Sin embargo, no va a ser de varianza mínima ya que la matriz de varianza y covarianzas $\sigma^2(X'X)^{-1}$ obtenida en el tema 2 es mínima bajo las hipótesis básicas. Ahora, sin embargo, éstas no se cumplen: estamos considerando el supuesto de heterocedasticidad por tanto $E(u_t^2) \neq \sigma^2$, el Teorema de Gauss-Markov no se cumple y el estimador no es de varianza mínima. Ahora la matriz de varianzas y covarianzas de los coeficientes obtenida bajo este supuesto no vendrá dada por la expresión $\sigma^2(X'X)^{-1}$ y por tanto no será mínima. El estimador no es eficiente.

- **En los contrastes de hipótesis y la predicción:** Una forma sencilla de pensar en las consecuencias sobre los contrastes de hipótesis es pensar que dado que el estimador no es el mejor de los posibles la inferencia realizada con el mismo no será fiable.

  Formalmente lo que está ocurriendo es que el estimador de σ^2 propuesto en el tema 2 ahora no es insesgado por lo que los estadísticos de contraste habituales no tendrán las distribuciones t y F habituales. Por tanto, los contrastes no son válidos.

  Siguiendo el razonamiento anterior, la predicción realizada con un estimador ineficiente no será eficiente.

La existencia de heterocedasticidad en u_t tiene consecuencias en los estimadores MCO, en concreto ya no son los estimadores de menor varianza entre los estimadores lineales e insesgados. Existe otro estimador, el estimador de Mínimos Cuadrados Generalizados que es el de menor varianza entre los lineales e insesgados y para el cual la inferencia y predicción son válidas. Las consecuencias y soluciones del problema no forman parte del contenido de este curso. Sin embargo, en la siguiente sección vamos a aprender a detectar la existencia de heterocedasticidad con un estadístico de contraste sencillo y que aparece por defecto en los resultados de estimación MCO. En cursos más avanzados aprenderéis a solucionar el problema.

Detección de la heterocedasticidad

Sabemos que en presencia de heterocedasticidad el estimador MCO es ineficiente, y los contrastes de hipótesis no son válidos por ello es importante detectar la posible existencia de heterocedasticidad. La determinación de la existencia de heterocedasticidad sólo podremos conseguirla aplicando un test de contraste para heterocedasticidad, sin embargo podemos aproximarnos gráficamente al problema realizando un estudio visual de los residuos del modelo. Los residuos MCO son un estimador insesgado de u_t aún en presencia de heterocedasticidad. Usaremos el residuo al cuadrado como aproximación al comportamiento de la varianza de la perturbación. Para ver si puede existir un problema de heterocedasticidad podemos empezar por dibujar el

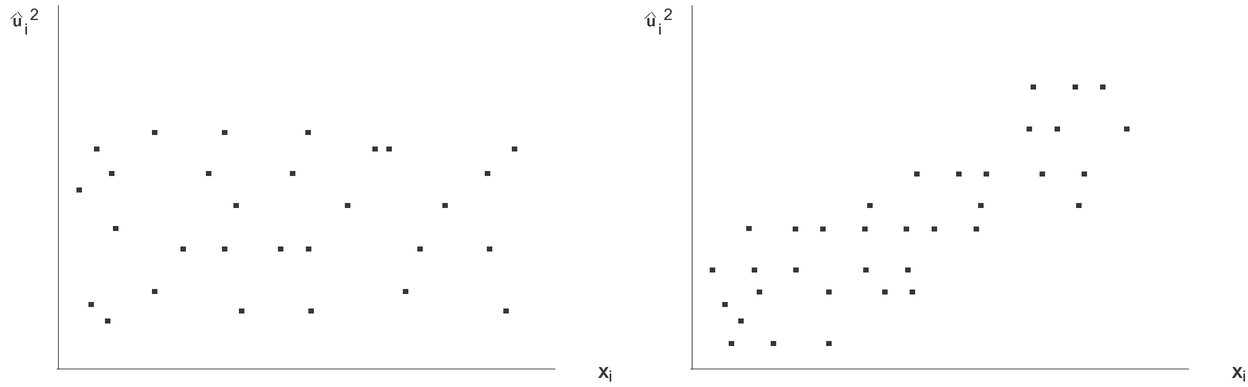


Gráfico 4.2: Perturbaciones homocedásticas versus heterocedásticas

cuadrado de los residuos MCO contra la variable de la cual sospechamos que depende σ^2, es decir, que sospechamos causa la heterocedasticidad

Por ejemplo si en el modelo:

$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 Z_i + u_i \quad i = 1, 2, \ldots, N \tag{4.4}$$

donde $E(u_i) = 0 \quad \forall i$ y $E(u_i u_j) = 0 \quad \forall i, j \quad i \neq j$
sospechamos que u_i es heterocedástica debido a la variable X_i, por ejemplo, su varianza es creciente con X_i. La forma correcta de proceder para detectar la existencia de heterocedasticidad en las perturbaciones del modelo sería estimar éste por MCO y estudiar el gráfico de los residuos MCO, $(\hat{u}_{MCO,i})$, frente a X_i. Si el gráfico es como el de la derecha de el Gráfico 4.2.1 pensaremos que los residuos $\hat{u}_{MCO,i}$ se incrementan con X_i y que el incremento es proporcional. Como hemos vamos a usar el cuadrado de los residuos como estimador de la varianza de la perturbación propondremos, por ejemplo:

$$V(u_i) = E(u_i^2) = \sigma^2 X_i$$

Si el gráfico de los residuos $\hat{u}_{MCO,i}$ y X_i fuera como en la derecha de el Gráfico 4.2.1 supondríamos que el aumento en la varianza de u_i es lineal a X_i y propondríamos:

$$E(u_i^2) = a + bX_i$$

En el caso de que no conozcamos cuál de las variables exógenas genera heterocedasticidad tendremos que estudiar los gráficos de los residuos de MCO, contraponiéndolos a cada una de las variables exógenas restantes. Así, si la gráfica entre $\hat{u}_{MCO,i}$ y X_i resultara como la izquierda de el Gráfico 4.2.1, en la que no se aprecia ningún patrón de comportamiento y parece que hay una distribución aleatoria de los pares $(X_i, \hat{u}_i^2)$, procederíamos a analizar los residuos frente a Z_i. Si suponemos que la dependencia es de un conjunto de regresores podemos dibujarlos contra $\hat{Y}_t$.

En el Gráfico 4.2.1 podemos observar otros patrones de comportamiento en los residuos que pueden indicar la existencia de heterocedasticidad en las perturbaciones.

Sin embargo el estudio gráfico de los residuos no es determinativo. Para determinar si existe o no heterocedasticidad tendremos que realizar un contraste de existencia de heterocedasticidad con un estadístico adecuado. Estadísticos de contraste de existencia de heterocedasticidad hay muchos y unos se adecúan más a unas situaciones que otros y en general necesitan suponer

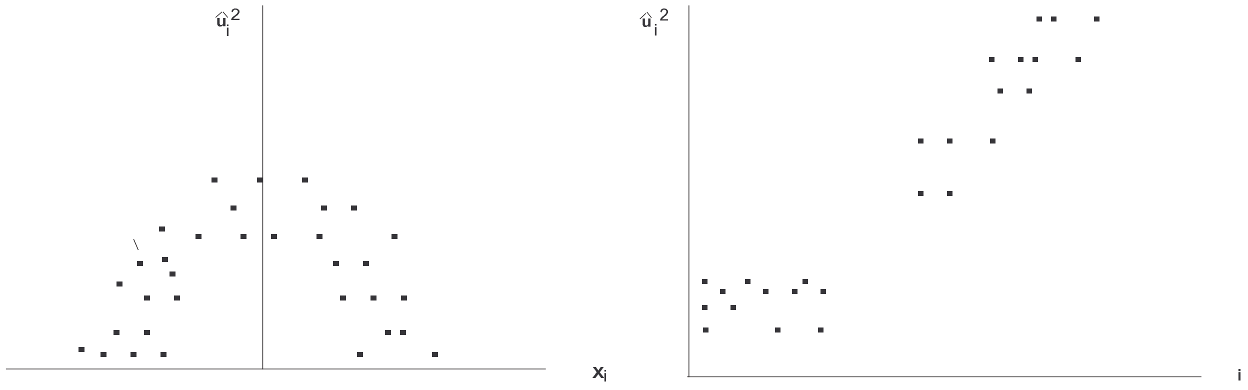


Gráfico 4.3: Perturbaciones heterocedásticas

una forma funcional para σ_t^2. El análisis gráfico no es una pérdida de tiempo ya que la relación entre X_{ki} y $\hat{u}_{MCO,i}$ nos indicará una posible forma funcional (de heterocedasticidad) para la varianza de la perturbación y puede indicarnos cuál es el test de contraste más adecuado. En este tema vamos a estudiar un único test de heterocedasticidad que tiene carácter general y no exige supuestos sobre el comportamiento de σ_t^2. Además Gretl lo proporciona directamente.

4.2.2. Contraste de White

El contraste de heterocedasticidad propuesto por White en 1980 es un contraste paramétrico, de carácter general, que no precisa especificar la forma que puede adoptar la heterocedasticidad. En este sentido puede calificarse de robusto. Antes de aplicar el contraste con Gretl vamos a desarrollar paso a paso el contraste para entender su mecanismo. Para la ilustración vamos a suponer que queremos contrastar la existencia de heterocedasticidad en el modelo:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t \tag{4.5}$$

Se procede de la forma siguiente:

a) Estimamos por MCO el modelo original y calculamos los residuos de MCO, $\hat{u}_{MCO,i}$.

b) Estimamos la regresión auxiliar: el cuadrado de los residuos mínimo-cuadráticos de la regresión anterior, sobre una constante, los regresores del modelo original, sus cuadrados y productos cruzados de segundo orden, evitando los redundantes:

$$\hat{u}_t^2 = \alpha_1 + \alpha_2 X_{2t} + \alpha_3 X_{3t} + \alpha_4 X_{2t}^2 + \alpha_5 X_{3t}^2 + \alpha_6 X_{2t} X_{3t} + \omega_t \tag{4.6}$$

Contrastar la hipótesis nula de homocedasticidad es equivalente a contrastar que todos los coeficientes de esta regresión, exceptuando el término independiente son cero. Es decir:

$$H_0 : \alpha_2 = \alpha_3 = \ldots = \alpha_6 = 0$$

c) El estadístico de contraste es $\lambda = NR^2$ donde R^2 es el coeficiente de determinación de la regresión auxiliar (4.6). Rechazamos H_0 si $NR^2 > \chi_\alpha(p)$ siendo p el número de coeficientes en la regresión auxiliar sin incluir el término independiente, en el ejemplo $p = 5$.

Observaciones:

a) Este contraste es muy flexible ya que no especifica la forma funcional de heterocedasticidad, pero por otro lado, si se rechaza la hipótesis nula de homocedasticidad no indica cuál puede ser la dirección a seguir.

b) A la hora de incluir los regresores de la regresión auxiliar debemos ser muy cuidadosos para no incurrir en multicolinealidad exacta, por ejemplo en el caso de las variables ficticias con valores 0 y 1, en este caso el cuadrado de la variable coincide con ella misma.

c) También pueden surgir problemas en modelos con un alto número de regresores que puede conllevar que en la regresión auxiliar el número de variables sea tal que no supere al número de observaciones y nos quedemos sin grados de libertad. Si éste es el caso podemos optar por regresar el cuadrado de los residuos MCO sobre $\hat{Y}_t$ y $\hat{Y}_t^2$ ya que $\hat{Y}_t$ es el ajuste de Y_t usando el estimador MCO con todos los regresores originales.

d) El contraste de White puede recoger otro tipo de problemas de mala especificación de la parte sistemática, omisión de variables relevantes, mala forma funcional etc. Esto es positivo si se identifica cuál es el problema, en caso contrario, la solución que se tome puede estar equivocada. Si la detección de heterocedasticidad se debe a un problema de mala especificación la solución pasa por especificar correctamente el modelo y no pro proponer un estimador alternativo insesgado y de varianza mínima.

4.2.3. Contraste de ausencia de correlación temporal

En el modelo de regresión, el término de perturbación engloba aquellos factores que determinando la variable endógena, no están recogidos en la parte sistemática del modelo. Estos factores pueden ser innovaciones, errores de medida en la variable endógena, variables omitidas, etc. Hasta el momento uno de los supuestos básicos del modelo de regresión lineal es que la covarianza entre perturbaciones de distintos periodos es cero. Sin embargo, si estos factores están correlacionados en el tiempo o en el espacio, entonces no se satisface la hipótesis de NO autocorrelación que escribíamos como $E(u_t u_s) = 0 \quad \forall t, s \quad t \neq s$. Este fenómeno se conoce con el nombre de **autocorrelación**: *correlación serial*, en el caso de series temporales y *correlación espacial* en el caso de datos de sección cruzada.

Concepto de autocorrelación:
Existe autocorrelación cuando el término de error de un modelo econométrico está correlacionado consigo mismo. Es decir, la covarianza entre las perturbaciones es distinta de cero para diferentes momentos del tiempo (o entre distintos individuos)[1] :

$$E(u_t u_s) \neq 0 \quad \forall t, s \quad t \neq s$$

Esta dinámica, aunque no sea relevante en media, refleja un patrón sistemático que tenemos que considerar a la hora de estimar el modelo.

La existencia de autocorrelación supone el incumplimiento de una de las hipótesis básicas sobre la perturbación de forma similar a la existencia de heterocedasticidad. Esta afecta a las varianzas mientras que la autocorrelación afecta a las covarianzas. En cualquier caso, las consecuencias

[1]No es preciso que u_t esté correlacionada consigo misma en cada dos instantes distintos del tiempo, sino que basta que la correlación se extienda a algunos periodos.

sobre el estimador MCO son las mismas: el estimador no es de varianza mínima, aunque sigue siendo lineal e insesgado. Los contrastes de hipótesis no son válidos por las mismas razones que en el supuesto de heterocedasticidad y la predicción no es eficiente. Si sospechamos que en un modelo la perturbación está autocorrelada, primero deberíamos cerciorarnos realizando un contraste y en el caso de que ésta exista es importante estimar el modelo bajo estos nuevos supuestos con un estimador alternativo a MCO que sea de varianza mínima y válido para hacer inferencia. Este estimador queda fuera del contenido de este tema, sin embargo, aprenderemos a detectar el problema mediante un contraste y estudiaremos un proceso sencillo para recoger el comportamiento de la perturbación bajo autocorrelación: el proceso autorregresivo de primer orden.

Causas de autocorrelación

Como decíamos al iniciar este capítulo el término de perturbación de un modelo engloba aquellos factores que, determinando la variable endógena, no están recogidos en la parte sistemática del modelo. Factores como variables omitidas, mala especificación de la forma funcional o errores de medida, entre otros, son causa de autocorrelación. Repasaremos algunos de ellos:

- **Shocks aleatorios prolongados**
  Sea el modelo:
  $$R_t = \beta_1 + \beta_2 RM_t + u_t \quad t = 1, 2, \ldots, T$$
  donde R_t es la rentabilidad de un activo en el periodo t y RM_t es la rentabilidad del mercado en dicho periodo t. Si en un momento dado se produce una caída del mercado, la rentabilidad del activo se verá afectada a la baja y como consecuencia la rentabilidad obtenida será menor que la esperada. Este efecto se prolongará en el tiempo hasta que poco a poco los inversores recuperen la confianza y el mercado vuelva a estabilizarse. El shock se recogerá en el término de perturbación. Si por ejemplo, la caída se produce en $(t-1)$, lo que estamos diciendo es que la perturbación en t dependerá de lo ocurrido en $(t-1)$ vía u_{t-1}.

- **Existencia de ciclos y tendencias**
  Si estamos analizando un modelo econométrico cuya variable endógena presenta ciclos y/o tendencias que no se explican a través de las variables exógenas, la perturbación recoge dichas estructuras, presentando un comportamiento de autocorrelación. En este caso, los residuos presentan rachas de desviaciones por encima del promedio (en la parte alta del ciclo) y rachas de desviaciones por debajo del promedio (parte baja del ciclo).

- **Relaciones no lineales**
  Supongamos que la verdadera relación entre los tipos de interés, r_t, y el stock de Deuda Pública, D_t, es cuadrática:
  $$r_t = \beta_1 + \beta_2 D_t + \beta_3 D_t^2 + u_t \quad t = 1, 2, \ldots, T \quad \beta_2 > 0, \beta_3 < 0$$
  Este modelo implica que los tipos de interés aumentan al crecer el stock de deuda pública, aunque menos que proporcionalmente, puesto que se tiene:
  $$\frac{\partial r_t}{\partial D_t} = \beta_2 + 2\beta_3 D_t < \beta_2$$
  tanto menor cuanto mayor es D_t. Pero sin embargo se especifica y se estima un modelo lineal:
  $$r_t = \beta_1 + \beta_2 D_t + u_t \quad t = 1, 2, \ldots, T$$

En este caso la curvatura de la parte sistemática pasa a ser recogida por la perturbación. Los residuos presentarán una racha de residuos negativos seguida de otra racha de residuos positivos para seguir con otra negativa.

- **Variables omitidas relevantes correlacionadas**
Si el modelo realmente se especifica como:

$$\begin{aligned} Y_t &= \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t \qquad t = 1, 2, \ldots, T \\ u_t &= Y_t - \beta_1 - \beta_2 X_{2t} - \beta_3 X_{3t} \\ \hat{u}_t &= Y_t - \hat{\beta}_1 - \hat{\beta}_2 X_{2t} - \hat{\beta}_3 X_{3t} \end{aligned}$$

Pero estimamos:

$$\begin{aligned} Y_t &= \beta_1 + \beta_2 X_{2t} + v_t \qquad t = 1, 2, \ldots, T \\ v_t &= Y_t - \beta_1 - \beta_2 X_{2t} = u_t + \beta_3 X_{3t} \\ \hat{v}_t &= Y_t - \tilde{\beta}_1 - \tilde{\beta}_2 X_{2t} = \hat{u}_t - (\tilde{\beta}_1 - \hat{\beta}_1) - (\tilde{\beta}_2 - \hat{\beta}_2) X_{2t} + \hat{\beta}_3 X_{3t} \end{aligned}$$

En este contexto de omisión recordemos que los estimadores MCO son sesgados en general. En consecuencia, tras un análisis de los residuos $\hat{v}_t$ tanto gráfico como mediante tests, es muy probable que si la variable omitida está correlacionada o presenta ciclos o tendencias el investigador llegue a la conclusión de que: $Cov(v_t, v_s) \neq 0$. De todas formas hay que tener en cuenta que no siempre que se omite una variable relevante se causa autocorrelación.

Proceso autorregresivo de primer orden, AR(1)

Existen numerosos procesos capaces de reproducir estructuras de autocorrelación en la perturbación, sin embargo el *proceso autorregresivo de primer orden* es el proceso de autocorrelación más sencillo y uno de los que mejor se suele ajustar a datos económicos[2]. Se especifica como:

$$u_t = \rho u_{t-1} + \epsilon_t \qquad t = 1, 2, \ldots, T$$

de forma que la perturbación en el periodo t depende de la perturbación del periodo anterior $(t-1)$ más un término aleatorio (o innovación) ϵ_t cuyas características son:

$$\left. \begin{array}{l} E(\epsilon_t) = 0 \quad \forall t \\ E(\epsilon_t^2) = \sigma_\epsilon^2 \quad \forall t \\ E(\epsilon_t \epsilon_s) = 0 \quad \forall t, s \quad t \neq s \end{array} \right. \qquad \epsilon_t \sim iid(0, \sigma_\epsilon^2)$$

y que habitualmente se le llama ruido blanco[3]. La especificación completa del MRLG cuando la perturbaciones presentan autocorrelación es:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \ldots + \beta_K X_{Kt} + u_t \qquad t = 1, 2, \ldots, T$$

$$u_t = \rho u_{t-1} + \epsilon_t \qquad \epsilon_t \sim iid(0, \sigma_\epsilon^2) \qquad |\rho| < 1$$

[2]El proceso autorregresivo de orden uno no es el único proceso que recoge dinámica en la perturbación. El proceso autorregresivo más general de todos es el proceso autorregresivo de orden p, AR(p):

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \ldots + \rho_p u_{t-p} + \epsilon_t$$

Además existen otros procesos alternativos a los autorregresivos.

[3]Si nos fijamos, ϵ_t cumple las hipótesis básicas sobre la perturbación, es por tanto homocedástica y no autocorrelada.

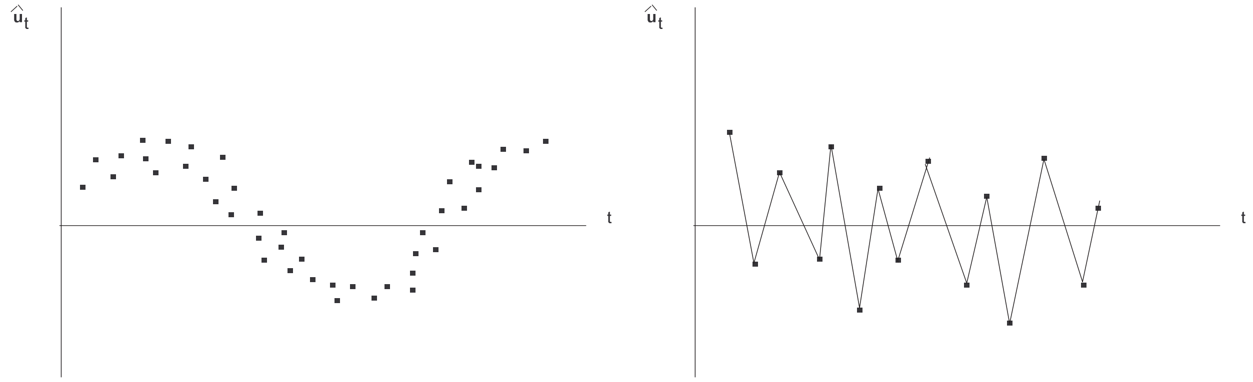


Gráfico 4.4: Proceso autorregresivo de orden uno

El coeficiente ρ mide la correlación entre u_t y u_{t-1} y debe cumplir que $|\rho| < 1$ para que el proceso no sea explosivo. Se le denomina coeficiente de autocorrelación de orden uno (o primer orden) ya que uno es el número de periodos entre u_t y u_{t-1}:

$$\rho = \frac{Cov(u_t, u_{t-1})}{\sqrt{Var(u_t)}\sqrt{Var(u_{t-1})}} \qquad -1 < \rho < 1$$

Si la covarianza es positiva se le denomina correlación positiva y si es negativa, correlación negativa. Dado que $u_t = Y_t - E(Y_t/\{X_{it}\}_{i=1}^K)$ la perturbación representa la diferencia entre el comportamiento observado y el comportamiento promedio. Dados los posibles valores de ρ tenemos que:

i) Si $\rho > 0$ entonces un valor elevado de u_t genera un valor de Y_t por encima del promedio y tendrá mayor probabilidad de ir seguido por un valor elevado de u_{t+1} y así sucesivamente.

ii) Si $\rho < 0$ un valor alto de u_t irá seguido por un valor bajo de u_{t+1} y éste por uno alto de u_{t+2} y así sucesivamente.

La relación entre la perturbación u_t y la innovación ϵ_t se recoge en el diagrama siguiente:

$$\begin{array}{llll} u_{t-2} \longrightarrow & u_{t-1} \longrightarrow & u_t \longrightarrow & u_{t+1} \\ \uparrow & \uparrow & \uparrow & \uparrow \\ \epsilon_{t-2} & \epsilon_{t-1} & \epsilon_t & \epsilon_{t+1} \end{array}$$

por lo que cada innovación influye sobre la perturbación en el mismo periodo o períodos posteriores, pero nunca sobre los valores anteriores, es decir: $E(\epsilon_t u_{t-s}) = 0 \quad s > 0$. Además en el diagrama se puede observar que u_t no depende directamente de u_{t-2} pero sí lo hace a través de u_{t-1}, por tanto u_t está correlado con todos las perturbaciones pasadas.

En la figura de la izquierda del Gráfico 4.2.3 se observa un proceso autorregresivo de primer orden con parámetro ρ positivo. En ella podemos observar una racha de residuos positivos seguidos de una racha de residuos negativos y así sucesivamente. En cambio, cuando el parámetro del proceso autorregresivo es negativo, los signos de los residuos se alternan como podemos ver en la figura de la derecha del mismo gráfico.

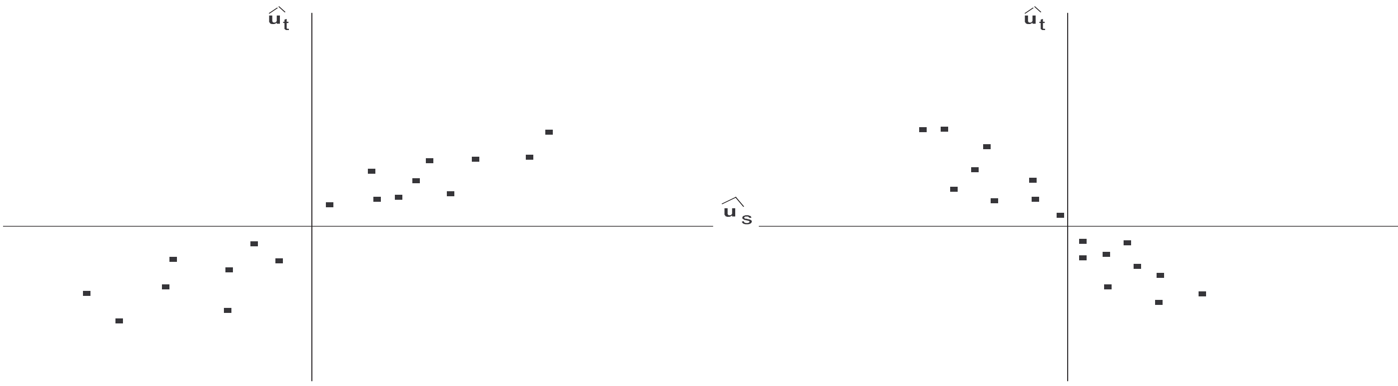


Gráfico 4.5: Perturbaciones AR(1) positivo versus AR(1) negativo

Detección de la autocorrelación

En la práctica no se conoce a priori si existe autocorrelación ni cuál es el proceso más adecuado para modelarla. Para determinar su existencia es necesario contrastar dicha hipótesis mediante un estadístico de contraste.

Si embargo, ningún contraste de autocorrelación debe excluir un examen riguroso de los residuos generados en la estimación del modelo. El gráfico de los mismos puede indicarnos la existencia de autocorrelación. Dado que los residuos son una aproximación a la perturbación, la existencia de patrones o comportamientos sistemáticos en los mismos indicaría la posible existencia de autocorrelación en u_t. Por ejemplo, podemos esperar que el gráfico de la evolución temporal de $\hat{u}_{t,MCO}$ se comporte de forma similar a lo mostrado por el Gráfico 4.2.3. Sin embargo, también podemos dibujar la evolución temporal de $\hat{u}_{t,MCO}$ contra la de $\hat{u}_{s,MCO}$ para $s = t-1$. Si encontramos que la mayoría de los puntos en dicho gráfico se hallan en el primer o tercer cuadrante, izquierda del Gráfico 4.2.3, ello es un indicio de autocorrelación positiva. Si se hallan en el segundo y cuarto cuadrante, derecha del Gráfico 4.2.3, indicará autocorrelación negativa. Tras el análisis gráfico, si sospechamos que existe autocorrelación debemos contrastarla con un estadístico de contraste. Existen varios estadísticos de contraste pero en este tema vamos a estudiar sólo uno, el estadístico de Durbin Watson, específico para contrastar la existencia de un proceso autorregresivo de primer orden.

Contraste de Durbin Watson

Durbin y Watson propusieron en 1951 un estadístico para contrastar la existencia de un proceso AR(1) en el término de perturbación. La hipótesis nula es la no existencia de autocorrelación:

$$H_0 : \rho = 0$$

frente a la alternativa

$$H_a : \rho \neq 0 \quad \text{en} \quad u_t = \rho u_{t-1} + \epsilon_t \quad \epsilon_t \sim (0, \sigma^2_\epsilon)$$

y se contrasta mediante el estadístico:

$$DW = \frac{\sum_{t=2}^{T}(\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^{T} \hat{u}_t^2}$$

donde $\hat{u}_t$ son los residuos mínimo-cuadráticos ordinarios de estimar el modelo original sin tener en cuenta la existencia de autocorrelación en las perturbaciones. Gretl proporciona el valor de este estadístico entre los resultados de la estimación MCO. Sin embargo antes de utilizarlo vamos a estudiar su interpretación y manejo sin recurrir a Gretl.

• Interpretación del estadístico DW:
Si el tamaño muestral es suficientemente grande podemos emplear las aproximaciones

$$\sum_{t=2}^{T} \hat{u}_t^2 \simeq \sum_{t=2}^{T} \hat{u}_{t-1}^2 \simeq \sum_{t=1}^{T} \hat{u}_t^2$$

con lo que

$$DW \simeq \frac{2\sum_{t=2}^{T} \hat{u}_t^2 - 2\sum_{t=2}^{T} \hat{u}_t \hat{u}_{t-1}}{\sum_{t=1}^{T} \hat{u}_t^2} \simeq 2 - 2\frac{\sum_{t=2}^{T} \hat{u}_t \hat{u}_{t-1}}{\sum_{t=2}^{T} \hat{u}_{t-1}^2} \simeq 2(1-\hat{\rho})$$

donde $\hat{\rho}$ es el estimador de ρ por MCO en el modelo $u_t = \rho u_{t-1} + \epsilon_t$, empleando como aproximación de u_t el residuo MCO, es decir, $\hat{u}_t = \rho \hat{u}_{t-1} + \epsilon_t$ y $\hat{\rho} = \frac{\sum \hat{u}_t \hat{u}_{t-1}}{\sum \hat{u}_{t-1}^2}$.

En base a la relación $DW \simeq 2(1-\hat{\rho})$ podemos establecer el siguiente comportamiento en los residuos:

- Si existe autocorrelación positiva de primer orden, valores positivos del término de error u_t tiendan a ir seguidos de valores positivos y asimismo, valores negativos tiendan a ir seguidos de valores negativos. Dado que la aproximación a la perturbación es el residuo, los patrones en la perturbación serán detectados en el residuo. Así, observaremos rachas de residuos positivos seguidas de rachas de residuos negativos. En estas circunstancias, generalmente $|\hat{u}_t - \hat{u}_{t-1}| < |\hat{u}_t| \Rightarrow (\hat{u}_t - \hat{u}_{t-1})^2 < \hat{u}_t^2$ y el numerador del estadístico "será pequeño" en relación al denominador, con lo que el estadístico "será pequeño". En consecuencia cuanto más cercano esté el parámetro ρ a la unidad más próximo a cero estará el DW. En el extremo positivo tenemos que $\rho \longrightarrow 1 \quad \Rightarrow DW \longrightarrow 0$.

- Si existe autocorrelación negativa de primer orden, valores positivos de $\hat{u}_t$ tienden a ir seguidos de valores negativos, en este caso $|\hat{u}_t - \hat{u}_{t-1}| > |\hat{u}_t| \Rightarrow (\hat{u}_t - \hat{u}_{t-1})^2 > \hat{u}_t^2$ con lo que el estadístico DW tenderá a tomar valores grandes. En el extremo negativo tenemos que $\rho \longrightarrow -1 \quad \Rightarrow DW \longrightarrow 4$.

A partir de la relación $DW \simeq 2(1-\hat{\rho})$ se puede establecer el rango de valores que puede tomar el estadístico DW.

$$\begin{array}{ll} 0 < \hat{\rho} < 1 & DW \in (0,2) \\ \hat{\rho} = 0 & DW \simeq 2 \\ -1 < \hat{\rho} < 0 & DW \in (2,4) \end{array}$$

La distribución del estadístico DW bajo H_0 depende de la matriz de regresores X por lo que los valores críticos del contraste también serán diferentes para cada posible X. Durbin y Watson tabularon los valores máximo (d_U) y mínimo (d_L) que puede tomar el estadístico independientemente de cuál sea X, y tal que $d_L < DW < d_U$. La distribución de d_L y d_U depende del tamaño de la muestra, T, y de K' que denota el número de variables explicativas del modelo exceptuando el término independiente.

• Contraste de existencia de autocorrelación positiva:
$H_0 : \rho = 0$
$H_a : \rho > 0 \quad$ en $\quad u_t = \rho u_{t-1} + \epsilon_t \quad |\rho| < 1 \quad \epsilon_t \sim iid(0, \sigma_\epsilon^2)$

a) Si $DW < d_L$ se rechaza la H_0 para un nivel de significatividad α dado, por tanto existe autocorrelación positiva.

b) Si $DW > d_U$ no se rechaza la H_0 para un nivel de significatividad α dado, por tanto no existe autocorrelación positiva.

c) Si $d_L < DW < d_U$ estamos en una zona de incertidumbre y no podemos concluir si existe o no autocorrelación positiva de primer orden.

• **Contraste de existencia de autocorrelación negativa:**
$H_0 : \rho = 0$
$H_a : \rho < 0$ en $u_t = \rho u_{t-1} + \epsilon_t \quad |\rho| < 1 \quad \epsilon_t \sim iid(0, \sigma^2_\epsilon)$

a) Si $DW < 4 - d_U$ no se rechaza la H_0 para un nivel de significatividad α dado, por tanto no existe autocorrelación negativa.

b) Si $DW > 4 - d_L$ se rechaza la H_0 para un nivel de significatividad α dado, por tanto existe autocorrelación negativa.

c) Si $4 - d_U < DW < 4 - d_L$ estamos en una zona de incertidumbre y como en el caso anterior, no podemos concluir si existe o no autocorrelación negativa de primer orden.

Gráficamente:

$$H_o : \rho = 0$$

Autocorrelación positiva — Autocorrelación negativa

$\longleftarrow - - - - H_a : \rho > 0 - - - - - \longrightarrow \quad \longleftarrow - - - - - - - - H_a : \rho < 0 - - - - - \longrightarrow$

Rechazar $\rho = 0$	Duda	Aceptar $\rho = 0$	Duda	Rechazar $\rho = 0$

$0 \quad d_L \quad d_U \quad 2 \quad 4 - d_U \quad 4 - d_L \quad 4$

Si el resultado del contraste es que existe autocorrelación y ésta no es debida a una mala especificación del modelo, éste no debe ser estimado por MCO ya que este estimador es ineficiente. Si la autocorrelación es originada por una mala especificación del modelo primero se ha de corregir esta especificación y una vez el modelo esté correctamente especificado analizar las propiedades de la perturbación y actuar en consecuencia.

• Observaciones sobre el contraste de Durbin Watson:

a) El contraste de Durbin Watson también se puede considerar un contraste de mala especificación del modelo. La omisión de variables relevantes correlacionadas, una forma funcional inadecuada, cambios estructurales no incluidos en el modelo, etc., pueden originar un estadístico DW significativo. Esto nos puede llevar a errores si consideramos que hay evidencia de autocorrelación y se modela un proceso AR(1). Por otro lado, si u_t sigue un proceso distinto de un AR(1), es probable que el estadístico DW lo detecte. Por lo tanto, el estadístico de Durbin Watson es útil porque nos indica la existencia de problemas en el modelo, pero a veces no nos ayuda a establecer cuál es la estructura real. En caso de no rechazar la H_0, podemos afirmar que no tenemos un AR(1), pero no sabemos si tenemos alguna otra estructura alternativa.

b) Por otro lado el estadístico DW sólo debe aplicarse cuando los regresores son **no estocásticos**, en presencia de regresores aleatorios como la variable endógena retardada no tiene validez.

c) Cuando el estadístico DW cae en zona de duda, y si no podemos llevar a cabo un contraste alternativo, no debemos concluir que no existe autocorrelación. El procedimiento conservador aconseja rechazar la hipótesis nula y estimar por un estimador alternativo a MCO ya que las consecuencias de ignorar su existencia cuando sí la hay son más graves que las correspondientes al caso contrario.

4.3. Validación en Gretl

4.3.1. Contraste de cambio estructural o Chow con Gretl

Utilizando Gretl una vez abierto el fichero de datos y estimado el modelo correspondiente por MCO, en la ventana de resultados de la estimación haríamos:

Contrastes ⟶ *Contraste de Chow*

A la pregunta *Observación en la cual dividir la muestra* contestaríamos *fecha correspondiente a* T_1 y automáticamente Gretl realiza el contraste y nos muestra el resultado.

Por ejemplo el fichero data7-19 del libro de Ramanathan contiene datos para 1960-1988 sobre la demanda de tabaco y sus determinantes en Turquía. Las variables de interés para el ejemplo son las siguientes:
Q: consumo de tabaco por adulto (en kg).
Y: PNB real per cápita en liras turcas de 1968.
P: precio real del kilogramo de tabaco, en liras turcas.
$D82$: variable ficticia que toma valor 1 a partir de 1982.

A mediados de 1981 el gobierno turco lanza una campaña de salud pública advirtiendo de los peligros de salud que conlleva el consumo de tabaco. Nuestro objetivo es determinar si existen cambios en la demanda de tabaco tras la campaña institucional en cuyo caso la especificación:

$$LnQ_t = \alpha + \beta LnY_t + \gamma LnP_t + u_t \qquad t = 1960, \ldots, 1988 \tag{4.7}$$

no es correcta para todo el período muestral y deberíamos especificar dos ecuaciones:

$$LnQ_t = \alpha_1 + \beta_1 LnY_t + \gamma_1 LnP_t + u_{1t} \qquad t = 1960, \ldots, 1981 \tag{4.8}$$
$$LnQ_t = \alpha_2 + \beta_2 LnY_t + \gamma_2 LnP_t + u_{2t} \qquad t = 1982, \ldots, 1988 \tag{4.9}$$

Si existe cambio estructural rechazaríamos $H_0 : \alpha_1 = \alpha_2, \beta_1 = \beta_2$ y $\gamma_1 = \gamma_2$

En este caso la contestación a la pregunta *Observación en la cual dividir la muestra* contestaríamos *1982* y el output de Gretl muestra lo siguiente:

Modelo 1: estimaciones MCO utilizando las 29 observaciones 1960-1988
Variable dependiente: lnQ

Variable	Coeficiente	Desv. típica	Estadístico t	valor p
const	−4,58987	0,724913	−6,332	0,00001∗∗∗
lnY	0,688498	0,0947276	7,268	0,00001∗∗∗
lnP	0,485683	0,101394	−4,790	0,00006∗∗∗

Media de la var. dependiente = 0,784827
Desviación típica de la var. dependiente. = 0,108499
Suma de cuadrados de los residuos = 0,0949108
Desviación típica de los residuos = 0,0604187
R-cuadrado = 0,712058
R-cuadrado corregido = 0,689908
Estadístico F (2, 26) = 32,148 (valor p < 0,00001)
Estadístico de Durbin-Watson = 1,00057
Coef. de autocorr. de primer orden. = 0,489867
Log-verosimilitud = 41,8214
Criterio de información de Akaike (AIC) = -77,6429
Criterio de información Bayesiano de Schwarz (BIC) = -73,541
Criterio de Hannan-Quinn (HQC) = -76,3582

Contraste de Chow de cambio estructural en la observación 1982 -
Hipótesis nula: no hay cambio estructural
Estadístico de contraste: F(3, 23) = 20,1355
con valor p = P(F(3, 23) > 20,1355) = 1,25619e-006

El estadístico calculado es $F_c = 20,135 > F_{0,05}(3,23)$ por lo que rechazamos H_0 para un nivel de significatividad del 5 %, es decir existe cambio estructural, la campaña institucional ha tenido efecto y la demanda de tabaco en Turquía de 1960 a 1988 queda especificada por las ecuaciones (4.8) y (4.9). Los resultados de la estimación mínimo cuadrática de estas ecuaciones son los siguientes:

$$\widehat{LnQ_t} = \underset{(-10,614)}{-5,024} + \underset{(11,587)}{0,735}\, LnY_t - \underset{(-4,227)}{0,381}\, LnP_t \qquad t = 1960, \ldots, 1981 \quad SCR_1 = 0,01654$$

$$\widehat{LnQ_t} = \underset{(2,170)}{8,837} + \underset{(-1,941)}{-0,953}\, LnY_t + \underset{(0,654)}{0,108}\, LnP_t \qquad t = 1982, \ldots, 1988 \quad SCR_2 = 0,00965$$

Cambio estructural utilizando variables ficticias

Alternativamente, el contraste anterior podríamos haberlo realizado mediante la variable ficticia $D82$ especificando el modelo con $t = 60, \ldots, 88$:

$$LnQ_t = \beta_1 + \beta_2 LnY_t + \beta_3 LnP_t + \beta_1^\star D82_t + \beta_2^\star D82_t \cdot LnY_t + \beta_3^\star D82_t \cdot LnP_t + u_t \qquad (4.10)$$

En el cual, si existe cambio estructural rechazaríamos $H_0 : \beta_1^\star = \beta_2^\star = \beta_3^\star = 0$. De nuevo el contraste puede realizarse con el estadístico F habitual de sumas residuales donde el modelo no restringido es el (4.10) y el modelo restringido es

$$LnQ_t = \beta_1 + \beta_2 LnY_t + \beta_3 LnP_t + u_t \qquad (4.11)$$

Utilizando Gretl el proceso después de abierto el fichero de datos, tomado logaritmos y construido las variables $D82 \cdot LnY$ y $D82 \cdot LnP$ sería: estimaríamos el modelo (4.10) por MCO y en la ventana de resultados de la estimación haríamos

Contrastes $\longrightarrow$ *Omitir variables*

elegiríamos $D82$, $D82 \cdot LnY$ y $D82 \cdot LnP$ y obtendríamos el siguiente resultado:

Modelo 1: estimaciones MCO utilizando las 29 observaciones 1960-1988
Variable dependiente: lnQ

Variable	Coeficiente	Desv. típica	Estadístico t	valor p
const	−4,58987	0,724913	−6,332	0,00001∗∗∗
lnY	0,688498	0,0947276	7,268	0,00001∗∗∗
lnP	0,485683	0,101394	−4,790	0,00006∗∗∗

Media de la var. dependiente = 0,784827
Desviación típica de la var. dependiente. = 0,108499
Suma de cuadrados de los residuos = 0,0949108
Desviación típica de los residuos = 0,0604187
R-cuadrado = 0,712058
R-cuadrado corregido = 0,689908
Estadístico F (2, 26) = 32,148 (valor p < 0,00001)
Estadístico de Durbin-Watson = 1,00057
Coef. de autocorr. de primer orden. = 0,489867
Log-verosimilitud = 41,8214
Criterio de información de Akaike (AIC) = -77,6429
Criterio de información Bayesiano de Schwarz (BIC) = -73,541
Criterio de Hannan-Quinn (HQC) = -76,3582

Comparación entre el modelo 10 y el modelo 11:
Hipótesis nula: los parámetros de regresión son cero para las variables
D82
D82Y
D82P
Estadístico de contraste: F(3, 23) = 20,1355, con valor p = 1,25619e-006
De los 3 estadísticos de selección de modelos, 0 han mejorado.

Dado el p-value rechazamos la hipótesis nula para un nivel de significatividad del 5 % y existe cambio estructural. La demanda de tabaco en Turquía de 1960 a 1988 queda especificada por las ecuaciones (4.8) y (4.9).

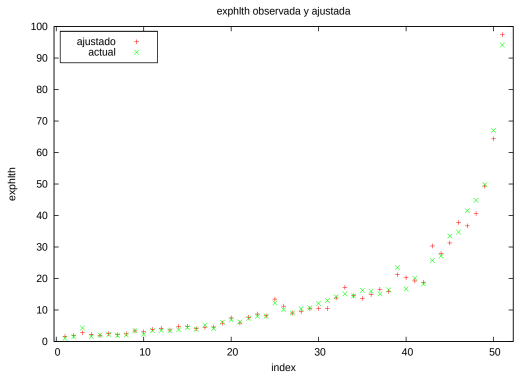
4.3.2. Contraste de heterocedasticidad con Gretl

Para ilustrar esta sección y lo que queda del tema vamos a utilizar el conjunto de datos *Data 3.2* del Ramanathan. En este conjunto de datos se dispone de 51 observaciones sobre renta personal, $INCOME$, y gasto sanitarios, $EXPLTH$ ambos en billones de dólares para el estado de Washington D.C. en 1993. Se trata por tanto de una muestra de sección cruzada. Queremos analizar la evolución del gasto en función de la renta, así especificamos el modelo:

$$EXPLTH_i = \alpha + \beta INCOME_i + u_i \qquad i = 1, \ldots, 51 \tag{4.12}$$

suponemos que se cumplen las hipótesis básicas y estimamos el modelo por MCO con los resultados siguientes[4]:

[4]Recordatorio de la secuencia de órdenes para obtener la estimación:
Archivo ⟶ *Abrir datos* ⟶ *Archivo de muestra* ⟶ *Data3.2*
Modelo ⟶ *Mínimos Cuadrados* ⟶ *seleccionar la variable endógena y exógenas*
Los resultados se muestran en una ventana llamada Gretl:modelo1



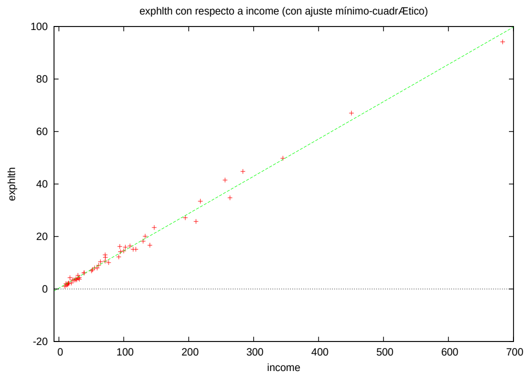


Gráfico 4.6: Gasto sanitario real y ajustado

```
Modelo 2: estimaciones MCO utilizando las 51 observaciones 1-51
Variable dependiente: exphlth

      VARIABLE      COEFICIENTE       DESV.TÍP.      ESTAD.T   2Prob(t > |T|)

   0)     const          0,325608          0,319742       1,018    0,313515
   2)    income          0,142099          0,00196623    72,270   < 0,00001 ***

  Media de la var. dependiente = 15,2649
  D.T. de la var. dependiente = 17,8877
  Suma de cuadrados de los residuos = 148,699
  Desviación típica de los residuos = 1,74203
  R-cuadrado = 0,990705
  R-cuadrado corregido = 0,990516
  Grados de libertad = 49
  Criterio de información de Akaike (AIC) = 203,307
  Criterio de información Bayesiano de Schwarz (BIC) = 207,17
```

Para un nivel de significatividad del 5 % la variable $INCOME$ es significativa, pero el término independiente no es significativamente distinto de cero. El ajuste es muy alto, 99,07 %. Gráficamente podemos ver si la forma funcional lineal elegida resulta adecuada para la relación gasto sanitario-renta. Para ello vamos a dibujar la relación gasto sanitario real y ajustado y además el ajuste MCO. Para ello dentro de la venta Gretl:Modelo1 pinchamos la secuencia:

Gráficos $\longrightarrow$ *Gráfico de ajustada-obervada* $\longrightarrow$ *por número de observación*

Gráficos $\longrightarrow$ *Gráfico de ajustada-observada* $\longrightarrow$ *contra income*

Las figuras obtenidas aparecen a la izquierda y derecha, respectivamente, en el Gráfico 4.3.2. Aparentemente el modelo está correctamente especificado y la forma lineal especificada es adecuada. Antes de seguir vamos a guardar los valores ajustados del gasto sanitario los residuos y sus cuadrados añadiéndolos al conjunto de datos para luego poder trabajar con ellos si es necesario. La secuencia de órdenes a realizar en la ventana Gretl:Modelo1 es:

Datos del modelo $\longrightarrow$ *Añadir al conjunto de datos* $\longrightarrow$ *valores ajustados*

Datos del modelo $\longrightarrow$ *Añadir al conjunto de datos* $\longrightarrow$ *residuos*

Datos del modelo $\longrightarrow$ *Añadir al conjunto de datos* $\longrightarrow$ *residuos al cuadrado*

Gretl los va a añadir al conjunto de datos con el que trabajamos y los denota respectivamente por yhat1, uhat1 e usq1 respectivamente. Además añade una leyenda explicativa de la variable. Una vez hecho esto seguimos con el ejercicio.

A pesar del buen ajuste encontrado, no resulta descabellado pensar que la varianza del gasto sanitario, $EXPHLTH$, probablemente dependerá de la renta $INCOME$. Hemos visto que estudiar el gráfico de residuos frente a $INCOME$ es un instrumento válido para ver indicios del problema. Para obtener el gráfico en la ventana Gretl: Modelo 1 pinchamos:

Gráficos $\longrightarrow$ *Gráfico de residuos* $\longrightarrow$ *contra income*

La figura obtenida se recoge en el Gráfico 4.7:

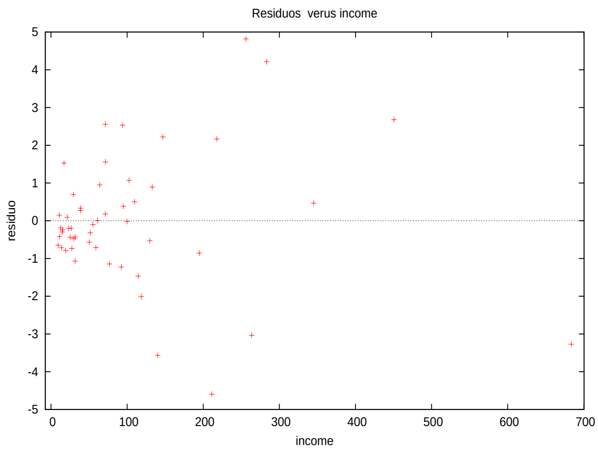


Gráfico 4.7: Residuos MCO versus RENTA

En el se aprecia que la dispersión de los residuos aumenta a medida que aumenta $INCOME$. Parece que la varianza de $EXPHLTH$ aumenta con $INCOME$

Para confirmarlo realizamos el contraste de White y los resultados del mismo son:

```
Contraste de heterocedasticidad de White estimaciones MCO
utilizando las 51 observaciones 1-51 Variable dependiente: uhat^2

      VARIABLE      COEFICIENTE      DESV.TÍP.      ESTAD.T   2Prob(t > |T|)

   0)     const       -1,40227          0,986956     -1,421     0,161839
   2)    income        0,0558410        0,0121349     4,602     0,000031 ***
   4) sq_incom        -5,87208E-05      2,10114E-05  -2,795     0,007445 ***

  R-cuadrado = 0,421177
```

```
Estadístico de contraste: TR^2 = 21,480039, con valor p  =
P(Chi-cuadrado(2) > 21,480039) = 0,000022
```

Los resultados confirman que efectivamente existe heterocedasticidad en la perturbaciones del modelo (4.12). El estimador MCO obtenido no es de varianza mínima y la inferencia realizada de acuerdo a él no es válida.

Esta introducción al problema de heterocedasticidad pretende que hayáis aprendido que nunca se da una especificación por correcta sin un análisis de residuos. Que a pesar de que durante todo el curso hemos trabajado suponiendo que se cumplen unas hipótesis básicas lo habitual es que no sea así y que estas situaciones hay que saber reconocerlas. Ampliaciones sobre este tema podéis encontrar en el Capítulo 8 del Ramanathan del que nosotros solo hemos hecho un esbozo de su introducción.

4.3.3. Contraste de ausencia de correlación con Gretl

Para mostrar cómo contrastar la ausencia de correlación utilizaremos el archivo de datos Ramanathan Data3-3. En este archivo de datos se dispone de 34 observaciones para el periodo 1960-1993 (serie temporal por tanto) sobre el número de resultados de patentes en miles, $PATENTES$, y sobre el gasto en $I + D$, en billones de dólares. La relación a estudiar es:

$$PATENTES_t = \alpha + \beta(I + D)_t + u_t \tag{4.13}$$

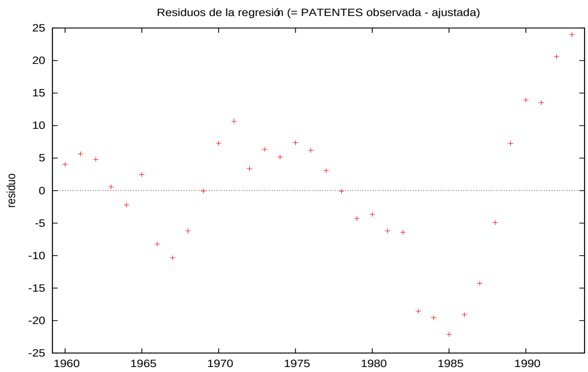
Los resultados de la estimación MCO son los siguientes:

```
 Estimaciones MCO utilizando las 34 observaciones 1960-1993

Variable dependiente: PATENTES
      VARIABLE      COEFICIENTE        DESV.TÍP.      ESTAD.T    2Prob(t > |T|)
   0)    const        34,5711            6,35787        5,438    < 0,00001 ***
   3)      I+D         0,791935          0,0567036     13,966    < 0,00001 ***
  Media de la var. dependiente = 119,238
  D.T. de la var. dependiente = 29,3058
  Suma de cuadrados de los residuos = 3994,3
  Desviación típica de los residuos = 11,1724
  R-cuadrado = 0,859065
  R-cuadrado corregido = 0,854661
  Grados de libertad = 32
  Estadístico de Durbin-Watson = 0,233951
  Coef. de autocorr. de primer orden = 0,945182
  Criterio de información de Akaike (AIC) = 262,541
  Criterio de información Bayesiano de Schwarz (BIC) = 265,593
  Contraste LM de autocorrelación hasta el orden 1 -
  Hipótesis nula: no hay autocorrelación
  Estadístico de contraste: LMF = 120,454
  con valor p  = P(F(1,30) > 120,454) = 5,01763e-012
```

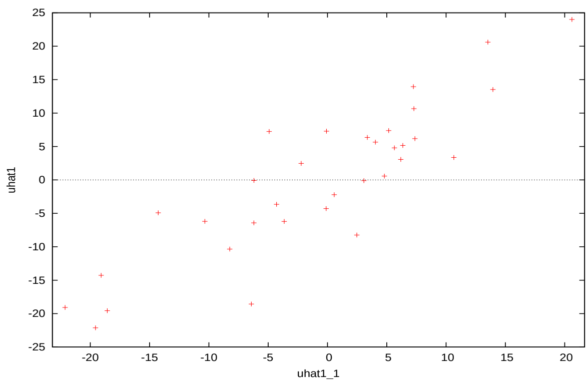
Los resultados muestran que para un nivel de significatividad del 5 % el término independiente es significativamente distinto de cero y el gasto en I+D es una variable significativa para explicar las aplicaciones de las patentes. Además existe un buen ajuste en términos del R^2 (85,9 %). Si analizamos los residuos MCO dibujándolos contra el tiempo obtenemos el Gráfico 4.8. En él podemos ver un primer grupo de residuos positivos que va seguido de un grupo de residuos negativos, otrora positivos y a continuación negativos. Este comportamiento puede indicar la posible existencia de un proceso autorregresivo de primer orden y signo positivo. También podíamos

Gráfico 4.8: Residuos versus tiempo



haber dibujado los pares (u_{t-1}, u_t), ver Gráfico 4.9, en este caso los puntos los encontramos en el primer y tercer cuadrante indicando autocorrelación de primer orden de signo positivo[5].

Gráfico 4.9: Residuos en t versus residuos en t-1



Una vez analizados los gráficos debemos realizar un contraste para cerciorarnos de la existencia del problema. Entre los resultados de la regresión se nos muestra:

```
Estadístico de Durbin-Watson = 0,233951
```

que utilizaremos para contrastar la existencia de un proceso autocorregresivo de primer orden y signo positivo en la perturbación, ya que $DW \in (0, 2)$.

$$H_0 : \rho = 0$$

[5]Para guardar los residuos en la ventana de resultados de la estimación pinchamos
Datos del modelo $\longrightarrow$ *Añadir al conjunto de datos* $\longrightarrow$ *residuos*
y para obtener su retardo u_{t-1} seleccionamos la variable residuos y pinchamos la secuencia
Datos $\longrightarrow$ *Añadir al conjunto de datos* $\longrightarrow$ *retardos de las variables seleccionadas.*

$$H_a : \rho > 0 \quad \text{en} \quad u_t = \rho u_{t-1} + \epsilon_t \quad \epsilon_t \sim (0, \sigma^2_\epsilon)$$

Para ello sólo necesitamos comparar el valor del estadístico con d_L y d_U obtenidos en las tablas correspondientes. Gretl nos proporciona estas tablas en la opción *Utilidades* que aparece en la primera pantalla una vez se abre el programa. La secuencia a pinchar es:

Utilidades $\longrightarrow$ *tablas estadísticas* $\longrightarrow$ *señalar la tabla deseada*

Gretl proporciona las tablas estadísticas de la normal, t (t-student), chi-cuadrado, F (F-snedercor) y DW (Durbin-Watson). En nuestro caso pinchamos en esta última y se nos despliega una ventana que nos solicita el tamaño de muestra. Se lo damos y pinchamos *Aceptar*. Como resultado Gretl nos devuelve una ventana con el valor de d_L y d_U para el tamaño de muestra dado y diferentes valores de K'. Para nuestro ejemplo obtenemos:

```
Valores críticos al 5% del estadístico de Durbin-Watson

                Número de variables explicativas (excluyendo la constante):

            1              2             3              4             5
        dL    dU     dL      dU     dL      dU     dL    dU     dL    dU

n=34  1,39  1,51  1,33  1,58  1,27  1,65  1,21 1,73  1,15 1,81
```

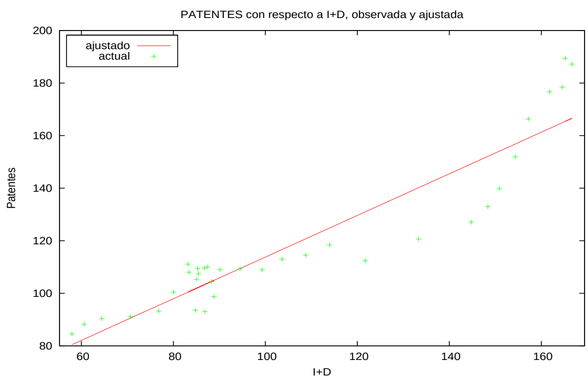
Por tanto para nuestro ejemplo $d_L(T = 34, K' = 1) = 1,39$ y $d_U(T = 34, K' = 1) = 1,51$. Como $DW = 0,233 < d_L$ se rechaza la H_0 y por tanto existe autocorrelación positiva de primer orden o bien

puede deberse a una mala especificación del modelo. Antes de buscar un estimador alternativo a MCO debemos explorar esta posibilidad e intentar especificar bien el modelo y volver a realizar un estudio de existencia de autocorrelación para el modelo correctamente especificado. Si analizamos la relación entre las variables exógena y endógena vemos que esta no parece ser lineal si no cuadrática al menos en los dos últimos tercios de la muestra,

por lo que vamos a proponer la siguiente relación cuadrática:

$$PATENTES_t = \alpha + \beta(I + D)_t + \gamma(I + D)^2_t + u_t \tag{4.14}$$

Gráfico 4.10: Variable endógena versus exógena

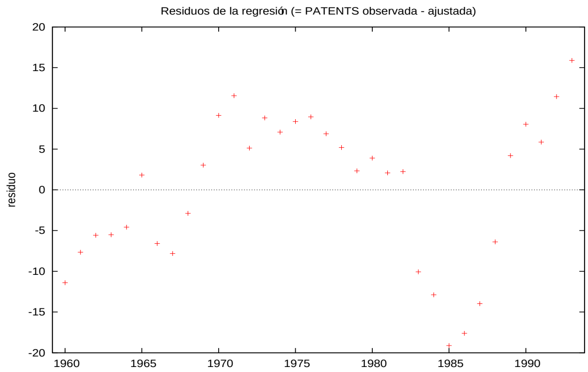


Los resultados de su estimación MCO son:

$$\underset{\text{(t-estad)}}{\widehat{PATENTES}_t} = \underset{(5,23)}{121,57} - \underset{(-1,98)}{0,85}\,(I+D)_t + \underset{(3,85)}{0,007}\,(I+D)_t^2 \quad R^2 = 0,90 \qquad DW = 0,28 \qquad (4.15)$$

Las variables son significativas para un nivel de significatividad del 5 % y el ajuste es bueno, 90 %. Sin embargo, para el modelo (4.15) sigue existiendo autocorrelación positiva de primer orden ya que $DW = 0,28 < d_L(T = 34, K' = 2) = 1,33$. Si miramos el gráfico de residuos de esta relación, Gráfico 4.11, encontramos las misma evolución cíclica de grupos de residuos

Gráfico 4.11: Residuos modelo (2) versus tiempo



positivos-negativos-positivos. Por tanto una vez especificado correctamente el modelo se sigue manteniendo la autocorrelación en las perturbaciones. El modelo (4.14) debe ser estimado por un estimador alternativo a MCO que sea de varianza mínima y permita realizar inferencia válida.

Bibliografía

[1] Alegre, J., J. Arcarons, C. Bolancé y L. Díaz, (1995), Ejercicios y problemas de Econometría, Ed. AC, Colección Plan Nuevo, Madrid.

[2] Alonso, A., F.J. Fernández e I. Gallastegui (2005), Econometría, Prentice Hall, Madrid.

[3] Aznar, A. y A. García (1984), Problemas de Econometría, Pirámide, Madrid.

[4] Belsley, D. A., E. Kuh y R. E. Welsch (1998), Regression Diagnostics: Identifying influential data and sources of collinearity, John Wiley, New York.

[5] Esteban, M.V. (2007), *Estadística Actuarial: Regresión*, Material docente. Servicio de Publicaciones.

[6] Esteban, M.V. (2007), *Colección de ejercicios y exámenes.* Material docente. Servicio de Publicaciones.

[7] Fernández, A., P. González, M. Regúlez, P. Moral y M. V. Esteban (2005), Ejercicios de Econometría, 2ª edn., MacGraw-Hill, serie Schaum, Madrid.

[8] Greene, W. (1998), Análisis Econométrico, 3ª edn., Prentice Hall, New Jersey.

[9] Gretl. *Paquete Econométrico*, disponible en http://gretl.sourceforge.net. Existe versión en castellano, inglés y euskera.

[10] Gretl. *Manual para el usuario*, disponible en http://gretl.sourceforge.net, Existe versión en castellano.

[11] Gujarati, D. (1990), Econometría, 2ª edn., MacGraw-Hill, Madrid.

[12] Johnston, J y J. Dinardo (2001), Métodos de Econometría, Vicens Vives, Barcelona.

[13] Novales, A. (1993), Econometría, Edición revisada, McGraw-Hill, Madrid.

[14] Ramanathan, R. (2002), Introductory Econometrics with applications, 5th. edition, Ed. South-Western, Mason, Ohio.

[15] Uriel, E., D. Contreras, L. Moltó y A. Peiro (1990), Econometría. El modelo lineal, Ed. AC, Madrid.

[16] Wooldridge, J. M. (2003), Introductory Econometrics: A modern Approach, 2nd. edition, Thomson Learning, Mason, Ohio.