Sarriko-On

Análisis de datos: un enfoque econométrico

ISBN: 978-84-693-8458-9

Maria Victoria Esteban González
Marta Regúlez Castillo

04-10

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

eman ta zabal zazu
Universidad del País Vasco
Euskal Herriko Unibertsitatea

Análisis de datos: un enfoque econométrico

Autores:

M. Victoria Esteban Gonzalez
Marta Regúlez Castillo

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

Presentación

El objetivo de este documento es presentar un conjunto de técnicas econométricas avanzadas para la estimación de modelos lineales en situaciones donde las hipótesis estadísticas de comportamiento habituales no se cumplen. Estas notas se estructuran en cinco temas más un tema introductorio y de contextualización del curso y un tema final con orientaciones dirigidas al desarrollo por parte de los alumnos de un proyecto final donde se muestre la evolución de un caso práctico de interés. A través de los temas se van relajando las hipótesis básicas sobre la perturbación aleatoria y sobre la matriz de regresores. El tema introductorio revisa los conceptos de Teoría Asintótica que los alumnos ya han visto en las asignaturas de Estadística. Muestra los diferentes conceptos de convergencia y el Teorema de Mann y Wald adiestrando al alumno en su utilidad para derivar las propiedades en muestras grandes y distribución asintótica de los diferentes estimadores que verán en el curso.

El tema uno introduce el concepto de perturbaciones esféricas y muestra las consecuencias en las propiedades del estimador Mínimo Cuadrático Ordinario de que las perturbaciones no cumplan las hipótesis básicas. Asimismo deriva el estimador Mínimo Cuadrático Generalizado. Los temas dos y tres analizan los problemas de heterocedasticidad y autocorrelación, respectivamente. Muestran como detectar perturbaciones no esféricas y como contrastar la existencia de heterocedasticidad y/o autocorrelación. Aplican el estimador Mínimo Cuadrático Generalizado en el caso de que sea necesario y enseñan cómo estimar cuando la matriz de varianzas y covarianzas de la perturbación es desconocida utilizando el estimador de Mínimos Cuadrados Generalizados Factibles.

En el tema cuatro se relaja la hipótesis básica sobre la matriz de regresores. Se aborda el escenario en que la matriz de datos es estocástica analizando los diferentes estadios de relación entre los regresores estocásticos y la perturbación aleatoria. Se deriva el estimador de Variables Instrumentales y se muestra la utilidad del contraste de Hausman. En el quinto tema se intenta relacionar todos los temas anteriores para lo cual se abordan modelos con dinámica en la parte sistemática y/o dinámica en la perturbación.

En cada tema se muestran ejemplos que ilustran los diferentes escenarios de trabajo así como se recomienda la realización de ejercicios. Se incluye también preguntas evaluativas de los contenidos desarrollados. Al término de cada tema se muestra la bibliografía correspondiente. Al final del documento aparece la bibliografía completa.

El Espacio Europeo de Educación Superior. Los créditos ECTS y los distintos tipos de docencia

El diseño de este curso y la organización de sus contenidos cumple con los criterios de la declaración de Bolonia, que tiene como ejes fundamentales el proceso de enseñanza-aprendizaje y la adquisición no sólo de conocimientos, sino también, y fundamentalmente, de destrezas. Esta adaptación al Espacio Europeo de Educación Superior (EEES) ha permitido basar la metodología docente del curso en clases magistrales (CM), las clases prácticas de aula (PA), las clases prácticas del Centro de Cálculo (PO), los seminarios (S) y los talleres (TA).
Cada una de estas clases conlleva una hora de trabajo presencial. Con respecto a las horas no presenciales por cada clase recibida, es decir de tiempo de trabajo no presencial del alumno, las equivalencias son las siguientes: cada clase magistral implica una hora de trabajo no presencial del alumno, mientras que cada clase práctica de aula, práctica de laboratorio informático, de taller o de seminario implica dos horas de trabajo no presencial del alumno. Este curso conlleva 50 horas de trabajo presencial y 70 horas de trabajo no presencial. Tiene asignados 5 créditos ECTS. En este curso cada tipo de docencia tiene las siguientes características:

- **Clases Magistrales:** se utilizan para transmitir conocimientos teóricos a grupos numerosos de alumnos. En estas clases el protagonista es el profesor y en ellas se exponen los contenidos teóricos junto a ilustraciones prácticas en pizarra o utilizando el ordenador.

- **Prácticas de Aula:** en estas clases se aborda el componente práctico de la asignatura. Son clases participativas en las que esperamos que el alumno acuda con los ejercicios realizados en tiempo no presencial. Con antelación se proporciona al alumno los enunciados de los ejercicios o casos a resolver y en la clase práctica se solucionan las dudas o escollos encontrados por los alumnos a la hora de resolver el ejercicio. El número de alumnos en estas prácticas de aula es la mitad que en las clases magistrales. Por ello, es fácil hacer que el alumno participe resolviendo parte del ejercicio en la misma clase consultando dudas directamente al profesor.

- **Prácticas de Ordenador:** son sesiones docentes en las que, en un aula informática, un grupo de alumnos, bajo la dirección de un profesor, realiza una actividad práctica programada que requiere el uso del ordenador. Dependiendo de la disponibilidad de ordenadores y de la naturaleza de las prácticas el grupo se encontrará subdividido. El software informático constituye en este tipo de práctica la herramienta de trabajo fundamental. Las prácticas de ordenador permiten al alumno un afianzamiento de los contenidos teóricos del curso de Econometría como la puesta en práctica de casos reales con la utilización del software gretl[1].
  gretl es software libre especialmente dirigido hacia la práctica de la econometría y la estadística, muy fácil de utilizar. Ha sido elaborado por Allin Cottrell (Universidad Wake Forest) y existen versiones en inglés, castellano y euskera, además de en otros idiomas. Junto con el programa se pueden cargar los datos utilizados como ejemplos de aplicaciones econométricas en los siguientes libros de texto Davidson y Mackinnon (2004), Greene (2008), Gujarati (1997), Ramanathan (2002), Stock y Watson (2003), Verbeek (2004), Wooldridge (2003). Al instalar

[1] Acrónimo de *Gnu Regression, Econometric and Time Series* (Biblioteca Gnu de Regresión Econometría y Series Temporales)

gretl automáticamente se cargan los datos utilizados en Ramanathan (2002) y Greene (2008). El resto se pueden descargar de la página:

$$http://gretl.sourceforge.net/gretl\_data.html$$

en la opción *textbook datasets*. Este curso se estructura sobre casos prácticos presentados en Ramanathan (2002) y en Wooldridge (2003) y ejercicios a resolver con ayuda de gretl.

También da acceso a bases de datos muy amplias, tanto de organismos públicos, como el Banco de España, como de ejemplos recogidos en textos de Econometría. En la página

$$http://gretl.sourceforge.net/gretl\_espanol.html$$

se encuentra la información en castellano relativa a la instalación y manejo del programa. También hay versiones de esta ayuda en euskera y en inglés.

Una página web interesante sobre las posibilidades del programa para el aprendizaje de Econometría es:

http://www.learneconometrics.com/gretl.html

- **Los Talleres (No Industriales):** en los talleres los alumnos trabajan en modo cooperativo un caso práctico bajo la tutela del profesor. Con anterioridad al taller se proporciona el enunciado del caso o situación a analizar con el fin de que de forma individual los alumnos puedan reflexionar sobre la información proporcionada y como enfrentar la solución al problema planteado. En el aula los miembros del grupo tendrán que ayudarse en la dirección de solucionar el caso propuesto, argumentando la adecuación de las decisiones tomadas y siendo capaces de defenderlas ante el resto del grupo al final de la sesión. para ello será necesario que cada equipo de alumnos nombre un portavoz que puede ir cambiando a lo largo del curso.

- **Los Seminarios y el Proyecto:** el desarrollo de un trabajo o proyecto por parte de un alumno o grupo de alumnos es un tipo de docencia esencial para facilitar la evaluación continua del alumno y conocer el rendimiento de autoaprendizaje. Sin ella no hay metodología docente orientada a la autoformación. Algunas de las más apreciadas habilidades que debe desarrollar el alumno: presentar y exponer un trabajo, resumir y realizar un análisis, trabajar en grupo... se consiguen precisamente gracias a la presentación de un proyecto. El proyecto permite el trabajo en equipo de un reducido grupo de alumnos. Al inicio del curso se fijará el tema del proyecto y se irá desarrollando a lo largo del mismo utilizando para ello los seminarios.
  Los seminarios permiten un tipo de docencia que facilita la interacción fluida entre un profesor y un reducido grupo de alumnos. Se utilizan para que el alumno dé cuenta al profesor, en presencia de los compañeros con los que ha compartido el trabajo, de las tareas prácticas que conlleva el desarrollo del proyecto. Algunos seminarios se pueden llevar a cabo en el Centro de Cálculo para que los alumnos puedan trabajar y consultar dudas sobre el desarrollo del proyecto directamente al profesor. Los seminarios finales serán reservados para presentar trabajos al resto del grupo. Sobre como llevar a cabo este proyecto hablaremos en una de las secciones siguientes.

Las competencias específicas de la asignatura y la evaluación

"Lo que escucho olvido, lo que veo recuerdo, lo que hago entiendo" (Proverbio Chino)

Toda la organización de la metodología docente junto con el diseño de los contenidos de los temas del curso van dirigidos a que los alumnos alcancen las siguientes competencias específicas de la asignatura:

1. Comprender la importancia de los supuestos empleados en la especificación de un modelo econométrico básico para poder proponer y emplear supuestos más realistas.
2. Diferenciar distintos métodos de estimación y evaluar su uso de acuerdo a las características de las variables económicas de interés para obtener resultados fiables.
3. Utilizar diversas fuentes estadísticas y adquirir destreza en el uso de un software econométrico para analizar relaciones entre variables económicas.
4. Elaborar en grupos de trabajo y exponer en público, un proyecto empírico donde se valore adecuadamente los resultados obtenidos del análisis de un modelo econométrico.

El sistema actual de docencia dentro del EEES tiene como ejes fundamentales el proceso de enseñanza-aprendizaje y la adquisición no sólo de conocimientos, sino también, y fundamentalmente, de destrezas implica directamente la valoración del trabajo diario del alumno y su evolución en la adquisición de las competencias. La utilización de la evaluación continua en la evaluación de los alumnos implica la realización en clase, en general con componente de sorpresa, es decir sin previo aviso, de test rápidos o de preguntas cortas en relación a todo lo visto en las clases, conceptos teóricos y ejercicios prácticos tanto de práctica de aula como de práctica de ordenador que permitan evaluar individualmente al alumno y saber si han aprendido los procedimientos adecuados y ver si han alcanzado así las competencias específicas, en nuestro caso las competencias (1), (2) y (3). La evaluación del proyecto permitirá juzgar la competencia (4). Por ello en estas notas se incluye al final de cada tema ejemplos de preguntas cortas evaluativas de los contenidos de las CM, PA y PO a modo de ejemplo de lo que el alumno podría encontrarse en su propia evaluación.

Como se indicaba anteriormente estas notas sirven de apoyo al estudio. Analizan los problemas en profundidad y permiten al alumno profundizar en los temas que conforman el contenido del curso. Así mismo tienen una fuerte vertiente práctica que permitirá al alumno no solo saber sino también saber hacer. En ningún caso deben utilizarse como sustituto de los libros incluidos en la bibliografía. De igual manera se recomienda 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 y alcanzar las competencias específicas de la misma.

Las notas tienen como objetivo servir de apoyo al proceso de aprendizaje de los estudiantes de la asignatura *Econometría* de los Grados en Economía, Administración y Dirección de Empresas, Marketing, Fiscalidad y Administración Pública, y Finanzas y Seguros así como de las Licenciaturas en Economía y Administración y Dirección de Empresas, ambas en extinción. Así mismo sirven de apoyo a estudiantes de master por ejemplo el Master Universitario en Economía: Instrumentos del Análisis económico o el Máster Universitario en Banca y Finanzas Cuantitativas.

Contenido

Figuras

Tablas

Tema 0

Introducción y Contextualización

0.1. Introducción

La asignatura de Econometría avanza en el estudio de la materia profundizando en tópicos como: la relajación de las hipótesis básicas sobre la perturbación, los regresores estocásticos y los modelos dinámicos. Todos ellos conceptos con dificultad relativa. Es por ello, que merece la pena dedicar un tiempo a contextualizar los conceptos que van a estudiarse en esta asignatura recordando algunos vistos en clases de estadística o en econometría no tan avanzada en relación a la estimación e inferencia del modelo de regresión lineal. Este tema introductorio pretende situarnos en los nuevos marcos a trabajar en el resto del curso. Le dedicaremos tres clases magistrales en las que recordamos los conceptos teóricos y los ilustraremos con ejemplos para su mejor comprensión. Al final del tema se proponen algunos ejercicios que sería interesante resolver para fijar los conceptos.

La inferencia estadística es el área que trata sobre los procedimientos que permiten utilizar la información contenida en los datos muestrales de forma eficaz, para obtener información sobre la población de la que provienen o sobre el proceso que los ha generado. Supongamos que existe un proceso desconocido que ha generado los datos muestrales descrito mediante una función de distribución que se caracteriza por un conjunto de parámetros. Sea una muestra de T variables aleatorias $Z = (Z_1, Z_2, \ldots, Z_T)$, cuya función de densidad conjunta $f(Z;\theta)$ depende de un vector de parámetros $\theta = (\theta_1, \theta_2, \ldots, \theta_k) \in \Theta$. Es habitual suponer que tanto la forma matemática de la función de densidad f como el espacio de parámetros Θ son conocidos, por lo que queda perfectamente determinado el modelo estadístico generador de las variables aleatorias Z. La inferencia estadística consiste en usar los datos muestrales de Z para inferir el valor del vector de parámetros θ. Para ello utilizaremos un **estimador** o regla.

Un **estimador por punto** es un estadístico calculado a partir de la muestra que nos proporciona un único valor para cada parámetro desconocido del vector de parámetros. Un **estimador por intervalo** proporciona un rango de valores que contiene el verdadero parámetro con una probabilidad determinada. Un estimador es una variable aleatoria ya que es una función de variables aleatorias. Sustituidos los valores muestrales en el estadístico obtenemos una **estimación**, o valor en la muestra del estimador. Es importante distinguir entre estimador, una función luego variable aleatoria y estimación, un número o realización en la muestra del estimador.

0.2. Propiedades de un estimador. Muestras finitas versus muestras grandes

Estimadores de los parámetros desconocidos contenidos en θ podemos encontrar muchos, pero solo algunos serán estimadores adecuados por lo que necesitaremos criterios de comparación entre ellos. En general compararemos estimadores a partir de una variedad de atributos. Las **propiedades en muestras finitas** de los estimadores son aquellos atributos que pueden ser comparados independientemente del tamaño de muestra y por tanto se cumplen independientemente de éste. En ocasiones algunas características de los estimadores no serán conocidas para muestras de tamaño finito, en esta situación compararemos los estimadores mediante sus **propiedades asintóticas**, es decir cuando el tamaño de muestra crece y tiende a infinito.

¿Qué estimadores nos interesan? En principio querremos estimadores que con alta probabilidad proporcionen estimaciones que estén cerca del verdadero valor del parámetro. Las propiedades en muestras finitas que en general se desean en un estimador son: que sea insesgado, es decir que en media coincida con el verdadero valor del parámetro y que tenga varianza mínima.

- **Insesgadez:** Se dice que un estimador $\hat{\theta}$ es insesgado si $E(\hat{\theta}) = \theta$.
- **Eficiencia:** Un estimador $\hat{\theta}$ es eficiente dentro de la clase de estimadores insesgados si su varianza es la menor entre todas las varianzas de los estimadores insesgados.

Al reconocer la eficiencia como una propiedad deseable en un estimador estamos dejando excluidos a estimadores sesgados incluso cuando su varianza sea muy pequeña y en ocasiones podríamos tolerar un cierto sesgo a cambio de una varianza reducida. Un criterio que nos permite tener en cuenta estas situaciones de un sesgo tolerable unido a varianza pequeña es el criterio del **Error Cuadrático Medio (ECM)**.

- El **ECM** de un estimador se define:

$$ECM(\hat{\theta}) = E(\hat{\theta} - \theta)^2 = V(\hat{\theta}) + [\text{Sesgo}(\hat{\theta})]^2$$

y elegimos el estimador que lo minimice. Utilizando el ECM vemos que entre dos estimadores insesgados preferimos aquel con menor varianza. Sin embargo, no es cierto que entre un estimador insesgado y otro sesgado tenga que ser mejor el insesgado. El sesgado puede tener menor ECM y ser el preferido si el criterio elegido es el de minimizar el ECM. El ECM permite comparar entre sí estimadores insesgados, sesgados o sesgados con insesgados.

La distribución de un estimador puede cambiar con el tamaño muestral. En ocasiones no es posible obtener cuantitativamente el valor medio de un estimador para saber si es insesgado o no. Lo mismo puede ocurrir con su varianza para un tamaño de muestra dado. En estas situaciones determinar las propiedades analíticas del estimador en muestras finitas es muy complicado y se pasa a estudiar las **propiedades asintóticas**. El conocimiento del comportamiento en el límite de la distribución de un estimador, puede utilizarse para inferir una distribución aproximada para el estimador obtenido en una muestra finita. Para ello necesitaremos conceptos de teoría asintótica.

La teoría asintótica analiza el comportamiento aproximado de un estimador o estadístico cuando $T \rightarrow \infty$. ¿Por qué analizamos esto? Es deseable poder obtener estimadores y estadísticos y conozcamos sus propiedades tal que para muestras finitas: insesgadez, eficiencia y sus distribuciones; pero en muchas ocasiones estas propiedades se conocen a costa de tener que hacer supuestos muy particulares y restrictivos que pueden no satisfacerse y/o ser difíciles de contrastar. Otras veces no es posible encontrar estimadores que tengan propiedades deseables para muestras finitas, las que se satisfacen para un tamaño muestral dado, pero que tienen propiedades deseables en muestras grandes, es decir cuando $T \rightarrow \infty$, cuando el tamaño muestral es suficientemente grande. A estas propiedades las llamaremos propiedades asintóticas o propiedades en muestras grandes.

Algunas propiedades asintóticas deseables en un estimador son la consistencia y la eficiencia asintótica. A continuación definiremos una serie de conceptos y propiedades deseables en un estimador que serán válidos asintóticamente, es decir, cuando el tamaño muestral sea suficientemente grande ($T \rightarrow \infty$).

- **Consistencia:** Se dice que un estimador $\hat{\theta}_T$ es un estimador consistente del parámetro desconocido θ, función de una muestra de tamaño T, es consistente si a medida que el tamaño muestral T aumenta, la sucesión $\{\hat{\theta}_T\} = \{\hat{\theta}_1, \hat{\theta}_2, \ldots\}$ converge en probabilidad al verdadero valor del parámetro θ. Esto es,

$$\forall \epsilon > 0 \qquad \lim_{T \rightarrow \infty} Pr\{|\hat{\theta}_T - \theta| < \epsilon\} = 1$$

  y lo denotamos.

$$\hat{\theta}_T \xrightarrow{p} \theta \qquad \text{ó como} \qquad \text{plim}_T\, \hat{\theta} = \theta$$

  Estas condiciones se corresponden con la Proposición 2.2 Novales, pp.38.

La consistencia es una propiedad asintótica deseable en un estimador ya que analizamos si la sucesión de estimadores $\{\hat{\theta}_T\}$ converge en probabilidad al verdadero valor del parámetro. Intuitivamente: Supongamos que $\hat{\theta}$ es un estimador del parámetro θ, variable aleatoria que es función de la muestra de un tamaño T. Si consideramos la sucesión de variables aleatorias:

$$\begin{array}{ll}
\hat{\theta}_1 \text{ estimador función de la muestra tamaño} & T_1 \\
\hat{\theta}_2 \text{ estimador función de la muestra tamaño} & T_2,\ T_2 > T_1 \\
\vdots & \\
\hat{\theta}_N \text{ estimador función de la muestra tamaño} & T_N,\ T_N > T_{N-1}
\end{array}$$

y así sucesivamente a medida que consideramos muestras de mayor tamaño ($T \longrightarrow \infty$) obtendríamos una sucesión de variables aleatorias $\{\hat{\theta}_T\}$. Si la sucesión $\{\hat{\theta}_T\}$ converge en probabilidad al verdadero valor (desconocido) del parámetro θ se dice que el estimador $\hat{\theta}$ es un estimador consistente.

- **Insesgadez Asintótica:** Diremos que el estimador $\hat{\theta}$ de θ es un estimador insesgado asintóticamente si

$$\lim_{T \rightarrow \infty} E(\hat{\theta}_T) = \theta$$

Es importante notar que consistencia no implica insesgadez asintótica ni viceversa, un estimador consistente no tiene porqué ser insesgado asintóticamente. De igual forma un estimador insesgado asintóticamente no tiene porque ser consistente. Sin embargo, si un estimador es insesgado asintóticamente y además su varianza en el límite es cero el estimador es consistente. La unión de estas dos condiciones, insesgadez asintótica y varianza asintótica que tienda a cero cuando el tamaño de muestra tiende a infinito conforman las condiciones suficientes, aunque no necesarias, de consistencia:

- **Condiciones suficientes de consistencia:** Sea $\{E(\hat{\theta}_T)\}$ una sucesión del momento de primer orden de $\hat{\theta}_T$ cuando el tamaño muestral T aumenta hasta infinito. Sea $\{Var(\hat{\theta}_T)\}$ una sucesión del momento centrado de orden dos de $\hat{\theta}_T$ cuando el tamaño muestral T aumenta hasta infinito. Entonces si:

$$\left.\begin{array}{ll} (1) & \lim\limits_{T \to \infty} E(\hat{\theta}_T) = \theta \\ (2) & \lim\limits_{T \to \infty} Var(\hat{\theta}_T) = 0 \end{array}\right\} \quad \hat{\theta}_T \xrightarrow{p} \theta \Longrightarrow \text{plim}\hat{\theta}_T = \theta$$

  Estas condiciones se corresponden con la Proposición 2.9 Novales pp.38.

En ocasiones demostrar la consistencia de un estimador mediante la definición formal es complicado y resulta más sencillo utilizar las condiciones suficientes de consistencia. Es importante notar que para que un estimador sea consistente no necesitamos conocer los momentos de $\hat{\theta}_T$, es decir $E(\hat{\theta}_T)$ ó $E(\hat{\theta}_T^2)$, ó $E[\hat{\theta}_T^2 - E(\hat{\theta}_T)]$ etc ... ni necesitamos que existan para cada tamaño muestral. Sin embargo, si decidimos probar la consistencia de un estimador mediante las condiciones suficientes de consistencia si necesitamos conocer $E(\hat{\theta}_T)$ y $Var(\hat{\theta}_T)$.

- **Eficiencia Asintótica dentro de una clase:** Si nos limitamos a la clase de estimadores consistentes y asintóticamente normales, diremos que un estimador de esa clase es eficiente asintóticamente, si y sólo si su varianza asintótica es la menor de todas las varianzas asintóticas de los estimadores de esa clase.

Distribución Asintótica: La distribución asintótica es una distribución que se utiliza para aproximar la verdadera distribución muestral de una variable aleatoria, que puede ser perfectamente un estimador. Cuando es difícil o imposible derivar la distribución muestral de un estimador o estadístico estudiaremos su distribución de probabilidad cuando la muestra tiende a infinito. Si a medida que el tamaño de muestra aumenta la distribución del estimador se aproxima a una distribución específica conocida, para tamaños de muestra grandes podemos usar esta distribución como una aproximación a la verdadera distribución del estimador. Esta es la idea que transmite la siguiente definición de convergencia:

- **Convergencia en Distribución:** Sea $Z_1, Z_2, Z_3 \ldots, Z_T$ una sucesión de variables aleatorias definidas conjuntamente con sus respectivas funciones de distribución $F_1(Z_1), F_2(Z_2)$, $F_3(Z_3)$ $\ldots, F_T(Z_T)$.

  Supongamos $F_T(z) \to F(z)$ en todos los puntos de continuidad de la función $F(z)$ y que en ellos, $F(z)$ es una función de distribución. Entonces se dice que la sucesión de variables aleatorias Z_T converge en distribución a Z, donde Z es una variable aleatoria con función de distribución F, y se representa por $Z_T \xrightarrow{d} Z$ (o también como $Z_T \overset{a}{\sim} F(z)$). La distribución $F(z)$ se conoce como Distribución Asintótica o Distribución Límite.

Estas condiciones se corresponden con la Definición 2.3 Novales pp.40.

A continuación se van a enunciar dos teoremas de gran utilidad para demostrar las propiedades asintóticas de un estimador: El Teorema de Mann y el teorema de Cramer. Ambos teoremas serán de utilidada en el marco del MRLG para derivar propiedades asintóticas de ciertos estimadores, en particular consistencia y distribución asintótica.

- **Teorema de Mann y Wald:** Sean X una matriz de orden $T \times K$ y u un vector de dimensión $T \times 1$ tales que:

  i) $E(u) = 0, \quad E(uu') = \sigma_u^2 I_T$

  ii) $E(X_i'u) = 0 \quad i = 1, 2, \ldots, K$, donde X_i es la columna i-ésima de la matriz X. $(\Rightarrow E(X'u) = 0)$

  iii) $\text{plim}\frac{X'X}{T} = Q$ matriz finita, simétrica y definida positiva.

  Si i), ii) y iii) se verifican, entonces tenemos dos resultados:

  a) $\text{plim}\dfrac{X'u}{T} = 0$

  b) $\dfrac{X'u}{\sqrt{T}} \xrightarrow{d} N\left(0, \sigma_u^2 Q\right)$

  Este teorema aparece en Novales pp. 43.

- **Teorema de Cramer:** Sea Y_T una matriz aleatoria de dimensión fija $p \times q$ y sea X_T un vector aleatorio de dimensión $q \times 1$. Entonces, si:

$$\left.\begin{array}{ll} (1) & Y_T \xrightarrow{p} A \quad \text{(matriz de constantes)} \\ (2) & X_T \xrightarrow{d} X \end{array}\right\} \text{se tiene que} \quad Y_T X_T \xrightarrow{d} AX$$

  Este teorema se corresponde con la Proposición 2.16 Novales pp. 41.

0.3. El Modelo de Regresión Lineal General. Estimador Mínimo Cuadrático Ordinario (MCO)

Escribimos el Modelo de Regresión Lineal General:

$$Y_t = \beta_1 X_{1t} + \beta_2 X_{2t} + \ldots + \beta_K X_{Kt} + u_t \quad t = 1, 2, \ldots, T \qquad \Longleftrightarrow \qquad Y_t = X_t'\beta + u_t \tag{1}$$

donde explicamos el comportamiento de la variable endógena Y con un conjunto de K variables exógenas o independientes, $X_{1t}, X_{2t}, X_{3t}, \ldots, X_{KT}$ donde $X_{1t} = 1 \quad \forall t$, con lo que el modelo tiene término constante. La variable aleatoria u se denomina perturbación o error y no es observable. Recoge todo aquello del comportamiento de la variable endógena que no es recogido por las variables exógenas. Los coeficientes $\beta_1, \beta_2, \ldots, \beta_K$ son los parámetros desconocidos de la relación y son los que queremos determinar. Para obtener valores factibles del promedio de Y a partir de una muestra de tamaño T, $t = 1, \ldots, T$ podemos utilizar una estimación por punto o una estimación por intervalo. Un estimador del vector de parámetros $\beta' = (\beta_1, \beta_2, \ldots, \beta_K)$ es una función de la muestra luego es una variable aleatoria.

En general se supone que se cumplen las siguientes hipótesis:

- *Sobre los regresores:*
  (1) Suponemos que X es una matriz de regresores que no son realizaciones de variables aleatorias (en adelante v.a.), es decir, son regresores no estocásticos. Este supuesto se puede entender como un análisis condicionado a unos valores dados de las variables explicativas[1]. Además, sobre los regresores suponemos también que la matriz X es de rango completo por columnas, $rg(X) = K$.
- *Sobre las perturbaciones u*, se las supone:
  (2) media cero, $E(u_t) = 0 \quad \forall t$,
  (3) varianza constante $E(u_t^2) = \sigma^2 \quad \forall t$, es decir, son homocedásticas
  (4) covarianzas cero, $E(u_t u_s) = 0 \quad \forall t, s \quad t \neq s$ luego son no autocorreladas.
  (5) además se supone que siguen una distribución normal, $u_t \sim N(0, \sigma^2)$.
- *Sobre los parámetros* se supone que son constantes durante todo el periodo muestral.
- *Sobre la relación* se la supone lineal en parámetros, o linealizable, además de correctamente especificada.

Las K-ecuaciones recogidas en (1) dan lugar a la siguiente expresión matricial del MRLG:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_T \end{bmatrix} = \begin{bmatrix} 1 & X_{21} & X_{31} & \ldots & X_{K1} \\ 1 & X_{22} & X_{32} & \ldots & X_{K2} \\ 1 & X_{23} & X_{33} & \ldots & X_{K3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{2T} & X_{3T} & \ldots & X_{KT} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \vdots \\ \beta_K \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ \vdots \\ u_T \end{bmatrix}$$

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

[1]En ocasiones este tipo de análisis no será posible y tendremos que considerar que X es una matriz estocástica.

Las hipótesis establecidas sobre la perturbación nos definen el vector de medias y la matriz de varianzas y covarianzas siguientes:

$$E(u) = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \qquad E(uu') = \sigma^2 \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} = \sigma^2 I_T$$

Matricialmente escribimos el modelo junto con los supuestos sobre la perturbación:

$$Y = X\beta + u \qquad u \sim N(0, \sigma^2 I_T)$$

Para lograr nuestro objetivo de estimar los coeficientes $\beta_1, \ldots, \beta_K$ desconocidos proponemos utilizar el criterio de estimación Mínimo Cuadrático Ordinario, MCO, donde se minimiza la Suma de Cuadrados Residual del modelo. Matricialmente el criterio se escribe:

$$\underset{\hat{\beta}}{Min}\; \hat{u}'\, \hat{u} = \underset{\hat{\beta}}{Min}\; (Y - X\hat{\beta})'\, (Y - X\hat{\beta})$$

Las ecuaciones normales que se obtienen de las condiciones de primer orden son:

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

a partir las cuales derivamos el estimador de los parámetros β:

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

Además, en general, la varianza de la perturbación será desconocida. Como estimador de σ^2 proponemos:

$$\hat{\sigma}^2_{MCO} = \frac{\hat{u}'\hat{u}}{T-k}$$

Bajo las hipótesis establecidas el estimador MCO del vector de parámetros desconocidos β es un estimador lineal en la perturbación, insesgado y de varianza mínima en muestras finitas. En muestras grandes o asintóticas es consistente. El estimador propuesto para σ^2, $\hat{\sigma}^2$ es insesgado y consistente.

Ejemplo 0.1 En el Modelo de Regresión Lineal General:

$$Y = X\beta + u$$

donde X es no estocástica y $u \sim NIID(0, \sigma^2 I_T)$ se van a obtener las propiedades y distribución en muestras finitas del estimador MCO del vector de coeficientes β.

Propiedades en muestras finitas:

1. **Linealidad:** Dado que X es no estocástica el estimador MCO es lineal en u, ya que se puede escribir como una combinación lineal del vector de perturbaciones u y la matriz de constantes $D = (X'X)^{-1}X'$, siendo u lo único aleatorio de su expresión:

$$\hat{\beta}_{MCO} = \beta + (X'X)^{-1}X'u = \beta + D\,u$$

   Para demostrar esta propiedad hemos utilizado la hipótesis recogida en (1).

2. **Insesgadez:**

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

   $E(\hat{\beta}_{MCO}) = \beta$, luego el estimador es insesgado.
   Para demostrar esta propiedad hemos utilizado las hipótesis recogidas en (1) y (2).

3. **Matriz de varianzas y covarianzas:**

$$\begin{aligned} V(\hat{\beta}_{MCO}) &= E[(\hat{\beta}_{MCO} - E(\hat{\beta}_{MCO}))(\hat{\beta}_{MCO} - E(\hat{\beta}_{MCO}))'] = \\ &= E((X'X)^{-1}X'(uu')X(X'X)^{-1}) \\ &= (X'X)^{-1}X'E(uu')X(X'X)^{-1} = \sigma^2(X'X)^{-1} \end{aligned}$$

   Para obtener esta matriz de varianzas y covarianzas del estimador MCO hemos utilizado las hipótesis recogidas en (1), (2) y (3).
   El Teorema de Gauss-Markov garantiza que el estimador de MCO tiene la menor varianza entre todos los estimadores lineales e insesgados, luego es **eficiente**.

Distribución en muestras finitas o distribución exacta: Dado que el estimador MCO es lineal en la perturbación sigue la misma distribución que ésta. Como la distribución de u es normal la distribución del estimador MCO es normal. Su distribución exacta o en muestras finitas es:

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

Ejercicio 0.1 En el Modelo de Regresión Lineal General:

$$Y = X\beta + u$$

donde X es no estocástica y $u \sim IID(0, \sigma^2 I_T)$. Obtén las propiedades del estimador MCO del vector de coeficientes β. ¿Es conocida su distribución en muestras finitas? ¿Puedes hacer contraste de hipótesis en muestras finitas?

Ejemplo 0.2 En el Modelo de Regresión Lineal General:

$$Y = X\beta + u$$

donde X es no estocástica y $u \sim NID(0, \sigma^2 I_T)$ vamos a demostrar la consistencia del estimador MCO del vector de coeficientes β y a obtener su distribución asintótica.

Dado que se cumplen las condiciones i), ii) del teorema de Mann y Wald y que suponemos que iii) también se cumple, vamos a utilizar este teorema para demostrar la consistencia del estimador[2]. A continuación derivamos su distribución asintótica.
Consistencia: Mann y Wald proporciona el resultado a) útil para demostrar la consistencia del estimador:

$$\text{resultado } a) \quad \text{plim}\, \frac{X'u}{T} = 0$$

luego

$$\hat{\beta}_{MCO} = \beta + (X'X)^{-1}X'u = \beta + \left(\frac{1}{T}X'X\right)^{-1}\left(\frac{1}{T}X'u\right)$$

$$\text{plim}\, \hat{\beta}_{MCO} = \text{plim}\, \beta + \text{plim}\, \left(\frac{1}{T}X'X\right)^{-1} \text{plim}\, \left(\frac{1}{T}X'u\right) = \beta + Q^{-1} \times 0 = \beta$$

luego el estimador es consistente.

Distribución Asintótica: Dado que el estimador es consistente

$$\hat{\beta}_{MCO} \xrightarrow{p} \beta \rightarrow (\hat{\beta}_{MCO} - \beta) \xrightarrow{p} 0$$

luego la distribución del estimador en el límite es degenerada, es decir, tiene toda la masa probabilística acumulada en un punto, β, su media, mientras que su varianza asintótica es cero. Por ello no vamos a buscar la distribución en el límite de $\hat{\beta}_{MCO}$ si no de una transformación del mismo que no tiene distribución límite degenerada. Esta transformación es $\sqrt{T}(\hat{\beta}_{MCO} - \beta)$. Vamos a comenzar desarrollando esta expresión:

$$\begin{array}{rrl} \hat{\beta}_{MCO} = \beta + (X'X)^{-1}X'u & \Rightarrow \quad \hat{\beta}_{MCO} - \beta & = (X'X)^{-1}X'u \\ & \sqrt{T}(\hat{\beta}_{MCO} - \beta) & = \left(\frac{1}{T}X'X\right)^{-1}\left(\frac{1}{\sqrt{T}}X'u\right) \end{array}$$

y para ella obtenemos la distribución asintótica ayudándonos del resultado b) de Mann y Wald y del Teorema de Cramer.

$$\text{resultado } b) \quad \left(\frac{1}{\sqrt{T}}X'u\right) \xrightarrow{d} N(0, \sigma^2 Q)$$

Por Cramer,
1) Tenemos una variable aleatoria Z que converge en distribución, $Z = \frac{1}{\sqrt{T}}X'u \xrightarrow{d} N(0, \sigma^2 Q)$.
2) Tenemos la matriz Q que converge en probabilidad, luego $\left(\frac{1}{T}X'X\right) \xrightarrow{p} Q$.

[2]En el Anexo del tema se muestra cómo demostrar la consistencia del estimador MCO utilizando la definición de consistencia y sin recurrir al uso del Teorema de Mann y Wald.

Usando 1) y 2) donde $Z \sim N(0, \sigma^2 Q)$, obtenemos $Q^{-1}Z \sim N(0, \sigma^2 Q^{-1})$ luego podemos demostrar también que[3]:

$$
\begin{aligned}
Q^{-1} Z = \left(\frac{1}{T}X'X\right)^{-1}\left(\frac{1}{\sqrt{T}}X'u\right) &\xrightarrow{d} Q^{-1}Z \\
\underbrace{\left(\frac{1}{T}X'X\right)^{-1}\left(\frac{1}{\sqrt{T}}X'u\right)} &\xrightarrow{d} N(0, \sigma^2 Q^{-1}) \\
\sqrt{T}(\hat{\beta}_{MCO} - \beta) &\xrightarrow{d} N(0, \sigma^2 Q^{-1})
\end{aligned}
$$

Notar que a pesar de que por el enunciado del ejercicio se conoce que la perturbación es normal a lo largo de la demostración no se ha utilizado este conocimiento. Luego si desconocemos la distribución de u el resultado de consistencia y la distribución asintótica permanecen invariantes. Este resultado nos va a permitir hacer inferencia sobre β basándonos en esta distribución asintótica de $\sqrt{T}(\hat{\beta}_{MCO} - \beta)$ cuando no conozcamos la distribución exacta para un tamaño de muestra dado (del que nosotros dispondremos) y que será una aproximación mejor cuanto mayor sea el tamaño muestral.

Ejemplo 0.3 Sea el modelo $Y = X\beta + u$ donde X es una matriz de variables no estocásticas, $u \sim (0, \sigma^2 I)$ y $\lim_{T \to \infty} = \left(\frac{1}{T}X'X\right) = Q$ finita, simétrica y definida positiva. Definimos el estimador:

$$\tilde{\beta} = \hat{\beta}_{MCO} + \frac{1}{T}\,\mathbf{c}$$

donde $\mathbf{c} = (c_1, c_2, \ldots, c_K)$ es un vector de constantes no aleatorias, finito, que no depende de T.

1. Obtén la esperanza matemática de $\tilde{\beta}$ y su matriz de varianzas y covarianzas. ¿Cuáles son las propiedades en muestras finitas del estimador $\tilde{\beta}$?
2. Deriva el sesgo del estimador $\tilde{\beta}$. ¿Cómo es el sesgo cuando aumenta T?
3. ¿Son los estimadores $\hat{\beta}_{MCO}$ y $\tilde{\beta}$ consistentes?
4. Busca la distribución asintótica del estimador $\tilde{\beta}$.

Solución:

1.

$$
\begin{aligned}
E(\tilde{\beta}) &= E(\hat{\beta}_{MCO} + \tfrac{1}{T}\,\mathbf{c}) = \beta + \tfrac{1}{T}\,\mathbf{c} \\
V(\tilde{\beta}) &= E[\tilde{\beta} - E(\tilde{\beta})]^2 = E[\tilde{\beta} - (\beta + \tfrac{1}{T}\,\mathbf{c})]^2 = E[\hat{\beta}_{MCO} + \tfrac{1}{T}\,\mathbf{c} - \beta - \tfrac{1}{T}\,\mathbf{c}]^2 \\
&= E(\hat{\beta}_{MCO} - \beta)^2 = \sigma^2 (X'X)^{-1}
\end{aligned}
$$

[3]

$$
\begin{aligned}
E(Q^{-1}Z) &= Q^{-1}E(Z) = 0. \\
V(Q^{-1}Z) &= E((Q^{-1}Z)(Q^{-1}Z)') = Q^{-1}E(ZZ')Q^{-1} = Q^{-1}(\sigma^2 Q)Q^{-1} = \sigma^2 Q^{-1}QQ^{-1} = \sigma^2 Q^{-1}.
\end{aligned}
$$

Luego si $u \sim (0, \sigma^2 I_T)$ el estimador $\tilde{\beta}$ es un estimador lineal en u, es sesgado y su matriz de varianzas y covarianzas coincide con la del estimador de MCO.

2. Sesgo$(\tilde{\beta}) = E(\tilde{\beta}) - \beta = \beta + \frac{1}{T}\,\mathbf{c} - \beta = \frac{1}{T}\,\mathbf{c}$. Si $T \longrightarrow \infty$ entonces el sesgo se hace nulo:

$$\lim_{T \to \infty} \frac{1}{T}\,\mathbf{c} = \begin{bmatrix} \lim\limits_{T \to \infty} \frac{1}{T}\, c_1 \\ \lim\limits_{T \to \infty} \frac{1}{T}\, c_2 \\ \vdots \\ \lim\limits_{T \to \infty} \frac{1}{T}\, c_k \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Además $ECM(\tilde{\beta}) = \sigma^2 (X'X)^{-1} + \left[\frac{1}{T}\,\mathbf{c}\right]^2 > \sigma^2 (X'X)^{-1} = ECM(\hat{\beta}_{MCO})$.

3. Como se ha probado en el Ejemplo 0.3 el estimador $\hat{\beta}_{MCO}$ es consistente. A continuación vamos a probar que el estimador $\tilde{\beta}$ es consistente utilizando las condiciones suficientes de consistencia:

$$\left.\begin{array}{lcl} (1) \lim\limits_{T \to \infty} E(\tilde{\beta}) & = & \lim\limits_{T \to \infty} \beta + \lim\limits_{T \to \infty} \frac{1}{T}\,\mathbf{c} = \beta + 0 = \beta \\ (2) \lim\limits_{T \to \infty} V(\tilde{\beta}) & = & \lim\limits_{T \to \infty} V(\hat{\beta}_{MCO}) = 0 \end{array}\right\} \quad \tilde{\beta} \xrightarrow{p} \beta \Longleftrightarrow \operatorname{plim} \tilde{\beta} = \beta$$

Luego el estimador es consistente.
Además en el límite ambos estimadores coinciden $\tilde{\beta} \xrightarrow{p} \hat{\beta}_{MCO} \xrightarrow{p} \beta$.

4. Distribución asintótica:
En primer lugar vamos a buscar la expresión para $\sqrt{T}(\tilde{\beta} - \beta)$

$$\begin{aligned} \sqrt{T}(\tilde{\beta} - \beta) &= \sqrt{T}(\hat{\beta}_{MCO} + \tfrac{1}{T}\,\mathbf{c} - \beta) = \\ &= \sqrt{T}(\hat{\beta}_{MCO} - \beta) + \sqrt{T}\tfrac{1}{T}\,\mathbf{c} = \\ &= \sqrt{T}(\hat{\beta}_{MCO} - \beta) + \tfrac{1}{\sqrt{T}}\,\mathbf{c} \end{aligned}$$

A continuación buscamos la distribución asintótica de $\sqrt{T}(\hat{\beta}_{MCO} - \beta) + \frac{1}{\sqrt{T}}\,\mathbf{c}$. Dado que:

$$\left.\begin{array}{c} \sqrt{T}(\hat{\beta}_{MCO} - \beta) \xrightarrow{d} N(0, \sigma^2 Q^{-1}) \\ \text{y} \\ \frac{1}{\sqrt{T}}\,\mathbf{c} \longrightarrow 0 \end{array}\right\} \quad \sqrt{T}(\tilde{\beta} - \beta) \xrightarrow{d} N(0, \sigma^2 Q^{-1})$$

Ejercicio 0.2 En el Modelo de Regresión Lineal General:

$$Y = X\beta + u$$

donde X es no estocástica y $u \sim IID(0, \sigma^2 I_T)$. Demuestra la consistencia del estimador MCO utilizando las condiciones suficientes de consistencia.

0.4. Contraste de hipótesis

Si conocemos la distribución de la perturbación y esta es normal, el estimador MCO en muestras finitas tiene distribución conocida normal y es posible derivar estadísticos de contraste válidos, los habituales estadísticos t y F con distribución conocida t-Student y F-Snedecor respectivamente.

Así contrastamos q restricciones lineales bajo $H_0 : R\beta = r$ versus $H_a : R\beta \neq 0$ con el estadístico:

$$(R\hat{\beta} - r)' \, [\, R \, (\widehat{V}(\hat{\beta}) \, R' \,]^{-1} \, (R\hat{\beta} - r) \, / \, q \overset{H_0}{\sim} F_{(q,T-K)}$$

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

Si $q = 1$ podemos escribir los estadísticos y distribuciones anteriores como:

$$\frac{(\, R\hat{\beta} - r)}{\sqrt{R\,\widehat{V}(\hat{\beta})R'}} \overset{H_0}{\sim} t_{(T-K)}$$

Las distribuciones tabuladas de estos estadísticos nos permiten, estadísticamente, distinguir a un nivel de significación elegido si aceptar o no la hipótesis nula dado el valor del estadístico obtenido en la muestra. Si F $> F_{(q,T-K)|\alpha}$ se rechaza la hipótesis nula para un nivel de significatividad α dado. O si t $> t_{(T-K)|\frac{\alpha}{2}}$ se rechaza la hipótesis nula para un nivel de significatividad α dado.

Sin embargo, cuando el estimador no es lineal en u o cuando aún siéndolo la distribución de u es desconocida no es posible derivar la distribución en muestras finitas del estimador MCO. Es en este escenario cuando es verdaderamente útil la teoría asintótica ya que si podemos obtener la distribución asintótica del estimador podremos derivar estadísticos de contraste válidos, aunque asintóticos.

Estamos interesados es realizar contraste de hipótesis de la forma $H_0 : R\beta = r$ en el modelo de regresión lineal $Y = X\beta + u$ bajo los supuestos:

$$\begin{array}{l} E(u) = 0 \\ E(uu') = \sigma^2 I_T \\ \text{plim}\frac{X'X}{T} = Q \ \text{ finita, simétrica, definida positiva y no singular} \end{array}$$

pero no especificamos la función de distribución de u (en particular no suponemos normalidad).

- R es una matriz de constantes conocidas $(q \times K)$.
- r es un vector de constantes conocidas $(q \times 1)$.

Contrastamos q restricciones lineales. Si no suponemos normalidad de u, el estadístico F bajo la $H_0 : R\beta = r$ es:

$$\text{F} = \frac{(R\hat{\beta}_{MCO} - r)'[R(X'X)^{-1}R']^{-1}(R\hat{\beta}_{MCO} - r)/q}{\hat{\sigma}^2_{MCO}}$$

no tiene porqué distribuirse como una F de Snedecor con $(q, T-K)$ grados de libertad, ni tampoco sabemos como se distribuye para un tamaño de muestra dado T.

Sin embargo dado que conocemos la distribución asintótica del estimador MCO:

$$\sqrt{T}(\hat{\beta}_{MCO} - \beta) \xrightarrow{d} N(0, \sigma^2 Q^{-1})$$

podemos derivar el resultado siguiente:

$$\sqrt{T}(R\hat{\beta}_{MCO} - r) \xrightarrow{d} N(0, \sigma^2(RQ^{-1}R'))$$

o lo que es igual

$$(R\hat{\beta}_{MCO} - r)'[\,R\hat{V}(\hat{\beta}_{MCO})R'\,]^{-1}(R\hat{\beta}_{MCO} - r) \xrightarrow{d} \chi^2_{(q)}$$

equivalente a

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

y hacer inferencia asintótica válida con este estadístico. Por lo tanto si el tamaño de muestra es suficientemente grande podemos utilizar este estadístico y aproximar su distribución por la distribución asintótica $\chi^2_{(q)}$. Rechazaremos la hipótesis nula si el valor del estadístico obtenido para la muestra utilizada es mayor que un valor crítico, elegido un valor de significación α.

• Para el caso en que $q = 1$ si no suponemos normalidad podemos utilizar:

$$\frac{R\hat{\beta}_{MCO} - r}{\hat{\sigma}\sqrt{R(X'X)^{-1}R'}} \xrightarrow{d} N(0,1)$$

Luego para un tamaño de muestra T dado si éste es suficientemente grande podemos aproximar la distribución exacta de este estadístico por la distribución asintótica $N(0,1)$. Por tanto, para un nivel de significación elegido α, rechazaremos la hipótesis nula si el valor obtenido dada nuestra muestra de este estadístico es mayor que el valor crítico $N(0,1)_{|\frac{\alpha}{2}}$.

Ejemplo 0.4 En el modelo $Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t$ donde X_{2t}, X_{3t} son no estocásticas y $u_t \sim NIID(0, \sigma^2)$. Utilizando el estimador MCO, ¿cómo contrastamos $H_0 : \beta_3 = 0$ contra $H_a : \beta_3 \neq 0$?

En este modelo los regresores son no estocásticos y la perturbación es homocedástica y no autocorrelada el estimador de MCO de los parámetros desconocidos es lineal en u, insesgado y de varianza mínima. Además dado que conocemos que la perturbación se comporta como una normal tenemos que

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

luego

$$\hat{\beta}_3 \sim N(\beta_3, \sigma^2 a_{33})$$

donde $V(\hat{\beta}_3) = \sigma^2 a_{33}$ y a_{33} es el elemento correspondiente en la diagonal de $(X'X)^{-1}$. Dado que en muestras finitas el estimador tiene una distribución exacta conocida el contraste puede ser realizado en muestras finitas con el estadístico y distribución siguientes:

$$t = \frac{\hat{\beta}_{i,MCO}}{\hat{\sigma}\sqrt{a_{ii}}} \overset{H_0}{\sim} t_{(T-K)}$$

donde $\widehat{des}(\hat{\beta}_3) = \hat{\sigma}\sqrt{a_{33}}$, $\hat{\sigma}^2 = \frac{\hat{u}'\hat{u}}{T-K}$ y a_{33} es el elemento correspondiente en la diagonal de $(X'X)^{-1}$. Para un nivel de significación α elegiremos el valor crítico con el que comparar el valor muestral obtenido del estadístico, mirando las tablas de la t-Student. No aceptamos H_0 al nivel de significación del 5 % si el valor absoluto del estadístico obtenido con la muestra utilizada es mayor que el valor crítico $t_{(T-K)|0,025}$.

Ejemplo 0.5 En el modelo $Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t$ donde X_{2t}, X_{3t} son no estocásticas y $u_t \sim ID(0, \sigma^2)$. Utilizando el estimador MCO, ¿cómo contrastamos $H_0 : \beta_3 = 0$ contra $H_a : \beta_3 \neq 0$? En este caso dado que los regresores son no estocásticos y la perturbación es homocedástica y no autocorrelada el estimador de MCO de los parámetros desconocidos es lineal en u, insesgado y de varianza mínima. Sin embargo, como no conocemos la distribución de la perturbación desconocemos la distribución exacta del estimador. No es posible hacer inferencia en muestras finitas. Por otro lado, se cumple el Teorema de Mann y Wald por lo que podemos hacer inferencia asintótica con el estadístico y distribución siguientes:

$$t = \frac{\hat{\beta}_{3,MCO}}{\hat{\sigma}\sqrt{a_{33}}} \xrightarrow{d} N(0,1)$$

Para un nivel de significación α elegiremos el valor crítico con el que comparar el valor muestral obtenido del estadístico, mirando las tablas de la normal $N(0,1)$. Por ejemplo, a un nivel $\alpha = 5\,\%$ $(\alpha = 0,05)$ dado que la hipótesis alternativa está definida a dos colas tomamos $\frac{\alpha}{2} = 0,025$ luego $\Phi(1,96) = 1 - 0,025 = 0,975$. No aceptamos H_0 al nivel de significación del 5 % si el valor absoluto del estadístico obtenido con la muestra utilizada es mayor que el valor crítico $N(0,1)_{|0,025} = 1,96$.

Ejercicio 0.3 En el modelo $Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t$ donde X_{2t}, X_{3t} son no estocásticas y $u_t \sim NID(0, \sigma^2)$. Utilizando el estimador MCO, ¿cómo contratarías la hipótesis nula de significatividad conjunta? ¿Y en ausencia del supuesto de normalidad en la perturbación?

0.5. ¿Qué vamos a aprender en este curso?

Decíamos al inicio de este tema que la asignatura de Econometría avanza en el estudio de la materia profundizando en tópicos como: la relajación de las hipótesis básicas sobre la perturbación, los regresores estocásticos y los modelos dinámicos. Para abordar estos tópicos es necesario que relajemos algunas o todas la hipótesis establecidas sobre el comportamiento de las variables del modelo de regresión lineal general enunciadas en la sección 0.3.

Comenzaremos el curso relajando hipótesis sobre el comportamiento de la perturbación. Si en vez de suponer que u es homocedástica y no autocorrelada consideramos la posibilidad de que su varianza no sea constante y/o que sus covarianzas no sean todas cero nuestro estimador de MCO no será de varianza mínima. En este marco más general derivaremos las propiedades del estimador MCO y

las consecuencias sobre el contraste de hipótesis. Estudiaremos el estimador de Mínimos Cuadrados Generalizados (MCG) y de Mínimos Cuadrados Generalizados Factibles (MCGF) cuando MCG no se pueda obtener dado que requiere el conocimiento de $E(uu')$. Analizaremos las propiedades de ambos y como realizar inferencia.
En los Temas 2 y 3 nos emplearemos a fondo en el diseño de $E(uu')$ bajo supuestos concretos del comportamiento de la perturbación que nos permitan considerar que o bien $E(u_t^2)$ no es constante y/o $E(u_t u_s)$ no es cero en todos los períodos. En este tema abordaremos por tanto contrates de existencia de heterocedasticidad y de existencia de autocorrelación.

Una situación diferente se produce si relajamos el supuesto de que los regresores son no estocásticos y suponemos que al menos una de las variables X_{it} incluidas en la matriz de regresores X es estocástica. En este marco de trabajo el estimador MCO del vector de coeficientes desconocidos β no es lineal en la perturbación ya que es una combinación no lineal de la matriz de regresores estocástica junto con el vector de variables aleatorias u. En este caso la media y varianza del estimador $\hat{\beta}$ dependen de la relación entre X y u, de los supuestos recogidos en su distribución conjunta. Bajo ciertos supuestos el estimador seguirá siendo insesgado y de varianza mínima pero su distribución en muestras finitas será desconocida ya que al ser $\hat{\beta}_{MCO}$ no lineal en u no podemos garantizar que tenga una distribución normal incluso en el caso en que u lo fuera. En esta situación debemos centrarnos en las propiedades asintóticas del estimador y si este es consistente y tiene distribución asintótica conocida podremos hacer inferencia asintótica.
Cómo trabajar en esta situación es lo que veremos en el Tema 4: Regresores Estocásticos. En este tema nuestro trabajo se centrará en el cumplimiento del Teorema de Mann-Wald. Como hemos dicho anteriormente si los requisitos i), ii) y iii) de Mann y Wald se cumplen el estimador de MCO es consistente y tiene distribución asintótica conocida con la que realizar inferencia asintótica. Sin embargo hay muchas situaciones en que estos requisitos no se cumplen. Por ejemplo si X_{it} y u_t estan correladas, $E(X'u) \neq 0$ y el estimador MCO no será consistente. En este tema mostraremos el estimador de Variables Instrumentales de utilidad en esta situación.

En el Tema 5 y último del curso daremos entrada a los Modelos Dinámicos. Consideraremos la posibilidad de que haya dinámica tanto en los regresores como en la perturbación y aplicaremos lo aprendido en los cuatro temas anteriores.

Durante el curso trabajaremos con un buen número de estimadores. El estimador de MCO será nuestra referencia, cuando mantenga sus propiedades en muestras finitas y/o sus propiedades asintóticas estimaremos por este criterio. Sin embargo en aquellas circunstancias en que las hipótesis estadísticas establecidas no se cumplan quizá debamos de cambiar de estimador, de ahí que aprendamos nuevos estimadores como Mínimos Cuadrados Generalizados, Mínimos Cuadrados Generalizados Factibles, Variables Instrumentales o Mínimos Cuadrados en dos Etapas. Las propiedades estadísticas de los estimadores dependen tanto del procedimiento de estimación utilizado como de las hipótesis estadísticas establecidas. Por ello será fundamental conocer el funcionamiento y propiedades de los estimadores anteriores para utilizarlos en el marco adecuado; eso es lo que se ha de aprender tal como se refieren las dos primeras competencias específicas de la asignatura a adquirir:

1. Comprender la importancia de los supuestos empleados en la especificación de un modelo econométrico básico para poder proponer y emplear supuestos más realistas.

2. Diferenciar distintos métodos de estimación y evaluar su uso de acuerdo a las características de las variables económicas de interés para obtener resultados fiables.

Las dos competencias restantes básicamente tienen el mismo objetivo, pero por una lado requieren que adaptéis esos conocimientos aprendiendo a utilizar un software estadístico para aplicar los diferentes estimadores y por otro que seáis capaces de aunar todo lo aprendido y darlo forma en un trabajo final.

3. Utilizar diversas fuentes estadísticas y adquirir destreza en el uso de un software econométrico para analizar relaciones entre variables económicas.
4. Elaborar en grupos de trabajo y exponer en público, un proyecto empírico donde se valore adecuadamente los resultados obtenidos del análisis de un modelo econométrico.

La docencia de la asignatura se basará en clases magistrales (M), clases prácticas de aula (PA), clases prácticas del Centro de Cálculo (PO), seminarios (S) y talleres (TA). Cada hora de estas clases equivale a una hora presencial. En el cronograma de la asignatura podéis ver la distribución de las mismas.

Con respecto a las horas no presenciales por cada clase recibida, es decir de tiempo de trabajo no presencial del alumno, las equivalencias son las siguientes, cada clase magistral implica una hora de trabajo no presencial del alumno, mientras que cada clase práctica de aula, práctica de laboratorio informático, de taller o de seminario implica dos horas de trabajo no presencial del alumno. Este curso conlleva 50 horas de trabajo presencial y 70 horas de trabajo no presencial.

En cada uno de los temas siguientes encontraréis la exposición de la materia en las Clases Magistrales, una muestra de los ejercicios que se realizan en estas así como ejemplos de los ejercicios que se deben realizar en las prácticas de aula y talleres. Los enunciados están recogidos en la Colección de Ejercicios Recomendados y Exámenes de Econometría cuyo enlace podéis encontrar en la página web de la asignatura. En este documento para cada tema se han seleccionado una colección de ejercicios recomendados que deberíais resolver, solo algunos de ellos se resolverán en clase. En la sección siguiente incluimos algunos enunciados para practicar los contenidos de este tema introductorio.

En las notas siguientes y al final de cada tema se incluyen preguntas evaluativas de los contenidos del tema a modo de ejemplo de las que podamos utilizar para evaluaros de forma continuada. Se os muestras preguntas cortas evaluativas de las clases magistrales, de las prácticas de aula y de las prácticas de ordenador. Cada tema cierra con la bibliografía correspondiente al mismo.

0.6. Ejercicios para practicar

Ejercicio 0.4 En el modelo $Y_t = \beta_1 + \beta_2 X_t + \beta_3 Z_t + u_t \quad t = 1, \ldots, T$ donde X_t y Z_t son variables no estocásticas y $u_t \sim iid(0, \sigma_u^2)$. se pide:

1. Obtener la propiedades del estimador MCO tanto en muestras finitas como asintóticas.

2. Obtener su distribución asintótica.
3. Indica como contrastarías $H_0 : \beta_2 = \beta_3$

Ejercicio 0.5 En el modelo $Y_t = \beta X_t + u_t \quad t = 1, 2, \ldots, T \quad u_t \sim iid(0, \sigma^2)$, X_t no estocástica. Consideramos los siguientes estimadores del parámetro desconocido β:

$$\hat{\beta}_{MCO} = \frac{\sum X_t Y_t}{\sum X_t^2} \qquad \tilde{\beta} = \frac{\sum Y_t}{\sum X_t} \qquad \beta^\star = \frac{\sum X_t Y_t}{\sum Y_t^2}$$

1. Demostrar las propiedades en muestras finitas de los tres estimadores.
2. Demostrar que los tres estimadores son consistentes.

Ejercicio 0.6 Comentar la siguiente afirmación: "Los estimadores consistentes son asintóticamente insesgados y viceversa".

Ejercicio 0.7 Comentar la siguiente afirmación: "Un estimador sesgado no puede ser consistentes porque insesgadez asintótica y consistencia son términos equivalentes".

Ejercicio 0.8 Sea modelo: $Y_t = \alpha + \beta X_t + u_t \quad u_t \sim iid(0, \sigma^2)$, se define el siguiente estimador del parámetro desconocido β:

$$\hat{\beta} = \frac{Y_T - Y_1}{X_T - X_1} \qquad u_t \sim (0, \sigma^2)$$

Comprueba si $\hat{\beta}$ es insesgado, de mínima varianza y consistente.

Problema 0.1

Supongamos el modelo correctamente especificado:

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

X_{1t}, X_{2t} no estocásticas $\forall t$, $u_t \sim iid(0, \sigma^2)$. Pero especificamos el modelo:

$$Y_t = \beta_1 + \beta_2 X_{1t} + v_t \qquad t = 1, 2, \ldots, T$$

1. Demuestra que el estimador MCO de β_2 en el modelo estimado es inconsistente.
2. Demuestra que el estimador MCO de β_1 en el modelo estimado es inconsistente.

Ejercicio 0.9 Sea el modelo:

$$Y_t = \beta + u_t \qquad u_t \sim (0, \sigma^2)$$

Definimos el siguiente estimador:

$$\beta^\star = \frac{1}{T} \sum_{t=1}^{T} \frac{Y_t}{c_t}$$

donde $c_t \quad t = 1, 2, \ldots, T$ es una serie no aleatoria tal que

$$\bar{c} = \frac{1}{T} \sum_{t=1}^{T} \frac{1}{c_t} = 1$$

Se pide:

1. Demostrar las propiedades del estimador $\beta^\star$ en muestras finitas.
2. Demostrar que el estimador $\beta^\star$ es consistente.
3. Obtener su distribución asintótica.

Ejercicio 0.10 En el modelo: $Y_t = \beta + u_t \quad t = 1, 2, \ldots, T \qquad u_t \sim iid(0, \sigma^2)$ encontrar las propiedades en muestras finitas y propiedades asintóticas del estimador MCO. Obtener su distribución asintótica bajo el supuesto de normalidad y en la ausencia de este supuesto.

0.7. Anexo: Demostración de la consistencia del estimador MCO

Para demostrar la consistencia del estimador MCO podemos tomar diferentes vías. En el tema se ha desarrollado la demostración utilizando el Teorema de Mann y Wald y también por las condiciones suficientes de consistencia. Pero también podemos demostrar la consistencia del estimador utilizando la difinición de consistencia y buscando el límite en probabilidad del estimador. Vamos a hacerlo así para practicar los conceptos vistos.

A) Demostración de la consistencia del estimador MCO por las condiciones suficientes de consistencia.

$$\left.\begin{array}{l} \lim\limits_{T\to\infty} E(\hat{\beta}_{MCO}) = \beta \\ \lim\limits_{T\to\infty} V(\hat{\beta}_{MCO}) = 0 \end{array}\right\}$$

condiciones suficientes a demostrar que se cumplen.

a.1) $E(\hat{\beta}) = E(\beta + (X'X)^{-1}X'u) = \beta + (X'X)^{-1}X'E(u) = \beta$ si X es no estocástica y $E(u) = 0$.

$$\Longrightarrow \lim_{T\to\infty} E(\hat{\beta}) = \lim_{T\to\infty} (\beta) = \beta$$

luego $\hat{\beta}_{MCO}$ es un estimador asintóticamente insesgado.

a.2) $Var(\hat{\beta}) = \sigma^2(X'X)^{-1} = \frac{\sigma^2}{T}\left(\frac{1}{T}X'X\right)^{-1}$ y se satisface para cualquier T dado si $u \sim N(0, \sigma^2 I)$.

$$\lim_{T\to\infty} Var(\hat{\beta}) = \lim_{T\to\infty}\left[\frac{\sigma^2}{T}\left(\frac{1}{T}X'X\right)^{-1}\right]$$

$$= \lim_{T\to\infty}\frac{\sigma^2}{T}\lim_{T\to\infty}\left(\frac{1}{T}X'X\right)^{-1} = 0 \times Q^{-1} = 0$$

si $\lim\limits_{T\to\infty}\left(\frac{1}{T}X'X\right)$ es una matriz finita q positiva definida. Luego por las condiciones suficientes

$$\text{plim}\hat{\beta}_{(T)MCO} = \beta$$

y $\hat{\beta}_{MCO}$ es un estimador consistente.

B) Demostración de la consistencia de $\hat{\beta}_{MCO}$ utilizando la definición de consistencia.
Para demostrarlo aplicando plim$\hat{\beta}_{MCO}$ no necesitamos conocer $E(\hat{\beta}_{(T)MCO})$ ni $V(\hat{\beta}_{(T)MCO})$.

$$\hat{\beta}_{MCO} = \beta + (X'X)^{-1}X'u = \beta + \left(\frac{1}{T}X'X\right)^{-1}\left(\frac{1}{T}X'u\right)$$

$$\text{plim}\hat{\beta}_{MCO} = \text{plim}\beta + \text{plim}\left[\left(\frac{1}{T}X'X\right)^{-1}\left(\frac{1}{T}X'u\right)\right]$$

$$= \beta + \text{plim}\left(\frac{1}{T}X'X\right)^{-1}\text{plim}\left(\frac{1}{T}X'u\right)$$

Sabemos que:

- $\text{plim}\left(\frac{1}{T}X'X\right) = Q \Longrightarrow \text{plim}\left(\frac{1}{T}X'X\right)^{-1} = Q^{-1}$
- y necesitamos encontrar:

$$\text{plim}\left(\frac{1}{T}X'u\right) = \begin{bmatrix} \text{plim}\frac{1}{T}\sum_{t=1}^{T} u_t \\ \text{plim}\frac{1}{T}\sum_{t=1}^{T} X_{2t}' u_t \\ \vdots \\ \text{plim}\frac{1}{T}\sum_{t=1}^{T} X_{kt}' u_t \end{bmatrix} = ?$$

Dado que se cumple Mann y Wald sabemos que $\text{plim}\left(\frac{1}{T}X'u\right) = 0$. Sin embargo, como estamos haciendo el ejercicio sin utilizar el teorema de Mann y Wald vamos a buscar el resultado aplicamos las condiciones suficientes de consistencia:

(1) $E\left(\frac{1}{T}X'u\right) = \frac{1}{T}X'E(u) = 0 \quad \forall t \Longrightarrow \lim_{T\to\infty} E\left(\frac{1}{T}X'u\right) = 0$

(2) $Var\left(\frac{1}{T}X'u\right) = \frac{\sigma^2}{T}\frac{X'X}{T} \Longrightarrow \lim_{T\to\infty} Var\left(\frac{1}{T}X'u\right) = \lim_{T\to\infty}\frac{\sigma^2}{T}\lim_{T\to\infty}\frac{X'X}{T} = 0 \times Q = 0$

por (1) y (2)

$$\text{plim}\left(\frac{1}{T}X'u\right) = 0$$

y por tanto

$$\text{plim}\hat{\beta}_{(T)MCO} = \beta + Q^{-1} \times 0 = \beta$$

luego $\hat{\beta}_{MCO}$ es un estimador consistente.

Tema 1

Mínimos Cuadrados Generalizados

Clases Magistrales

Las clases magistrales son clases en las cuales el profesor desarrolla la parte teórica de la asignatura, pero también tienen una vertiente práctica, ya que en ellas se resuelven ejercicios en pizarra. A pesar de que evidentemente el peso de la clase lo lleva el profesor, es necesaria la participación del alumno en los siguientes términos: es importante leer con antelación a la clase el material distribuido. Además, resulta conveniente que el alumno intente hacer por sí mismo los ejercicios propuestos para resolver apoyándose en las notas tomadas en clase.

De esta forma el alumno puede darse cuenta de aquellos puntos que ha entendido o no, le surgirán dudas y en resumen avanzará en su aprendizaje mucho más que si se limita a ser notario de lo que el profesor realiza en la pizarra. Por ello de antemano, en clase y en la plataforma docente, se indicará qué ejercicios se proponen para resolver y en qué clase magistral serán resueltos en la confianza de que el alumno invertirá tiempo de trabajo personal en ellos. Por cada clase magistral el alumno debe invertir en casa una hora de trabajo personal.

En el tema de Mínimos Cuadrados Generalizados habitualmente se dispone de tres clases magistrales. Son clases de carácter fundamentalmente teórico dedicadas a analizar las consecuencias de relajar las hipótesis básicas sobre la matriz de varianzas y covarianzas de la perturbación en las propiedades del estimador MCO. Se introducen los problemas de heterocedasticidad y/o autocorrelación. Finalmente, se presentan los estimadores de Mínimos Cuadrados Generalizados y Mínimos Cuadrados Generalizados Factibles, sus propiedades y brevemente, cómo hacer inferencia con estos estimadores. Es un tema claramente instrumental.
A continuación en los dos temas siguientes, Heterocedasticidad y Autocorrelación, se estudian a fondo estos dos problemas. Por ello, el tiempo dedicado a resolver ejercicios es muy reducido.

Competencias a trabajar en estas sesiones:

1. Comprender la importancia de los supuestos empleados en la especificación de un modelo econométrico básico para poder proponer y emplear supuestos más realistas.

2. Diferenciar distintos métodos de estimación y evaluar su uso de acuerdo a las características de las variables económicas de interés para obtener resultados fiables.

Al final de este tema deberíais ser capaces de:

1. Explicar las implicaciones que tiene para la estimación Mínimo Cuadrática Ordinaria el hecho de que la matriz de varianzas y covarianzas del vector de perturbaciones de un modelo no sea escalar.
2. Derivar y comparar las propiedades en muestras finitas y grandes del estimador de Mínimos Cuadrados Ordinarios y el estimador de Mínimos Cuadrados Generalizados bajo perturbaciones no esféricas.
3. Explicar las consecuencias en el estimador de Mínimos Cuadrados Generalizados de que la matriz de varianzas y covarianzas del vector de perturbaciones de un modelo sea desconocida. Derivar un estimador alternativo en este caso, el estimador de Mínimos Cuadrados Generalizados Factibles, y conocer sus propiedades.
4. Conocer la forma de realizar contrastes válidos sobre los parámetros de interés.

Bibliografía Recomendada:

Al final del tema tenéis recogida la bibliografía correspondiente. En particular os recomendamos leer los capítulos correspondientes a la bibliografía básica detallados a continuación :

- Greene, W. (1998), cap. 11 y cap. 15.
- Ramanathan, R. (2002), cap. 8 y cap. 9.
- Wooldridge, J. M. (2003), cap. 8, cap. 11 y cap. 12.

Y para profundizar, podéis leer los capítulos detallados a continuación correspondientes a la bibliografía complementaria:

- Alonso, A. et al. (2005), cap. 8.
- Gujarati, D. (2004), cap. 11 y cap. 12.
- Johnston, J. (1984), cap. 7.
- Maddala, G. S. (1996), cap. 5 y cap. 6.
- Novales, A. (1993), cap. 5 y cap. 8.

1.1. Modelo de regresión con perturbaciones no esféricas

En el tema de Mínimos Cuadrados Generalizados vamos a relajar dos de las hipótesis básicas sobre la perturbación. Hasta ahora hemos supuesto que la perturbación era homocedástica, $Var(u_t) = \sigma^2 \quad \forall t$, y no autocorrelada, $Cov(u_t, u_s) = 0 \quad \forall t, s \;\; t \neq s$, lo que en la literatura econométrica se conoce como perturbaciones esféricas. En este tema vamos a relajar estas dos hipótesis básicas, permitiremos que exista heterocedasticidad y/o autocorrelación. Escribimos estas hipótesis como:

- heterocedasticidad: $Var(u_t) = \sigma_t^2$
- autocorrelación: $Cov(u_t, u_s) \neq 0, \quad \forall t, s \;\; t \neq s$

Para el propósito de estimación distinguir entre heterocedasticidad y/o autocorrelación no es necesario ya que en ambos casos el modelo se estima de la misma manera, por ello en este tema presentaremos el estimador Mínimo Cuadrático Generalizado y sus propiedades, comunes a ambos casos. En los dos temas siguientes veremos los problemas de heterocedasticidad y autocorrelación por separado particularizando en cada uno de ellos.

Nuestro marco de trabajo queda definido como sigue:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \ldots + \beta_K X_{Kt} + u_t \quad t = 1, 2, \ldots, T \qquad \Longleftrightarrow \qquad Y_t = X_t' \beta + u_t$$

mantenemos que los regresores $\mathbf{X}_t' = [\,1 \quad \mathbf{X}_{2t} \, \ldots \, \mathbf{X}_{Kt}\,]$ son no estocásticos y sobre las perturbaciones suponemos:

$$\begin{array}{l} E(u_t) = 0 \quad \forall t, \text{ media cero} \\ E(u_t^2) = \sigma_t^2 \quad t = 1, 2, \ldots, T, \text{ varianza no constante y/o} \\ E(u_t u_s) \neq 0 \;\; \forall t, s \quad t \neq s, \text{ covarianzas no nulas,} \end{array}$$

En términos matriciales el modelo se escribe:

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

y la estructura de **la matriz de varianzas y covarianzas del vector de perturbaciones** que recoge los supuestos anteriores sobre u sería una matriz no escalar como la siguiente:

$$E(uu') = \quad \Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1T} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2T} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{T1} & \sigma_{T2} & \cdots & \sigma_T^2 \end{bmatrix}$$

$$= \sigma^2 \Omega = \sigma^2 \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1T} \\ w_{21} & w_{22} & \cdots & w_{2T} \\ \vdots & \vdots & \ddots & \vdots \\ w_{T1} & w_{T2} & \cdots & w_{TT} \end{bmatrix}$$

donde

$$\begin{aligned} Var(u_t) &= \sigma_t^2 = \sigma^2 w_{tt}, \quad t = 1, ..., T \\ Cov(u_t, u_s) &= \sigma_{ts} = \sigma_{st} \ = \ \sigma^2 w_{ts}, \quad t \neq s \end{aligned}$$

Es decir, la varianza de la perturbación puede cambiar en cada momento y puede existir autocorrelación entre perturbaciones de distintos momentos del tiempo. Además, podemos suponer que existe un factor de escala común a todos los elementos, el parámetro σ^2, que perfectamente puede tomar el valor de la unidad.

Vamos a empezar viendo ejemplos en los que la hipótesis de perturbaciones esféricas no se cumple:

Ejemplo 1.1 Supongamos una muestra de observaciones de sección cruzada relativas a gastos de consumo familiares, C, y renta disponible, R, de un colectivo de N familias. La perturbación mide la diferencia entre el consumo de una familia y el consumo medio de todas las familias que poseen la misma renta, $u_i = C_i - E(C_i/R_i)$, y σ^2 mide la dispersión de estas observaciones. En familias con rentas bajas, las posibilidades de consumo están restringidas por la renta. Sin embargo, a medida que aumenta la renta se amplían las posibilidades y podrán decidir cuanto consumir y cuanto ahorrar. Así, podemos encontrarnos con familias de rentas altas ahorrativas, con bajo consumo, y otras con alto consumo y poco ahorro. En este caso hay una gran dispersión del consumo y σ^2 será grande mientras que para las rentas bajas σ^2 será pequeña ya que tienen menos posibilidades de variar su consumo dad su renta. En este supuesto la varianza de la perturbación cambia según la renta de las familias, existe **heterocedasticidad** y podemos escribirla por ejemplo como $E(u_i^2) = \sigma^2 R_i$.

En el caso de datos de sección cruzada, y si suponemos que no existe correlación entre perturbaciones de distintos individuos, la matriz de varianzas y covarianzas de la perturbación sería:

$$E(uu') = \quad \Sigma = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_N^2 \end{bmatrix}$$

donde suponemos $E(u_i) = 0 \; \forall i \quad i = 1, \ldots, N, \quad E(u_i^2) = \sigma_i^2$, y $E(u_i u_j) = 0 \quad \forall i, j \;\; i \neq j$.

Ejemplo 1.2 Supongamos que analizamos la función de formación de precios de una acción de un banco con fuerte participación en el negocio de la construcción en los Emiratos Árabes. Creemos que los precios de las acciones dependen de las reservas del banco, de su volumen de negocio y de los beneficios obtenidos, entre otros factores que afectan en la media del nivel de precios. Una quiebra no esperada del negocio en dichos Emiratos afectará negativamente a sus resultados y a la confianza que los agentes del mercado tienen depositada en el banco. Este shock no predecible será recogido por la perturbación del modelo, u_t. Resulta razonable pensar que el shock no va a afectar únicamente a un periodo t, si no que los agentes recuperaran la confianza lentamente por lo que parte de ese shock se mantendrá en el futuro. Si esto es así, lo que ocurre en un período actual depende de lo ocurrido en el periodo pasado y será difícil mantener $E(u_t u_{t-1}) = 0$, es decir, que las perturbaciones de los momentos t y $t-1$ estén incorrelacionadas. Ocurrirá lo contrario, que las perturbaciones están correlacionadas. Para el caso general podemos recoger la existencia de autocorrelación denotándola como :

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

En este caso, para datos de serie temporal, y suponiendo que la varianza es constante, escribimos la matriz de varianzas y covarianzas de la perturbación bajo el supuesto de **autocorrelación** como:

$$E(uu') = \Sigma = \begin{bmatrix} \sigma^2 & \sigma_{12} & \cdots & \sigma_{1T} \\ \sigma_{21} & \sigma^2 & \cdots & \sigma_{2T} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{T1} & \sigma_{T2} & \cdots & \sigma^2 \end{bmatrix}$$

Ejemplo 1.3 Supongamos que queremos estimar la demanda de automóviles Y como una función de la renta X, utilizando datos microeconómicos sobre gastos de las familias en dos núcleos geográficos distintos, núcleo urbano y núcleo rural. La función de demanda para las familias del núcleo urbano es:

$$Y_{1i} = \alpha_1 + \beta_1 X_{1i} + u_{1i} \qquad i = 1, \ldots, N_1 \qquad u_{1i} \sim N(0, \sigma_1^2 I_{N_1})$$

La función de demanda para las familias del núcleo rural es:

$$Y_{2j} = \alpha_2 + \beta_2 X_{2j} + u_{2j} \qquad j = 1, \ldots, N_2 \qquad u_{2j} \sim N(0, \sigma_2^2 I_{N_2})$$

Supongamos que la propensión marginal a consumir de ambos núcleos es la misma, $\beta_1 = \beta_2$. En este caso deberíamos estimar la función de demanda en el siguiente modelo conjunto:

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} i_{N_1} & 0 & X_1 \\ 0 & i_{N_2} & X_2 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \beta \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \Longleftrightarrow \underset{((N_1+N_2)\times 1)}{Y} = \underset{((N_1+N_2)\times 3)}{X} \underset{(3\times 1)}{\beta} + \underset{((N_1+N_2)\times 1)}{u}$$

Suponiendo ambas muestras independientes, el vector de medias y la matriz de varianzas y covarianzas del vector de perturbaciones del sistema de ecuaciones a estimar es:

$$E(u) = \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \vec{0}_{(N_1+N_2)\times 1} \qquad E(uu') = \begin{bmatrix} \sigma_1^2 I_{N_1} & 0 \\ 0 & \sigma_2^2 I_{N_2} \end{bmatrix} = \Sigma_{(N_1+N_2)\times(N_1+N_2)}$$

Puede ocurrir que la variación en el gasto en automóviles dada una renta, sea mayor en el núcleo urbano que en el rural, esto es $\sigma_1^2 > \sigma_2^2$, en este caso, en el modelo a estimar la perturbación es **heterocedástica** ya que $\sigma_1^2 \neq \sigma_2^2$.

Además, las muestras no tienen porque ser independientes podría ocurrir que aún en el caso de que la variación en el gasto de automóviles fuese la misma en la zona urbana que en la rural, hubiese factores que siendo no significativos en media en las respectivas demandas implicasen dependencia entre u_1 y u_2, tal que $\sigma_{12} \neq 0$. En este caso las ecuaciones estarían relacionadas a través del término de perturbación y $E(uu') = \Sigma \neq I$ tal que[1].

$$E(uu') = \begin{bmatrix} \sigma^2 I_{N_1} & \sigma_{12} I_N \\ \sigma_{12} I_N & \sigma^2 I_{N_2} \end{bmatrix} = \Sigma_{(N_1+N_2)\times(N_1+N_2)}$$

[1]Estas no son las únicas posibilidades. Por ejemplo podemos pensar en $\sigma_1^2 > \sigma_2^2$ y $\sigma_{12} \neq 0$ o cualquier otra combinación de posibilidades.

Como se puede observar hay muchas y diversas situaciones en que pueden aparecer los problemas de heterocedasticidad y/o autocorrelación. En general la heterocedasticidad aparece con mayor probabilidad en datos de sección cruzada mientras que la autocorrelación es más fácil encontrarla con datos de serie temporal. Dado que la existencia de heterocedasticidad y/o autocorrelación implica que algunas de las hipótesis básicas utilizadas para derivar las propiedades del estimador MCO no se cumplen, debemos empezar viendo que implicaciones tiene en el estimador de MCO y sus propiedades el hecho de que la matriz de varianzas y covarianzas de la perturbación no sea ya una matriz escalar.

1.2. Propiedades del estimador de MCO

Sea el Modelo de Regresión Lineal General, $Y = X\beta + u$, donde se mantienen las hipótesis básicas salvo que[2]:

$$E(uu') = \Sigma = \sigma^2\Omega, \text{ donde } \Omega \neq I \tag{1.1}$$

El estimador MCO de β, $\hat{\beta}_{MCO} = (X'X)^{-1}X'Y$, en muestras finitas sigue siendo lineal e insesgado, pero no es varianza mínima. En muestras grandes o asintóticas es consistente. Prueba:

- **Lineal:** dado que X es no estocástica el estimador MCO es lineal en u, ya que se puede escribir como una combinación lineal del vector de perturbaciones u y la matriz de constantes $D = (X'X)^{-1}X'$, siendo u lo único aleatorio de su expresión:

$$\hat{\beta}_{MCO} = \beta + (X'X)^{-1}X'u = \beta + D\,u$$

- **Insesgado:** dado que X es no estocástica y $E(u) = 0$ el estimador MCO es insesgado.

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

- **Matriz de varianzas y covarianzas:**

$$\begin{aligned} V(\hat{\beta}_{MCO}) &= E[(\hat{\beta} - E(\hat{\beta}))(\hat{\beta} - E(\hat{\beta}))'] = E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'] \\ &= E[(X'X)^{-1}X'uu'X(X'X)^{-1}] \\ &= (X'X)^{-1}X'\,E(uu')\,X(X'X)^{-1} \\ &= \sigma^2(X'X)^{-1}X'\Omega X(X'X)^{-1} \end{aligned}$$

  Bajo heterocedasticidad y/o autocorrelación el estimador MCO **no** es el estimador con varianza mínima entre los estimadores lineales e insesgados. Si $\Omega \neq I$ veremos que existe un estimador que alcanza la cota de varianza mínima, y que es el eficiente entre los estimadores lineales e insesgados en esta situación.

[2]Para el propósito de la demostración es indiferente que en la matriz de varianzas y covarianzas exista o no factor de escala distinto de la unidad. En lo que sigue supondremos que $E(uu') = \sigma^2\Omega$.

- **Distribución en muestras finitas**[3]**:** Si las perturbaciones tienen una distribución normal $u \sim N(0, \sigma^2\Omega)$
$$\hat{\beta}_{MCO} \sim N(\beta, \sigma^2(X'X)^{-1}X'\Omega X(X'X)^{-1})$$

- **Consistente:** El estimador MCO bajo perturbaciones no esféricas es consistente. Vamos a demostrar la consistencia del estimador utilizando las condiciones suficientes de consistencia. Sean:
$$\begin{aligned} \lim_{T\to\infty} \frac{X'X}{T} &= Q \quad \text{finita, definida positiva} \\ \lim_{T\to\infty} \frac{X'\Omega X}{T} &= Z \quad \text{finita, definida positiva} \end{aligned}$$
Entonces:
$$\begin{aligned} \lim_{T\to\infty} E(\hat{\beta}_{MCO}) &= \beta \\ \lim_{T\to\infty} V(\hat{\beta}_{MCO}) &= \lim_{T\to\infty} \frac{\sigma^2}{T}\left(\frac{X'X}{T}\right)^{-1}\left(\frac{X'\Omega X}{T}\right)\left(\frac{X'X}{T}\right)^{-1} = \\ &= 0\cdot Q^{-1}\cdot Z\cdot Q^{-1} = 0, \end{aligned}$$
por lo que el estimador es consistente: $\text{plim}\,\hat{\beta}_{MCO} = \beta$.

1.2.1. Estimador de la matriz de varianzas y covarianzas de $\hat{\beta}_{MCO}$

El estimador MCO de β es lineal e insesgado en muestras finitas. Su matriz de varianzas y covarianzas se define $V(\hat{\beta}_{MCO}) = \sigma^2(X'X)^{-1}X'\Omega X(X'X)^{-1}$. Un estimador de la matriz de varianzas y covarianzas del estimador MCO es $\widehat{V}(\hat{\beta}_{MCO}) = \hat{\sigma}^2(X'X)^{-1}$ siendo $\hat{\sigma}^2 = \frac{\hat{u}'\hat{u}}{T-K}$. Utilizar este estimador para hacer inferencia cuando la matriz de varianzas y covarianzas de la perturbación del modelo no es escalar, $E(uu') = \sigma^2\Omega,\ \Omega \neq I$, no es adecuado. Los estadísticos t y F habituales para hacer inferencia sobre β definidos en base a este estimador de la matriz de varianzas y covarianzas de $\hat{\beta}_{MCO}$ son inapropiados ya que $\widehat{V}(\hat{\beta}_{MCO}) = \hat{\sigma}^2(X'X)^{-1}$ es un estimador sesgado e inconsistente. Dichos estadísticos ahora no tienen las distribuciones t-Student y F-Snedecor habituales por lo tanto, la inferencia en base a ellos no es válida.

Por todo ello parece más adecuado buscar un nuevo estimador del vector de coeficientes β que tuviera en cuenta que $E(uu') = \sigma^2\Omega$, con Ω definida bien para el caso de heterocedasticidad, bien para el caso de autocorrelación, bien para ambos, y que fuese lineal, insesgado, de varianza mínima y consistente. Este estimador es el de Mínimos Cuadrados Generalizados y a su vez permite proponer un estimador insesgado para σ^2 y realizar inferencia válida. El único requisito para poder obtenerlo es que Ω sea conocida. En el caso de que no exista factor de escala $\sigma^2\Omega = \Sigma$ y requerimos que Σ sea conocida.

[3] Es importante recordar que $\hat{\beta}_{MCO}$ es lineal en u. Esto implica que su distribución en muestras finitas coincide con la distribución de u, independientemente de cuál sea ésta. En nuestro caso como suponemos a u normal el estimador MCO tendrá distribución normal.

1.3. Método de Mínimos Cuadrados Generalizados (MCG)

Supongamos el MRLG $Y = X\beta + u$ donde $E(u) = 0$ y $E(uu') = \sigma^2\Omega$, siendo Ω conocida y se cumplen el resto de hipótesis básicas. Si lo que queremos es estimar los coeficientes β desconocidos de forma que el estimador sea eficiente, una manera adecuada de proceder es transformar el modelo tal que sus perturbaciones sean esféricas, es decir, de media cero, varianza constante y covarianzas cero. El estimador de MCO en el modelo así transformado será lineal, insesgado, de varianza mínima o eficiente y consistente. Veámoslo.

Dado que $\Sigma = \sigma^2\Omega$ es simétrica y semidefinida positiva, y hemos supuesto que Ω es conocida, existe una matriz no singular P tal que $\Omega = PP'$. Además, P es conocida y no estocástica. La inversa de la matriz P se utiliza como matriz de transformación del modelo original dado que:

$$\begin{aligned} \Omega &= PP' \\ \Omega^{-1} &= (PP')^{-1} = (P')^{-1}P^{-1} \\ P^{-1}\Omega(P')^{-1} &= P^{-1}PP'(P')^{-1} = I \end{aligned}$$

Premultiplicando el modelo por P^{-1} obtenemos el siguiente modelo transformado:

$$\underbrace{P^{-1}Y}_{Y^\star} = \underbrace{P^{-1}X}_{X^\star}\beta + \underbrace{P^{-1}u}_{u^\star} \Longleftrightarrow Y^\star = X^\star\beta + u^\star$$

Este modelo transformado tiene perturbaciones, $u^\star$, esféricas, es decir, de media cero, varianza constante y covarianzas cero:

$$\begin{aligned} E(u^\star) &= E(P^{-1}u) = P^{-1}E(u) = 0 \\ E(u^\star u^{\star\prime}) &= E(P^{-1}uu'(P^{-1})') = P^{-1}E(uu')(P^{-1})' = \\ &= \sigma^2 P^{-1}\,\Omega\, P^{-1\prime} = \sigma^2 P^{-1}PP'(P')^{-1} = \sigma^2 I. \end{aligned}$$

Por lo tanto, en el modelo transformado se cumplen las hipótesis básicas, y el estimador MCO tendrá las propiedades habituales de linealidad, insesgadez y varianza mínima. Si definimos el estimador de MCO en el modelo transformado, y sustituimos las matrices transformadas por su expresión en términos de las variables originales del modelo, obtenemos la expresión del estimador de Mínimos Cuadrados Generalizados, MCG, en el modelo de interés. Así:

$$\begin{aligned} \hat\beta_{MCO} &= (X^{\star\prime}X^\star)^{-1}X^{\star\prime}Y^\star = && (1.2) \\ &= (X'(P')^{-1}P^{-1}X)^{-1}X'(P')^{-1}P^{-1}Y = \\ &= (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y = \hat\beta_{MCG} && (1.3) \end{aligned}$$

Si Ω (ó Σ en su caso) es conocida el estimador es inmediatamente calculable.

• Criterio de estimación y función objetivo:

El estimador de MCG dado por la expresión (1.3) puede ser obtenido minimizando la siguiente función objetivo:

$$\underset{\hat\beta}{Min}\ \hat u'\,\Omega^{-1}\,\hat u = \underset{\hat\beta}{Min}\ (Y - X\hat\beta)'\,\Omega^{-1}\,(Y - X\hat\beta) = \underset{\hat\beta}{Min}\,(Y^\star - X^\star\hat\beta)'(Y^\star - X^\star\hat\beta)$$

Utilizando las condiciones de primer orden podemos derivar el sistema de ecuaciones normales $X'\Omega^{-1}Y = \hat\beta(X'\Omega^{-1}X)^{-1}$ de donde obtenemos $\hat\beta_{MCG} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y$.

1.3.1. Propiedades del estimador de MCG

Como ya se ha dicho el estimador de MCG es un estimador lineal, insesgado y de varianza mínima en muestras finitas, en muestras grandes es consistente y eficiente asintóticamente. Las propiedades del estimador MCG podemos demostrarlas alternativamente en el modelo transformado utilizando la expresión (1.2) o en el modelo original utilizando la expresión (1.3).

- **Lineal:** dado que X y Ω son no estocásticas el estimador MCG es lineal en la perturbación u, ya que se puede expresar como una combinación lineal del vector de perturbaciones u y la matriz de constantes no estocástica $C = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}$, siendo u lo único aleatorio de su expresión:

$$\hat{\beta}_{MCG} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y = \beta + (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}u = \beta + C\,u$$

- **Insesgado:** Dado que X y Ω son no estocásticas y $E(u) = 0$ el estimador MCG es insesgado:

$$\begin{aligned} E(\hat{\beta}_{MCG}) &= \beta + E[\,(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}u\,] = \\ &= \beta + (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}E(u) = \beta \end{aligned}$$

- **Matriz de varianzas y covarianzas:**

$$\begin{aligned} V(\hat{\beta}_{MCG}) &= E[(\hat{\beta} - E(\hat{\beta}))(\hat{\beta} - E(\hat{\beta}))'] = E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'] \\ &= E[(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\,uu'\,\Omega^{-1}X(X'\Omega^{-1}X)^{-1}] \\ &= (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\,E(uu')\,\Omega^{-1}X(X'\Omega^{-1}X)^{-1} \\ &= \sigma^2(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\,\Omega\,\Omega^{-1}X(X'\Omega^{-1}X)^{-1} \\ &= \sigma^2(X'\Omega^{-1}X)^{-1} \end{aligned}$$

  Cuando $E(uu') = \sigma^2\Omega \neq I$ y Ω es conocida, el estimador MCG de β es el de menor varianza entre los estimadores lineales e insesgados.

- **Distribución en muestras finitas:** Bajo el supuesto de normalidad de las perturbaciones, $u \sim N(0,\, \sigma^2\Omega)$, tenemos que

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

- **Consistente:** Podemos demostrar la consistencia del estimador mediante las condiciones suficientes de consistencia:
  Sea $\operatorname{plim} \dfrac{X'\Omega^{-1}X}{T} = G$, finita, semidefinida positiva y no singular, entonces

$$\begin{aligned} \lim_{T\to\infty} E(\hat{\beta}_{MCG}) &= \beta \\ \lim_{T\to\infty} \mathrm{V}(\hat{\beta}_{MCG}) &= \lim_{T\to\infty} \frac{\sigma^2}{T}\left(\frac{X'\Omega^{-1}X}{T}\right)^{-1} = 0 \times G^{-1} = 0 \end{aligned}$$

  luego $\operatorname{plim} \hat{\beta}_{MCG} = \beta$, con lo que el estimador es consistente.

- **Distribución Asintótica:** Dado que

$$\text{plim}\,\hat{\beta}_{MCG} = \beta \Longrightarrow \hat{\beta}_{MCG} \xrightarrow{p} \beta \Longrightarrow (\hat{\beta}_{MCG} - \beta) \xrightarrow{p} 0$$

el estimador tiene una distribución degenerada en el límite, por lo que buscamos la distribución asintótica para $\sqrt{T}(\hat{\beta}_{MCG} - \beta)$

$$\sqrt{T}(\hat{\beta}_{MCG} - \beta) = \left(\frac{X'\Omega^{-1}X}{T}\right)^{-1} \frac{X'\Omega^{-1}u}{\sqrt{T}}$$

En el modelo original el teorema de Mann-Wald no se puede aplicar, por lo que demostrar la consistencia es un poco más costoso que si lo hacemos en el modelo transformado. En el modelo transformado las perturbaciones son esféricas y $X^\star$ es no estocástica por lo que podemos aplicar el Teorema de Mann-Wald. Vamos a mostrar como utilizarlo en la obtención de la distribución asintótica de $\hat{\beta}_{MCG}$.

i) Sea $u^\star \sim iid(0, \sigma^2 I)$

ii) Sea $E(X^{\star\prime} u^\star) = X^{\star\prime} E(u^\star) = 0$ ya que $X^\star$ es no estocástica,

iii) Sea $\text{plim}\left(\frac{X^{\star\prime}X^\star}{T}\right) = Q^\star = G$ finita y no singular

Entonces se cumplen los dos resultados siguientes:

1. $\text{plim}\left(\frac{X^{\star\prime}u^\star}{T}\right) = 0$
2. $\frac{X^{\star\prime}u^\star}{\sqrt{T}} \xrightarrow{d} N(0,\, \sigma^2 G)$

por lo que de Mann-Wald, y utilizando el teorema de Cramer, tenemos

$$\sqrt{T}(\hat{\beta}_{MCO} - \beta) = \left(\frac{X^{\star\prime}X^\star}{T}\right)^{-1} \frac{X^{\star\prime}u^\star}{\sqrt{T}} \quad \xrightarrow{d} \quad N(0,\, \sigma^2 G^{-1})$$

de donde

$$\sqrt{T}(\hat{\beta}_{MCG} - \beta) = \left(\frac{X'\Omega^{-1}X}{T}\right)^{-1} \frac{X'\Omega^{-1}u}{\sqrt{T}} \xrightarrow{d} N(0,\, \sigma^2 G^{-1})$$

con $G = \text{plim}\,\frac{X'\Omega^{-1}X}{T}$.

1.3.2. Estimador de la matriz de varianzas y covarianzas de $\hat{\beta}_{MCG}$

El estimador de MCG es lineal, insesgado y de varianza mínima. Cuando $E(uu') = \sigma^2\Omega$, a pesar de que Ω sea conocida, en general σ^2 será desconocida y su matriz de varianzas y covarianzas habrá de ser estimada. Un estimador de $V(\hat{\beta}_{MCG})$ insesgado y consistente es

$$\widehat{V}(\hat{\beta}_{MCG}) = \hat{\sigma}^2_{MCG}(X'\Omega^{-1}X)^{-1} = \hat{\sigma}^2_{MCG}(X^{\star\prime}X^\star)^{-1}$$

siendo

$$\hat{\sigma}^2_{MCG} = \frac{\hat{u}'_{MCG}\,\Omega^{-1}\,\hat{u}_{MCG}}{T-K} = \frac{\hat{u}^{\star\prime}\hat{u}^{\star}}{T-K}$$

Puede resultar útil tener las expresiones anteriores en términos de las variables del modelo tal que:

$$\begin{aligned}
\hat{\sigma}^2_{MCG} &= \frac{\hat{u}^{\star\prime}\hat{u}^{\star}}{T-K} = \frac{(Y^\star - X^\star\hat{\beta}_{MCG})'(Y^\star - X^\star\hat{\beta}_{MCG})}{T-K} = \frac{Y^{\star\prime}Y^\star - \hat{\beta}'_{MCG}X^{\star\prime}Y^\star}{T-K} \\
\hat{\sigma}^2_{MCG} &= \frac{\hat{u}'_{MCG}\,\Omega^{-1}\,\hat{u}_{MCG}}{T-K} = \\
&= \frac{(Y - X\hat{\beta}_{MCG})'\Omega^{-1}(Y - X\hat{\beta}_{MCG})}{T-K} = \frac{Y'\Omega^{-1}Y - \hat{\beta}'_{MCG}X'\Omega^{-1}Y}{T-K}
\end{aligned}$$

Ejercicio 1.1 En el modelo $Y = X\beta + u$ donde $E(u) = 0,\ E(uu') = \sigma^2\Omega$ con Ω conocida y X es no estocástica. Demostrar la consistencia del estimador MCG utilizando el modelo transformado y el Teorema de Mann-Wald.

Ejercicio 1.2 En el modelo $Y = X\beta + u$ donde $E(u) = 0, \quad E(uu') = \Sigma,\ \Sigma$ conocida y X es no estocástica. Se pide:

1. Escribir la función objetivo. Obtener las ecuaciones normales y derivar el estimador MCG.
2. Demostrar sus propiedades en muestras finitas y asintóticas.
3. Obtener su distribución en muestras finitas y asintóticas.

Ejercicio 1.3 En el modelo $Y_i = \beta_1 + \beta_2 X_i + u_i \quad i = 1, \ldots, 20$ donde X es no estocástica, $E(u_i) = 0\ \forall i$, $Var(u_i) = \sigma^2 X_i$, y $Cov(u_i u_j) = 0 \quad \forall i \neq j$. Se pide:

1. Escribir la función objetivo.
2. Obtener las ecuaciones normales y derivar el estimador MCG de β_1 y β_2.

1.4. Método de Mínimos Cuadrados Generalizados Factibles (MCGF)

Hasta ahora hemos supuesto que conocíamos Ω o Σ en $E(uu') = \sigma^2\Omega = \Sigma$, pero en la práctica la mayoría de las veces Ω (o Σ) son desconocidas. En este caso el estimador MCG no es directamente calculable ya que en su expresión aparecen estas matrices. La solución habitual es sustituir Ω (o Σ) por una estimación suya en la expresión del estimador de MCG. Este es el estimador de Mínimos Cuadrados Generalizados Factibles, MCGF:

$$\hat{\beta}_{MCGF} = (X'\widehat{\Omega}^{-1}X)^{-1}X'\widehat{\Omega}^{-1}Y \tag{1.4}$$

Notar que sustituir Ω por una estimación suya implica hacer un supuesto de comportamiento, es decir, suponer una forma funcional para los elementos de Ω que son desconocidos. Una vez supuesta una forma funcional, que en general dependerá de un conjunto de parámetros desconocidos y variables exógenas, es necesario estimar esos parámetros desconocidos[4]. Cómo proponer una forma funcional para los elementos de Ω y cómo estimar los parámetros desconocidos de esta forma funcional es algo que se verá en los temas siguientes, con cada caso en concreto, heterocedasticidad y/o autocorrelación.

Ahora vamos a limitarnos a estudiar las propiedades del estimador MCGF una vez supuesto que hemos sido capaces de proponer una forma funcional sensata a la información disponible sobre el comportamiento de las varianzas y covarianzas del vector de perturbaciones. Además, hemos sido capaces de estimar de forma consistente los parámetros desconocidos de los que dependen estos momentos. Es decir, disponemos de $\widehat{\Omega}$ estimador consistente de Ω.

1.4.1. Propiedades del estimador de MCGF

Bajo el supuesto de que las varianzas de las perturbaciones se han modelado correctamente, el estimador MCGF en muestras finitas es un estimador no lineal y sesgado en general, su distribución en muestras finitas no es conocida. En muestras grandes, bajo ciertas condiciones de regularidad, y si $\widehat{\Omega}$ es un estimador consistente de Ω, el estimador MCGF es consistente, asintóticamente eficiente y tiene distribución asintótica conocida.

- **Linealidad:** El estimador MCGF no es lineal en u.

$$\hat{\beta}_{MCGF} = \beta + (X'\widehat{\Omega}^{-1}X)^{-1}(X'\widehat{\Omega}^{-1}u)$$

  En la expresión del estimador MCGF aparecen u y $\widehat{\Omega}$, ambas son variables aleatorias, ya que $\widehat{\Omega}$ que ahora es un estimador, luego es una matriz estocástica. Por tanto, $\hat{\beta}_{MCGF}$ es una función no lineal de u y $\widehat{\Omega}$.

- **Insesgadez y matriz de varianzas y covarianzas:** En general el estimador de MCGF es sesgado:

$$E(\hat{\beta}_{MCGF}) = \beta + E[(X'\widehat{\Omega}^{-1}X)^{-1}(X'\widehat{\Omega}^{-1}u)] \neq \beta$$

  Para determinar $E[(X'\widehat{\Omega}^{-1}X)^{-1}(X'\widehat{\Omega}^{-1}u)]$ necesitamos conocer la distribución conjunta de las variables aleatorias en $\widehat{\Omega}$ y en u. En general $E[(X'\widehat{\Omega}^{-1}X)^{-1}(X'\widehat{\Omega}^{-1}u)] \neq 0$ y por tanto $E(\hat{\beta}_{MCGF}) \neq \beta$. Además, $V(\hat{\beta}_{MCGF}) = E[\hat{\beta}_{MCGF} - E(\hat{\beta}_{MCGF})][\hat{\beta}_{MCGF} - E(\hat{\beta}_{MCGF})]'$ es una expresión difícil de obtener analíticamente. Las propiedades en muestras finitas y la distribución exacta del estimador MCGF son desconocidas. Por lo tanto, en muestras finitas su comportamiento es difícil de comparar con el de MCO e incluso puede ser peor.

- **Consistencia:** Bajo ciertas condiciones de regularidad, y si $\widehat{\Omega}$ es un estimador consistente de Ω, el estimador MCGF es consistente, $\text{plim}\,\hat{\beta}_{MCGF} = \beta$.

[4] El número de parámetros desconocidos es limitado ya que en otro caso no se podrían estimar los K coeficientes β del modelo y los $T(T+1)/T$ elementos desconocidos en la matriz de varianzas y covarianzas de la perturbación con T observaciones.

- **Distribución asintótica:** En general si las condiciones anteriores se satisfacen el estimador MCGF es asintóticamente equivalente al estimador MCG y su distribución asintótica coincide y por tanto es asintóticamente eficiente dentro de esta clase. Así, se puede demostrar que

$$\sqrt{T}(\hat{\beta}_{MCGF} - \beta) \xrightarrow{d} N(0, \sigma^2 G^{-1})$$

con $G = \text{plim}\, \frac{X'\Omega^{-1}X}{T}$.

1.4.2. Estimador de la matriz de varianzas y covarianzas de $\hat{\beta}_{MCGF}$

Cuando $E(uu') = \sigma^2\Omega$ y decimos que Ω es desconocida en realidad nos estamos refiriendo a que $\sigma^2\Omega$ es desconocido. Parece complicado ser capaces de establecer que $E(uu')$, que es desconocida, tiene un elemento común, o factor de escala, a todos sus elementos que también es desconocido y distinto de Ω. Lo lógico es pensar en $E(uu') = \Sigma$, Σ desconocida. Un estimador consistente de $V(\hat{\beta}_{MCGF})$ es $\widehat{V}(\hat{\beta}_{MCGF}) = (X'\widehat{\Sigma}^{-1}X)^{-1}$.

Sin embargo, y por si fuera el caso de que, siendo Ω desconocida existe este factor de escala σ^2, también desconocido en $E(uu')$, un estimador consistente de la matriz de varianzas y covarianzas asintótica del estimador de MCGF es $\widehat{V}(\hat{\beta}_{MCGF}) = \hat{\sigma}^2_{MCGF}(X'\widehat{\Omega}^{-1}X)^{-1}$ siendo

$$\hat{\sigma}^2_{MCGF} = \frac{\hat{u}'_{MCGF}\,\widehat{\Omega}^{-1}\,\hat{u}_{MCGF}}{T-K} \quad \text{y} \quad \hat{u}_{MCGF} = Y - X\hat{\beta}_{MCGF}$$

Es importante recalcar que la propiedad de consistencia es invariante a tener en el denominador $(T-K)$ o T.

Ejercicio 1.4 En el modelo $Y = X\beta + u$ donde $E(u) = 0$, $E(uu') = \Sigma$, Σ desconocida y X es no estocástica. Se pide:

1. Escribir la expresión del estimador MCGF.
2. Demostrar sus propiedades en muestras finitas y asintóticas.
3. Obtener su distribución asintótica.

1.5. Contrastes de restricciones lineales

Una vez hemos estimado los coeficientes del modelo de manera apropiada estaremos interesados en realizar contraste de hipótesis. Vamos a ver cómo realizar contrastes de restricciones lineales sobre el vector β en el MRLG con perturbaciones no esféricas.

Sean las hipótesis nula y alternativa para el contraste de q restricciones lineales:

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

donde R es una matriz $(q \times K)$ y r es un vector de dimensión $(q \times 1)$, siendo q el número de restricciones lineales a contrastar.

• Estadísticos basados en el estimador de $\boldsymbol{\beta}$ por MCG.
En este caso

$$Y = X\beta + u \qquad u \sim N(0, \sigma^2\Omega) \quad \text{y} \quad \Omega \quad \text{es conocida}$$

el estimador MCG es lineal en u, insesgado y de varianza mínima, su distribución en muestras finitas es:

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

Estadísticos de contraste y distribución asociada:

Para $q = 1$

$$\frac{R\hat{\beta}_{MCG} - r}{\sqrt{R\,\widehat{V}(\hat{\beta}_{MCG})\,R'}} \overset{H_0}{\sim} t_{(T-K)}$$

luego

$$\frac{R\hat{\beta}_{MCG} - r}{\hat{\sigma}\sqrt{R\,(X'\Omega^{-1}X)^{-1}\,R'}} \overset{H_0}{\sim} t_{(T-K)}$$

Si el valor muestral del estadístico, t, es tal que t $< t_{(T-K)|\frac{\alpha}{2}}$ no rechazamos la H_0 para un nivel de significación α.

Para $q > 1$

$$(R\hat{\beta} - r)'\,[R\,\widehat{V}(\hat{\beta}_{MCG})\,R'\,]^{-1}\,(R\hat{\beta} - r)\frac{1}{q} \overset{H_0}{\sim} F_{(q,T-K)}$$

luego

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

Si el valor muestral del estadístico, F, es tal que F $< F_{(q,T-K)|\alpha}$ no rechazamos la H_0 para un nivel de significación α.

Ejercicio 1.5 Sea el modelo:

$$Y = X\beta + u \qquad u \sim N(0, \sigma^2\Omega) \qquad \Omega \text{ conocida} \quad \text{y} \quad X \text{ no} \quad \text{estocástica}$$

Escribe el estadístico general y distribución asociada para el contraste de q restricciones lineales en el modelo transformado. Detalla cada uno de sus elementos y la regla de decisión.

Ejercicio 1.6 En el modelo $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \quad i = 1, \ldots, 500$ donde X_2 y X_3 son no estocásticas, $E(u_i) = 0\, \forall i$, $Var(u_i) = \sigma^2 X_{2i}$, y $Cov(u_i u_j) = 0\, \forall i \neq j$. Se pide:

1. ¿Cómo estimarías eficientemente β_1, β_2 y β_3? Describe claramente el método de estimación y escribe la forma matricial del estimador completando todos los elementos de las matrices implicadas.
2. ¿Como contrastarías la significatividad individual de la variable X_{2i}? Escribe la hipótesis nula, la alternativa y el estadístico de contraste junto con su distribución.
3. ¿Como contrastarías $H_0 : \beta_2 + \beta_3 = 2$? Escribe la hipótesis nula, la alternativa y el estadístico de contraste junto con su distribución.
4. ¿Como contrastarías la hipótesis compuesta $H_0 : \beta_2 = 1$ y $\beta_3 = 2$? Escribe la hipótesis nula, la alternativa y el estadístico de contraste junto con su distribución.

• Estadísticos basados en el estimador de $\boldsymbol{\beta}$ por MCGF.

En este caso

$$Y = X\beta + u \qquad u \sim N(0, \Sigma) \qquad \Sigma \text{ desconocida, pero estimable consistentemente}$$

El estimador de MCGF, $\hat{\beta}_{MCGF} = (X'\widehat{\Sigma}^{-1}X)^{-1}(X'\widehat{\Sigma}^{-1}Y)$, es un estimador no lineal y en general sesgado, cuya distribución en muestras finitas no es conocida. En muestras grandes es un estimador consistente si $\widehat{\Sigma}$ es un estimador consistente de Σ. Su distribución asintótica es:

$$\sqrt{T}(\hat{\beta}_{MCGF} - \beta) \xrightarrow{d} N(0, G^{-1})$$

Los estadísticos de contraste y distribución asociada utilizando como estimador de $G^{-1} = \text{plim}\left[\frac{1}{T}X'\Sigma X\right]^{-1}$ a $T\widehat{V}(\hat{\beta}_{MCGF}) = T(X'\widehat{\Sigma}^{-1}X)^{-1}$ son,

para $q = 1$

$$\frac{R\hat{\beta}_{MCGF} - r}{\sqrt{R\,(X'\widehat{\Sigma}^{-1}X)^{-1}\,R'}} \xrightarrow{d,H_0} N(0,1)$$

para $q > 1$

$$(R\hat{\beta}_{MCGF} - r)'[\,R\,(X'\widehat{\Sigma}^{-1}X)^{-1}\,R'\,]^{-1}(R\hat{\beta}_{MCGF} - r) \xrightarrow{d,H_0} \chi^2_q$$

Las reglas de decisión son las habituales.

• Estadísticos basados en el estimador de $\boldsymbol{\beta}$ por MCO y un estimador robusto a heterocedasticidad y/o autocorrelación de $\boldsymbol{V(\hat{\beta}_{MCO})}$

En presencia de perturbaciones no esféricas, el estimador Mínimo Cuadrático Ordinario es lineal, insesgado y consistente, pero no es de varianza mínima. Su matriz de varianzas y covarianzas se define $V(\hat{\beta}_{MCO}) = \sigma^2(X'X)^{-1}X'\Omega X(X'X)^{-1}$. Un estimador de la matriz de varianzas y covarianzas del estimador MCO es $\widehat{V}(\hat{\beta}_{MCO}) = \hat{\sigma}^2(X'X)^{-1}$. Utilizar este estimador para hacer inferencia cuando hay heterocedasticidad y/o autocorrelación no es adecuado. Los estadísticos t y F habituales para hacer inferencia sobre β definidos en base a este estimador de la matriz de varianzas y covarianzas de $\hat{\beta}_{MCO}$ son inapropiados ya que es un estimador sesgado e inconsistente.

La dificultad que entraña el conocimiento de Ω, o Σ en su caso, hace interesante el poder contar con un estimador de $V(\hat{\beta}_{MCO})$ consistente y robusto a la posible existencia de heterocedasticidad y/o autocorrelación para de esta forma derivar estadísticos válidos, al menos asintóticamente, y contrastar hipótesis sobre el vector de coeficientes β utilizando $\hat{\beta}_{MCO}$.

Supongamos que $\widehat{V}(\hat{\beta}_{MCO})_{Consistente}$ es la estimación consistente de $V(\hat{\beta}_{MCO})$ de la que hablamos. Dado que el estimador MCO es consistente podemos utilizarlos conjuntamente para hacer inferencia asintótica válida. En este caso el estadístico válido para realizar contraste de hipótesis de la forma $H_0 : R\beta = r$ y su distribución asintótica asociada bajo la hipótesis nula son:

$$(R\hat{\beta}_{MCO} - r)'(R\,\widehat{V}(\hat{\beta}_{MCO})_{Consistente}\,R')^{-1}(R\hat{\beta}_{MCO} - r) \xrightarrow{d,H_0} \chi^2_q$$

Siendo q el número de restricciones de contraste. Para $q = 1$ podemos escribir el estadístico anterior como:

$$\frac{R\hat{\beta}_{MCO} - r}{\sqrt{R\,\widehat{V}(\hat{\beta}_{MCO})_{Consistente}\,R'}} \xrightarrow{d,H_0} N(0,1)$$

Las reglas de decisión son las habituales. En los temas siguientes veremos que esta matriz robusta a heterocedasticidad y/o autocorrelación existe y puede ser calculada.

1.6. Ejemplo: Sistemas de Ecuaciones

En ocasiones necesitamos estimar un sistema de ecuaciones. Vamos a ver diferentes posibilidades de estimación de un sistema como ilustración del tema de perturbaciones no esféricas y a la vez, mostraremos como realizar el contraste de cambio estructural o de Chow.

Ilustraremos esta sección utilizando uno de los ejemplos vistos al comienzo del tema. Supongamos que queremos estimar la demanda de automóviles Y como una función de la renta X, utilizando datos microeconómicos sobre gastos de las familias en dos núcleos geográficos distintos, núcleo urbano y núcleo rural.

La función de demanda para las familias del núcleo urbano es:

$$Y_{1i} = a_1 + b_1 X_{1i} + u_{1i} \quad \Longleftrightarrow \quad Y_1 = X_1\beta_1 + u_1 \quad \text{siendo} \quad u_1 \sim N(0, \sigma_1^2 I_{N_1}) \quad i = 1, \ldots, N_1$$

La función de demanda para las familias del núcleo rural es:

$$Y_{2j} = a_2 + b_2 X_{2j} + u_{2j} \quad \Longleftrightarrow \quad Y_2 = X_2\beta_2 + u_2 \quad \text{siendo} \quad u_2 \sim N(0, \sigma_2^2 I_{N_2}) \quad j = 1, \ldots, N_2$$

Podemos escribir el siguiente sistema de ecuaciones:

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \quad \Longleftrightarrow \quad Y = X\beta + u \qquad (1.5)$$

$$\underset{((N_1+N_2)\times 1)}{Y} = \underset{((N_1+N_2)\times 4)}{X} \underset{(4\times 1)}{\beta} + \underset{((N_1+N_2)\times 1)}{u}$$

Suponemos que X es no estocástica. El vector de medias y la matriz de varianzas y covarianzas del vector de perturbaciones u son:

$$\underset{((N_1+N_2)\times 1)}{E(u)} = \begin{bmatrix} E(u_1) \\ E(u_2) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\underset{((N_1+N_2)\times(N_1+N_2))}{E(uu')} = \begin{bmatrix} E(u_1u_1') & E(u_1u_2') \\ E(u_2u_1') & E(u_2u_2') \end{bmatrix}$$

En el sistema de ecuaciones anterior no hay relación entre los coeficientes de las dos ecuaciones. En cuanto a la estructura de $E(uu')$, podemos distinguir tres situaciones:

1. $E(u_1u_1') = \sigma_1^2 I_{N_1}$, $E(u_2u_2') = \sigma_2^2 I_{N_2}$ con $\sigma_1^2 = \sigma_2^2$, es decir, homocedasticidad entre ecuaciones y $E(u_1u_2') = E(u_2u_1') = 0$ lo que implica que no hay relación entre las perturbaciones de las dos ecuaciones.

2. $E(u_1u_1') = \sigma_1^2 I_{N_1}$, $E(u_2u_2') = \sigma_2^2 I_{N_2}$ con $\sigma_1^2 \neq \sigma_2^2$, es decir, heterocedasticidad entre ecuaciones, y $E(u_1u_2') = E(u_2u_1') = 0$.

3. Ecuaciones aparentemente no relacionadas $E(u_1u_2') = E(u_2u_1') \neq 0$.

Dado que no existe relación entre los coeficientes de las dos ecuaciones, el método más adecuado para estimar un modelo $Y = X\beta + u$ depende de la estructura de $E(uu')$.

1.6.1. Ecuaciones no relacionadas con varianza común

Sean las ecuaciones:

$$Y_1 = X_1\beta_1 + u_1 \quad N_1 \text{ obs.}$$

$$Y_2 = X_2\beta_2 + u_2 \quad N_2 \text{ obs.}$$

En principio β_1 y β_2 son distintos, como nada relaciona a las dos ecuaciones podríamos pensar que ganamos en eficiencia utilizando toda la información conjuntamente y por ello deberíamos estimar conjuntamente el modelo. Dado que $\sigma_1^2 = \sigma_2^2$ y $E(u_1u_2') = E(u_2u_1') = 0$ tenemos que

$$E(uu') = \begin{bmatrix} \sigma_1^2 I_{N_1} & 0 \\ 0 & \sigma_2^2 I_{N_2} \end{bmatrix} = \begin{bmatrix} \sigma^2 I_{N_1} & 0 \\ 0 & \sigma^2 I_{N_2} \end{bmatrix} = \sigma^2 I_{N_1+N_2}$$

El modelo puede ser estimado por MCO que será lineal, insesgado y de varianza mínima. Se puede probar que dado que no hay información común entre las ecuaciones, es decir, dado que X es diagonal por bloques y $E(u_1u_2') = E(u_2u_1') = 0$, la estimación del modelo conjunto por MCO es equivalente a estimar cada ecuación por separado por MCO. La estimación conjunta no gana en eficiencia, por tanto estimaremos las ecuaciones por separado por MCO.

Además, $\hat{\sigma}^2$ es insesgado y consistente:

$$\hat{\sigma}^2 = \frac{\hat{u}'\hat{u}}{N-K} = \frac{\hat{u}_1'\hat{u}_1 + \hat{u}_2'\hat{u}_2}{N-K} \quad \text{donde} \quad N = N_1 + N_2 \quad \text{y} \quad K = K_1 + K_2$$

• **Contraste de cambio estructural o de Chow:** Se llama contraste de cambio estructural al contraste de que todos o algunos de los coeficientes que corresponden a las mismas variables en las dos ecuaciones son iguales. Supongamos que queremos contrastar la igualdad de ordenadas y pendientes o lo que es igual cambio estructural total:

$$\left\{ \begin{array}{ll} H_0: & a_1 = a_2 \text{ y } b_1 = b_2 \\ H_a: & a_1 \neq a_2 \text{ y/o } b_1 \neq b_2 \end{array} \right. \Longleftrightarrow \left\{ \begin{array}{ll} H_0: & \beta_1 = \beta_2 \\ H_a: & \beta_1 \neq \beta_2 \end{array} \right. \longrightarrow q = 2$$

Hay dos formas alternativas de realizar el contraste:

- Alternativa 1: Con el estadístico:

$$\mathrm{F} = \frac{(R\hat{\beta}_{MCO} - r)' \left[R(X'X)^{-1}R'\right]^{-1} (R\hat{\beta}_{MCO} - r)/q}{\hat{\sigma}^2} \overset{H_0}{\sim} \mathcal{F}(q, N-K)$$

  donde

$$R\beta - r = \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} a_1 \\ b_1 \\ a_2 \\ b_2 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix} \qquad q = 2 \qquad K = 4$$

$$\hat{\sigma}^2 = \frac{\hat{u}'\hat{u}}{N-K} = \frac{\hat{u}_1'\hat{u}_1 + \hat{u}_2'\hat{u}_2}{N-K}$$

  Regla de decisión:

  - Si F $< F_{(q,N-K)|\alpha}$ no rechazamos la H_0 para un nivel de significación α y concluimos que no existe cambio estructural.
  - Si F $> F_{(q,N-K)|\alpha}$ rechazamos la H_0 para un nivel de significación α y concluimos que existe cambio estructural.

- Alternativa 2: Con el estadístico

$$\frac{(\hat{u}_r'\hat{u}_r - \hat{u}'\hat{u})/q}{\hat{u}'\hat{u}/N-K} \overset{H_0}{\sim} \mathcal{F}(q, N-K)$$

  donde $\hat{u}'\hat{u}$ es la SCR del modelo no restringido (1.5) tal que $\hat{u}'\hat{u} = \hat{u}_1'\hat{u}_1 + \hat{u}_2'\hat{u}_2$ y $\hat{u}_r'\hat{u}_r$ es la SCR del modelo restringido siguiente:

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \beta + \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \tag{1.6}$$

  Dado que hemos supuesto que $E(uu') = \sigma^2 I_{N_1+N_2}$ el modelo restringido se estima por MCO. Esta estimación no es equivalentemente a estimar por MCO ecuación por ecuación, ya que su matriz de regresores no es diagonal por bloques.

1.6.2. Ecuaciones no relacionadas con varianzas distintas

Sean las ecuaciones:

$$Y_1 = X_1\beta_1 + u_1 \quad N_1 \text{ obs.}$$

$$Y_2 = X_2\beta_2 + u_2 \quad N_2 \text{ obs.}$$

Dado que $\sigma_1^2 \neq \sigma_2^2$ y $E(u_1u_2') = E(u_2u_1') = 0$ tenemos que

$$E(uu') = \begin{bmatrix} \sigma_1^2 I_{N_1} & 0 \\ 0 & \sigma_2^2 I_{N_2} \end{bmatrix} = \Sigma_{(N_1+N_2)\times(N_1+N_2)}$$

Suponiendo σ_1^2 y σ_2^2 conocidas, el modelo debiera ser estimado por MCG. Sin embargo, este procedimiento coincide con estimar por MCO ecuación por ecuación, ya que no existe relación entre los coeficientes de ambas ecuaciones, X es diagonal por bloques, $E(uu')$ también lo es y hay homocedasticidad dentro de cada ecuación.

$$\hat{\beta}_{MCG} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}Y = \begin{bmatrix} (X_1'X_1)^{-1}X_1'Y_1 \\ (X_2'X_2)^{-1}X_2'Y_2 \end{bmatrix} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix}_{MCO}$$

$$V(\hat{\beta}_{MCG}) = (X'\Sigma^{-1}X)^{-1} = \begin{bmatrix} \sigma_1^2(X_1'X_1)^{-1} & 0 \\ 0 & \sigma_2^2(X_2'X_2)^{-1} \end{bmatrix}$$

En este caso la estimación del modelo conjunto por MCG no mejora la eficiencia con respecto a estimar cada ecuación por separado por MCO, por tanto estimaremos las ecuaciones por separado. Además, el resultado de estimar β es independiente de que conozcamos o no σ_1^2 y σ_2^2.

• **Contraste de cambio estructural o de Chow:**

$$\left\{ \begin{array}{ll} H_0: & a_1 = a_2 \text{ y } b_1 = b_2 \\ H_a: & a_1 \neq a_2 \text{ y/o } b_1 \neq b_2 \end{array} \right. \Longleftrightarrow \left\{ \begin{array}{ll} H_0: & \beta_1 = \beta_2 \\ H_a: & \beta_1 \neq \beta_2 \end{array} \right. \longrightarrow q = 2$$

Una forma de realizar el contraste es utilizar el estadístico:

$$(R\hat{\beta}_{MCG} - r)' \left[R(X'\Sigma^{-1}X)^{-1}R'\right]^{-1} (R\hat{\beta}_{MCG} - r) \overset{H_0}{\sim} \chi_q^2$$

donde

$$R\beta - r = \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} a_1 \\ b_1 \\ a_2 \\ b_2 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix} \qquad q = 2 \qquad K = 4$$

con la regla de decisión habitual.

• Si σ_1^2 y σ_2^2 son desconocidas es necesario estimarlas. El estimador de los coeficientes en el modelo conjunto sería el de MCGF tal que:

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

Estimar Σ implica estimar σ_1^2 y σ_2^2; un estimador consistente de $\sigma_i^2 \quad i = 1, 2$ sería:

$$\hat{\sigma}_1^2 = \frac{\hat{u}_1'\hat{u}_1}{N_1} \qquad \hat{u}_1'\hat{u}_1 = Y_1'Y_1 - \hat{\beta}_{1,MCO}'X_1'Y_1 \qquad \hat{u}_1 = Y_1 - X_1\hat{\beta}_{1,MCO}$$

$$\hat{\sigma}_2^2 = \frac{\hat{u}_2'\hat{u}_2}{N_2} \qquad \hat{u}_2'\hat{u}_2 = Y_2'Y_2 - \hat{\beta}_{2,MCO}'X_2'Y_2 \qquad \hat{u}_2 = Y_2 - X_2\hat{\beta}_{2,MCO}$$

Notar que en cada ecuación por separado hay homocedasticidad y no autocorrelación.

1.6.3. Ecuaciones aparentemente no relacionadas

Si un conjunto de ecuaciones se relacionan únicamente por los términos de perturbación reciben el nombre de Ecuaciones Aparentemente no Relacionadas. Sea el sistema de ecuaciones

$$\left\{ \begin{array}{lll} Y_1 & = X_1\beta_1 + u_1 & N \text{ obs} \\ Y_2 & = X_2\beta_2 + u_2 & N \text{ obs} \end{array} \right.$$

Un supuesto sencillo acerca de la estructura de la matriz de varianzas y covarianzas sería:

- $E(u_1u_1') = \sigma_1^2 I_N \quad E(u_2u_2') = \sigma_2^2 I_N$, homocedasticidad y no autocorrelación dentro de cada ecuación.
- $E(u_1u_2') = E(u_2u_1') = \sigma_{12} I_N$, correlación contemporánea entre las perturbaciones de las ecuaciones.

$$E(uu') = \begin{bmatrix} \sigma_1^2 I_N & \sigma_{12} I_N \\ \sigma_{12} I_N & \sigma_2^2 I_N \end{bmatrix} = \Sigma_{2N\times 2N}$$

En este caso ganamos en eficiencia si estimamos el modelo conjunto (1.5). Si σ_1^2, σ_2^2, σ_{12} son conocidas estimaremos por MCG. Si son desconocidas el modelo conjunto debiera estimarse por MCGF. Podemos encontrar estimadores consistentes de estos parámetros utilizando los residuos MCO de estimar cada ecuación por separado:

$$\hat{\sigma}_1^2 = \frac{\hat{u}_1'\hat{u}_1}{N} \qquad \hat{u}_1'\hat{u}_1 = Y_1'Y_1 - \hat{\beta}_{1,MCO}'X_1'Y_1 \qquad \hat{u}_1 = Y_1 - X_1\hat{\beta}_1$$

$$\hat{\sigma}_2^2 = \frac{\hat{u}_2'\hat{u}_2}{N} \qquad \hat{u}_2'\hat{u}_2 = Y_2'Y_2 - \hat{\beta}_{2,MCO}'X_2'Y_2 \qquad \hat{u}_2 = Y_2 - X_2\hat{\beta}_2$$

$$\hat{\sigma}_{12} = \frac{\hat{u}_1'\hat{u}_2}{N}$$

1.7. Ejercicios a resolver

Además, de los ejercicios que se resuelven en clase tenéis a vuestra disposición una colección de ejercicios de examen que se actualiza cada año. No deberíais dar por acabado el trabajo de cada tema hasta que están hechos los ejercicios recomendados en clase.

Ejercicio M-MCG.1:
Sea el modelo $\quad Y_t = \alpha + \beta X_t + u_t \quad$ donde $\quad E(u_t^2) = tX_t^2$

1. Conociendo cinco observaciones de Y_t y X_t obtén por MCO y en forma matricial, las estimaciones de α y β del modelo anterior.

t	1	2	3	4	5
Y_t	1	1	0	-1	-1
X_t	1	-1	1	1	1

2. Ahora además, se conoce:

$$E(u_1u_3) = E(u_3u_1) = 1 \text{ y } E(u_1u_2) = E(u_2u_1) = E(u_2u_3) = E(u_3u_2) = 0$$
$$E(u_1u_4) = E(u_4u_1) = -1 \text{ y } E(u_1u_5) = E(u_5u_1) = E(u_2u_5) = E(u_5u_2) = 0$$
$$E(u_3u_5) = E(u_5u_3) = 1 \text{ y } E(u_3u_4) = E(u_4u_3) = E(u_4u_5) = E(u_5u_4) = 0$$
$$E(u_2u_4) = E(u_4u_2) = 1$$

   Dadas las observaciones de Y_t e X_t y la información sobre $E(u_tu_s)$, calcula la matriz de varianzas y covarianzas del estimador MCO.

3. Dada la información anterior, ¿qué propiedades tiene el estimador Mínimo Cuadrático Ordinario?

4. ¿Conoces un estimador con mejores propiedades? ¿Cuál es? ¿Qué propiedades tiene? Escribe su matriz de varianzas y covarianzas. No la estimes, escribe solo su fórmula y explica que son cada uno de sus elementos.

Ejercicio M-MCG.2:
Considera el siguiente modelo de regresión general:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t \quad t = 1, \ldots, 500$$

donde X_2 y X_3 son no estocásticas y $u_t \sim NID\,(0, \sigma_t^2)$ con $\sigma_t^2 = \sigma^2 t^2$.

1. Escribe $E(u)$ y $E(uu')$.

2. Obtén la matriz de varianzas y covarianzas de Y.

3. Se ha estimado el modelo por Mínimos Cuadrados Generalizados, obteniéndose las siguientes estimaciones:

$$\hat{\beta}_{MCG} = \begin{bmatrix} 1 \\ 3 \\ -1 \end{bmatrix} \quad \widehat{V}(\hat{\beta}_{MCG}) = \begin{bmatrix} 3 & -1 & 0 \\ -1 & 6 & -2 \\ 0 & -2 & 2 \end{bmatrix}$$

Realiza los contrastes de las siguientes hipótesis:

a) $\beta_1 = 0$

b) $\beta_2 + 2\beta_3 = 1$

c) $\beta_1 = 0$ y $\beta_2 + 2\beta_3 = 1$

Ejercicio M-MCG.3:

En el modelo $Y_t = \alpha + \beta X_t + u_t \quad t = 1, \ldots, 4$ se cumple $E(u_t^2) = tX_t^2$ y $E(u_t u_s) = 0 \quad \forall t, s \quad t \neq s$, tal que:

t	1	2	3	4
Y_t	2	1	-1	-3
X_t	1	-1	0	1

1. Escribe la matriz de varianzas y covarianzas de la perturbación $E(uu')$.
2. Escribe el modelo transformado con perturbaciones esféricas y demuestra que sus perturbaciones tienen varianza constante.
3. En el modelo transformado, escribe la matriz de regresores y el vector de valores de la variable endógena.
4. Estima, utilizando cálculo matricial, los parámetros del modelo transformado.

1.8. Bibliografía del tema

Referencias Bibliográficas Básicas:

• Teórica:

[1] Greene, W. (1998), Análisis Econométrico, ed. Prentice Hall, 3ª edition.

[2] Ramanathan, R. (2002), Introductory Econometrics with applications, ed. South-Western, 5th edition.

[3] Wooldridge, J. M. (2003), Introductory Econometrics: A modern Approach, ed. South-Western, 2nd edition.

• Ejercicios:

[1] Fernández, A., González, P., Regúlez, M., Moral, P., Esteban, V. (2005), Ejercicios de Econometría, ed. McGraw-Hill, 2 ª edición.

[2] Recopilación de ejercicios recomendados y exámenes de Econometría. Departamento de Economía Aplicada III (Econometría y Estadística). Mimeo Febrero 2009.

• Ejercicios con gretl:

[1] Ramanathan, R. (2002), Instructor's Manual to accompany, del libro Introductory Econometrics with applications, ed. South-Western, 5th edition, Harcourt College Publishers.

[2] Wooldridge, J. M. (2003), Student Solutions Manual, del libro Introductory Econometrics: A modern Approach, ed. South-Western, 2nd edition.

Referencias Bibliográficas Complementarias:

• Teórica:

[1] Alonso, A., Fernández, J. y Gallastegui, I. (2005), Econometría, ed. Pearson: Prentice Hall.

[2] Gujarati, D. (2004), Econometría, ed. McGraw-Hill, 4ª edición.

[3] Johnston, J. (1984), Métodos de Econometría, ed. Vicens Vivens, 4ª edición.

[4] Maddala, G. S. (1996), Introducción a la Econometría, ed. Pearson: Prentice Hall.

[5] Novales, A. (1993), Econometría, ed. McGraw-Hill, 2ª edición.

Tema 2

Heterocedasticidad

Clases Magistrales

En las clases magistrales del tema de heterocedasticidad vamos a abordar uno de los problemas mencionados en el tema anterior: cómo estimar eficientemente los coeficientes desconocidos de un modelo cuando la varianza de la perturbación no es constante y cómo realizar inferencia de forma adecuada.

En el tema anterior se han mostrado las consecuencias en el estimador de Mínimos Cuadrados Ordinarios, MCO, de no tener perturbaciones esféricas y cómo estimar por Mínimos Cuadrados Generalizados, MCG, obteniendo estimadores eficientes de los coeficientes desconocidos. Por ello los instrumentos ya se conocen. En este tema vamos a concentrarnos en mostrar situaciones en las que la varianza de la perturbación no es constante y trabajar de forma práctica con diferentes supuestos sobre el comportamiento de la varianza de la perturbación.

Si bien buena parte del instrumental de estimación y contraste no es nuevo, si que lo serán los estadísticos de contraste para la hipótesis de homocedasticidad. Junto con el análisis gráfico de los residuos mínimo-cuadráticos se mostraran dos estadísticos de contraste, el contraste de Goldfeld y Quandt y el contraste de Breusch y Pagan. También se mostrará como utilizar el estimador de Mínimos Cuadrados Generalizados Factibles, MCGF, bajo el supuesto de varianza no constante, desconocida y estimable. Finalmente, se aprenderá a obtener un estimador consistente de la matriz de varianzas y covarianzas del estimador MCO.

En el tema de heterocedasticidad habitualmente se dispone de siete clases magistrales de las cuales, aproximadamente, tres horas se dedican a resolver ejercicios. Además, se resolverán las dudas surgidas y se realizaran preguntas al alumno sobre los contenidos vistos, en muchas ocasiones en forma de preguntas cortas que se recogen y evalúan convenientemente.

Competencias a trabajar en estas sesiones:

1. Comprender la importancia de los supuestos empleados en la especificación de un modelo econométrico básico para poder proponer y emplear supuestos más realistas.

2. Diferenciar distintos métodos de estimación y evaluar su uso de acuerdo a las características de las variables económicas de interés para obtener resultados fiables.

Al final de este tema deberíais ser capaces de:

1. Explicar que se entiende por un modelo de regresión lineal con heterocedasticidad, cómo modelizar esta característica en el término de perturbación del modelo y sus implicaciones en su matriz de varianzas y covarianzas.
2. Saber analizar gráficamente la posible existencia de heterocedasticidad y saber contrastarla utilizando los estadísticos de Goldfeld-Quandt y/o Breusch-Pagan.
3. Describir y comparar las propiedades de los estimadores MCO y MCG bajo heterocedasticidad.
4. Transformar un modelo con perturbación heterocedástica en un modelo con perturbación homocedástica.
5. Estimar por MCG o MCGF según la especificación de la varianza de la perturbación del modelo.
6. Razonar por qué resulta conveniente disponer de un estimador consistente de la matriz de varianzas y covarianzas de los estimadores de MCO en presencia de heterocedasticidad. Conocer, saber obtener y utilizar dicho estimador en el contexto adecuado.

Bibliografía Recomendada:

Al final del tema tenéis recogida la bibliografía correspondiente. En particular os recomendamos leer los capítulos correspondientes a la bibliografía básica detallados a continuación:

- Greene, W. (1998), cap. 12.
- Ramanathan, R. (2002), cap. 8.
- Wooldridge, J. M. (2003), cap. 8.

Y para profundizar, podéis leer los capítulos detallados a continuación correspondientes a la bibliografía complementaria:

- Alonso, A. et al. (2005), cap. 9.
- Gujarati, D. (2004), cap. 11.
- Johnston, J. (1984), cap. 7.
- Maddala, G. S. (1996), cap. 5.
- Novales, A. (1993), cap. 5, cap. 6.

2.1. Concepto de heterocedasticidad. Naturaleza y consecuencias. Ejemplos

Se dice que la varianza del término de perturbación del modelo de regresión lineal es heterocedástica cuando no es constante para todas las observaciones. En este tema nuestro marco de trabajo queda definido como sigue:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \ldots + \beta_K X_{Ki} + u_i \qquad i = 1, 2, \ldots, N \Longleftrightarrow Y_i = \mathbf{X}_i' \beta + u_i$$

donde los regresores contenidos en la matriz $\mathbf{X}_i' = [1 \quad \mathbf{X}_{2i} \ \ldots \ \mathbf{X}_{Ki}]$ son no estocásticos y las perturbaciones tienen media cero, varianza no constante y covarianzas cero:

$$\begin{array}{l} E(u_i) = 0 \quad \forall i \\ E(u_i^2) = \sigma_i^2 \quad i = 1, 2, \ldots, N \\ E(u_i u_j) \neq 0 \ \forall i, j \quad i \neq j \end{array}$$

En términos matriciales el modelo se escribe:

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

y la estructura de **la matriz de varianzas y covarianzas del vector de perturbaciones** sería la siguiente:

$$\underset{(N\times N)}{E(uu')} = \begin{bmatrix} \sigma_1^2 & 0 & 0 & \ldots & 0 \\ 0 & \sigma_2^2 & 0 & \ldots & 0 \\ 0 & 0 & \sigma_3^2 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & \sigma_N^2 \end{bmatrix} = \sigma^2 \begin{bmatrix} w_{11} & 0 & 0 & \ldots & 0 \\ 0 & w_{22} & 0 & \ldots & 0 \\ 0 & 0 & w_{33} & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & w_{NN} \end{bmatrix}$$

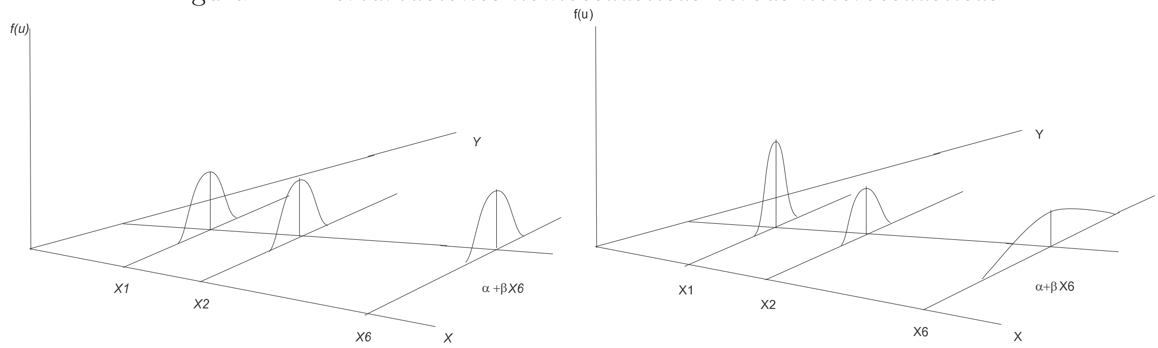
donde $Var(u_i) = \sigma_i^2$ puede cambiar para cada individuo y suponemos que no existe correlación entre perturbaciones de distintos individuos, $E(u_i u_j) = 0 \ \forall i, j \quad i \neq j$. Es decir, sólo consideramos la existencia de heterocedasticidad. Esta matriz es una matriz diagonal de dimensión $(N \times N)$ donde los elementos de la diagonal no son todos iguales. El parámetro σ^2 es un factor de escala común a todos los elementos de la matriz, que perfectamente puede tomar el valor de la unidad.

Para diferenciar entre el concepto de homocedasticidad y el concepto de heterocedasticidad podemos considerar la Figura 2.1. En la izquierda se puede observar que la varianza condicional de Y_i a las X_i permanece igual sin importar los valores que tome la variable X. Hay que recordar que la varianza condicional de Y_i es la misma que la de u_i, por tanto, en el gráfico estamos observando como la varianza de la perturbación permanece constante independientemente del valor que tome el regresor. En la derecha se muestra que la varianza de Y_i aumenta a medida que X aumenta[1]. Hay heterocedasticidad, y lo denotamos:

$$Var(u_i) = E(u_i^2) = \sigma_i^2.$$

[1] En este caso la varianza de u_i se relaciona con X_i y lo hace de forma creciente, pero podría ser función de otra u otras variables que no sean regresores del modelo. La forma funcional también puede ser diferente.

Figura 2.1: *Perturbaciones homocedásticas versus heterocedásticas*



Homocedásticas *Heterocedásticas*

La existencia de heterocedasticidad puede aparecer en numerosas aplicaciones económicas sin embargo, es más habitual en datos de sección cruzada. A continuación veremos algunas situaciones en las cuales las varianzas de u_i pueden no ser constantes.

- **En datos de sección cruzada.**

Ejemplo 2.1 Supongamos que tenemos datos para diferentes comunidades autónomas españolas en el año 2005 sobre gasto sanitario agregado, GS, renta personal disponible, R, el porcentaje de población que supera los 65 años, SEN y población, POP, con los que estimar el siguiente modelo:

$$GS_i = \beta_1 + \beta_2 R_i + \beta_3 SEN_i + \beta_4 POP_i + u_i \qquad i = 1, \ldots, N \tag{2.1}$$

Las comunidades con más población y/o mayor porcentaje de población con edad superior a 65 años tendrán mayor gasto sanitario que aquellas con menor población o más joven. En esta situación suponer que la dispersión de los gastos sanitarios es la misma para todas las comunidades con distinto nivel de población y composición de la misma no es realista, y se debería proponer que la varianza de la perturbación sea heterocedástica $Var(u_i) = \sigma_i^2$, permitiendo por ejemplo que varíe en función creciente con la población, es decir, $\sigma_i^2 = \sigma^2 POP_i$. Incluso podemos pensar que varíe en función creciente con el porcentaje de población mayor de 65 años, en cuyo caso propondríamos $Var(u_i) = \sigma^2 SEN_i$ o con ambas variables, por lo que la forma funcional pudiera ser $Var(u_i) = \sigma^2(a\, POP_i + b\, SEN_i)$.

Ejemplo 2.2 Un ejemplo recurrente para mostrar la heterocedasticidad es el estudio de la relación entre consumo y renta. Este ejemplo se va a seguir detalladamente a lo largo del tema utilizándolo como ejemplo magistral. Será completamente descrito en la siguiente sección, pero reflexionemos un momento sobre él. Supongamos que tenemos datos sobre renta, R, y gasto en consumo, C, para N familias, con los que estimar el modelo:

$$C_i = \beta_1 + \beta_2 R_i + u_i \qquad i = 1, \ldots, N \tag{2.2}$$

Las familias con mayor renta, una vez satisfechas sus necesidades primordiales tienen mayores posibilidades de decidir cuánto ahorrar y cuánto consumir, por lo que es habitual encontrar una mayor variabilidad en el gasto realizado por familias de renta alta que por familias de renta baja. En esta situación suponer que la dispersión de los gastos de consumo es la misma para todas las familias con distinto nivel de renta no es realista y se debería proponer que la varianza de la perturbación sea heterocedástica $Var(u_i) = \sigma_i^2$, permitiendo por ejemplo que varíe en función creciente con la renta de las familias, es decir, $\sigma_i^2 = \sigma^2 R_i$.

Ejemplo 2.3 Un fenómeno parecido ocurre con las empresas que deben decidir qué porcentaje de sus beneficios, B, deben repartir como dividendos, D. Las empresas con mayores beneficios tienen un margen de decisión muy superior al fijar su política de dividendos. Al estimar el modelo:

$$D_i = \beta_1 + \beta_2 B_i + u_i \qquad i = 1, \ldots, N \tag{2.3}$$

cabría esperar que la varianza de u_i dependa del nivel de beneficios de la empresa i-ésima y podríamos proponer que por ejemplo, $E(u_i^2) = \sigma_i^2 = \sigma^2 B_i$.

- La heterocedasticidad también puede aparecer **como consecuencia de la agregación de datos**. En este caso la varianza puede depender del número de observaciones del grupo.

Ejemplo 2.4 Supongamos un investigador que desea estimar los coeficientes del siguiente modelo:

$$Y_j = \beta_1 + \beta_2 X_j + u_j \qquad j = 1, \ldots, N \tag{2.4}$$

donde $u_j \sim N(0, \sigma^2)$, es decir, la varianza de la perturbación es homocedástica. Supongamos que el número de observaciones N es tal que aconseja agrupar las observaciones en m-grupos de n_i observaciones cada uno. Supongamos que como observación del grupo i-ésimo se toma la media aritmética dentro del grupo. El modelo a estimar sería:

$$\bar{Y}_i = \beta_1 + \beta_2 \bar{X}_i + \bar{u}_i \qquad i = 1, \ldots, m \tag{2.5}$$

y la nueva perturbación $\bar{u}_i$ seguirá teniendo media cero, pero su varianza no será constante ya que dependerá del número de observaciones dentro del grupo,

$$Var(\bar{u}_i) = \frac{\sigma^2}{n_i} \qquad i = 1, \ldots, m.$$

Si el número de observaciones dentro del grupo es el mismo en todos los grupos la varianza de la perturbación $\bar{u}_i$ es homocedástica.

- También encontraremos heterocedasticidad en un **modelo con coeficientes aleatorios** en el cual se permite cierta heterogeneidad entre individuos en los efectos de una variable explicativa.

Ejemplo 2.5 Sea el modelo:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_{3i} X_{3i} + u_i \qquad i = 1, \ldots, N \tag{2.6}$$

donde:

- las variables exógenas X_{2i} y X_{3i} son no estocásticas.
- $u_i \sim iid(0, \sigma_u^2)$
- $\beta_{3i} = \beta + \epsilon_i$ siendo β una constante a estimar. La variable aleatoria ϵ_i recoge la heterogeneidad tal que, aunque en media el efecto es común a todos los individuos, recogido en β, hay cierta variabilidad recogida por σ_ϵ^2.
- $\epsilon_i \sim iid(0, \sigma_\epsilon^2)$ y $Cov(u_i, \epsilon_i) = 0$

En este caso la ecuación a estimar sería:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta X_{3i} + v_i \qquad i = 1, \ldots, N \tag{2.7}$$

donde $v_i = u_i + \epsilon_i X_{3i}$ con distribución:

$$v_i \sim (0, \sigma_u^2 + X_{3i}^2 \sigma_\epsilon^2)$$

y vemos que la varianza del término de perturbación del modelo a estimar no es constante ya que depende de la variable exógena X_{3i} que varía con i.

- Otro caso sería la **existencia de un cambio estructural en varianza** recogido por una variable ficticia en la varianza de la perturbación.

Ejemplo 2.6 Supongamos que se desea estudiar la relación entre producción, Y, y mano de obra, X, para un conjunto de 20 trabajadores de los cuales 10 son mujeres y el resto hombres. Si suponemos que la variabilidad de la producción es distinta para los hombres que para las mujeres nuestro modelo a estimar sería:

$$Y_i = \beta_1 + \beta_2 X_i + u_i \qquad i = 1, \ldots, 20 \tag{2.8}$$

donde $u_i \sim (0, \alpha_1 + \alpha_2 D_i)$ siendo D_i una variable ficticia que toma valor la unidad si la observación corresponde a una mujer y cero en el caso contrario. En este caso:

$$\begin{array}{ll} Var(u_i) = \alpha_1 + \alpha_2 & \text{para las observaciones correspondientes a las mujeres} \\ Var(u_i) = \alpha_1 & \text{para las observaciones correspondientes a los hombres} \end{array}$$

Suponiendo que las primeras diez observaciones corresponden a mujeres, la matriz de varianzas y covarianzas del vector de perturbaciones sería la siguiente:

$$E(uu') = \begin{bmatrix} (\alpha_1 + \alpha_2) I_{10} & 0 \\ 0 & \alpha_1 I_{10} \end{bmatrix}$$

Antes de pasar a abordar el problema de hereterocedasticidad más en detalle vamos a recordar cuáles son las consecuencias de la heterocedasticidad sobre el estimador de MCO y su matriz de varianzas y covarianzas:

- Si $u \sim N(0, \sigma^2\Omega)$ entonces:

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

  el estimador MCO será lineal e insesgado, pero no será de varianza mínima si $\Omega \neq I$.

- En el caso de un modelo con perturbaciones heterocedásticas la variabilidad de las observaciones de la variable endógena dada la exógena no es la misma para todas las observaciones. El criterio MCO prescinde de esta información ya que concede a todas las observaciones el mismo peso.

- En estas circunstancias parecería más adecuado ponderar más las realizaciones muestrales que sistemáticamente se desvían menos del valor promedio de la variable endógena, que es sobre lo que queremos inferir. Es decir, desearíamos dar mayor peso a aquellas observaciones que surgen de poblaciones con menor variabilidad que a aquellas que provienen de poblaciones con mayor variabilidad.

- Debemos estimar por MCG (cuando Ω es conocida). El criterio de estimación Mínimo Cuadrático Generalizado que estudiábamos en el tema anterior es:

$$\underset{\hat{\beta}}{Min} \left[(Y - X\hat{\beta})'\, \Omega^{-1}\, (Y - X\hat{\beta})\right]$$

- El estimador de MCG que se obtiene de la función objetivo anterior se define como:

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

  El estimador MCG es lineal, insesgado y de varianza mínima. Su matriz de varianzas y covarianzas es:

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

  Si σ^2 es desconocida, que es lo habitual, la estimamos con la expresión $\hat{\sigma}^2 = \frac{\hat{u}'\Omega^{-1}\hat{u}}{N-K}$ siendo $\hat{u} = Y - X\hat{\beta}_{MCG}$.

Ejemplo 2.7 Para el modelo $Y_i = \beta X_i + u_i$ siendo X una variable no estocástica y $u_i \sim NID(0, \sigma_i^2)$. El criterio MCG (o función objetivo) se escribiría:

$$\underset{\hat{\beta}}{Min} \sum_{i=1}^{N} \frac{1}{\sigma_i^2}\, (Y_i - X_i\hat{\beta})^2$$

de donde

$$\hat{\beta}_{MCG} = \left[\sum_{i=1}^{N} \frac{1}{\sigma_i^2}\, X_iX_i\right]^{-1} \left[\sum_{i=1}^{N} \frac{1}{\sigma_i^2}\, X_iY_i\right] = \frac{\sum_{i=1}^{N} \frac{1}{\sigma_i^2} X_iY_i}{\sum_{i=1}^{N} \frac{1}{\sigma_i^2} X_iX_i}$$

En la práctica, generalmente no sabemos de antemano si hay o no problemas de heterocedasticidad en las perturbaciones, por lo que se ha desarrollado un gran número de procedimientos para contrastar la hipótesis nula de igualdad de varianzas u homocedasticidad. Esta gran variedad se debe a que la especificación de la hipótesis alternativa de heterocedasticidad no suele ser conocida y puede ser más o menos general. En la siguiente sección vamos a abordar algunos de estos contrastes.

2.2. Contrastes de heterocedasticidad

Nuestro objetivo es claro: **Detectar la existencia de heterocedasticidad en las perturbaciones de un modelo**. La primera aproximación al objetivo es el estudio de los gráficos de residuos y de las variables del modelo.

2.2.1. Detección gráfica.

La aplicación del estimador de MCG y algunos contrastes de heterocedasticidad requieren conocer la forma funcional de la varianza de la perturbación. Si suponemos que la varianza de la perturbación depende de uno o más regresores, u otras variables conocidas, un instrumento adecuado para aproximarnos a la misma sería llevar a cabo un análisis de los residuos MCO donde no hemos tenido en cuenta la existencia de heterocedasticidad. Aunque $\hat{u}_{MCO,i}$ no es lo mismo que u_i la detección de patrones sistemáticos en la variabilidad de los residuos MCO nos indicará la posible existencia de heterocedasticidad en las perturbaciones. Además, puede indicarnos una posible forma funcional de la misma.

Consideramos el modelo (2.1) recogido en el Ejemplo 2.1:

$$GS_i = \beta_1 + \beta_2 R_i + \beta_3 SEN_i + \beta_4 POP_i + u_i \qquad i = 1, \ldots, N$$

donde suponemos $E(u_i) = 0 \quad \forall i$ y $E(u_i u_j) = 0 \quad \forall i, j \quad i \neq j$. Si sospechamos que u_i es heterocedástica debido a la variable POP, podemos intentar detectar la existencia de heterocedasticidad en las perturbaciones del modelo ayudándonos del gráfico de los residuos MCO, ($\hat{u}_{MCO,i}$), frente a la variable POP_i.

Si el gráfico es como el recogido en la Figura 2.2 pensaremos que la variabilidad de los residuos $\hat{u}_{MCO,i}$ se incrementan con POP_i y que el incremento es directamente proporcional. Así, podríamos proponer, por ejemplo:

$$E(u_i^2) = \sigma^2 POP_i \qquad i = 1, 2, \ldots, N$$

Si el gráfico de los residuos MCO frente a POP *hubiera* sido como el recogido en la Figura 2.3 supondríamos que el aumento en la varianza de u_i es inversamente proporcional a POP_i y propondríamos:

$$E(u_i^2) = \sigma^2 POP_i^{-1} \qquad i = 1, 2, \ldots, N$$

También podemos optar por dibujar la serie de los residuos al cuadrados MCO frente a la variable que creemos causa la heterocedasticidad como se muestra en la Figura 2.4. En el gráfico de la izquierda se muestran los pares $(SEN_i, \hat{u}_{MCO,i})$, en el gráfico de la derecha se muestran los pares

Figura 2.2: *Residuos MCO versus POP*

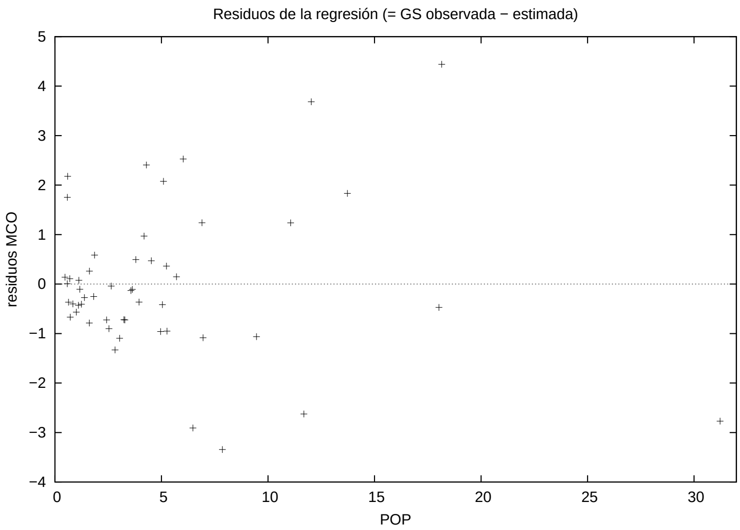
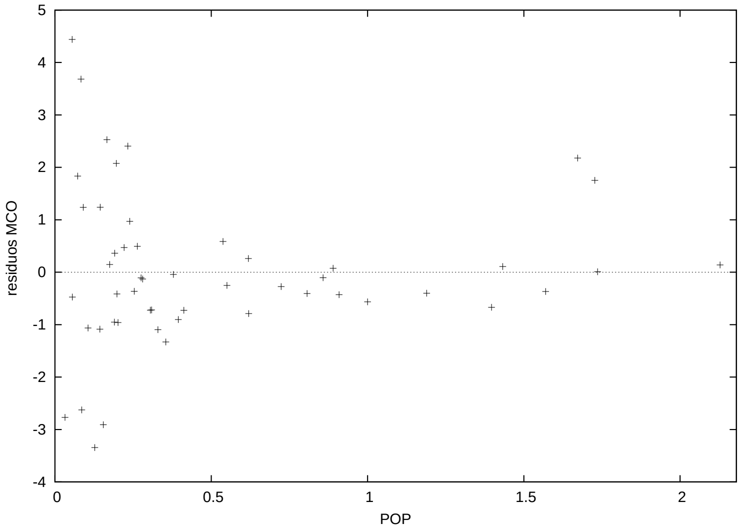


Figura 2.3: *Residuos MCO versus POP*



$(SEN_i, \hat{u}^2_{MCO,i})$. Ambos gráficos muestran la misma información, muestran que la variabilidad de los residuos se incrementa con SEN y podríamos proponer, por ejemplo $Var(u_i) = E(u_i^2) = \sigma^2 SEN_i$.

En general a priori no se conocerá cuál de las variables exógenas genera la heterocedasticidad por lo que resulta aconsejable estudiar los gráficos de los residuos de MCO, contraponiéndolos a cada una de las variables exógenas del modelo, como estamos haciendo al estudiar los residuos frente a POP_i y frente a SEN_i. Notar que ambas variables parecen afectar a la varianza de la perturbación, por ello estaría justificado proponer $Var(u_i) = (a\, POP_i + b\, SEN_i)$, donde a y b son desconocidos y el factor de escala es la unidad, $\sigma^2 = 1$.

Figura 2.4: *Residuos MCO y sus cuadrados versus SEN*

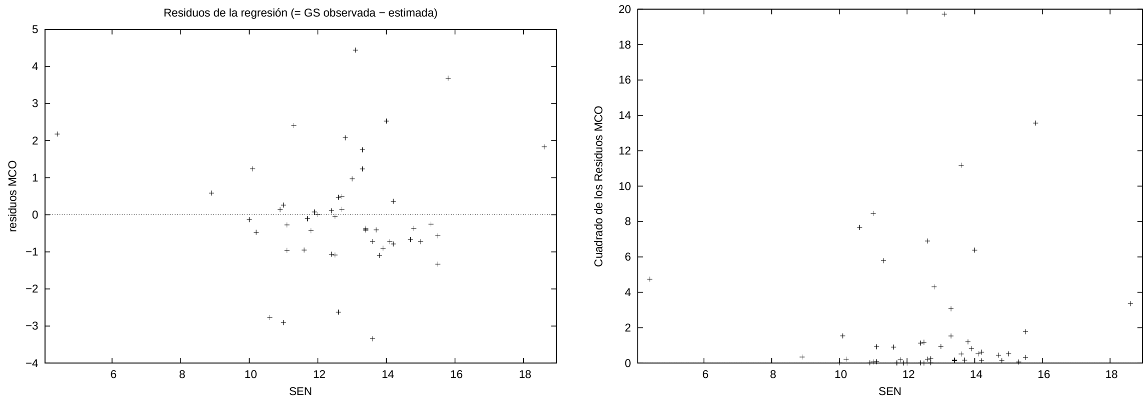
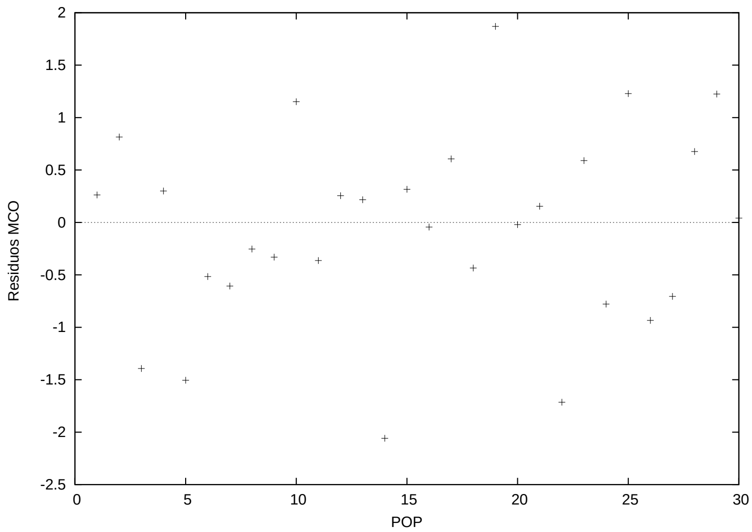


Figura 2.5: *Perturbaciones homocedásticas*

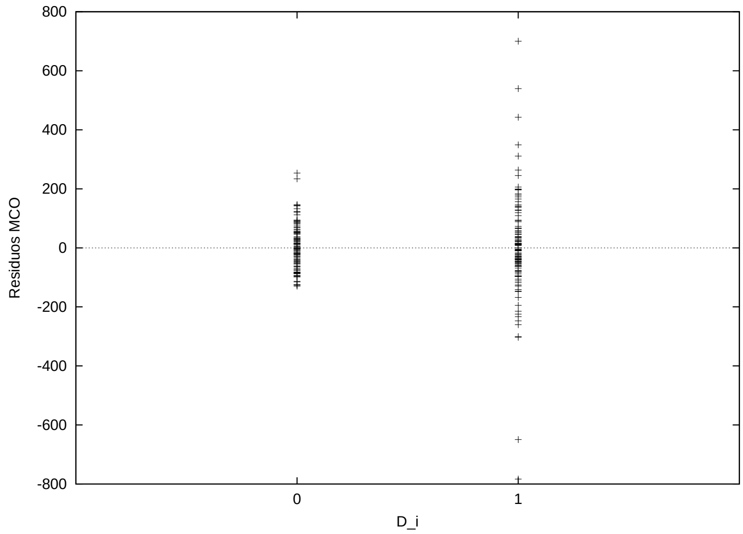


Si la gráfica entre $\hat{u}_{MCO,i}$ y POP_i *hubiera* resultado como la de la Figura 2.5, concluiríamos que la varianza de la perturbación no depende de POP_i ya que no se aprecia ningún patrón de comportamiento y parece que hay una distribución aleatoria de los pares $(POP_i, \hat{u}_i)$. En esta situación procede analizar los residuos frente al resto de regresores del modelo.

Las formas anteriores no son las únicas. Si recordamos, en el Ejemplo 3.6 se suponía una situación donde hombres y mujeres en una empresa tenían diferente productividad y se suponía que $Var(u_i) = \alpha_1 + \alpha_2 D_i$ siendo D_i una variable ficticia que toma valor uno si la observación corresponde a una mujer y cero en caso contrario. En esta situación esperaríamos un gráfico como el recogido en la Figura 2.6 donde claramente la dispersión de los residuos para las mujeres es mucho mayor que para los hombres.

Como conclusión diremos que al analizar los gráficos de la relación residuos MCO, o sus cuadrados, con cada uno de los regresores lo que intentaremos detectar visualmente es un crecimiento o

Figura 2.6: *Residuos MCO frente a una variable ficticia*



decrecimiento en la variabilidad de los residuos con respecto a la variable en cuestión.

• Ejemplo a seguir en las clases magistrales del resto del tema:

En lo que sigue del tema resulta muy interesante disponer de un ejercicio sobre el que ir trabajando y viendo resultados, y sobre todo para que podáis replicar los mismos. Por ello os proponemos realizar un estudio sobre la relación entre Consumo, C, y Renta, R, para 40 familias. En la plataforma de apoyo a la docencia, eKASI, tenéis disponible un archivo llamado ejemHETERO.gdt que contiene las observaciones de consumo y renta del ejercicio para que podáis replicarlo por vuestra cuenta. En el Anexo 1.1 que se encuentra al final de este tema podéis encontrar los resultados de gretl que se van mostrando.

Objetivo: Analizar el comportamiento de la varianza de la perturbación en un análisis de la relación entre consumo y renta.

Datos: Observaciones sobre gasto semanal en alimentos y renta semanal para 40 familias. Ambos medidos en dólares. Se muestran en la Tabla A.3 recogida en el Apéndice[2]. Modelo a estimar:

$$C_i = \beta_1 + \beta_2 R_i + u_i \qquad i = 1, 2, \ldots, 40$$

donde:

$$E(u_i) = 0 \quad \forall i, \qquad E(u_i u_j) = 0 \qquad \forall i, j \quad i \neq j$$

y se cumplen el resto de hipótesis básicas. Nos preguntamos: **¿Es constante la varianza de la perturbación?**

[2]Con la tabla también tenéis suficiente información para introducir vosotros mismos los datos en gretl. El procedimiento aparece descrito en el Anexo 1.2 en el que se muestran instrucciones básicas de gretl.

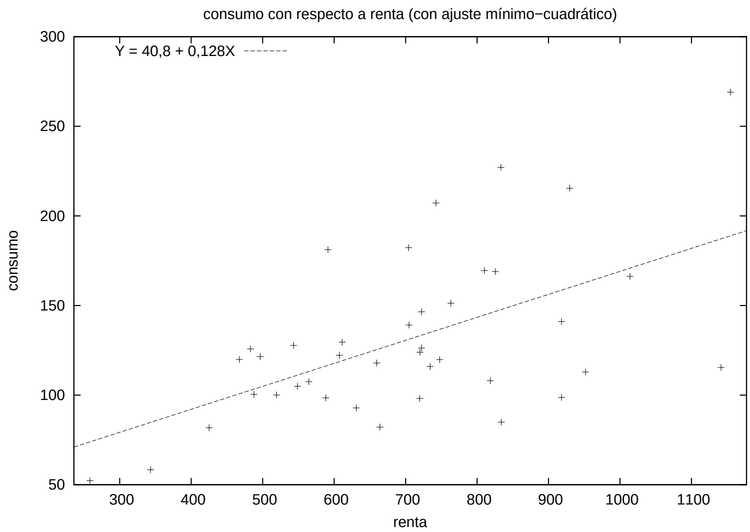
Lo primero que vamos a hacer es estimar el modelo propuesto por MCO y analizar el gráfico de la relación entre consumo y renta, Figura 2.7. Los resultados de la regresión son los siguientes[3]:

Estimaciones MCO utilizando las 40 observaciones 1–40
Variable dependiente: consumo

	Coeficiente	Desv. típica	estadístico t	valor p
const	40,7676	22,1387	1,8415	0,0734
renta	0,1282	0,0305	4,2008	0,0002

Suma de cuadrados de los residuos	54311,3
Desviación típica de la regresión ($\hat{\sigma}$)	37,8054
R^2	0,3171
$\bar{R}^2$ corregido	0,2991

Figura 2.7: *Consumo versus Renta*

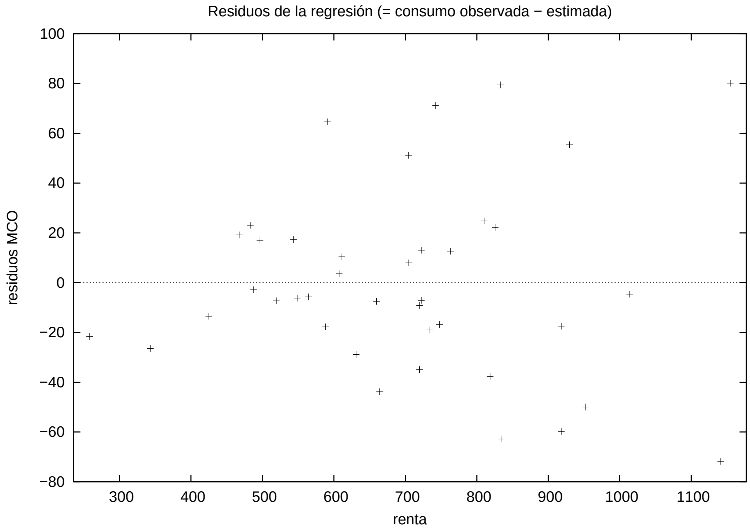


En la Figura 2.7 se muestra el gráfico de las observaciones de la variable Consumo, C, sobre la variable explicativa Renta, R, junto con la recta de ajuste mínimo-cuadrático. El gráfico muestra que existe una relación lineal y positiva entre consumo y renta. A mayor renta mayor consumo. Además, se puede observar como no solo que las familias con mayor renta tienen un consumo en media mayor que las de menor renta, sino que además, las familias de rentas altas tienen una mayor variación en su consumo relativamente a las familias de rentas bajas. Lo mismo podemos concluir de los resultados de la estimación. Aunque el ajuste no es muy bueno, ya que el R^2 es bajo, si la perturbación fuese esférica y normal, la renta sería una variable significativa al 5 % para explicar el consumo.

La Figura 2.8 muestra los pares $(R_i, \hat{u}_{i,MCO})$. El gráfico permite dudar sobre el comportamiento homocedástico de la varianza de la perturbación ya que se puede ver claramente como a medida que aumenta la variable renta aumenta la dispersión en los residuos.

[3]Ver Anexo 1.1, Resultado 1.

Figura 2.8: *Residuos MCO versus Renta*



La apreciación gráfica de una posible existencia de heterocedasticidad debe ser refrendada mediante un contraste.

2.2.2. Test de contraste para heterocedasticidad

A continuación veremos algunos de los test de contraste para heterocedasticidad más importantes. Todos ellos contrastan la existencia de heterocedasticidad suponiendo:

H_0**: ausencia de heterocedasticidad.**
H_a**: existencia de heterocedasticidad.**

Existe una gran variedad de test de contraste de heterocedasticidad que se diferencian entre sí en su generalidad a la hora de modelizar la heterocedasticidad en la hipótesis alternativa y su potencia para detectarla. En este curso aprenderemos dos test de contraste, el test de Goldfeld y Quandt y el test de Breusch y Pagan. El primero de ellos parte del supuesto de que la varianza de la perturbación depende monótonamente de los valores de una única variable y es razonablemente potente si somos capaces de identificar bien esa variable. Por contra, el segundo supone que varianza de la perturbación varía en función de un conjunto de variables independientes. Cuanto mejor seamos capaces de identificar la naturaleza de la heterocedasticidad mejor podremos contrastarla. Pero incluso en el caso en que no seamos capaces de identificar cual es su naturaleza podemos implementar un contraste de tipo general como es el test de White que no requiere hacer supuestos sobre la naturaleza de la misma. El test de White es un test de carácter muy general, y por lo mismo de baja potencia, que en gretl puede ser obtenido directamente como se muestra en el Anexo 1.2.

- **Test de Goldfeld y Quandt (1965):**

Parte del supuesto de que la varianza de la perturbación, σ_i^2 depende monótonamente de los valores de una variable Z_i, que puede ser o no uno de los regresores del modelo. En cualquier caso debe ser una variable observable. Para contrastar la hipótesis nula de ausencia de heterocedasticidad:

$$H_0 : \sigma_1^2 = \sigma_2^2 = \ldots = \sigma_N^2$$

contra la alternativa de existencia de heterocedasticidad:

$$H_a : \sigma_i^2 = \sigma^2 g(Z_i)$$

donde $g(\cdot)$ es una función monótona, supongamos que creciente con Z_i, podemos proceder de la manera siguiente:

1. Ordenar las observaciones de todas las variables del modelo en la muestra según un ordenamiento de los valores de Z_i de menor a mayor.
2. Dividir la muestra en dos bloques de tamaño muestral N_1 y N_2 respectivamente, pudiendo dejar fuera p observaciones centrales para hacer más independientes los dos grupos. El número de observaciones de cada grupo ha de ser similar y mayor que el número de parámetros a estimar[4].
3. Estimar por MCO el modelo de regresión separadamente para cada grupo de observaciones. Guardar la Suma de Cuadrados Residual (SCR) de cada regresión.
4. Construir el siguiente estadístico de contraste, que bajo la hipótesis nula de ausencia de heterocedasticidad y suponiendo que la perturbación sigue una distribución normal, de media cero y no está serialmente correlacionada, sigue una distribución F-Snedecor.

$$GQ = \frac{\hat{\sigma}_2^2}{\hat{\sigma}_1^2} = \frac{\hat{u}_2'\hat{u}_2}{\hat{u}_1'\hat{u}_1}\frac{N_1 - K}{N_2 - K} \overset{H_0}{\sim} F_{(N_2-K, N_1-K)}$$

donde:
$\hat{u}_2'\hat{u}_2$ es la SCR de la regresión de Y sobre X en el segundo grupo de observaciones.
$\hat{u}_1'\hat{u}_1$ es la SCR de la regresión de Y sobre X en el primer grupo de observaciones.

• Interpretación del contraste y regla de decisión:
Si existe homocedasticidad las varianzas han de ser iguales, pero si existe heterocedasticidad del tipo propuesto, con la ordenación de la muestra de menor a mayor, la varianza del término de error será mayor al final de la muestra. Entonces $\hat{u}_2'\hat{u}_2$ debería ser sensiblemente mayor que $\hat{u}_1'\hat{u}_1$.

Cuanto más diverjan las sumas de cuadrados, mayor será el valor del estadístico y mayor será la evidencia contra la H_0, rechazaremos H_0, a un nivel de significación α si:

$$GQ > F_{(N_1-K, N_2-K)|\alpha}$$

Observaciones:

[4]En el caso de tener pocas observaciones, p puede ser cero.

- Si se sospecha que la varianza del término de error depende inversamente de los valores que toma una variable Z_i, entonces se debería ordenar la muestra de acuerdo a un ordenamiento de mayor a menor de los valores decrecientes de dicha variable y proceder del modo descrito anteriormente.

- ¿Cómo elegir p?
Anteriormente se ha propuesto dividir la muestra en dos partes. Elegir el valor de p es relevante ya que cuanto mayor sea p más grados de libertad se pierden y por tanto, perdemos potencia del contraste. Si p es demasiado pequeño no habrá independencia entre grupos y se prima la homocedasticidad frente a la posibilidad de heterocedasticidad. Harvey y Phillips (1974) sugieren fijar p a $\frac{1}{3}$ de la muestra.

- El contraste se puede utilizar para detectar, en principio, heterocedasticidad de forma general, aunque está pensado para alternativas específicas donde se supone un crecimiento de las varianzas en función de una determinada variable. Si en realidad el problema no es ese, sino que existe otra forma de heterocedasticidad, el estadístico puede no capturarla y no ser significativo.

- Por otro lado si no se rechaza la H_0 también puede deberse a una mala especificación de σ_i^2, que quizá depende de una variable diferente a la supuesta. Por ello puede ser necesario repetir el contraste para otras variables de las que podamos sospechar a priori.

Aplicación del test de Goldfeld y Quandt al ejercicio propuesto. Contrastamos:

$$\begin{array}{l} H_0 : \sigma_1^2 = \sigma_2^2 \\ H_a : \sigma_1^2 < \sigma_2^2 \end{array}$$

Estadístico de contraste:

$$GQ = \frac{\hat{\sigma}_2^2}{\hat{\sigma}_1^2} = \frac{\hat{u}_2'\hat{u}_2}{\hat{u}_1'\hat{u}_1}\frac{N_1 - K}{N_2 - K} \overset{H_0}{\sim} F_{(N_2-K, N_1-K)}$$

donde:
$\hat{u}_2'\hat{u}_2$ es la SCR de la regresión de C sobre R, incluyendo un término constante, en el segundo grupo de observaciones, $i = 21, \ldots, 40$.
$\hat{u}_1'\hat{u}_1$ es la SCR de la regresión de C sobre R, incluyendo un término constante, en el primer grupo de observaciones, $i = 1, \ldots, 20$.
En el Anexo 1.1, Resultado 2 se recogen los resultados obtenidos de gretl para resolver el contraste y que se muestran a continuación. Los resultados de la estimación en el primer grupo de observaciones son:

$$\begin{array}{llll} \widehat{C}_i & = 12,9388 + 0,1823 R_i & SCR = 12284,2 & R^2 = 0,4052 \\ (\widehat{des}(\hat{\beta}_{i,MCO})) & \quad (28{,}96) \qquad (0{,}52) & & \end{array}$$

Los resultados de la estimación en el segundo grupo de observaciones son:

$$\begin{array}{llll} \widehat{C}_i & = 48,1767 + 0,1176 R_i & SCR = 41146,9 & R^2 = 0,1036 \\ (\widehat{des}(\hat{\beta}_{i,MCO})) & \quad (70{,}24) \qquad (0{,}08) & & \end{array}$$

Dada la muestra

$$GQ = \frac{41146,9}{12284,2} = 3{,}34 \qquad F_{(18,18)|0{,}05} = 2{,}19$$

como 3,34 > 2,19 rechazo la H_0 para $\alpha = 5\,\%$ y podemos concluir que existe heterocedasticidad explicada por la variable renta. Una forma funcional razonable para la varianza de la perturbación sería:

$$Var(u_i) = \sigma^2 R_i \qquad i = 1, 2, \ldots, 40$$

pero también pudieran ser otras como por ejemplo $Var(u_i) = \sigma^2 R_i^2$ tal que se recoja una relación creciente de la varianza de u_i con la renta.

• Contraste de Breusch y Pagan (1979):

Breusch y Pagan en 1979 derivan un contraste de heterocedasticidad donde la hipótesis alternativa es bastante general:

$$H_0 : E(u_i^2) = \sigma^2 \quad \forall i \quad u_i \sim NID(0, \sigma^2)$$
$$H_a : \sigma_i^2 = \sigma^2 g(\alpha_0 + \alpha_1 Z_{1i} + \alpha_2 Z_{2i} + \ldots + \alpha_p Z_{pi})$$

las variables Z_{pi} pueden o no ser variables explicativas del modelo de interés, en cualquier caso deben ser observables y $g(\cdot)$ no se especifica. Si todos los coeficientes de la combinación lineal $Z_i'\alpha$ fuesen cero excepto α_0, la varianza sería homocedástica, $\sigma_i^2 = \sigma^2 g(\alpha_0)$. Por tanto un contraste de la hipótesis nula de homocedasticidad vendría dado por la siguiente hipótesis:

$$H_0 : \alpha_1 = \alpha_2 = \ldots = \alpha_p = 0$$

donde se contrastan p restricciones lineales. El proceso de contraste es el siguiente:

1. Estimar por MCO el modelo de interés obteniendo los residuos correspondientes, $\hat{u}_{MCO,i}$.
2. Se obtiene la siguiente serie de residuos normalizados:

$$\hat{e}_i^2 = \frac{\hat{u}_{MCO,i}^2}{\hat{u}'\hat{u}/N} \qquad i = 1, 2, \ldots, N$$

3. Se estima por MCO y se guarda la Suma de Cuadrados Explicada, SCE, de la siguiente regresión auxiliar:

$$\hat{e}_i^2 = \alpha_0 + \alpha_1 Z_{1i} + \alpha_2 Z_{2i} + \ldots + \alpha_p Z_{pi} + \eta_i \quad i = 1, 2, \ldots, N$$

4. Se utiliza como estadístico de contraste el siguiente:

$$BP = \frac{SCE}{2} \xrightarrow{d, H_0} \chi_p^2$$

siendo p los grados de libertad, igual al número de variables Z_{pi}. Rechazamos la H_0 de homocedasticidad a un nivel de significatividad α, si el valor muestral del estadístico excede del valor crítico esto es, $BP > \chi^2_{p|\alpha}$.

• Interpretación del contraste y regla de decisión: si las perturbaciones fuesen homocedásticas, las variables Z_{pi} no deberían tener poder explicativo acerca de los residuos transformados y por tanto la SCE debería ser pequeña. Si $SCE/2$ es grande, rechazaremos la H_0 y existiría heterocedasticidad.

Aplicación del test Breusch y Pagan al ejercicio propuesto. Contrastamos:

$$\begin{array}{l} H_0 : E(u_i^2) = \sigma^2 \quad \forall i \\ H_a : \sigma_i^2 = \sigma^2 g(\alpha_0 + \alpha_1 R_i) \end{array}$$

En este caso la regresión auxiliar es:

$$\hat{e}_i^2 = \alpha_0 + \alpha_1 R_i + \eta_i \qquad i = 1, \ldots, 40$$

En el Anexo 1.1, Resultado 3 se recogen los resultados obtenidos de gretl para resolver el contraste y que se muestran a continuación. Los resultados de la estimación por MCO de la regresión auxiliar son:

$$\begin{array}{lcccl} \widehat{\hat{e}_i^2} & = -1,6788 + & 0,0038 & R_i & SCR = 52,3940 \quad R^2 = 0,3010 \\ (\widehat{des}(\hat{\beta}_{i,MCO})) & (0{,}6876) & (0{,}0009) & & \end{array}$$

El valor muestral del estadístico de contraste es:

$$\frac{SCE}{2} = 11{,}28 \qquad \chi^2_{1|0{,}05} = 3{,}84$$

como 11,28 > 3,84, rechazo la H_0 para $\alpha = 5\,\%$ y podemos concluir que existe heterocedasticidad.

2.3. El estimador MCG bajo heterocedasticidad. Mínimos Cuadrados Ponderados

Una vez detectada estadísticamente la existencia de heterocedasticidad debemos abordar la estimación del modelo teniendo en cuenta esta información. Una posibilidad es obtener el estimador de Mínimos Cuadrados Generalizados, MCG, que tiene en cuenta la estructura existente en la matriz de varianzas y covarianzas. Sin embargo, implica conocer o proponer una forma funcional para la varianza de la perturbación en cada i. Si la estructura de la matriz de varianzas y covarianzas se conoce al menos excepto por un factor de escala, $E(uu') = \sigma^2\Omega$, y Ω es conocida, el vector de coeficientes β se puede estimar por MCG resolviendo el siguiente problema de minimización:

$$\underset{\hat{\beta}}{Min} (Y - X\hat{\beta})'\Omega^{-1}(Y - X\hat{\beta}) = \underset{\hat{\beta}}{Min} \sum_{i=1}^{N} \frac{1}{\sigma_i^2}(Y_i - X_i'\hat{\beta})^2$$

a resultas del cual podemos definir el estimador de MCG como:

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

• **Veamos una aplicación del procedimiento:**
En el ejemplo que estamos desarrollando el modelo a estimar es el siguiente:

$$C_i = \beta_1 + \beta_2 R_i + u_i \qquad i = 1, 2, \ldots, 40 \tag{2.9}$$

donde $E(u_i) = 0 \; \forall \, i, \; E(u_i u_j) = 0 \quad \forall \, i, j \quad i \neq j$ y hemos concluido que un supuesto razonable para el comportamiento de la varianza de la perturbación es:

$$Var(u_i) = \sigma^2 R_i \qquad i = 1, 2, \ldots, 40$$

La estructura de la matriz de varianzas y covarianzas del vector de perturbaciones es la siguiente:

$$E(uu') = \sigma^2 \begin{bmatrix} R_1 & 0 & 0 & \ldots & 0 \\ 0 & R_2 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ldots & \vdots \\ 0 & 0 & 0 & \ldots & R_{40} \end{bmatrix} = \sigma^2 \Omega$$

luego Ω es conocida. Si queremos obtener estimadores lineales, insesgados y de varianza mínima estimaremos por MCG. El vector de coeficientes β se puede estimar por MCG resolviendo el siguiente problema de minimización:

$$\underset{\hat{\beta}}{Min} \sum_{i=1}^{N} \frac{1}{\sigma_i^2} (C_i - \hat{\beta}_1 - \hat{\beta}_2 R_i)^2 = \underset{\hat{\beta}}{Min} \sum_{i=1}^{N} \frac{1}{R_i} (C_i - \hat{\beta}_1 - \hat{\beta}_2 R_i)^2 \tag{2.10}$$

Minimizar esta función es equivalente a estimar por MCO el siguiente modelo transformado[5]:

$$\frac{C_i}{\sqrt{R_i}} = \beta_1 \frac{1}{\sqrt{R_i}} + \beta_2 \sqrt{R_i} + \frac{u_i}{\sqrt{R_i}} \qquad i = 1, 2, \ldots, 40 \tag{2.11}$$

donde la perturbación es homocedástica, no autocorrelada y de media cero:

$$E\left(\frac{u_i}{\sqrt{R_i}}\right) = \frac{E(u_i)}{\sqrt{R_i}} = 0 \quad \forall i$$

$$Var\left(\frac{u_i}{\sqrt{R_i}}\right) = E\left(\frac{u_i}{\sqrt{R_i}} - E\left(\frac{u_i}{\sqrt{R_i}}\right)\right)^2 = E\left(\frac{u_i}{\sqrt{R_i}}\right)^2 = \frac{E(u_i^2)}{R_i} = \frac{\sigma^2 R_i}{R_i} = \sigma^2 \quad \forall i$$

$$Cov\left(\frac{u_i}{\sqrt{R_i}} \frac{u_j}{\sqrt{R_j}}\right) = E\left(\frac{u_i}{\sqrt{R_i}} \frac{u_j}{\sqrt{R_j}}\right) = \frac{E(u_i u_j)}{\sqrt{R_i}\sqrt{R_j}} = 0 \quad \forall i, j \quad i \neq j$$

A la vista de las propiedades de la perturbación del modelo transformado el estimador de MCO de este modelo transformado esto es, el de MCG, es lineal, insesgado y de varianza mínima. Si

[5] Para el modelo transformado podemos escribir la función a minimizar, equivalente a las anteriores, como: $\underset{\hat{\beta}}{Min} \sum_{i=1}^{N} \left(\frac{C_i}{\sqrt{R_i}} - \hat{\beta}_1 \frac{1}{\sqrt{R_i}} - \hat{\beta}_2 \sqrt{R_i}\right)^2$. Notar que el criterio y el estimador son invariantes al factor de escala σ^2.

conocemos la distribución de la perturbación podemos hacer inferencia en muestras finitas con este estimador. Lo veremos más adelante.

Vamos a formalizar matricialmente el modelo transformado (2.11):

$$P^{-1}Y = P^{-1}X\beta + P^{-1}u \Leftrightarrow Y^\star = X^\star\beta + u^\star$$

donde P es la matriz de transformación tal que $PP' = \Omega$ y $(P^{-1})'P^{-1} = \Omega^{-1}$:

$$P = \begin{bmatrix} \sqrt{R_1} & 0 & 0 & \dots & 0 \\ 0 & \sqrt{R_2} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & \sqrt{R_{40}} \end{bmatrix} \qquad P^{-1} = \begin{bmatrix} \frac{1}{\sqrt{R_1}} & 0 & 0 & \dots & 0 \\ 0 & \frac{1}{\sqrt{R_2}} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & \frac{1}{\sqrt{R_{40}}} \end{bmatrix};$$

Las matrices transformadas $X^\star = P^{-1}X$ y $Y^\star = P^{-1}Y$ son las siguientes:

$$X^\star = P^{-1}X = \begin{bmatrix} \frac{1}{\sqrt{R_1}} & \sqrt{R_1} \\ \frac{1}{\sqrt{R_2}} & \sqrt{R_2} \\ \vdots & \vdots \\ \frac{1}{\sqrt{R_{40}}} & \sqrt{R_{40}} \end{bmatrix}; \qquad Y^\star = P^{-1}Y = \begin{bmatrix} \frac{C_1}{\sqrt{R_1}} \\ \frac{C_2}{\sqrt{R_2}} \\ \vdots \\ \frac{C_{40}}{\sqrt{R_{40}}} \end{bmatrix};$$

La expresión matricial del estimador MCG para el modelo sería:

$$\begin{aligned} \hat\beta_{MCO} &= (X^{\star\prime}X^\star)^{-1}X^{\star\prime}Y^\star = \\ &= ((P^{-1}X)'(P^{-1}X))^{-1}((P^{-1}X)'(P^{-1}Y)) = \\ &= (X'\Omega^{-1}X)^{-1}(X'\Omega^{-1}Y) = \hat\beta_{MCG} \end{aligned}$$

luego:

$$\hat\beta_{MCG} = \begin{bmatrix} \sum_1^{40}\frac{1}{R_i} & N \\ N & \sum_1^{40} R_i \end{bmatrix}^{-1} \begin{bmatrix} \sum_1^{40}\frac{C_i}{R_i} \\ \sum_1^{40} C_i \end{bmatrix}$$

La matriz de varianzas y covarianzas del estimador de MCG sería:

$$V(\hat\beta_{MCG}) = \sigma^2(X'\Omega^{-1}X)^{-1} = \sigma^2 \begin{bmatrix} \sum_1^{40}\frac{1}{R_i} & N \\ N & \sum_1^{40} R_i \end{bmatrix}^{-1}$$

Un estimador insesgado de dicha matriz de varianzas y covarianzas sería:

$$\widehat{V}(\hat\beta_{MCG}) = \hat\sigma^2_{MCG}\,(X'\Omega^{-1}X)^{-1}$$

donde:

$$\hat\sigma^2_{MCG} = \frac{\hat u'_{MCG}\,\Omega^{-1}\,\hat u_{MCG}}{40-K} = \frac{(Y - X\hat\beta_{MCG})'\,\Omega^{-1}\,(Y - X\hat\beta_{MCG})}{40-K} = \frac{Y'\,\Omega^{-1}\,Y - \hat\beta'_{MCG}X'\,\Omega^{-1}\,Y}{40-K}$$

siendo en este caso $Y' \Omega^{-1} Y = \sum_1^{40} \frac{C_i^2}{R_i}$.

Resultados de la estimación del modelo[6]:

$$C_i = \beta_1 + \beta_2 R_i + u_i \qquad i = 1, 2, \ldots, 40$$

- por MCO, luego sin tener en cuenta la existencia de heterocedasticidad

$$\begin{array}{ccccccc} \widehat{C}_i & = & 40,7676 & + & 0,1282 & R_i & R^2 = 0,3171 \\ (\widehat{des}(\hat{\beta}_{i,MCO})) & & (22{,}1387) & & (0{,}0305) & & \end{array} \tag{2.12}$$

- por MCG, luego teniendo en cuenta la heterocedasticidad y suponiendo $Var(u_i) = \sigma^2 R_i$:

$$\begin{array}{ccccccc} \hat{C}_i & = & 31,9243 & + & 0,1409 & R_i & R^2 = 0,4177 \\ (\widehat{des}(\hat{\beta}_{i,MCG})) & & (17{,}9860) & & (0{,}0269) & & \end{array} \tag{2.13}$$

- Con respecto a los resultados de la estimación, en (2.12) no podemos decir nada ya que hemos detectado la existencia de heterocedasticidad y por lo tanto, las estimaciones $\hat{\beta}_{i,MCO}$ mostradas no son las mejores posibles ya que han sido obtenidas con un estimador no eficiente y en cuanto a $\widehat{des}(\hat{\beta}_{i,MCO})$, corresponden a un estimador sesgado por lo que no son adecuadas para hacer inferencia.
  Los resultados mostrados en (2.13) tienen en cuenta la existencia de heterocedasticidad. Si la forma funcional propuesta es correcta, el estimador utilizado, $\hat{\beta}_{MCG}$, es un estimador eficiente y las estimaciones $\widehat{des}(\hat{\beta}_{i,MCG})$ corresponden a un estimador insesgado y son válidas para hacer inferencia. Los R^2 no son comparables, ni tampoco las desviaciones estimadas de los coeficientes estimados, $\widehat{des}(\hat{\beta}_i)$.

Al estimador de MCG bajo heterocedasticidad también se le conoce con el nombre de Mínimos Cuadrados Ponderados, MCP, porque pondera las observaciones inversamente al peso de la varianza de la perturbación. En la suma de cuadrados de (2.10) se ponderan más las desviaciones o residuos $[C_i - \widehat{E(C_i)}]$ con menor varianza que las de mayor varianza o dispersión. Esto es muy fácil de entender en el ejemplo desarrollado si observamos la función objetivo, la forma de la matriz P^{-1} y en definitiva, las variables del modelo transformado.

Ejercicio 2.1 Sea el modelo:

$$C_i = \beta_1 + \beta_2 R_i + u_i$$

donde $E(u_i) = 0 \quad \forall i$; $E(u_i^2) = a + bR_i$; $\quad E(u_i u_j) = 0 \quad \forall i, j \quad i \neq j$. Escribe el correspondiente modelo transformado y demuestra las propiedades de la perturbación de dicho modelo suponiendo que a y b son conocidas.

[6] Ver Anexo 1.1, Resultados 1 y 4.

2.4. Estimador de Mínimos Cuadrados Generalizados Factibles. Especificación de un modelo para la heterocedasticidad

En la sección anterior la matriz de varianzas y covarianzas de la perturbación es conocida excepto por un factor de escala σ^2. Cuando los elementos de $E(uu')$ no son conocidos, $E(uu') = \sigma^2\Omega = \Sigma$, no es posible estimar N varianzas más K coeficientes de regresión con N observaciones[7]. Una forma de abordar el problema es modelar la varianza de la perturbación en función de un conjunto de variables observables, Z_i, que pueden ser o no regresores del modelo, y de un vector de parámetros desconocido θ, cuya dimensión es estimable y no crece con el tamaño muestral N:

$$\sigma_i^2 = g(Z_i, \theta), \quad \forall i \quad \text{de forma que} \quad \Sigma = \Sigma(\theta)$$

Una vez obtenido un estimador de θ, $\hat{\theta}$, se puede definir un estimador de $\Sigma, \widehat{\Sigma} = \Sigma(\hat{\theta})$, y estimar el vector de coeficientes β por el método de Mínimos Cuadrados Generalizados Factibles, MCGF. Bajo ciertas condiciones de regularidad, si el estimador $\widehat{\Sigma}$ es consistente, el estimador de MCGF tiene buenas propiedades asintóticas.

El estimador de MCGF es un estimador en dos etapas. En la primera etapa utilizamos los residuos de MCO para estimar θ. Dado que el residuo es una aproximación a la perturbación,

$$\hat{u}_i = Y_i - X_i'\hat{\beta}_{MCO} = Y_i - X_i'\beta - X_i'(\hat{\beta}_{MCO} - \beta) = u_i + \text{error}$$

y que hemos supuesto $E(u_i^2) = \sigma_i^2 = g(Z_i, \theta)$, entonces podemos definir la ecuación $u_i^2 = g(Z_i, \theta) +$ error. Si $g(Z_i, \theta)$ es lineal en θ, por ejemplo $\theta' Z_i$, se puede considerar la siguiente regresión auxiliar para estimar los parámetros θ:

$$\hat{u}_i^2 = \theta' Z_i + \zeta_i \qquad i = 1, \ldots, N$$

Bajo ciertas condiciones de regularidad, se puede demostrar que el estimador MCO de θ es consistente. Una vez obtenido el estimador consistente de θ, y por lo tanto, de $\hat{\sigma}_i^2 = g(Z_i, \hat{\theta})$, en la segunda etapa, se sustituye en la función suma de cuadrados ponderada y se minimiza con respecto a β:

$$\underset{\hat{\beta}}{Min}\ (Y - X\hat{\beta})'\widehat{\Sigma}^{-1}(Y - X\hat{\beta}) = \underset{\hat{\beta}}{Min} \sum_{i=1}^{N} \frac{1}{\hat{\sigma}_i^2}(Y_i - X_i'\hat{\beta})^2$$

obteniendo el estimador de MCGF definido como:

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

• **Veamos una aplicación de este procedimiento utilizando el ejercicio magistral:**
El modelo a estimar es:

$$C_i = \beta_1 + \beta_2 R_i + u_i \qquad i = 1, 2, \ldots, 40 \tag{2.14}$$

[7] Notar que cuando los elementos de Ω no son conocidos difícilmente será conocida σ^2, lo lógico es pensar que nada en las varianzas es conocido, ni tan siquiera que exista una parte común en todas ellas, así es habitual denotar a $E(uu')$ como Σ.

donde $E(u_i) = 0 \quad \forall i$ y $E(u_i u_j) = 0 \quad \forall i,j \quad i \neq j$. Suponemos que $Var(u_i) = a + bR_i$ siendo a y b constantes desconocidas. La forma funcional que hemos supuesto para la varianza de la perturbación es razonable con respecto a la información disponible, existencia de heterocedasticidad y dependencia creciente con respecto a R_i y depende de un número de parámetros pequeño que son estimables de forma consistente como veremos a continuación.

La estructura de la matriz de varianzas y covarianzas del vector de perturbaciones, que evidencia que Σ es desconocida, es la siguiente:

$$E(uu') = \begin{bmatrix} a+bR_1 & 0 & 0 & \dots & 0 \\ 0 & a+bR_2 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & a+bR_{40} \end{bmatrix} = \Sigma$$

¿Cómo estimamos los parámetros desconocidos del modelo anterior?
Nuestro problema ahora es cómo estimar los parámetros desconocidos β_1, β_2, a y b del modelo. El modelo transformado para la forma funcional propuesta para la varianza de u_i es:

$$\frac{C_i}{\sqrt{a+bR_i}} = \beta_1 \frac{1}{\sqrt{a+bR_i}} + \beta_2 \frac{R_i}{\sqrt{a+bR_i}} + \frac{u_i}{\sqrt{a+bR_i}} \qquad i = 1, \dots, 40 \tag{2.15}$$

En este caso en el modelo transformado tenemos dos constantes desconocidas a y b que deben ser previamente estimadas para poder aplicar el estimador de MCO al modelo (2.15). Para obtener estimadores consistentes de a y b podemos proceder de la forma siguiente:

1. Estimamos por MCO el modelo (2.9) y guardamos los residuos de mínimos cuadrados ordinarios.

2. Estimamos por MCO la siguiente regresión auxiliar:

$$\hat{u}^2_{MCO,i} = a + bR_i + \epsilon_i \qquad i = 1, 2, \dots, 40$$

   De esta regresión obtenemos $\hat{a}_{MCO}$ y $\hat{b}_{MCO}$ estimados consistentemente.

Una vez estimados consistentemente $\hat{a}_{MCO}$ y $\hat{b}_{MCO}$ podemos estimar por MCO, o lo que es lo mismo MCGF, el siguiente modelo transformado:

$$\frac{C_i}{\sqrt{\hat{a}+\hat{b}R_i}} = \beta_1 \frac{1}{\sqrt{\hat{a}+\hat{b}R_i}} + \beta_2 \frac{R_i}{\sqrt{\hat{a}+\hat{b}R_i}} + \frac{u_i}{\sqrt{\hat{a}+\hat{b}R_i}} \qquad i = 1, \dots, 40 \tag{2.16}$$

Los estimadores $(\hat{\beta}_1, \hat{\beta}_2)_{MCGF}$ así obtenidos son consistentes dado que los estimadores de a y b a su vez lo son. Además, serán asintóticamente eficientes. En muestras pequeñas no se conocen sus propiedades porque MCGF es un estimador no lineal. No obstante, al ser un estimador consistente y tener distribución asintótica conocida podemos utilizarlo para hacer inferencia asintótica válida, como veremos más adelante.

A la hora de implementar el estimador de MCGF es fundamental tener en cuenta que Σ es una matriz de varianzas y covarianzas, luego la variable de ponderación es $\hat{\sigma}_i^2$ y ésta debe ser positiva $\forall i$. Es necesario comprobarlo, ya que si a la hora de estimar σ_i^2 no se impone esta restricción no tiene porqué satisfacerse.

En el caso del ejemplo $\hat{\sigma}_i^2 = \hat{a} + \hat{b}R_i$ no es positiva para las 40 familias de la muestra, por lo que la forma funcional elegida para la varianza $Var(u_i) = a + bR_i$ a pesar de ser razonable para la muestra, no es implementable. Debemos probar otras, que también sean razonables claro está, por ejemplo $Var(u_i) = (a + bR_i)^2$. Para esta forma funcional la regresión auxiliar es:

$$\hat{u}_{MCO,i}^2 = a + bR_i + cR_i^2 + \epsilon_i \qquad i = 1, 2, \ldots, 40$$

El correspondiente modelo transformado una vez estimados consistentemente a, b y c es:

$$\frac{C_i}{\sqrt{\hat{a} + \hat{b}R_i + \hat{c}R_i^2}} = \beta_1 \frac{1}{\sqrt{\hat{a} + \hat{b}R_i + \hat{c}R_i^2}} + \beta_2 \frac{R_i}{\sqrt{\hat{a} + \hat{b}R_i + \hat{c}R_i^2}} + \frac{u_i}{\sqrt{\hat{a} + \hat{b}R_i + \hat{c}R_i^2}}$$
$$i = 1, \ldots, 40$$

La estimación del modelo por MCO, o lo que igual MCGF para esta última forma funcional de la varianza de la perturbación proporciona los siguientes resultados[8]:

- Modelo estimado por MCGF:

$$\begin{array}{ccccc} \hat{C}_i & = & 34,2386 & + & 0,1410 \;\; R_i \\ (\widehat{des}(\hat{\beta}_{i,MCGF})) & & (17{,}6042) & & (0{,}0289) \end{array}$$

  Si observamos los resultados de la estimación vemos como cuantitativamente las realizaciones muestrales de los estimadores MCG y MCGF son muy similares. Sin embargo, es importante recalcar que las varianzas de los estimadores MCG y MCGF no son comparables entre sí. El marco de trabajo es distinto. Con el estimador MCG estamos trabajando en términos de propiedades en muestras finitas de un estimador y bajo un supuesto concreto en la varianza de la perturbación y en base a su certeza se demuestran las propiedades del estimador. Con el estimador de MCGF al ser no lineal buscamos propiedades asintóticas bajo el supuesto de que la forma funcional propuesta para la varianza de la perturbación es adecuada y además, se ha estimado consistentemente. Pero si el supuesto sobre $Var(u_i)$ es adecuado con MCGF, éste es asintóticamente equivalente a MCG, en el sentido de que ambos estimadores tienen la misma distribución asintótica.

2.5. MCO: Estimador de $V(\hat{\beta}_{MCO})$ robusto a heterocedasticidad

En presencia de heterocedasticidad el estimador Mínimo Cuadrático Ordinario es lineal, insesgado y consistente, pero no es de varianza mínima. Su matriz de varianzas y covarianzas se define $V(\hat{\beta}_{MCO}) = \sigma^2(X'X)^{-1}X'\Omega X(X'X)^{-1}$. Un estimador de la matriz de varianzas y covarianzas del

[8] Ver Anexo 1.1, Resultado 5.

estimador MCO es $\widehat{V}(\hat{\beta}_{MCO}) = \hat{\sigma}^2(X'X)^{-1}$. Utilizar este estimador para hacer inferencia cuando hay heterocedasticidad no es adecuado. Los estadísticos t y F habituales para hacer inferencia sobre β definidos en base a este estimador de la matriz de varianzas y covarianzas de $\hat{\beta}_{MCO}$ son inapropiados ya que es un estimador sesgado e inconsistente.

La dificultad que entraña el conocimiento de Ω, o Σ en su caso, hace interesante el poder contar con un estimador de $V(\hat{\beta}_{MCO})$ consistente y robusto a la posible existencia de heterocedasticidad y de esta forma derivar estadísticos válidos, al menos asintóticamente, para contrastar hipótesis sobre el vector de coeficientes β utilizando $\hat{\beta}_{MCO}$.

White (1980) demuestra que un estimador consistente de la matriz de varianzas y covarianzas asintótica de $\hat{\beta}_{MCO}$ en presencia de heterocedasticidad es:

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

donde $S = diag(\hat{u}_1^2, \hat{u}_2^2, \ldots, \hat{u}_N^2)$. Esta matriz de varianzas y covarianzas consistente asintóticamente puede ser utilizada para hacer inferencia válida al menos asintóticamente utilizando $\hat{\beta}_{MCO}$ sin tener que especificar a priori la estructura de heterocedasticidad.

- Aplicación del estimador de White de la matriz de varianzas y covarianzas del estimador MCO al ejercicio propuesto[9]:

$$\begin{array}{rcccc} \widehat{C}_i & = & 40,7676 & + & 0,1282 \;\; R_i \\ \widehat{des}(\hat{\beta}_{i,MCO}) & & (22,1387) & & (0,0305) \\ \widehat{des}(\hat{\beta}_{i,MCO})_{White} & & (27,3213) & & (0,0439) \end{array} \qquad (2.17)$$

  Notar que los coeficientes han sido estimados por MCO, estimador lineal, insesgado, consistente y no de varianza mínima dado que $E(u_i^2) = \sigma_i^2$. El estimador, $\widehat{des}(\hat{\beta}_{i,MCO})$ se obtiene de $\hat{\sigma}^2(X'X)^{-1}$ estimador sesgado e inconsistente bajo heterocedasticidad. Sin embargo, $\widehat{des}(\hat{\beta}_{i,MCO})_{White}$ es un estimador consistente de $des(\hat{\beta}_{i,MCO})$ y ha sido obtenido con la expresión $(X'X)^{-1}(X'SX)(X'X)^{-1}$. En este ejercicio en concreto se observa que con el estimador $\widehat{des}(\hat{\beta}_{i,MCO}) = \hat{\sigma}^2(X'X)^{-1}$, sesgado e inconsistente se estaba subestimando la desviación típica poblacional de $\hat{\beta}_{MCO}$, que es la raíz cuadrada del elemento i-ésimo de la diagonal principal de $\hat{\sigma}^2(X'X)^{-1}$.

- El problema de estimación de los parámetros de interés como una cuestión de objetivos e información disponible:

Si conocemos la forma funcional de la varianza de la perturbación, es decir, si conocemos Ω y si nuestro objetivo es obtener estimadores **lineales, insesgados y de varianza mínima** estimaremos por MCG. Si no conocemos Ω habremos de estimarla y una vez estimada podremos sustituirla en la expresión del estimador. En este caso nuestro estimador es MCGF, definido como:

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

[9] Anexo 1.1, Resultado 6.

Bajo ciertas condiciones suficientes el estimador es consistente y podremos obtener una distribución asintótica conocida que nos permitirá realizar inferencia asintótica.

En ocasiones modelizar Ω, o estimarla si es desconocida no resulta fácil. En estas ocasiones puede resultar preferible quedarnos con el estimador de los coeficientes β por MCO y ocuparnos de cómo realizar inferencia válida con este estimador en estas circunstancias. En este caso una solución podría ser obtener un estimador de la matriz de varianzas y covarianzas de los estimadores MCO con adecuadas propiedades, en concreto consistente y robusto a la posible existencia de heterocedasticidad. La utilización del estimador MCO, consistente en presencia de heterocedasticidad y la matriz de varianzas y covarianzas propuesta nos permitiría realizar inferencia asintótica sin tener que establecer la forma funcional de la heterocedasticidad.

Hay que notar que el estimador de White permite a los investigadores no arriesgarse a manipular los datos a la búsqueda de una forma funcional de la heterocedasticidad cuando esta no está muy clara y realizar inferencia asintótica. Sin embargo, no soluciona la cuestión de la ineficiencia del estimador.

2.6. Contraste de restricciones lineales

Una vez hemos estimado adecuadamente los coeficientes del modelo de interés es probable que nuestro objetivo sea realizar contraste de hipótesis. En el tema anterior ya se vio con detalle cómo contrastar restricciones lineales[10]. Sean las hipótesis nula y alternativa para el contraste de q restricciones lineales:

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

donde R es una matriz $(q \times K)$ y r es un vector de dimensión $(q \times 1)$, siendo q el número de restricciones lineales a contrastar.

• **Estadísticos basados en el estimador de $\boldsymbol{\beta}$ por MCG o MCP.**
En este caso

$$Y = X\beta + u \qquad u \sim N(0, \sigma^2\Omega) \quad \text{y} \quad \Omega \quad \text{es conocida}$$

el estimador MCG es lineal en u, insesgado y de varianza mínima, su distribución en muestras finitas es:

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

Estadísticos de contraste y distribución asociada:

Para $q = 1$

$$\frac{R\hat{\beta}_{MCG} - r}{\sqrt{R\,\hat{\sigma}^2(X'\Omega^{-1}X)^{-1}\,R'}} \overset{H_0}{\sim} t_{(N-K)}$$

[10]La notación utilizada para el tamaño muestral en ese tema era T en lugar de N.

para $q > 1$

$$\frac{(R\hat{\beta}_{MCG} - r)' \, [R \, (X'\Omega^{-1}X)^{-1} \, R']^{-1} \, (R\hat{\beta}_{MCG} - r)/q}{\hat{\sigma}^2} \overset{H_0}{\sim} F_{(q,N-K)}$$

siendo un estimador insesgado de σ^2, $\hat{\sigma}^2_{MCG} = \dfrac{(Y - X\hat{\beta}_{MCG})' \, \Omega^{-1}(Y - X\hat{\beta}_{MCG})}{N-K}$.
Las reglas de decisión son las habituales.

Ejemplo 2.8 En el modelo $C_i = \beta_1 + \beta_2 R_i + u_i \quad u_i \sim N(0, \sigma^2 R_i)$ contrastamos la significatividad de la variable R_i con el estadístico y distribución siguientes:

$$H_0 : \beta_2 = 0 \quad \text{versus} \quad H_a : \beta_2 \neq 0 \qquad \frac{\hat{\beta}_{2,MCG}}{\widehat{des}(\hat{\beta}_{2,MCG})} \overset{H_0}{\sim} t_{(40-2)}$$

Rechazamos la hipótesis nula de no significatividad de la variable renta para $\alpha = 5\,\%$ si el valor muestral del estadístico t es tal que t $> t_{(40-2)|0{,}025}$.

Ejercicio 2.2 Sea el modelo:

$$Y = X\beta + u \qquad u \sim N(0, \Sigma) \qquad \Sigma \;\text{ conocida}$$

Deriva la distribución del estimador MCG de β y el estadístico de contraste y su distribución asociada para contrastar $H_0 : R\beta = r$.

Ejercicio 2.3 Sea el modelo

$$Y = X\beta + u \qquad u \sim N(0, \sigma^2\Omega) \quad \text{y} \quad \Omega \quad \text{es conocida}$$

En el modelo transformado correspondiente, deriva la distribución del estimador MCO de β y el estadístico de contraste y su distribución asociada para contrastar $H_0 : R\beta = r$.

• Estadísticos basados en el estimador de $\boldsymbol{\beta}$ por MCGF

En este caso

$$Y = X\beta + u \qquad u \sim N(0, \Sigma) \qquad \Sigma \;\text{ desconocida, pero estimable consistentemente}$$

El estimador de MCGF, $\hat{\beta}_{MCGF} = (X'\widehat{\Sigma}^{-1}X)^{-1}(X'\widehat{\Sigma}^{-1}Y)$, es un estimador no lineal y en general sesgado, cuya distribución en muestras finitas no es conocida. En muestras grandes es un estimador consistente si $\widehat{\Sigma}$ es un estimador consistente de Σ. Su distribución asintótica es:

$$\sqrt{N}(\hat{\beta}_{MCGF} - \beta) \xrightarrow{d} N(0, G^{-1})$$

Los estadísticos de contraste y distribución asociada utilizando como estimador de $G^{-1} = \text{plim}\left[\frac{1}{N}X'\Sigma X\right]^{-1}$ a $N\widehat{V}(\hat{\beta}_{MCGF}) = N(X'\widehat{\Sigma}^{-1}X)^{-1}$ son:

para $q = 1$

$$\frac{R\hat{\beta}_{MCGF} - r}{\sqrt{R\,(X'\widehat{\Sigma}^{-1}X)^{-1}\,R'}} \xrightarrow{d,H_0} N(0,1)$$

para $q > 1$

$$(R\hat{\beta}_{MCGF} - r)'[\;R\,(X'\widehat{\Sigma}^{-1}X)^{-1}\;R'\;]^{-1}(R\hat{\beta}_{MCGF} - r) \xrightarrow{d,H_0} \chi^2_q$$

Las reglas de decisión son las habituales.

Ejemplo 2.9 En el modelo $C_i = \beta_1 + \beta_2 R_i + u_i \quad u_i \sim N(0, a + bR_i)$ contrastamos la significatividad de la variable R_i con el estadístico y distribución siguientes:

$$H_0 : \beta_2 = 0 \quad \text{versus} \quad H_a : \beta_2 \neq 0$$

$$\frac{\hat{\beta}_{2,MCGF}}{\widehat{des}(\hat{\beta}_{2,MCGF})} \xrightarrow{d,H_0} N(0,1)$$

Rechazamos la hipótesis nula de no significatividad de la variable renta para $\alpha = 5\,\%$ si el valor muestral del estadístico t es tal que t $> N(0,1)_{|\frac{0,05}{2}}$.

• **Estadísticos basados en el estimador de β por MCO y un estimador robusto a heterocedasticidad de $V(\hat{\beta}_{MCO})$**

En este caso

$$Y = X\beta + u \qquad u \sim N(0, \Sigma) \qquad \Sigma \text{ desconocida y difícil de modelar y estimar consistentemente}$$

Dado que el estimador MCO es consistente y $\widehat{V}(\hat{\beta}_{MCO})_{White}$ también lo es bajo heterocedasticidad podemos utilizarlos conjuntamente para hacer inferencia asintótica válida. En este caso el estadístico válido para realizar contraste de hipótesis de la forma $H_0 : R\beta = r$ y su distribución asintótica asociada son:

$$(R\hat{\beta}_{MCO} - r)'(R\,\widehat{V}(\hat{\beta}_{MCO})_{White}\,R')^{-1}(R\hat{\beta}_{MCO} - r) \xrightarrow{d,H_0} \chi^2_q$$

Siendo q el número de restricciones de contraste. Para $q = 1$ podemos escribir el estadístico anterior como:

$$\frac{R\hat{\beta}_{MCO} - r}{\sqrt{R\,\widehat{V}(\hat{\beta}_{MCO})_{White}\,R'}} \xrightarrow{d,H_0} N(0,1)$$

Ejemplo 2.10 En el modelo $C_i = \beta_1 + \beta_2 R_i + u_i \quad u_i \sim N(0, \sigma_i^2)$, siendo σ_i^2 desconocida, contrastamos la significatividad de la variable R_i utilizando el estimador MCO de β_2 con el estadístico y distribución siguientes:

$$H_0 : \beta_2 = 0 \quad \text{versus} \quad H_a : \beta_2 \neq 0$$

$$\frac{\hat{\beta}_{2,MCO}}{\widehat{des}(\hat{\beta}_{2,MCO})_{White}} \overset{d,H_0}{\longrightarrow} N(0,1)$$

Rechazamos la hipótesis nula de no significatividad de la variable renta para $\alpha = 5\,\%$ si el estadístico calculado t es tal que t $> N(0,1)_{|\frac{0,05}{2}}$.

2.7. Resumen de los resultados obtenidos en el ejercicio magistral

En esta sección vamos a resumir los resultados obtenidos en este tema. Para ello vamos utilizar el ejercicio magistral. En el ejemplo nos proponíamos estudiar la relación entre consumo y renta para lo cual se disponía de 40 observaciones de ambas variables.

$$C_i = \beta_1 + \beta_2 R_i + u_i \qquad i = 1, \ldots, 40$$

Los resultados de la estimación MCO son:

$$\begin{array}{lcccc} \widehat{C}_i & = & 40,7676 & + & 0,1282\ R_i \\ (\widehat{des}(\hat{\beta}_{i,MCO})) & & (22{,}1387) & & (0{,}0305) \end{array}$$

$$R^2 = 0,317118 \qquad GQ = 3,34 \qquad BP = 11,28$$

Dados lo valores muestrales de los estadísticos GQ y BP, para ambos rechazamos la hipótesis nula de homocedasticidad. La perturbación del modelo es heterocedástica y el estimador MCO utilizado para estimar, aunque es un estimador lineal, insesgado y consistente no es de varianza mínima. Las desviaciones típicas estimadas para los parámetros se han calculado en base a la expresión $\widehat{V}(\hat{\beta}_{MCO}) = \hat{\sigma}^2(X'X)^{-1}$, estimador sesgado e inconsistente de $V(\hat{\beta}_{MCO})$ por lo que no deben ser utilizadas para hacer inferencia. Los estadísticos t y F habituales para realizar inferencia no siguen las distribuciones t-Student y F-Snedecor que les corresponden.

A partir de la Figura 2.8 y de los resultados de los contrastes de heterocedasticidad se ha supuesto que $Var(u_i) = \sigma_i^2 = \sigma^2 R_i$ y se ha reestimado el modelo por MCP con los siguientes resultados:

$$\begin{array}{lcccccc} \hat{C}_i & = & 31,9243 & + & 0,1409\ R_i & & R^2 = 0,417757 \\ (\widehat{des}(\hat{\beta}_{i,MCG})) & & (17{,}9860) & & (0{,}0269) & & \end{array}$$

Si la forma funcional propuesta para σ_i^2 es correcta, el estimador MCP es un estimador lineal, insesgado y de varianza mínima. Si la perturbación es normal, el estimador MCP sigue una distribución normal en muestras finitas, $\hat{\beta}_{MCG} \sim N(\beta, \sigma^2(X'\Omega^{-1}X)^{-1})$ válida para hacer inferencia. El estadístico calculado para la $H_0 : \beta_2 = 0$ frente a $H_a : \beta_2 \neq 0$ es t $= 4,20 > 2,02 = t_{(40-2)|0,025}$, luego para un nivel de significatividad $\alpha = 5\,\%$ la variable es significativa para explicar el consumo de los individuos.

Respecto a la bondad del ajuste no podemos decir nada. Bajo heterocedasticidad el coeficiente R^2 calculado en el modelo original no tiene la interpretación que conocemos. Además, hay que tener mucho cuidado con su interpretación cuando trabajamos en el modelo transformado ya que en muchos casos tras la transformación, el modelo no tiene término independiente.

Una alternativa a la forma funcional propuesta para $Var(u_i)$ y coherente tanto con la Figura 2.8 como con los resultados de los estadísticos de contraste utilizados para contrastar la existencia de heterocedasticidad es $Var(u_i) = (a + bR_i)^2$. En este caso hay dos coeficientes desconocidos en la forma funcional propuesta por lo que el estimador MCG (o MCP) no es aplicable. El modelo debe ser estimado por MCGF y los resultados de dicha estimación son:

$$\begin{array}{ccccc} \hat{C}_i & = & 34,2386 & + & 0,1410 \;\; R_i \\ (\widehat{des}(\hat{\beta}_{i,MCGF})) & & (17{,}6042) & & (0{,}0289) \end{array}$$

Si la forma funcional propuesta es correcta el estimador MCGF es un estimador no lineal y consistente si $\hat{\sigma}_i^2$ a su vez es consistente. Además, es eficiente asintóticamente y tiene distribución asintótica conocida y válida para hacer inferencia asintótica. En este caso el estadístico calculado para la $H_0 : \beta_2 = 0$ frente a $H_a : \beta_2 \neq 0$ es t $= 4,87 > 1,96 = N(0,1)_{|0,025}$, luego para un nivel de significatividad $\alpha = 5\,\%$ la variable renta es significativa asintóticamente para explicar el consumo de los individuos.

Los estimadores MCG y MCGF propuestos son correctos y adecuados para las formas funcionales propuestas para la varianza de la perturbación, cada uno en su contexto. Sin embargo, si proponemos $Var(u_i) = \sigma^2 R_i$ y la forma funcional correcta es otra distinta, por ejemplo $Var(u_i) = (a + bR_i)^2$ el estimador MCG no posee las propiedades indicadas. Y viceversa. Es por ello, que ante esta dificultad, en muchas ocasiones sea más adecuado no realizar ningún supuesto sobre $Var(u_i)$, y utilizar el estimador de MCO combinado con un estimador consistente de su matriz de varianzas y covarianzas bajo heterocedasticidad y realizar inferencia con ellos. Para el caso de heterocedasticidad el estimador robusto a heterocedasticidad de $V(\hat{\beta}_{MCO})$ es el estimador de White. Si optamos por esta solución para estimar y realizar inferencia, los resultados obtenidos son:

$$\begin{array}{ccccc} \widehat{C}_i & = & 40,7676 & + & 0,1282 \;\; R_i \\ (\widehat{des}(\hat{\beta}_{i,MCO})_{White}) & & (27{,}3213) & & (0{,}0439) \end{array}$$

El estimador MCO utilizado para estimar los coeficientes del modelo es lineal, insesgado y no es de varianza mínima. Sin embargo, es consistente, al igual que lo son las desviaciones típicas estimadas con el estimador de White. Por ello podemos utilizar el estadístico t-habitual con distribución asintótica $N(0,1)$ para contrastar la significatividad de la variable renta, $H_0 : \beta_2 = 0$ frente a $H_a : \beta_2 \neq 0$. En este caso t $= 2,91 > 1,96 = N(0,1)_{|0,025}$, luego para un nivel de significatividad $\alpha = 5\,\%$ la variable es significativa para explicar el consumo de los individuos.

2.8. Ejercicios a resolver

Además, de los ejercicios que se resuelven en clase tenéis a vuestra disposición una colección de ejercicios de examen que se actualiza cada año, la "Recopilación de exámenes de Econometría". No deberíais dar por acabado el trabajo de cada tema hasta que estén hechos los ejercicios recomendados en clase. A continuación se proponen varios ejercicios, algunos de esa colección, cuyas dudas se resolverán a lo largo de las clases magistrales.

Ejercicio M-H.1:
Sea el modelo:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \qquad u_i \sim (0, \sigma_i^2) \qquad i = 1, \ldots, 10$$

con X_2, y X_3 no estocásticas. Indica para cada uno de los siguientes modelos transformados cual sería la correspondiente forma funcional que se supone para σ_i^2 y escribe la correspondiente matriz de varianzas y covarianzas de u_i bajo el supuesto de que $Cov(u_i, u_j) = 0 \quad \forall i \neq j$:

1. $\frac{Y_i}{X_{2i}} = \beta_1 \frac{1}{X_{2i}} + \beta_2 + \beta_3 \frac{X_{3}i}{X_{2i}} + \frac{u_i}{X_{2i}}$
2. $Y_i X_{3i} = \beta_1 X_{3i} + \beta_2 X_{2i} X_{3i} + \beta_3 X_{3i}^2 + u_i X_{3i}$
3. $\frac{Y_i}{\sqrt{X_{3i}}} = \beta_1 \frac{1}{\sqrt{X_{3i}}} + \beta_2 \frac{X_{2i}}{\sqrt{X_{3i}}} + \beta_3 \sqrt{X_{3i}} + \frac{u_i}{\sqrt{X_{3i}}}$

Ejercicio M-H.2:
Con una muestra de 15 países se desea estimar el efecto que un aumento en las cotizaciones de la Seguridad Social tendría sobre la parte de las cotizaciones a cargo de los trabajadores. La información, correspondiente al año 1982, de las cotizaciones a la Seguridad Social (CSS) y la parte correspondiente a los trabajadores ($CSST$), en ambos casos como porcentaje del total de ingresos fiscales, se presenta en las dos primeras columnas de la siguiente tabla:

	CSS	$CSST$	$\hat{u}$
Austria	31,9	13,5	
Bélgica	29,8	10,1	-0,08327
Dinamarca	2,8	1,5	-2,97434
Francia	43,2	11,5	
Alemania	36,2	16,1	
Irlanda	15,0	5,4	-1,65393
Italia	47,2	7,1	
Japón	30,4	10,7	0,38986
Luxemburgo	28,0	11,2	1,39732
Países Bajos	41,6	18,0	
Portugal	28,5	10,8	0,89160
España	46,5	10,3	
Suiza	31,0	10,2	-0,23700
Reino Unido	16,9	7,6	0,14433
EE.UU.	27,7	10,8	1,06076

Consideramos el siguiente modelo:

$$CSST_i = \beta_1 + \beta_2 CSS_i + u_i \qquad i = 1, \ldots, 15$$

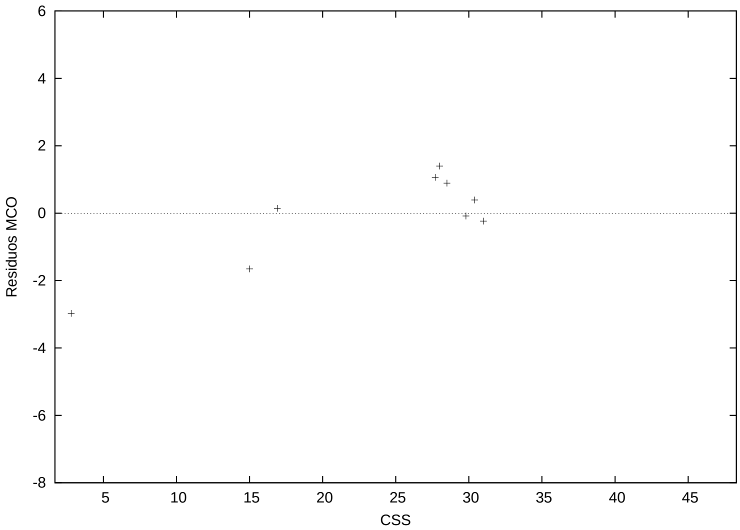
Los resultados de la estimación del modelo anterior por MCO con la muestra de los 15 países son los siguientes:

$$\underset{\text{(t-estad.)}}{\widehat{CSST}_i} = \underset{(1,69)}{3,8823} + \underset{(3,01)}{0,211442}\ CSS_i \tag{2.18}$$

$$\bar{R}^2 = 0,365 \qquad SCR = 132,7767$$

1. **Fíjate en la tabla**, en la tercera columna se muestran los residuos MCO, $\hat{u}_i$. Indica la forma general de obtener $\hat{u}_i$. A continuación completa los que faltan en la misma tabla y en la Figura 2.9.

Figura 2.9: CSS_i *versus residuos MCO*



2. Una vez completado el gráfico comenta si crees que puede existir algún problema razonando tu respuesta.

3. Con la siguiente información lleva a cabo el contraste de Goldfeld y Quandt. Debes de completar la información que falta y señalar claramente todos los elementos del contraste incluidas la hipótesis nula y la alternativa.

   - Primera submuestra:

$$\widehat{CSST}_i = 0,463351 + 0,374431\ CSS_i \tag{2.19}$$

$CSST_i$	1,5					
CSS_i	2,8					
$\hat{u}_1$	-0,011759		0,808758		0,25257	

- Segunda submuestra:

$$\widehat{CSST}_i = 28,9928 - 0,395203\, CSS_i \tag{2.20}$$

$CSST_i$	13,5					
CSS_i	31,9					
$\hat{u}_2$		1,413507		-0,420075		-3,239264

4. Dada la evidencia obtenida en los apartados anteriores y **con la siguiente información**, estima **eficientemente** los coeficientes del modelo. Explica cómo se obtiene este estimador y qué supuestos se están haciendo para que este estimador sea de varianza mínima.

	$CSST_i/CSS_i$	$1/CSS_i$	$Constante_i = 1$
$CSST_i/CSS_i$	2,12814	0,3672255	5,47296
$1/CSS_i$		0,1463262	0,8374455
$Constante_i = 1$			15

donde por ejemplo $\sum CSST_i/CSS_i = 5,47296$.

5. Con el estimador que has propuesto en el apartado anterior **contrasta** la hipótesis nula de que un aumento en las cotizaciones de la Seguridad Social recaería totalmente sobre los trabajadores esto es, $H_0 : \beta_2 = 1$. Indica todos los supuestos necesarios para que sea válido el contraste.

Ejercicio M-H.3:

La sección de estudios de mercado de la empresa Lydia Pinkham quiere analizar la influencia de la publicidad en sus ventas[11]. Para ello dispone de observaciones anuales para el periodo de 1907 a 1960 sobre las ventas de su producto, V, y los gastos en publicidad, G, ambas en millones de dólares. Se propone la siguiente relación:

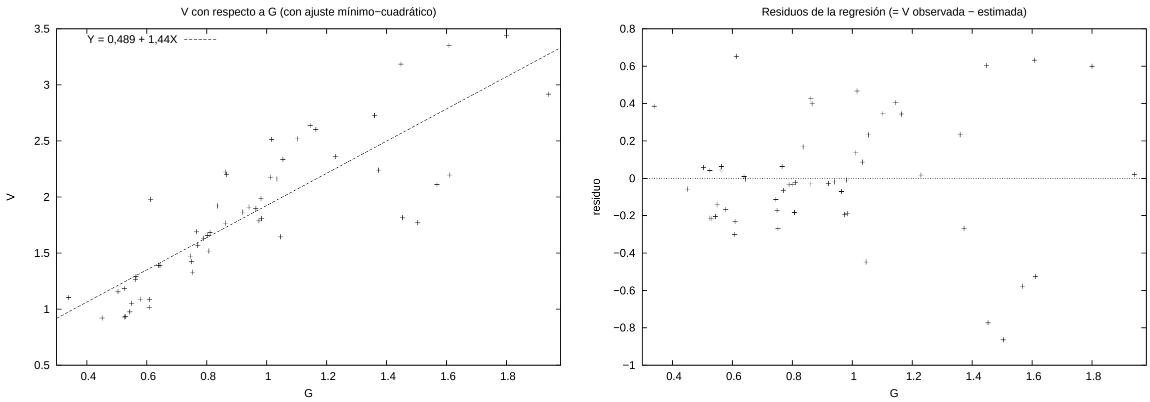
$$V_t = \beta_1 + \beta_2 G_t + \beta_3 G_t^2 + u_t \tag{2.21}$$

[11] En el Apéndice tenéis recogidos los datos en la Tabla A.4

El jefe de la sección presenta al gerente de la empresa los resultados de la estimación MCO junto con los gráficos mostrados en la Figura 2.10:

$$\begin{array}{rl} \widehat{V}_t = & \underset{(0{,}317)}{-0{,}163} + \underset{(0{,}635)}{2,825}\; G_t - \underset{(0{,}288)}{0,0642}\; G_t^2 \\ & R^2 = 0,7366 \qquad SCR = 5,589 \end{array} \tag{2.22}$$

Figura 2.10: *Variables y Residuos MCO del modelo*



Ventas versus Gastos *Residuos MCO versus Gastos*

1. El gerente no está satisfecho con estos resultados, ¿qué problemas crees que reflejan los gráficos anteriores?

El jefe de sección propone dos posibles vías para mejorar el estudio. La primera consiste en estimar por MCO la siguiente ecuación:

$$\frac{V_t}{\sqrt{G_t}} = \beta_1 \frac{1}{\sqrt{G_t}} + \beta_2 \sqrt{G_t} + \beta_3 \frac{G_t^2}{\sqrt{G_t}} + \frac{u_t}{\sqrt{G_t}} \tag{2.23}$$

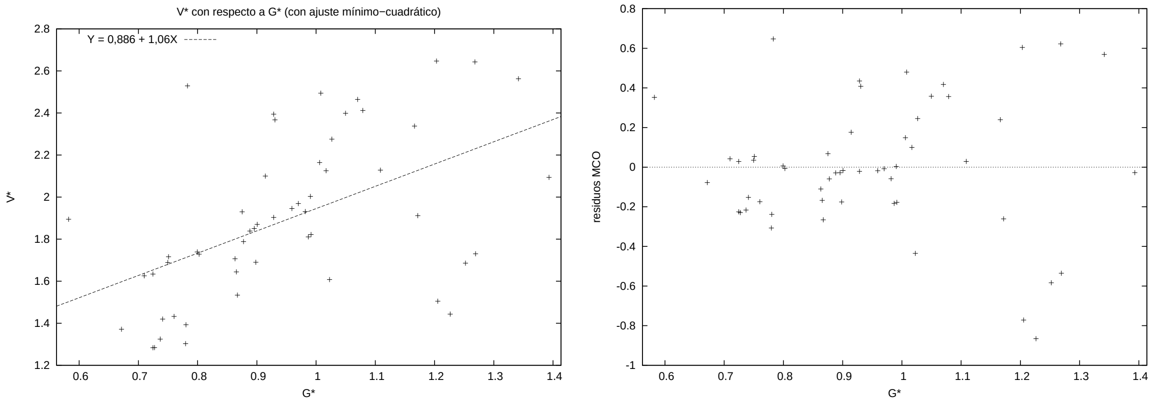
2. ¿Cuál es la hipótesis básica que se debe incumplir en el modelo (2.21) para utilizar el modelo (2.23)? ¿Qué solución se está proponiendo? ¿En qué se espera mejorar con respecto a la estimación MCO (2.22)?

3. A la vista de la Figura 2.11 en la que se representan los residuos del ajuste MCO del modelo (2.23) y la variable $V^\star = \frac{V_t}{\sqrt{G_t}}$ sobre $G^\star = \sqrt{G_t}$ ¿crees que se está resolviendo correctamente el problema?

La segunda alternativa es que la relación entre ventas y gastos en publicidad no sea lineal. Se estima por MCO el siguiente modelo log-lineal para la relación publicidad-ventas:

$$LnV_t = \gamma_1 + \gamma_2 LnG_t + v_t \tag{2.24}$$

Figura 2.11: *Variables y residuos MCO del modelo transformado*



Ventas versus Gastos**   *Residuos MCO* versus Gastos**

con los siguientes resultados:

$$\begin{array}{rcccc} \widehat{LnV}_t & = & 0,656 & + & 0,778 \;\; LnG_t \qquad R^2 = 0,7603 \qquad SCR = 1,5452 \\ (\widehat{des}(\hat{\beta}_{i,MCO})) & & (0{,}025) & & (0{,}060) \end{array} \tag{2.25}$$

$$\begin{array}{rcccc} \dfrac{\widehat{\hat{v}_t^2}}{0,028} & = & 1,052 & + & 0,367 \;\; LnG_t \qquad R^2 = 0,0117 \qquad SCR = 92,1022 \\ & & (0{,}193) & & (0{,}467) \end{array} \tag{2.26}$$

4. Interpreta los coeficientes estimados de este modelo.

5. ¿Crees que el modelo (2.24) presenta el mismo problema de incumplimiento de hipótesis que el modelo (2.21)? Justifica tu respuesta mediante un contraste. Explica detalladamente lo que haces y por qué lo haces.

6. ¿Alguna de las dos soluciones propuestas te parece mejor que la otra? Razona tu respuesta.

Ejercicio M-H.4:

Se desea analizar la siguiente la relación entre los gastos agregados en sanidad, Y y la renta agregada, X, ambos en billones de dólares, para 51 estados norteamericanos[12]:

$$Y_i = \beta_1 + \beta_2 X_i + u_i \tag{2.27}$$

[12]Fichero data3-2.gdt, disponible en gretl pestaña Ramanathan. Fuente: Statistical Abstract of U.S. (1995), recogida en Ramanathan, R. (2002), *Introductory econometrics with applications*, 5th. Ed., South-Western.

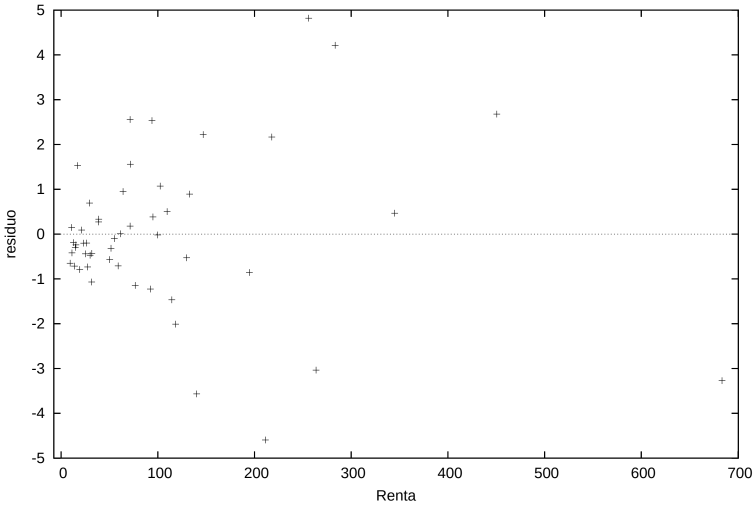
Los resultados de la estimación por Mínimos Cuadrados Ordinarios son los siguientes:

$$\begin{array}{rcllll} \hat{Y}_i & = & 0,3256 & + & 0,1420 & X_i \qquad R^2 = 0,999 \\ (\widehat{des}(\hat{\beta}_{i,MCO})) & & (0,3197) & & (0,0019) & \\ (\widehat{des}(\hat{\beta}_{i,MCO})_{White}) & & (0,2577) & & (0,0031) & \end{array}$$

$$\frac{\widehat{\hat{u}_i^2}}{\frac{\hat{u}'\hat{u}}{N}} = 0,113 + 0,008X_i + \hat{\epsilon}_i \qquad R^2 = 0,3269 \qquad SCE = 55,89 \tag{2.28}$$

Posteriormente, se dibujan los residuos frente a la renta agregada (Figura 2.12).

Figura 2.12: *Residuos MCO frente a Renta Agregada*



1. Explica cómo crees que se han calculado los residuos y para qué crees que se ha dibujado la Figura 2.12. Interprétala.

2. Teniendo en cuenta la Figura 2.12 realiza el contraste que consideres oportuno.

3. Explica, razonando tu respuesta, qué estadístico utilizarías para contrastar la significatividad de la variable renta. Realiza el contraste detallando todos sus elementos.

4. A la vista de los resultados de la estimación del modelo (2.27) el investigador estima de nuevo el modelo suponiendo la siguiente estructura para la varianza de la perturbación: $Var(u_i) = \sigma^2 X_i$.

   Se obtienen los siguientes resultados:

Estimaciones MC.Ponderados utilizando las 51 observaciones 1–51
Variable dependiente: gasto sanitario
Variable utilizada como ponderación: 1/renta

	Coeficiente	Desv. típica	estadístico t	valor p
const	0,1045	0,1624	0,6432	0,5231
renta	0,1442	0,0025	55,5126	0,0000

Estadísticos basados en los datos ponderados:

Suma de cuadrados de los residuos	1,145344	R^2	0,984348
Desviación típica de los residuos	0,152887	Adjusted R^2	0,984029

a) Razona la forma funcional escogida para la varianza de la perturbación. Explica cómo crees que se han obtenido las estimaciones.

b) Suponiendo normalidad en la perturbación, contrasta la significatividad de la variable renta.

5. El investigador no se siente conforme con la forma funcional escogida para $Var(u_i)$ y propone reestimar el modelo (2.27) suponiendo que $Var(u_i) = a + bX_i$, donde a y b son desconocidos.

   a) Explica detalladamente cómo estimarías los coeficientes del modelo (2.27) bajo este supuesto.

   b) Suponiendo $\hat{\sigma}_i^2 = \hat{a} + \hat{b}X_i$. Realiza dicha estimación con la siguiente información muestral:

   $$\begin{array}{lll} \sum \hat{u}_i^2 = 148,699 & \sum \hat{u}_i^2 X_i = 34945,67 & \sum (X_i/\hat{\sigma}_i)^2 = 196420,998 \\ \sum (1/\hat{\sigma}_i)^2 = 34,738 & \sum (Y_i/\hat{\sigma}_i{}^2) = 236,139 & \sum (Y_i X_i/\hat{\sigma}_i{}^2) = 28484,578 \\ \sum (X_i/\hat{\sigma}_i{}^2) = 1608,337 & \sum (Y_i^2/\hat{\sigma}_i{}^2) = 4168,919 & \end{array}$$

   c) Contrasta la significatividad de la variable explicativa.

6. ¿Qué comentarías sobre la validez de los contrastes realizados en los apartados 3), 4.b) y 5.c)?

2.9. Prácticas de Aula

Las prácticas de aula son clases participativas, ello quiere decir que el alumno debe acudir a clase con el ejercicio realizado en su tiempo de trabajo personal. Previamente y con anticipación suficiente se le habrá realizado la debida propuesta. Durante la clase se preguntará aleatoriamente a los alumnos sobre lo realizado.

En el tema de heterocedasticidad es habitual disponer de dos prácticas de aula, lo que equivale a dos horas de clase presencial y cuatro horas de trabajo personal. Si el ejercicio ha sido realizado previamente por el alumno, el tiempo es suficiente para su corrección y solución de las dudas existentes. A continuación se van a proponer los enunciados correspondientes a las dos prácticas citadas.

Competencias a trabajar en estas sesiones.

1. Comprender la importancia de los supuestos empleados en la especificación de un modelo econométrico básico para poder proponer y emplear supuestos más realistas.
2. Diferenciar distintos métodos de estimación y evaluar su uso de acuerdo a las características de las variables económicas de interés para obtener resultados fiables.

Práctica de Aula PA-H.1:

Se dispone de una base de datos sobre el precio de venta y distintas características de 224 viviendas pertenecientes a dos áreas residenciales del condado de *Orange* en California (USA), *Dove Canyon* y *Coto de Caza*[13]. Dove Canyon es una zona de viviendas relativamente pequeñas construidas alrededor de un campo de golf. Coto de Caza es un área de mayor nivel de vida, aunque más rural con viviendas más grandes. Las variables que se consideran son:

salepric	:	precio de venta de la vivienda en miles de dólares
sqft	:	tamaño de la vivienda en pies cuadrados
age	:	edad de la vivienda en años
city	:	1 si está en Coto de Caza, 0 si está en Dove Canyon

A continuación se muestran los resultados de la estimación por Mínimos Cuadrados Ordinarios (MCO) de un modelo para el precio de venta de la vivienda utilizando esa base de datos:

RESULTADOS A: Estimaciones MCO utilizando las 224 observaciones 1–224
Variable dependiente: salepric

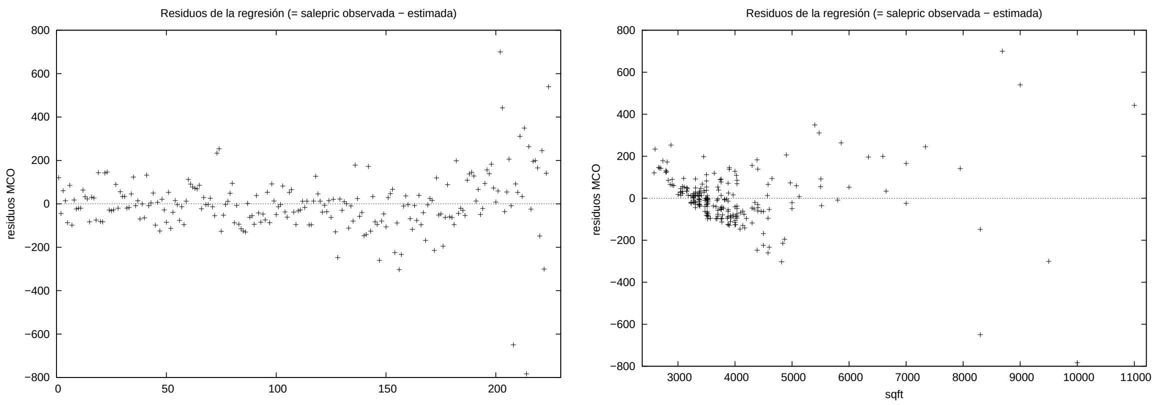
	Coeficiente	Desv. típica	estadístico t	valor p
const	−440,31	35,3203	−12,4663	0,0000
sqft	0,2520	0,0081	30,9047	0,0000
age	3,6980	3,0241	1,2228	0,2227
city	91,8038	21,7494	4,2210	0,0000

[13] Fichero data7-24.gdt, disponible en gretl pestaña Ramanathan. Recogida en Ramanathan, Ramu (2002), *Introductory econometrics with applications*, 5th. edn., South-Western.

Media de la var. dependiente	642,929
D.T. de la variable dependiente	371,376
Suma de cuadrados de los residuos	4,27804e+06
Desviación típica de la regresión ($\hat{\sigma}$)	139,4480
R^2	0,8609
$\bar{R}^2$ corregido	0,8590
$F(3, 220)$	453,8840

1. Escribe el modelo teórico que se ha estimado y comenta los resultados obtenidos en términos de bondad de ajuste, significatividad y signos de los coeficientes estimados.

Figura 2.13: *Gráfico de residuos MCO sobre las observaciones $i = 1, \ldots, 224$ y sobre la variable sqft*



2. Analiza de forma razonada la información que te proporcionan los gráficos recogidos en la Figura 2.13 y la regresión auxiliar. Si realizas algún contraste, indica todos los elementos del mismo. ¿Cuál de los gráficos es más informativo y por qué?

$$\widehat{\frac{\hat{u}_i^2}{SCR_A/224}} = -\underset{(-10{,}387)}{5,94184} + \underset{(12{,}727)}{0,00172457}\,\text{sqft}_i$$
$$N = 224 \quad R^2 = 0,421826 \quad SCR = 1478,52$$

A continuación se muestran los resultados de la estimación por MCO utilizando un estimador de la matriz de varianzas y covarianzas de los coeficientes consistente, aunque exista heterocedasticidad.

RESULTADOS B: Estimaciones MCO utilizando las 224 observaciones 1–224
Variable dependiente: salepric
Desviaciones típicas robustas ante heterocedasticidad, variante HC3

	Coeficiente	Desv. típica	estadístico t	valor p
const	−440,31	110,8800	−3,9711	0,0001
sqft	0,2520	0,0279	9,0120	0,0000
age	3,6980	5,1672	0,7157	0,4750
city	91,8038	26,3997	3,4774	0,0006

Media de la var. dependiente	642,9290
D.T. de la variable dependiente	371,3760
Suma de cuadrados de los residuos	4,27804e+06
Desviación típica de la regresión ($\hat{\sigma}$)	139,4480
R^2	0,8609
$\bar{R}^2$ corregido	0,8590
$F(3, 220)$	161,8190

1. ¿En qué varían los resultados mostrados ahora (RESULTADOS B) con los primeros (RESULTADOS A)? ¿Por qué? ¿Cuáles son fiables y para qué? Explica razonadamente.

Por último, se muestran los resultados de la estimación por Mínimos Cuadrados Generalizados o Ponderados utilizando como variable de ponderación el inverso del cuadrado del tamaño de la vivienda esto es, $\frac{1}{sqft^2}$.

RESULTADOS C: Estimaciones MC.Ponderados utilizando las 224 observaciones 1–224
Variable dependiente: salepric
Variable utilizada como ponderación: $\dfrac{1}{sqft^2}$

	Coeficiente	Desv. típica	estadístico t	valor p
const	−285,20	37,2121	−7,6643	0,0000
sqft	0,2155	0,0095	22,4752	0,0000
age	−0,5492	2,2800	−0,2409	0,8098
city	110,7800	15,6896	7,0607	0,0000

Estadísticos basados en los datos ponderados:

Suma de cuadrados de los residuos	0,15074
Desviación típica de la regresión ($\hat{\sigma}$)	0,02617
R^2	0,79881
$\bar{R}^2$ corregido	0,79607
$F(3, 220)$	291,177

Estadísticos basados en los datos originales:

Media de la var. dependiente	642,929
D.T. de la variable dependiente	371,376
Suma de cuadrados de los residuos	4,73514e+06
Desviación típica de la regresión ($\hat{\sigma}$)	146,708

1. ¿Qué se quiere decir con datos ponderados y datos originales? ¿Por qué se utiliza como variable de ponderación el inverso de $sqft^2$? Explica razonadamente.

2. ¿Qué resultados de los tres A,B, ó C te parecen mejores? ¿Por qué?

Práctica de Aula PA-H.2:

Una agencia de viajes de Chicago quiere analizar si hay diferencias significativas entre las familias en la elección del destino de vacaciones, más o menos alejados de su lugar de residencia, en función del número de hijos pequeños en la familia. Para ello dispone de una muestra de 200 familias de esta ciudad entrevistadas en el año 2007 y se especifica el siguiente modelo[14]:

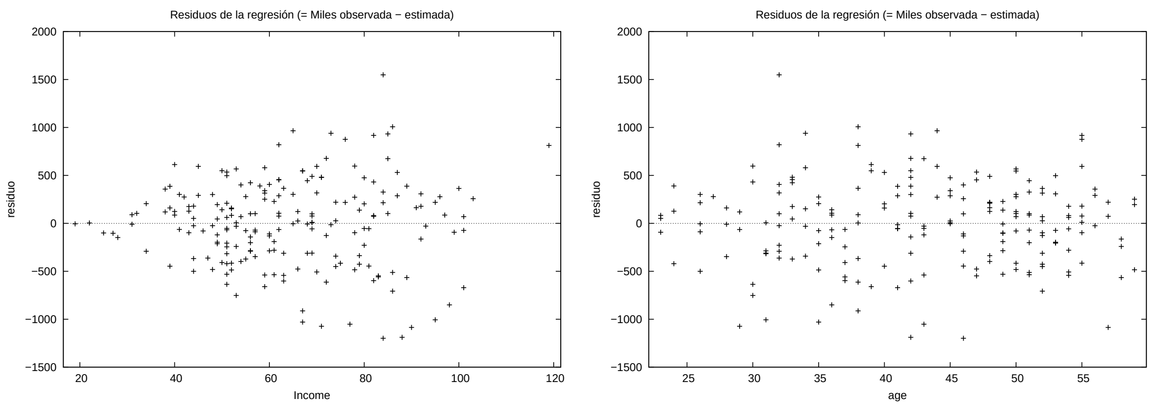
$$Miles_i = \beta_1 + \beta_2 Income_i + \beta_3 age_i + \beta_4 kids_i + u_i \qquad i = 1, \ldots, 200 \tag{2.29}$$

donde $Miles$ son las millas recorridas por una familia en las vacaciones de ese año, $Income$ es la renta familiar anual en miles de dólares, age es la edad media de los adultos en la familia y $kids$ el número de hijos menores de 16 años existentes en la familia.

Una primera estimación del modelo por MCO produce los siguientes resultados:

$$\begin{array}{rcl} \widehat{Miles}_i & = & -391,55 + 14,201\,Income_i + 15,741\; age_i - 81,826\,kids_i \\ (\widehat{des}(\hat{\beta}_{MCO})) & & \quad (169,8) \qquad (1,80) \qquad\qquad (3,757) \qquad (27,13) \\ R^2 & = & 0,340605 \qquad\qquad SCR = 40099000 \end{array} \tag{2.30}$$

Figura 2.14: *Gráfico de residuos MCO sobre la variable Income y sobre la variable age*



1. ¿Qué te sugieren los gráficos recogidos en la Figura 2.14? Comenta detalladamente cada uno de ellos.

Después de agrupar las observaciones de todas las variables en dos grupos en función de un ordenamiento decreciente de la variable $Income$ y estimar el modelo (2.29) anterior por MCO separadamente para cada grupo, se obtienen las siguientes resultados:

[14]Fichero vacation.dat. Recogido en Hill, R. C., Griffiths, W. e. y G. C. Judge (2001), *Undergraduate Econometrics*, 2ª edn., John Wiley and Sons, Inc.

Primera submuestra: Estimaciones MCO utilizando las 80 observaciones 1–80
Variable dependiente: Miles

Variable	Coeficiente	Desv. típica	Estadístico t	valor p
const	−129,22	615,6100	−0,2099	0,8343
Income	13,1490	6,1456	2,1396	0,0356
age	13,3666	7,5921	1,7606	0,0823
kids	−114,1800	52,9888	−2,1549	0,0343

Suma de cuadrados de los residuos	2,42765e+07
R^2	0,116112

Segunda submuestra: Estimaciones MCO utilizando las 80 observaciones 121–200
Variable dependiente: Miles

Variable	Coeficiente	Desv. típica	Estadístico t	valor p
const	−339,64	220,1600	−1,5427	0,1271
Income	9,6880	4,0104	2,4157	0,0181
age	18,6511	3,8740	4,8143	0,0000
kids	−66,0260	29,8963	−2,2085	0,0302

Suma de cuadrados de los residuos	7,04816e+06
R^2	0,308962

2. Realiza un contraste para verificar si lo que sugieren los gráficos es estadísticamente significativo. Debes señalar claramente todos los elementos del contraste incluidas la hipótesis nula y la alternativa.

3. Si el contraste realizado te diera que rechazas la hipótesis nula, ¿qué cambiarías de los resultados presentados en (2.30) si no quisieras cambiar el método de estimación de los coeficientes? ¿Por qué y para qué lo harías? Explica detalladamente.

Se ha utilizado un método de estimación alternativo a MCO para mejorar en términos de eficiencia la estimación de los coeficientes β. Utilizando el software gretl se han obtenido los siguientes resultados:

Estimaciones MC.Ponderados utilizando las 200 observaciones 1–200
Variable dependiente: Miles
Variable utilizada como ponderación: $\frac{1}{Income}$

Variable	Coeficiente	Desv. típica	Estadístico t	valor p
const	−408,37	145,7170	−2,8025	0,0056
Income	13,9705	1,6482	8,4762	0,0000
age	16,3483	3,4222	4,7771	0,0000
kids	−78,3630	24,7355	−3,1680	0,0018

Estadísticos basados en los datos ponderados:

Suma de cuadrados de los residuos	580616,00
Desviación típica de los residuos ($\hat{\sigma}$)	54,4272
R^2	0,3907
$\bar{R}^2$ corregido	0,3813
$F(3, 196)$	41,8975

Estadísticos basados en los datos originales:

Media de la var. dependiente	1054,2300
D.T. de la variable dependiente	552,7990
Suma de cuadrados de los residuos	4,01134e+07
Desviación típica de los residuos ($\hat{\sigma}$)	452,3940

4. Completa las siguientes expresiones sobre el término de perturbación del modelo y el método de estimación utilizado en la obtención de estos resultados.

$E(u_i) =$ $\qquad$ $E(u_i^2) =$ $\qquad$ $E(u_i u_j) =$

$$\underbrace{E(uu')}_{(\ldots\ldots\ldots \times \ldots\ldots\ldots)} = \begin{bmatrix} \\ \\ \\ \\ \end{bmatrix}$$

`Criterio de estimación:`........ $SCR = \sum_{i=\ldots.}^{i=\ldots.} (Y_i^* - \hat{\beta}_1 X_{1i}^* - \hat{\beta}_2 X_{2i}^* - \hat{\beta}_3 X_{3i}^* - \hat{\beta}_4 X_{4i}^*)^2$

$Y_i^* =$ ....; $\qquad$ $X_{1i}^* =$ ....; $\qquad$ $X_{2i}^* =$ ....;

$X_{3i}^* =$ ....; $\qquad$ $X_{4i}^* =$ ....;

$$\hat{\beta}_{\ldots\ldots} = \begin{bmatrix} \\ \\ \\ \\ \end{bmatrix}^{-1} \begin{bmatrix} \\ \\ \\ \\ \end{bmatrix}$$

5. Si tuvieras que contrastar $H_0 : \beta_2 = 10$ ¿Cómo lo harías? Razona y explica tu respuesta.

2.10. Prácticas de Ordenador

Las prácticas de ordenador se realizan en el Laboratorio Informático donde cada alumno dispone de una terminal. Son clases prácticas en las que es importante la colaboración del alumno para que se progrese en el aprendizaje. En general se proporcionará el enunciado de la tarea a trabajar con anticipación por si el alumno quiere investigar con el software gretl por su cuenta.

En el tema de heterocedasticidad es habitual disponer de dos prácticas de ordenador que equivalen a dos horas presenciales y cuatro horas de trabajo no presencial. A continuación se va a proponer un ejercicio para resolver en el centro de cálculo. El enunciado cubre todo lo aprendido en el tema y se va a marcar la división entre las dos horas de clase que conlleva, las prácticas de ordenador PO-H.1 y PO-H.2.

Es conveniente que una vez acabado el ejercicio y en vuestro tiempo de trabajo personal deis contenido al mismo. Es decir, en clase únicamente nos da tiempo a aprender como ir obteniendo los resultados con gretl. Aunque el profesor va comentando y explicando los resultados. Vosotros y de forma personal debéis redactar convenientemente las respuestas de cada apartado. Notar que en algunos de ellos se incluye la coletilla "A realizar en casa" por esa razón. Al final de este tema en el Anexo 1.2 tenéis un resumen de las instrucciones básicas de gretl para heterocedasticidad que podéis consultar para realizar las tareas.

Competencias a trabajar en estas sesiones.

2. Diferenciar distintos métodos de estimación y evaluar su uso de acuerdo a las características de las variables económicas de interés para obtener resultados fiables.

3. Utilizar diversas fuentes estadísticas y adquirir destreza en el uso de un software econométrico para analizar relaciones entre variables económicas.

Descripción del fichero de datos a utilizar en las sesiones:
Los datos necesarios para la realización de este ejercicio se encuentran en el fichero `EAF01.gdt`. Los datos corresponden a 540 individuos y contienen información sobre educación, trabajo, ingresos y otras características personales[15]. De las variables disponibles, se consideran las siguientes:

- $EARNINGS$: Ingresos por hora trabajada, en dólares.

[15] Fichero EAEF01.gdt. Fuente: US National Longitudinal Survey of Youth 1979, recogida en Dougherty, Ch. (2006), *Introduction to Econometrics*, 3rd. Ed., Oxford University Press. Disponible en gretl, pestaña Dougherty.

- $FEMALE$: Variable ficticia con valor 1 si el individuo es mujer, 0 si es hombre.
- S: Años de escolarización.
- EXP: Experiencia laboral, en años.
- $HOURS$: Número de horas trabajadas, por semana.

Práctica de Ordenador PO-H.1

Trabajos previos: El fichero de datos original considera un gran número de variables, por ello sería aconsejable que creases uno nuevo donde únicamente aparezcan las variables que se consideran. Para ello puedes proceder de la forma siguiente: en la pantalla inicial de gretl y después de haber abierto el archivo de datos EAF01.gdt pulsa con el botón derecho del ratón en:

Guardar datos como → *formato estándar*

Señala las variables a mantener y pulsa → *Aceptar*

Da nombre al nuevo fichero de datos. Guárdalo convenientemente. Ya puedes trabajar con el fichero reducido. Por razones de operatividad es necesario que ordenes la muestra de forma creciente con la variable S_i. Realízalo.

1. Estima por MCO el siguiente modelo para los ingresos:

$$EARNINGS_i = \beta_1 + \beta_2 FEMALE_i + \beta_3 S_i + \beta_4 EXP_i + \beta_5 HOURS_i + u_i \tag{2.31}$$

Interpreta los resultados obtenidos en términos de significatividad de las variables, signos de los coeficientes estimados y bondad del ajuste ("A realizar en casa"). Guarda los residuos y sus cuadrados así como los valores de la variable $\widehat{EARNINGS}_i$.
Completa lo siguiente con la salida que se obtiene de gretl:

$$\underset{(\widehat{des}(\hat\beta_{MCO}))}{\widehat{EARNINGS}_i} = \underset{(\qquad)}{\ldots} \underset{(\qquad)}{\ldots} FEMALE_i \underset{(\qquad)}{\ldots} S_i \underset{(\qquad)}{\ldots} EXP_i \underset{(\qquad)}{\ldots} HOURS_i \tag{2.32}$$

$$R^2 = \qquad\qquad SCR = \qquad\qquad \widehat{Cov}(\hat\beta_2, \hat\beta_3) =$$

	$EARNINGS_i$	$\hat{u}_i$	$\widehat{EARNINGS}_i$
$i = 1$			
$i = 2$			
$i = 3$			

2. Dibuja y comenta el gráfico de la variable endógena observada y estimada así como los gráficos de residuos siguientes:
   a) Residuos ($\hat{u}_{i,MCO}$) frente a EXP_i
   b) Residuos al cuadrado ($\hat{u}^2_{i,MCO}$) frente a S_i
3. Realiza el contraste de Goldfeld y Quandt para el supuesto de que $Var(u_i) = f(S_i)$, siendo $f(\cdot)$ una función creciente. Para ello, selecciona, por un lado, los valores de S_i estrictamente menores a 13 y, por otro, los valores de S_i estrictamente mayores a 14.
4. Realiza el contraste de Breusch-Pagan para el supuesto de que $Var(u_i) = f(\alpha_0 + \alpha_1 S_i)$.
5. Obtén la estimación MCO con desviaciones típicas robustas al problema planteado. Comenta qué razones hay para utilizarlas y cómo se utilizan ("A realizar en casa").
6. Supongamos que $E(u_i)^2 = aS_i^2$ siendo a una constante ($a > 0$). Estima por Mínimos Cuadrados Ponderados los coeficientes del modelo (2.31). Explica detalladamente cómo se calcula el estimador de varianza mínima de los coeficientes $\beta_1, ..., \beta_5$ del modelo (2.31), sin olvidar el criterio de estimación ("A realizar en casa").
   Completa los resultados de la estimación:

$$\underset{(\widehat{des}(\hat{\beta}_{.....}))}{\widehat{EARNINGS}_i} = \underset{(\qquad)}{\qquad} \; ... \; \underset{(\qquad)}{\qquad} FEMALE_i \; ... \; \underset{(\qquad)}{\qquad} S_i$$

$$... \; \underset{(\qquad)}{\qquad} EXP_i \; ... \; \underset{(\qquad)}{\qquad} HOURS_i$$

Práctica de Ordenador PO-H.2

En esta sesión vamos a continuar trabajando con el fichero de datos reducido que creamos en la práctica de ordenador PO-H.1. La clase comienza revisando los resultados más importantes obtenidos en dicha práctica y continuamos con la solución de los siguientes apartados:

7. En este apartado vamos a comprobar que aplicar el estimador de Mínimos Cuadrados Generalizados al modelo original es equivalente a estimar por Mínimos Cuadrados Ordinarios el correspondiente modelo transformado. Para ello seguimos considerando el supuesto sobre el comportamiento de la varianza de la perturbación realizado en el apartado anterior. Escribe la ecuación del modelo transformado y obtén los datos ponderados correspondientes a las variables del modelo (2.31). Finalmente, estima por MCO el modelo con los datos ponderados y realiza la comprobación pedida.

8. Dibuja los residuos ($\hat{u}_{i,MCO}$) frente a $FEMALE_i$. Comenta el gráfico obtenido.

9. Si pensamos en todos los gráficos de residuos realizados podríamos pensar que una forma funcional adecuada para la varianza de la perturbación es $Var(u_i) = \alpha_1 + \alpha_2 FEMALE_i + \alpha_3 S_i + \alpha_4 EXP_i$. Estima el modelo bajo este nuevo supuesto. Completa los resultados de la estimación:

$$\begin{array}{cccccccc}
\widehat{EARNINGS_i} & = & \ldots & & FEMALE_i & \ldots & & S_i \\
(\widehat{des}(\hat{\beta}_{.....})) & & (\qquad\quad) & & (\qquad\quad) & & (\qquad\quad) & \\
 & & \ldots & & EXP_i & \ldots & & HOURS_i \\
 & & (\qquad\quad) & & (\qquad\quad) & & &
\end{array}$$

10. Contrasta la hipótesis $\beta_4 = -2\beta_5$.

2.11. Taller sobre todo lo trabajado en el tema

Los talleres son clases participativas, ello quiere decir que el alumno debe acudir a clase con el ejercicio realizado en su tiempo de trabajo personal. Previamente y con anticipación suficiente se le habrá realizado la debida propuesta. En los talleres se trabaja el aprendizaje cooperativo, luego el ejercicio debe resolverse en equipo, idealmente en el grupo en que se va a realizar el proyecto. Esto favorece la interrelación y conocimiento entre los miembros. En los últimos 15 minutos de clase, un representante de cada equipo deberá exponer al resto de la clase la decisión adoptada por el grupo junto a las razones consideradas. Durante la clase se preguntará aleatoriamente y de forma individual a los alumnos sobre lo realizado.

En el tema de heterocedasticidad en general se lleva a cabo un taller. El objetivo del taller es analizar una serie de resultados de la estimación de varias especificaciones de un modelo o diferentes estimaciones de la misma especificación y que el alumno pueda ir practicando la toma de decisiones y la redacción apropiada de conclusiones. Por ello vamos a proponer un taller utilizando lo trabajado en las prácticas de ordenador. De esta forma el alumno ya está familiarizado con el problema en concreto.

Una hora de trabajo en un taller en clase equivale a dos horas de trabajo personal sobre el mismo. Además, es necesario que en este tiempo se reflexione y redacten los argumentos y conclusiones a las que se haya llegado en clase.

Competencias a trabajar en esta sesión.

1. Comprender la importancia de los supuestos empleados en la especificación de un modelo econométrico básico para poder proponer y emplear supuestos más realistas.

2. Diferenciar distintos métodos de estimación y evaluar su uso de acuerdo a las características de las variables económicas de interés para obtener resultados fiables.

4. Elaborar en grupos de trabajo y exponer en público un proyecto empírico, donde se valore adecuadamente los resultados obtenidos del análisis de un modelo econométrico.

Enunciado del taller:

Objetivo: Especificar y estimar la función de ingresos por hora trabajada. Se propone la siguiente relación:

$$\text{EARNINGS}_i = \beta_1 + \beta_2 FEMALE_i + \beta_3 S_i + \beta_4 EXP_i + \beta_5 HOURS_i + u_i \tag{2.33}$$

donde suponemos que las perturbaciones siguen una distribución normal de media cero.

Información: Se dispone de información sobre tres especificaciones alternativas y sus correspondientes estimaciones, para especificar la relación entre las siguientes variables:

- Ingresos por hora trabajada, $EARNINGS$, en dólares.
- Una variable ficticia, $FEMALE$, que determina el sexo del individuo. Toma valor 1 si el individuo es mujer, 0 si es hombre.
- Los años de escolarización del individuo, S.
- La experiencia laboral, EXP, en años.
- El número de horas trabajadas por semana, $HOURS$.

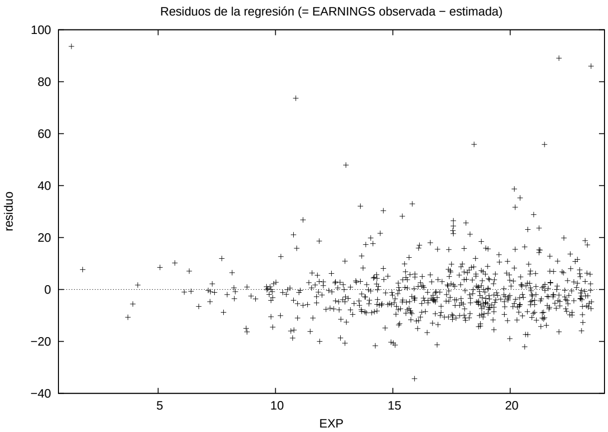
Procedimiento: Debéis analizar la información disponible y decidir cuál es la especificación más adecuada junto con su correcta estimación.

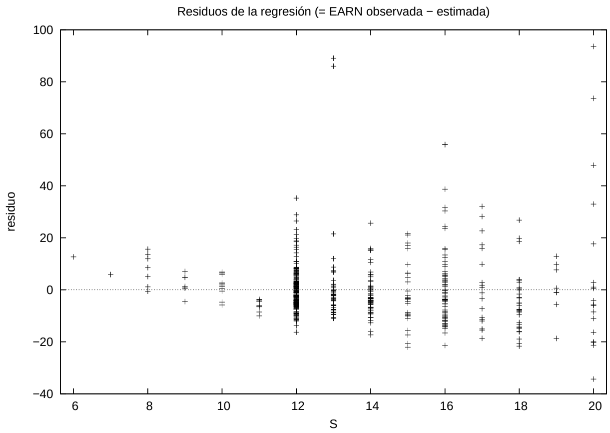
ESTIMACIÓN MCO:

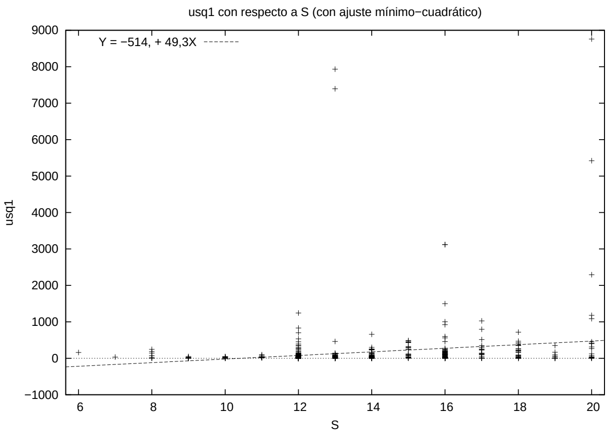
Los resultados de la estimación MCO son los siguientes:

$$
\begin{array}{lccccccc}
\widehat{\text{EARNINGS}}_i & = & -17,8361 & - & 6,74458 & \text{FEMALE}_i & + & 2,61304 \ \ \text{S}_i \\
(\widehat{des}(\hat\beta_{MCO})) & & (5,05978) & & (1,15363) & & & (0,22688) \\
(\widehat{des}(\hat\beta_{MCO})_{White}) & & (6,06795) & & (1,41171) & & & (0,28563) \\
 & & & + & 0,488736 & \text{EXP}_i & - & 0,07523 \ \ \text{HOURS}_i \\
 & & & & 0,13690 & & & (0,06531) \\
 & & & & (0,17326) & & & (0,10688)
\end{array}
$$

$R^2 = 0,24746 \qquad SCR = 86479,7 \qquad \widehat{Cov}(\hat\beta_2, \hat\beta_3) = 0,0096428 \qquad T = 540$







Además, se dispone de las siguientes regresiones auxiliares:

$$\hat u_i^2 = -514,322 + 49,33S_i + \hat\xi_{1i} \qquad SCR = 2,46476 + 008 \qquad R^2 = 0,03360 \quad (A)$$

$$\frac{\hat u_i^2}{\hat u'\hat u} = -0{,}0059 + 0,00057S_i + \hat\xi_{2i} \qquad SCR = 0,0329 \qquad R^2 = 0,0336 \quad (B)$$

$$\frac{\hat u_i}{(\hat u'\hat u/540)} = -3,211 + 0,308S_i + \hat\xi_{3i} \qquad SCR = 9610,24 \qquad R^2 = 0,0336 \quad (C)$$

donde $\hat u_i$ son los residuos MCO.

1. Debéis decidir si esta especificación junto con su método de estimación es la más adecuada. Analizar la información proporcionada por la regresión (2.33) junto con los gráficos y las posibilidades, **(TODAS)**, que os ofrecen las regresiones A a C.

ESTIMACIÓN MCG:
Dados los resultados analizados en el apartado anterior se propone estimar la ecuación (2.33) por Mínimos Cuadrados Generalizados y se obtienen los siguientes resultados:

Estimaciones MC.Ponderados utilizando las 540 observaciones 1–540
Variable dependiente: EARNINGS
Variable utilizada como ponderación: S^2

Variable	Coeficiente	Desv. típica	Estadístico t	valor p
const	−13,3666	6,0182	−2,2210	0,0268
FEMALE	−8,4725	1,3163	−6,4366	0,0000
S	2,7514	0,2560	10,7466	0,0000
EXP	0,3994	0,1640	2,4344	0,0152
HOURS	−0,1746	0,0715	−2,4429	0,0149

Estadísticos basados en los datos ponderados:

Suma de cuad. residuos	21920498	D.T. de la regresión	202,4176
R^2	0,244486	$F(4, 535)$	43,28177

1. Debéis decidir si esta especificación junto con su método de estimación es la más adecuada. Analizar la información proporcionada y en particular deberíais razonar sobre la adecuación de la ponderación utilizada. ¿Qué podéis decir acerca de la fiabilidad de los resultados mostrados? ¿A qué conclusiones llegáis?

ESTIMACIÓN MCGF:
Dados los siguientes resultados:

Estimaciones MCO utilizando las 540 observaciones 1–540
Variable dependiente: σ_i^2

Variable	Coeficiente	Desv. típica	Estadístico t	valor p
const	−1,3610	0,8162	−1,6675	0,0960
FEMALE	−0,6939	0,1995	−3,4780	0,0005
S	0,2524	0,0406	6,2137	0,0000
EXP	0,0771	0,0244	3,1584	0,0017

Estimaciones MC.Ponderados utilizando las 540 observaciones 1–540
Variable dependiente: EARNINGS
Variable utilizada como ponderación: $\dfrac{1}{\hat{\sigma}_i^2}$

Variable	Coeficiente	Desv. típica	Estadístico t	valor p
const	−12,4753	3,9112	−3,1896	0,0015
FEMALE	−5,8871	1,0521	−5,5951	0,0000
S	2,1640	0,1866	11,5961	0,0000
EXP	0,3798	0,1090	3,4831	0,0005
HOURS	−0,0219	0,0561	−0,3903	0,6964

1. Analizar la información proporcionada por los resultados de la estimación. Consejo: empezar por escribir la función de regresión poblacional que se está estimando e indicar cuáles son los supuestos sobre la perturbación que se han asumido. Analizar su coherencia dada **TODA** la información disponible hasta este momento. Finalmente, tomad una decisión sobre si esta especificación junto con su método de estimación es adecuada.

2. Si tuvieseis que escoger entre las alternativas de estimación empleadas para estimar el modelo (2.33), ¿cuál escogeríais? Razonar la respuesta. ("A realizar en casa")

2.12. Evaluativas - Preguntas Cortas

Dado que el curso contempla la evaluación continua es necesario que a lo largo del tema se evalúe en el día a día a los alumnos. Por ello se realizan pruebas en forma de preguntas cortas. Se pueden realizar tanto en las clases magistrales como en las prácticas de aula u ordenador. Se llevan a cabo en clase y de manera individual. A modo de ejemplo se incluyen las siguientes.

• Preguntas cortas en Clase Magistral:
En el modelo:

$$Y_i = \beta_1 + \beta_2 X_{2i} + u_i \qquad u_i \sim NID\left(0\,,\,\sigma^2 \frac{1}{X_{2i}}\right) \qquad i = 1, \ldots, 224$$

donde la matriz de regresores X es no estocástica.

Pc1. Completa las siguientes matrices:

$$E(u) = \begin{bmatrix} \cdots \\ \cdots \\ \cdots \\ \vdots \\ \cdots \end{bmatrix} \qquad E(uu') = \begin{bmatrix} \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix}$$

Pc2. ¿Es constante la varianza de la perturbación a lo largo de la muestra? ¿De qué depende?

Pc3. ¿Por qué método estimarías los coeficientes del modelo? Razona tu decisión en base a las propiedades del estimador.

Pc4. Si quisieras estimar el modelo por MCP, ¿cómo debes ponderar las observaciones? Escribe la ponderación que utilizarías.

Pc5. Escribe el correspondiente modelo transformado con perturbaciones esféricas y obtén la varianza de la perturbación de dicho modelo.

Pc6. Escribe explícitamente la fórmula del estimador eficiente e indica cómo son cada uno de sus componentes.

• Preguntas cortas en Práctica de Aula:
Se realiza un estudio de la función de ahorro familiar para lo cual se especifica el siguiente modelo:

$$SAVE_i = \beta_1 + \beta_2\, INCOME_i + \beta_3\, SIZE_i + u_i, \qquad i = 1, \ldots, 100 \tag{2.34}$$

donde

- $SAVE$: ahorro familiar en dólares en el año 1970.
- $INCOME$: renta familiar total en dólares en el año 1970.
- $SIZE$: número de miembros de la familia.

Los resultados de la estimación del modelo (2.34) por MCO son

$$\underset{(\widehat{des}(\hat{\beta}_{i,MCO}))}{\widehat{SAVE}_i} = \underset{(1200{,}88)}{-256{,}831} + \underset{(0{,}058)}{0{,}148}\; INCOME_i + \underset{(217{,}292)}{82{,}572}\; SIZE_i \qquad R^2 = 0{,}0635 \tag{2.35}$$

donde $\widehat{des}(\hat{\beta}_{i,MCO})$ ha sido obtenida en el estimador $\widehat{V}(\hat{\beta}_{MCO}) = \hat{\sigma}^2(X'X)^{-1}$.

Se dispone de los siguientes gráficos de los residuos MCO en función de $INCOME_i$ y $SIZE_i$ junto las regresiones auxiliares (2.36) y (2.37), donde $\hat{\sigma}^2 = \dfrac{\widehat{u}'\widehat{u}}{97}$.

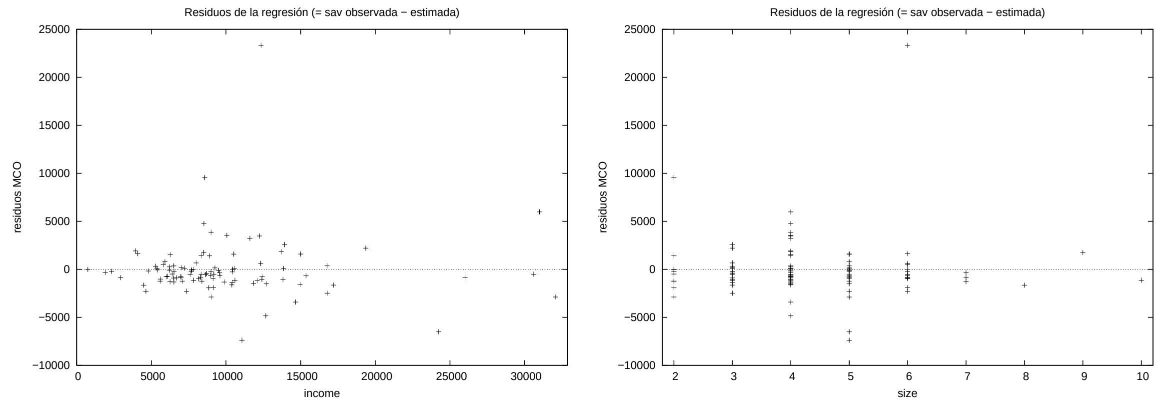
$$\frac{\widehat{u}_i^2}{\hat{\sigma}^2} = 0{,}0546 + 0{,}000095\, INCOME_i \qquad SCE = 27{,}91 \tag{2.36}$$

$$\frac{\widehat{u}_i^2}{\hat{\sigma}^2} = -0{,}3589 + 0{,}3123\, SIZE_i \qquad SCE = 21{,}54 \tag{2.37}$$

Pc1. Utilizando los gráficos y las regresiones (2.36) y (2.37), contrasta la posible existencia de heterocedasticidad en el modelo (2.34) y propón una forma funcional razonable para $Var(u_i)$. **Justifica** tu elección en base a toda la información disponible.

Pc2. ¿Cómo podrías realizar un contraste válido sobre la significatividad de la variable $INCOME_i$ utilizando $\hat{\beta}_{MCO}$? ¿Qué resultados mostrados en (2.35) debes variar y por qué?

Figura 2.15: *Residuos MCO versus variables independientes*



Pc3. Supón que $Var(u_i) = \sigma^2 INCOME_i^2,\ Cov(u_i, u_j) = 0,\ \forall i \neq j.$

a) Completa la expresión del criterio de estimación utilizado en el apartado anterior:

$$\underset{\hat{\beta}}{Min} \sum_{i=1}^{100} \ldots\ldots (Y_i - \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots)^2$$

b) ¿Qué tiene en cuenta el criterio de estimación de MCG que no tiene en cuenta el criterio de MCO? Razona y explica tu respuesta.

• Preguntas cortas en Práctica de Ordenador:

Accede al conjunto de datos **data8-3.gdt** del libro de Ramanathan incluido en gretl[16].

En este fichero vas a encontrar datos de las siguientes variables: el gasto sanitario agregado en billones de dólares (*exphlth*), la renta personal disponible agregada también en billones de dólares (*income*), el porcentaje de población que supera los 65 años en el año 2005 (*seniors*) y la población en millones (*pop*).

Para el modelo:

$$exphlth_i = \beta_1 + \beta_2 income_i + \beta_3 seniors_i + u_i \qquad i = 1, \ldots, 51$$

Pc1. Contrasta la posibilidad de que $Var(u_i) = \sigma^2 pop_i^2$ utilizando el estadístico de Breusch-Pagan y completa:

a) H_0 : H_a :

b) Estadístico de contraste y distribución:

[16] Fichero data8-3.gdt, disponible en gretl pestaña Ramanathan. Recogido en Ramanathan, R. (2002), *Introductory Econometrics with Applications*, 5th. edn., South-Western.

c) Regresión auxiliar:

d) Regresión auxiliar **estimada**:

e) Computa el estadístico de contraste y lleva a cabo el contraste:

Pc2. Suponiendo $Var(u_i) = \sigma^2 pop_i^2$ estima adecuadamente el modelo y razona las propiedades del estimador utilizado.

Pc3. Contrasta utilizando el estimador MCO de los coeficientes y de forma válida la significatividad individual de las variables *income* y *seniors*. Contrasta la significatividad conjunta.

2.13. Bibliografía del tema

Referencias Bibliográficas Básicas:

• Teórica:

[1] Greene, W. (1998), Análisis Econométrico, ed. Prentice Hall, 3ª edition.

[2] Ramanathan, R. (2002), Introductory Econometrics with applications, ed. South-Western, 5th edition.

[3] Wooldridge, J. M. (2003), Introductory Econometrics: A modern Approach, ed. South-Western, 2nd edition.

• Ejercicios:

[1] Fernández, A., González, P., Regúlez, M., Moral, P., Esteban, V. (2005), Ejercicios de Econometría, ed. McGraw-Hill, 2 ª edición.

[2] Recopilación de ejercicios recomendados y exámenes de Econometría. Departamento de Economía Aplicada III (Econometría y Estadística). Mimeo Febrero 2009.

• Ejercicios con gretl:

[1] Ramanathan, R. (2002), Instructor's Manual to accompany, del libro Introductory Econometrics with applications, ed. South-Western, 5th edition, Harcourt College Publishers.

[2] Wooldridge, J. M. (2003), Student Solutions Manual, del libro Introductory Econometrics: A modern Approach, ed. South-Western, 2nd edition.

Referencias Bibliográficas Complementarias:

[1] Alonso, A., Fernández, J. y Gallastegui, I. (2005), Econometría, ed. Pearson: Prentice Hall.

[2] Breusch, T. S. y Pagan, A. R. (1979), A Simple Test for Heteroscedasticity and Random Coefficient Variation, *Econometrica*, 47, pp. 1287-1294.

[3] Dougherty, Ch. (2006), Introduction to Econometrics, 3rd. Ed., Oxford University Press.

[4] Goldfeld, S. M. y Quandt, R. E. (1965), Some Test for Homoscedasticity, *Journal of the American Statistical Association*, 60, pp. 539-547.

[5] Gujarati, D. (2004), Econometría, ed. McGraw-Hill, 4ª edición. [6] Harvey, A., y Phillips, G. (1974), A Comparison of the Power of Some Test for Heteroscedasticity in the General Linear Model, *Journal of Econometrics*, 2, pp. 307-316.

[7] Johnston, J. (1984), Métodos de Econometría, ed. Vicens Vivens, 4ª edición.

[8] Maddala, G. S. (1996), Introducción a la Econometría, ed. Pearson: Prentice Hall.

[9] Novales, A. (1993), Econometría, ed. McGraw-Hill, 2ª edición.

[10] White, H. (1980), A heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity, *Econometrica*, 48, pp. 817-838.

2.14. Anexo 1.1: Resultados de gretl utilizados en las clases magistrales

- **Resultado 1:** Estimación por MCO:

Estimaciones MCO utilizando las 40 observaciones 1–40
Variable dependiente: consumo

	Coeficiente	Desv. típica	estadístico t	valor p
const	40,7676	22,1387	1,8415	0,0734
renta	0,1282	0,0305	4,2008	0,0002

Media de la var. dependiente	130,3130
D.T. de la variable dependiente	45,1586
Suma de cuadrados de los residuos	54311,30
Desviación típica de la regresión ($\hat{\sigma}$)	37,8054
R^2	0,3171
$\bar{R}^2$ corregido	0,2991
Grados de libertad	38

- **Resultado 2:** Regresiones parciales para computar el estadístico de Goldfeld y Quandt:

Estimaciones MCO utilizando las 20 observaciones 1–20
Variable dependiente: consumo

	Coeficiente	Desv. típica	estadístico t	valor p
const	12,9388	28,9666	0,4467	0,6604
renta	0,1823	0,0520	3,5023	0,0025

Media de la var. dependiente	112,3040
D.T. de la variable dependiente	32,9714
Suma de cuadrados de los residuos	12284,2
Desviación típica de la regresión ($\hat{\sigma}$)	26,1238
R^2	0,4052
$\bar{R}^2$ corregido	0,3722
Grados de libertad	18

Estimaciones MCO utilizando las 20 observaciones 21–40
Variable dependiente: consumo

	Coeficiente	Desv. típica	estadístico t	valor p
const	48,1767	70,2419	0,6859	0,5015
renta	0,1176	0,0815	1,4425	0,1663

Media de la var. dependiente	148,322
D.T. de la variable dependiente	49,1527
Suma de cuadrados de los residuos	41146,9
Desviación típica de la regresión ($\hat{\sigma}$)	47,8115
R^2	0,1036
$\bar{R}^2$ corregido	0,0538
Grados de libertad	18

- **Resultado 3:** Regresión auxiliar para computar el estadístico de Breusch Pagan:

Estimaciones MCO utilizando las 40 observaciones 1–40
Variable dependiente: enorm

	Coeficiente	Desv. típica	estadístico t	valor p
const	−1,6788	0,6876	−2,4415	0,0194
renta	0,0038	0,0009	4,0461	0,0002

Media de la var. dependiente	1,0000
D.T. de la variable dependiente	1,3864
Suma de cuadrados de los residuos	52,3940
Desviación típica de la regresión ($\hat{\sigma}$)	1,1742
R^2	0,3010
$\bar{R}^2$ corregido	0,2827
Grados de libertad	38

- **Resultado 4:** Estimación por MCP:

Estimaciones MC.Ponderados utilizando las 40 observaciones 1–40
Variable dependiente: consumo
Variable utilizada como ponderación: 1/renta

	Coeficiente	Desv. típica	estadístico t	valor p
const	31,9244	17,9861	1,7749	0,0839
renta	0,1409	0,0269	5,2216	0,0000

Estadísticos basados en los datos ponderados:

Suma de cuadrados de los residuos	68,7020
Desviación típica de la regresión ($\hat{\sigma}$)	1,3446
R^2	0,4177
$\bar{R}^2$ corregido	0,4024
Grados de libertad	38

Estadísticos basados en los datos originales:

Media de la var. dependiente	130,313
D.T. de la variable dependiente	45,1586
Suma de cuadrados de los residuos	54557,3
Desviación típica de la regresión ($\hat{\sigma}$)	37,8909

- **Resultado 5:** Regresión auxiliar para el estimador de MCGF:

Estimaciones MCO utilizando las 40 observaciones 1–40
Variable dependiente: $\hat{u}^2_{i,MCO}$

	Coeficiente	Desv. típica	estadístico t	valor p
const	1923,6000	2365,4800	0,8132	0,4213
renta	−7,4202	6,6882	−1,1094	0,2744
sq_renta	0,0087	0,0045	1,9221	0,0623

Media de la var. dependiente	1357,78
D.T. de la variable dependiente	1882,48
Suma de cuadrados de los residuos	8,78226e+07
Desviación típica de la regresión ($\hat{\sigma}$)	1540,64
R^2	0,3645
$\bar{R}^2$ corregido	0,3302
$F(2,37)$	10,6133
valor p para $F()$	0,0002

- Estimación por MCGF:

Estimaciones MC.Ponderados utilizando las 40 observaciones 1–40
Variable dependiente: consumo
Variable utilizada como ponderación: $\hat{\sigma}^2_i$

	Coeficiente	Desv. típica	estadístico t	valor p
const	34,2386	17,6042	1,9449	0,0592
renta	0,1410	0,0289	4,8757	0,0000

Estadísticos basados en los datos ponderados:

Suma de cuadrados de los residuos	39,3094
Desviación típica de la regresión ($\hat{\sigma}$)	1,0170
R^2	0,3848
$\bar{R}^2$ corregido	0,3686
Grados de libertad	38

Estadísticos basados en los datos originales:

Media de la var. dependiente	130,3130
D.T. de la variable dependiente	45,1586
Suma de cuadrados de los residuos	54793,9
Desviación típica de la regresión ($\hat{\sigma}$)	37,9730

- **Resultado 6:** Estimación de White de $V(\hat{\beta}_{MCO})$:

Estimaciones MCO utilizando las 40 observaciones 1–40
Variable dependiente: consumo
Desviaciones típicas robustas ante heterocedasticidad, variante HC3

	Coeficiente	Desv. típica	estadístico t	valor p
const	40,7676	27,3213	1,4922	0,1439
renta	0,1282	0,0439	2,9168	0,0059

Media de la var. dependiente	130,3130
D.T. de la variable dependiente	45,1586
Suma de cuadrados de los residuos	54311,3
Desviación típica de la regresión ($\hat{\sigma}$)	37,8054
R^2	0,3171
$\bar{R}^2$ corregido	0,2991
Grados de libertad	38

2.15. Anexo 1.2: Instrucciones básicas de gretl para heterocedasticidad

En las clases del Centro de Cálculo o Laboratorio Informático nuestro objetivo es aprender el manejo de un software libre especialmente indicado y creado para el aprendizaje de la Econometría. Es un software muy sencillo en el que podéis ser autodidactas. Sin embargo, vamos a comenzar con un recordatorio de las nociones básicas que ya se han visto en Introducción a la Econometría. A continuación se pasará a mostrar las instrucciones específicas más usuales para el caso de heterocedasticidad.

Comienzo de la sesión. Conexión y lectura de datos:

- Encender el terminal → Hacer click cuando lo pide → Introducir login y password donde lo pide. Introducir el pendrive o diskette.
- Si pensáis guardar los resultados en un documento Word. Pulsar:

  Inicio → Todos los programas → Microsoft Office → Microsoft Office Word.

  Así estamos abriendo un documento en Word para ir guardando los resultados[17].

  Minimizaremos la ventana del documento .doc (ó .tex) si es el caso para usarlo cuando haya resultados que guardar.
- Pulsar *Inicio → Todos los programas → gretl*

  Ya estamos dentro de gretl y veremos una ventana con diferentes opciones que podemos utilizar.
- En el Centro de Cálculo de la Facultad los resultados van a una carpeta compartida que está previamente creada, pero en nuestro PC necesitaremos guardar los resultados en una carpeta donde previamente hemos abierto el documento Word, Tex, etc. Por lo tanto, lo primero que haremos será predeterminar el destino. Pulsar:
  Archivo → Preferencias → General

  En la ventana *Directorio gretl de usuario* buscaremos la situación de la carpeta citada → Pulsar *Aceptar*
- Para leer los datos de la tarea. Pulsar:
  Archivo ⟶ Abrir datos ⟶ Archivo de muestra ⟶ Nombre del "fichero de datos" por ejemplo → *Ramanathan → data7-24.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, en *Modelo* podremos realizar estimaciones.

[17]Si no dominamos Word podemos guardar los resultados en formato texto para pegar en el block de notas o "Notepad". En este caso no crearemos el documento .doc. gretl también permite guardar resultados en formato Latex previo crear el documento .tex de la misma forma que el documento Word con la opción adecuada.

Algunas etiquetas de la pantalla principal:

- La etiqueta *Datos*:
  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 seleccionar la etiqueta correspondiente y la variable o variables a estudiar. Por ejemplo, para ver la estructura del conjunto de datos pulsamos en la etiqueta *Estructura del conjunto de datos* y obtendremos una pantalla en la que aparecerá seleccionado el tipo de datos con el que estamos trabajando, en este caso *Sección Cruzada* seleccionamos *Adelante* y nos confirma que la muestra es una sección cruzada junto con su tamaño, *1 a 224 observaciones.*

  Si la muestra fuese de serie temporal hubiera indicado *Serie temporal*, seleccionaríamos aceptar y veríamos la frecuencia, *mensual*, y el inicio y final de la muestra *1968:1 a 1998:12*, por ejemplo. 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.

- La etiqueta *Ver:*
  Se obtienen gráficos de las variables y sus estadísticos principales entre otros. Para obtener los estadísticos principales de las variables de la muestra podemos hacerlo pulsando 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, para una única serie o para el conjunto de ellas seleccionándolo previamente.

- La etiqueta *Variable*:
  Sirve 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

- La etiqueta *Añadir*:
  Con esta etiqueta podemos añadir variables o transformaciones de las existentes al conjunto de datos original, para ello tras pulsar en *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:
    index $i = 1, \ldots, N$, index $\equiv i$
    time $t = 1, \ldots, T$, time $\equiv t$
  - 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.
  - Para crear un conjunto de datos:

Datos no incluidos en gretl:
En ocasiones debemos trabajar con un fichero de datos que no está incluido en gretl. Podemos importarlo y trabajar con él con solo que esté en alguno de los formatos compatibles con gretl. Para ello pulsamos en *Archivo* $\rightarrow$ *Abrir datos* $\rightarrow$ *importar* y seleccionamos el formato adecuado siguiendo la secuencia de órdenes que se nos pida. Esta secuencia va dirigida a definir la muestra, tipo de datos, longitud, etc.

En otras ocasiones debemos crear directamente la muestra en gretl. Para ello debemos crear un nuevo conjunto de datos como sigue:

Archivo $\rightarrow$ *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*. Pulsar *Aceptar*
- En la hoja de cálculo situarnos en la primera celda y teclear la observación correspondiente, a continuación pulsar *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 → insertar obs*. Una vez introducidas todas las variables pulsar *Aplicar*.
- Para guardar los datos: en menú *Archivo → 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 sólo habrá que pulsar el botón izquierdo del ratón dos veces para que se active.

Un repaso a lo más básico:

• **Estimación MCO:** *Modelo ⟶ Mínimos Cuadrados Ordinarios*

Seleccionar la variable endógena y exógenas mediante el siguiente proceso:

1. Variable endógena, pulsar *nombre de la variable dependiente → Elegir*
2. Elegir los regresores, pulsar *Añadir* con cada una. Por defecto tendremos predeterminada una constante que se puede eliminar si es necesario. Para realizar la regresión pulsar *Aceptar*.

Se muestran los resultados de la estimación y diferentes estadísticos. Las desviaciones típicas son calculadas con la expresión $\hat{\sigma}^2(X'X)^{-1}$.

Notar que en la ventana abierta por MCO, abajo a la izquierda aparece una casilla con la leyenda estimaciones típicas robustas. En principio no debe estar activada. Corresponden a la estimación consistente de la matriz de varianzas y covarianzas de los estimadores MCO como veremos más adelante. En la ventana de resultados de la estimación MCO tenemos diferentes opciones, podemos hacer contrastes, gráficos etc.

• **Para guardar los resultados en formato word**: *Editar → Elegir formato RTF(Ms Word).* Abrir el documento Word creado anteriormente. Seleccionar:

Edición → Pegar → Guardar → Minimizar ventana[18]

• **Gráficos de residuos:**
Tenemos varias posibilidades para dibujar los residuos, podemos dibujar su evolución en la muestra o contra alguna de las variables exógenas de la muestra, dependiendo de lo que nos interese. Partimos de la ventana de resultados MCO:

1. Evolución de los residuos:

   a) si la muestra es de serie temporal:
   Gráficos ⟶ Gráfico de residuos ⟶ contra "el tiempo"

[18] Para guardarlo en Latex, .tex, o modo texto, .txt, proceder igual con la opción adecuada.

b) si la muestra es de sección cruzada:
Gráficos $\longrightarrow$ *Gráfico de residuos* $\longrightarrow$ *"por observación"*

2. Gráfico de residuos frente a alguno de los regresores:

   Gráficos $\longrightarrow$ *Gráfico de residuos* $\longrightarrow$ *contra "nombre del regresor"*

3. Para dibujar los residuos frente a una variable incluida en el fichero de datos, pero que no sea uno de los regresores debemos situarnos en la pantalla inicial de gretl y seguir la secuencia:

   Datos $\longrightarrow$ *Gráficos* $\longrightarrow$ *Gráfico X-Y scatter* $\longrightarrow$ *seleccionar X e Y adecuadamente*

• **Para guardar gráficos**: pulsar con el botón derecho del ratón en cualquier parte del gráfico y elegir la opción en que queremos que nos lo guarde, por ejemplo postcript (.eps) o cualquier otra que nos convenga. En la ventana que aparece indicarle donde queremos que nos lo guarde.

• **Para ver el gráfico variable ajustada-observada:**
Gráficos $\longrightarrow$ *Gráfico de ajustada-observada* $\longrightarrow$ *elegir por número de observación* (si es una muestra de sección cruzada) o *frente al tiempo* (si la muestra es de serie temporal)

• **Para ver el gráfico variable ajustada-observada frente a alguno de los regresores:**
Gráficos $\longrightarrow$ *Gráfico de ajustada-observada* $\longrightarrow$ *contra "nombre del regresor"*

• Podemos guardar los datos de la variable endógena estimada, los residuos y los residuos al cuadrado que posiblemente, necesitemos después. En la pantalla de resultados de la estimación:
Guardar $\longrightarrow$ *valores ajustados*

Guardar $\longrightarrow$ *residuos*

Guardar $\longrightarrow$ *residuos al cuadrado*

entre otros. gretl los va a añadir al conjunto de datos con el que trabajamos y los denota respectivamente por yhat1, uhat1 e usq1 respectivamente, donde 1 indica que corresponde al modelo 1, así que si lo buscáis para el modelo que habéis estimado en tercer lugar los llamara uhat3. Además, añade una leyenda explicativa de la variable. Como veis en la pestaña hay otros estadísticos que se pueden guardar de la misma forma por ejemplo la suma residual de cuadrados por ejemplo, ess1, etc.
También podemos hacer nuevas estimaciones o añadir variables explicativas a la anterior repitiendo los pasos anteriores.

Instrucciones de gretl específicas para heterocedasticidad

• **Contrastes de heterocedasticidad**:

gretl tiene implementados diferentes contrastes de heterocedasticidad. Por ejemplo el contraste de White o el contraste de Breusch-Pagan. Ambos se encuentran en la pantalla de resultados de la estimación MCO. Pulsar:

Contrastes → Heterocedasticidad → seleccionar el contraste deseado

Por supuesto ambos los podéis realizar siguiendo explícitamente todos los pasos aprendidos en clase. De la misma forma se puede llevar a cabo el contraste de Goldfeld y Quandt. Veamos como realizar el contraste de Breusch-Pagan.

• Contraste de **Breusch-Pagan**:
Para computar la regresión auxiliar para el contraste de Breusch-Pagan hemos de guardar en primer lugar, los cuadrados de los residuos MCO y la SCR del modelo de interés. A continuación hemos de definir la variable endógena de la regresión auxiliar, los residuos normalizados, que hemos llamado $e_i^2 = \hat{u}_{i,MCO}^2/\hat{\sigma}^2$, para ello en:

Variable ⟶ Definir una nueva variable

introducir la fórmula $e2 = usq1/(ess1/N)$ con los valores que previamente habéis guardado o anotado de la suma de cuadrados residual y el tamaño muestral[19]. A continuación estimáis la regresión auxiliar por MCO de la manera habitual y anotar los estadísticos necesarios para computar el test.

• Contraste de **Goldfeld y Quandt**:
En el contraste de Goldfeld y Quandt se supone que la varianza de la perturbación es monótonamente creciente (o decreciente) con una variable cuyos valores son conocidos. Es necesario ordenar la muestra conforme al crecimiento (decrecimiento) de esa variable. Para ello podemos seguir la siguiente secuencia desde la pantalla inicial:

Datos	⟶	*ordenar*
	⟶	*seleccionar la variable por cuyo crecimiento o decrecimiento queremos ordenar la muestra*
	⟶	*seleccionar el criterio, creciente o decreciente*

Además, es necesario que la muestra se divida en dos submuestras dejando un número de observaciones centrales que de independencia a ambas submuestras. A continuación se realiza la regresión MCO en ambas submuestras y se aplica el estadístico de contraste. La única dificultad de este test es dividir la muestra. Para ello debemos seguir la siguiente secuencia de órdenes partiendo de la pantalla principal de gretl. Seleccionar, según el criterio adecuado:

Muestra	→	*seleccionar rango*
	→	*restringir a partir de un criterio*
	→	*definir a partir de una variable ficticia*

con ello hemos restringido la muestra y podemos realizar la regresión en dicha submuestra. Tomar nota de los estadísticos necesarios. A continuación hay que recuperar el rango completo antes de volver a restringir la muestra para crear la segunda submuestra. Pulsar:

Muestra → recuperar rango completo

volver a restringir la muestra para realizar la segunda regresión y tomar nota de los datos necesarios para realizar el contraste. Antes de seguir trabajando recordar recuperar el rango completo de la muestra.

[19]Se ha denotado $\hat{u}_{i,MCO}^2$ =usq1 y $\hat{\sigma}^2$ =ess1/N.

• Los contraste de Goldfeld- Quandt y Breusch-Pagan no son los únicos posibles. De hecho gretl ofrece la posibilidad de contrastar heterocedasticidad usando el test de White y una variante del mismo como habéis podido ver al desplegar la pestaña *Contrastes* ⟶ *Heterocedasticidad* sin embargo, en los contenidos del tema no se suele incluir este contraste ya que el tiempo disponible es limitado.

• **Estimador de White**:
El estimador de White es el estimador consistente de la matriz de varianzas y covarianzas de los parámetros estimados por MCO, $\widehat{V}(\hat{\beta}_{MCO})$. Lógicamente ha de encontrarse dentro de la estimación MCO y ya lo hemos abordado antes, estaba en la ventana abierta por MCO. Abajo a la izquierda aparece una casilla con la leyenda estimaciones típicas robustas. Corresponden a la estimación consistente de la matriz de varianzas y covarianzas de los estimadores MCO a la White. Si deseamos esta estimación haremos *click* en la casilla. Pulsar en *configurar* y elegir *HC3a* que es la correspondiente al estimador visto en clase.

• **Mínimos Cuadrados Generalizados o Ponderados (MCP)**
Bajo MCP gretl implementa el estimador de MCG vía MCO en el modelo transformado.

Lo primero que va a necesitar gretl es conocer la ponderación por ello, si la ponderación no es una variable previamente definida, debéis definirla de la forma habitual:

Variable ⟶ *Definir una nueva variable* ⟶ *introducir la fórmula*

Una vez que tenéis construida la ponderación para estimar por MCP seleccionamos:
Modelo ⟶ *otros modelos lineales* ⟶ *mínimos cuadrados ponderados*

Se eligen adecuadamente la variable dependiente, las independientes y la ponderación seleccionando y añadiendo de la forma habitual.

Con la definición de la variable de ponderación hay que tener un poco de cuidado ya que es necesario adecuarse a la forma que utiliza el programa para construir las variables en el modelo transformado. Así para diferentes versiones hace diferentes cosas y necesitamos saber a ciencia cierta cómo construye el modelo transformado a partir de la definición de la variable de ponderación. Lo sabemos en la pestaña Ayuda, dentro de Mínimos Cuadrados Ponderados dependiendo de la versión. Por ejemplo para la versión 1.7.1, incluye la siguiente leyenda en inglés[20]:

Weighted Least Squares

If “wtvar”is a dummy variable, WLS estimation is equivalent to eliminating all observations with value zero for wtvar.

Let “wtvar”denote the variable selected in the ”Weight variable”box. An OLS regression is run, where the dependent variable is the product of the positive square root of wtvar and the selected dependent variable, and the independent variables are also multiplied by the square root of wtvar. Statistics such as R-squared are based on the weighted data. If wtvar is a dummy variable, weighted least squares estimation is equivalent to eliminating all observations with value zero for wtvar.

[20]Las etiquetas de ayuda están en inglés ya que no se traducen como el resto del programa.

Donde "wtvar" denota la variable de ponderación y el modelo transformado se construye multiplicando la raíz cuadrada de "wtvar" por cada variable del modelo original incluida la constante. Por ejemplo si suponemos que $Var(u_i) = \sigma^2 X_{2i}^2$ tomaremos como ponderación a $\frac{1}{X_{2i}^2}$ y si $Var(u_i) = \sigma^2 X_{2i}$ tomaremos como ponderación a $\frac{1}{X_{2i}}$.

• **Mínimos Cuadrados Generalizados Factibles**

1. Primero hemos de pensar en la forma funcional de la varianza. Supongamos que la forma funcional de la varianza que vamos a proponer es:

$$Var(u_i) = \sigma_i^2 = \alpha_1 + \alpha_2 W_{1i} + \alpha_3 W_{2i}$$

   A continuación hemos de crear la regresión auxiliar y estimarla por MCO. Si la variable endógena son los residuos MCO al cuadrado del modelo de interés y los hemos guardado podemos utilizarlos directamente y si no lo hemos hecho elevaremos al cuadrado los residuos MCO guardados en la estimación MCO original. Estimamos la regresión auxiliar y guardamos las estimaciones de la variable endógena de esta regresión. Con esta serie construiremos la ponderación, es decir, vamos a obtener $\hat{\sigma}_i^2$ por lo que debemos comprobar que es siempre positiva.

2. Construimos la ponderación $1/\hat{\sigma}_i^2$ y la añadimos al conjunto de datos usando como habitualmente

   Variable $\longrightarrow$ *Definir una nueva variable* $\longrightarrow$ *introducir la fórmula*

3. A continuación seleccionamos:

   Modelo $\longrightarrow$ *Mínimos Cuadrados Ponderados*

   y la variable dependiente, las independientes y la ponderación seleccionando y añadiendo de la forma habitual. Este modo de proceder proporciona el estimador de MCGF.

Tema 3

Autocorrelación

Clases Magistrales

Habitualmente se dedican a este tema seis clases magistrales. En estas clases el profesor ilustra al alumno otra situación habitual en la práctica cuando en el análisis econométrico se dispone de datos de las variables observadas a lo largo del tiempo. El alumno ya está familiarizado en líneas generales con el problema econométrico, ya que se enclava dentro del marco del modelo de regresión lineal con perturbaciones no esféricas. Por lo tanto, en el Tema 1 sobre Mínimos Cuadrados Generalizados, el alumno ya ha visto las consecuencias de tener este marco de análisis en términos de estimación e inferencia. En el tema anterior ya han sido aplicadas estas técnicas en el caso concreto de heterocedasticidad y la estructura será bastante similar. Se comenzará con la introducción del concepto de autocorrelación y se estudiarán dos modelos concretos para modelar la autocorrelación: los procesos autorregresivos y los de medias móviles. Seguidamente se dedicarán varias clases a explicar y a ilustrar con varios ejemplos la utilización del análisis gráfico de los residuos y a dos contrastes de autocorrelación: el contraste de Durbin-Watson y el contraste de Breusch-Godfrey. Las últimas clases magistrales se dedicarán a ver aplicado a este contexto varios procedimientos de estimación e inferencia por Mínimos Cuadrados Generalizados Factibles junto a las consecuencias de estimar por Mínimos Cuadrados Ordinarios y una forma alternativa de realizar la inferencia con este estimador. A lo largo de las clases magistrales se pedirá al alumno además, de la lectura de las notas sobre el tema, la realización de diversos ejercicios que se resolverán en clase. Para conocer el grado de seguimiento del alumno, se podrán realizar en clase preguntas cortas a resolver.

Competencias a trabajar en estas sesiones:

1. Comprender la importancia de los supuestos empleados en la especificación de un modelo econométrico básico para poder proponer y emplear supuestos más realistas.

2. Diferenciar distintos métodos de estimación y evaluar su uso de acuerdo a las características de las variables económicas de interés para obtener resultados fiables.

Al final de este tema deberíais ser capaces de:

1. Explicar que se entiende por un modelo de regresión lineal con autocorrelación, cómo modelizar esta característica en el término de perturbación del modelo y sus implicaciones en su matriz de varianzas y covarianzas.
2. Saber analizar gráficamente la posible existencia de autocorrelación y otros problemas de especificación.
3. Utilizar lo contrastes de Durbin-Watson y Breusch-Godfrey para contrastar la posible existencia de autocorrelación.
4. Describir y comparar las propiedades de los estimadores MCO y MCG bajo autocorrelación.
5. Transformar un modelo con un término de error que sigue un AR(1) en otro cuyo término de perturbación es un ruido blanco.
6. Estimar y hacer inferencia por MCG o MCGF bajo el supuesto de que el término de error del modelo de regresión sigue un proceso AR(1).
7. Razonar por qué resulta conveniente disponer de un estimador consistente de la matriz de varianzas y covarianzas de los estimadores de MCO en presencia de autocorrelación. Saber utilizar dicho estimador en el contexto adecuado.

Bibliografía Recomendada:

Al final del tema tenéis recogida la bibliografía correspondiente. En particular os recomendamos leer los capítulos correspondientes a la bibliografía básica detallados a continuación:

- Greene, W. (1998), cap. 13.
- Ramanathan, R. (2002), cap. 9.
- Wooldridge, J. M. (2003), cap. 11 y cap. 12.

Y para profundizar, podéis leer los capítulos detallados a continuación correspondientes a la bibliografía complementaria:

- Alonso, A. y otros (2005), cap.10.
- Gujarati, D. (2004), cap.12.
- Johnston, J. (1984), cap. 8.
- Maddala, G. S. (1996), cap. 6.
- Novales, A. (1993), cap. 5 y cap. 7.
- Pindyck R. S. y Rubinfeld, D. L. (1998), cap. 6.

3.1. El concepto de autocorrelación y su modelización

3.1.1. Introducción

Este tema aborda el problema de autocorrelación, una característica que puede presentar el término de error de un modelo econométrico y que se manifiesta con mayor frecuencia al tratar con datos observados a lo largo del tiempo. Este tipo de datos, llamados datos de series temporales, son observaciones de una variable recogidas a lo largo del tiempo (días, meses, trimestres, años etc.) de cierta unidad económica (empresa, consumidor, región, país etc.). Estos datos presentan un ordenamiento natural, ya que las observaciones están ordenadas de acuerdo al momento del tiempo en que han sido observadas.

Es probable que a la hora de explicar el comportamiento de una variable observada en el tiempo, se tenga que tener en cuenta que sus observaciones puedan estar correlacionadas a lo largo del mismo. Por ejemplo, el consumo de una familia en un año es probable que esté correlacionada con el consumo del año anterior ya que en media, ceteris paribus, esperaríamos que una familia no cambiara mucho su comportamiento en su consumo de un año a otro. Esto implicaría posiblemente, introducir en el modelo para explicar el consumo de un año, por ejemplo, el consumo del año anterior. También podemos considerar que un cambio en el nivel de una variable explicativa, por ejemplo la renta de esa familia en un año, tenga efectos sobre el consumo en años sucesivos. Una forma de tener en cuenta esto sería introducir como variable explicativa del consumo en un año no sólo la renta de ese año, sino la renta de años anteriores.

Finalmente, además de factores observables cuya influencia en el consumo se distribuye a lo largo del tiempo, podemos tener factores no observables recogidos en el término de perturbación del modelo cuya influencia se mantiene durante varios periodos presentando lo que se llama autocorrelación o correlación serial en el tiempo.

Como vemos, hay distintas formas de introducir en un modelo efectos que se distribuyen a lo largo del tiempo. Podemos introducir retardos tanto de la variable endógena, por ejemplo Y_{t-1}, como de ciertas variables explicativas, $X_{t-j}, j = 1, \ldots, J$, o modelar el término de error u_t en función por ejemplo de su propio pasado $u_{t-j}, j = 1, \ldots, p$. Esto es, lo que se conoce como especificación dinámica del modelo. En muchos casos es necesario reflexionar sobre esta especificación y saber asignar a qué parte, sistemática o de perturbación, corresponde. Una mala especificación dinámica o simplemente funcional de la parte sistemática de un modelo, puede llevar a detectar autocorrelación en el término de perturbación.

En este tema estudiaremos cómo detectar la presencia de correlación serial en el término de error, analizar cuál puede ser la fuente de este problema y si no es un problema de mala especificación, cómo recoger en el modelo esta característica dinámica en el término de error. También estudiaremos las implicaciones de la existencia de autocorrelación en los métodos de estimación alternativos y en los contrastes de hipótesis sobre los parámetros de interés del modelo. En un tema posterior nos ocuparemos de la dinámica en la parte sistemática.

3.1.2. Procesos Autorregresivos y de Medias Móviles

Normalmente se considera como autocorrelación en el término de error de un modelo correctamente especificado en media, al incumplimiento de la hipótesis básica $E(u_t u_s) = 0 \quad t \neq s$. Esto implica que puede existir relación entre pares de perturbaciones u_t, u_s, de forma que no todas las covarianzas entre perturbaciones de distintos periodos de tiempo van a ser cero.

Una forma de recoger este comportamiento es proponer que las perturbaciones siguen algún tipo de proceso estocástico, tal que sus realizaciones están correlacionadas en el tiempo. Vamos a ver dos tipos de procesos, los llamados procesos autorregresivos (AR) y los de medias móviles (MA).

En general, un proceso o modelo Autorregresivo de orden p, que denotamos como AR(p), se define como un modelo de regresión lineal donde la variable dependiente es el término de perturbación y los regresores son sus propias realizaciones pasadas esto es:

$$u_t = \rho_1 u_{t-1} + \ldots + \rho_p u_{t-p} + \varepsilon_t$$

donde $\varepsilon_t \sim iid(0, \sigma_\varepsilon^2)$, y se conoce como un ruido blanco. Un proceso o modelo de medias móviles de orden q, que se denota como MA(q), también se representa como un modelo de regresión lineal donde la variable dependiente es el término de perturbación, pero los regresores son las realizaciones presente y pasadas de la innovación o ruido blanco ε_t.

$$u_t = \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

La combinación de ambos procesos sería un modelo ARMA(p,q).

La matriz de varianzas y covarianzas de las perturbaciones, Ω, es distinta para cada proceso y bajo ciertas condiciones depende de pocos parámetros, que serán los coeficientes del modelo AR o MA y la varianza de ε_t. Lo veremos con dos casos particulares: los procesos AR(1) y MA(1).

$$\underline{\textbf{Proceso AR(1):}} \qquad u_t = \rho u_{t-1} + \varepsilon_t \qquad \varepsilon_t \overset{iid}{\sim} (0, \sigma_\varepsilon^2) \qquad |\rho| < 1 \tag{3.1}$$

Es el proceso más habitual y simple para modelizar autocorrelación en el término de error de un modelo de regresión adecuadamente especificado, en especial cuando los datos son de frecuencia anual. Sustituyendo recursivamente podemos expresar u_t como una combinación lineal de los valores pasados y presente de ε:

$$u_t = \rho u_{t-1} + \varepsilon_t = \rho^2 u_{t-2} + \rho \varepsilon_{t-1} + \varepsilon_t = \cdots = \sum_{j=0}^{\infty} \rho^j \varepsilon_{t-j}$$

Las propiedades de este proceso son las siguientes:

$$\begin{aligned}
E(u_t) &= E(\sum_{j=0}^{\infty} \rho^j \varepsilon_{t-j}) = \sum_{j=0}^{\infty} \rho^j E(\varepsilon_{t-j}) = 0 \\
E(u_t^2) &= E(\sum_{j=0}^{\infty} \rho^j \varepsilon_{t-j})^2 = \sum_{j=0}^{\infty} \rho^{2j} E(\varepsilon_{t-j}^2) = \sigma_\varepsilon^2 (1 + \rho^2 + \rho^4 + \rho^6 + \ldots)
\end{aligned}$$

por lo que si $|\rho| < 1$:

$$Var(u_t) = \sigma_u^2 = \frac{\sigma_\varepsilon^2}{1-\rho^2} \qquad \forall t$$

Además,

$$\begin{aligned} E(u_t u_{t-1}) &= E[(\rho u_{t-1} + \varepsilon_t)u_{t-1}] = \rho E(u_{t-1}^2) + E(u_{t-1}\varepsilon_t) = \rho\sigma_u^2 \\ E(u_t u_{t-2}) &= E[(\rho u_{t-1} + \varepsilon_t)u_{t-2}] = \rho E(u_{t-1}u_{t-2}) + E(u_{t-2}\varepsilon_t) = \rho^2\sigma_u^2 \end{aligned}$$

En general, $\quad E(u_t u_{t-s}) = \rho^s \sigma_u^2$ y por lo tanto:

$$E(uu') = \frac{\sigma_\varepsilon^2}{1-\rho^2} \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{T-1} \\ \rho & 1 & \rho & \cdots & \rho^{T-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{T-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \rho^{T-3} & \cdots & 1 \end{bmatrix} \tag{3.2}$$

Es decir, bajo el modelo AR(1) con el parámetro $|\rho| < 1$, el número de parámetros contenidos en $E(uu')$ es solamente dos, σ_ε^2 y ρ. Esto es conveniente a la hora de tener que utilizar un método de estimación que requiera estimar esta matriz de varianzas y covarianzas.

Como se puede apreciar, la correlación entre perturbaciones no es cero para ningún retardo, aunque a medida que están más alejadas entre sí, dado que $|\rho| < 1$, esta autocorrelación va disminuyendo.

Si $0 < \rho < 1$, todas las covarianzas, y por tanto las correlaciones, son positivas. Esto indica que el efecto de un shock positivo se mantiene en el tiempo con el mismo signo, aunque cada vez será de menor magnitud. Si este signo cambia por un nuevo shock ε_t entonces también se mantendrá en el tiempo el signo negativo. Por lo tanto, las realizaciones de un proceso AR(1) con coeficiente positivo, mostrará una serie temporal con observaciones consecutivas agrupadas del mismo signo. Cuanto más cercano a uno sea el valor de ρ más tardará en general en cambiar el signo.

Si $-1 < \rho < 0$, las covarianzas, y por tanto las correlaciones, alternan el signo siendo positivas entre perturbaciones separadas un número de retardos par y negativas entre aquellas separadas un número de retardos impar. Esto indica que el efecto de un shock positivo tenderá a asociarse con un siguiente error de signo negativo y este con uno positivo, aunque cada vez de menor magnitud. Por lo tanto, las realizaciones de un proceso AR(1) con coeficiente negativo mostrarán una serie temporal con observaciones consecutivas que alternan el signo. Cuanto más cercano a uno en valor absoluto sea el valor de ρ, más se mantendrá la alternancia del signo y su magnitud.

Ejercicio 3.1 Escribe de forma general, sin restricciones sobre sus parámetros, la matriz de varianzas y covarianzas del vector de perturbaciones $E(uu')$ para $T = 10$. ¿Es una matriz simétrica? ¿Cuantas parámetros desconocidos tendrías que estimar? ¿Y si supones que la perturbación sigue un AR(1) con $|\rho| < 1$?

Los siguientes gráficos ilustran la diferencia entre las realizaciones en el tiempo de un proceso que es ruido blanco es decir, sin autocorrelación ($u_t = \varepsilon_t$), versus otras que siguen un AR(1) con coeficiente positivo ($u_t = 0,95u_{t-1} + \varepsilon_t$) o negativo ($u_t = -0,95u_{t-1} + \varepsilon_t$).

Figura 3.1: *Ruido blanco* $\rho = 0$

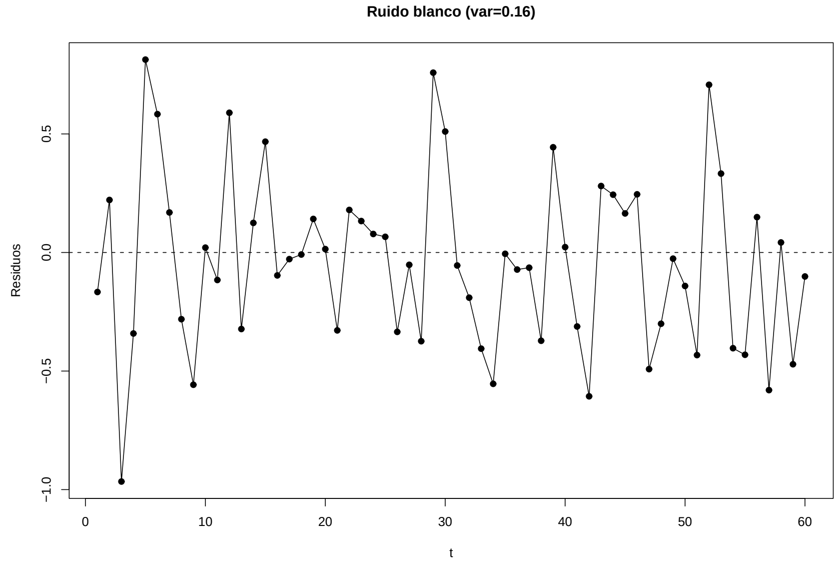


Figura 3.2: *Proceso AR(1) con* $\rho = 0,95$

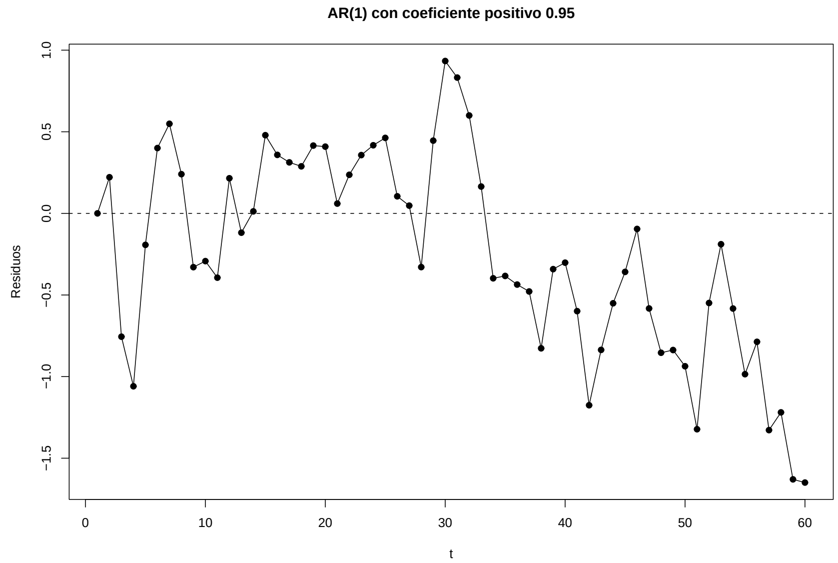
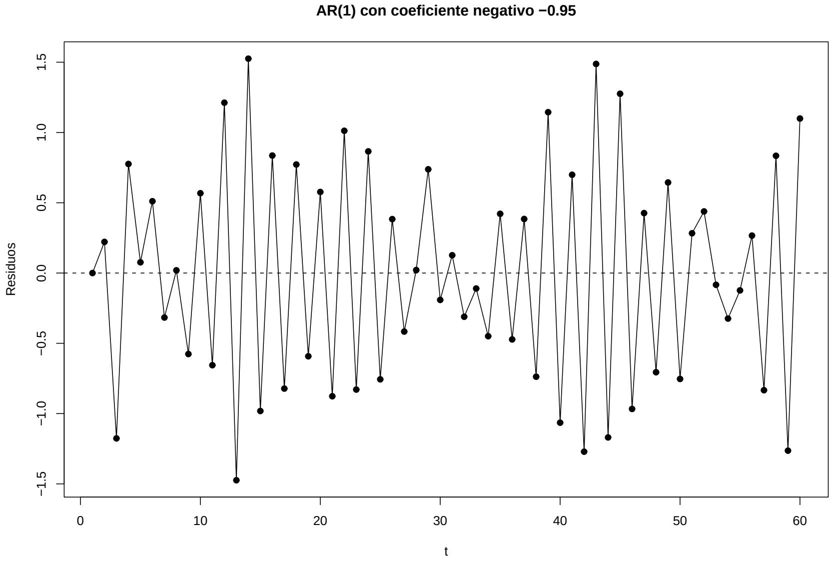


Figura 3.3: *Proceso AR(1) con* $\rho = -0,95$



$$\underline{\textbf{Proceso MA(1)}}: \qquad u_t = \theta\varepsilon_{t-1} + \varepsilon_t \qquad \varepsilon_t \overset{iid}{\sim} (0, \sigma^2_\varepsilon) \tag{3.3}$$

Las propiedades de este proceso son las siguientes:

$$\begin{aligned} E(u_t) &= \theta E(\varepsilon_{t-1}) + E(\varepsilon_t) = 0 \\ E(u_t^2) &= E(\theta\varepsilon_{t-1} + \varepsilon_t)^2 = \theta^2 E(\varepsilon^2_{t-1}) + E(\varepsilon^2_t) + 2\theta E(\varepsilon_t\varepsilon_{t-1}) = \sigma^2_\varepsilon(1+\theta^2) = \sigma^2_u \end{aligned}$$

Es decir, las perturbaciones que siguen un MA(1) tienen media cero y varianza constante. Además:

$$\begin{aligned} E(u_t u_{t-1}) &= E(\theta\varepsilon_{t-1} + \varepsilon_t)(\theta\varepsilon_{t-2} + \varepsilon_{t-1}) = \theta\sigma^2_\varepsilon \\ E(u_t u_{t-2}) &= E(\theta\varepsilon_{t-1} + \varepsilon_t)(\theta\varepsilon_{t-3} + \varepsilon_{t-2}) = 0 \\ E(u_t u_{t-s}) &= 0 \qquad \text{si } s \geq 2 \qquad \text{y por lo tanto:} \end{aligned}$$

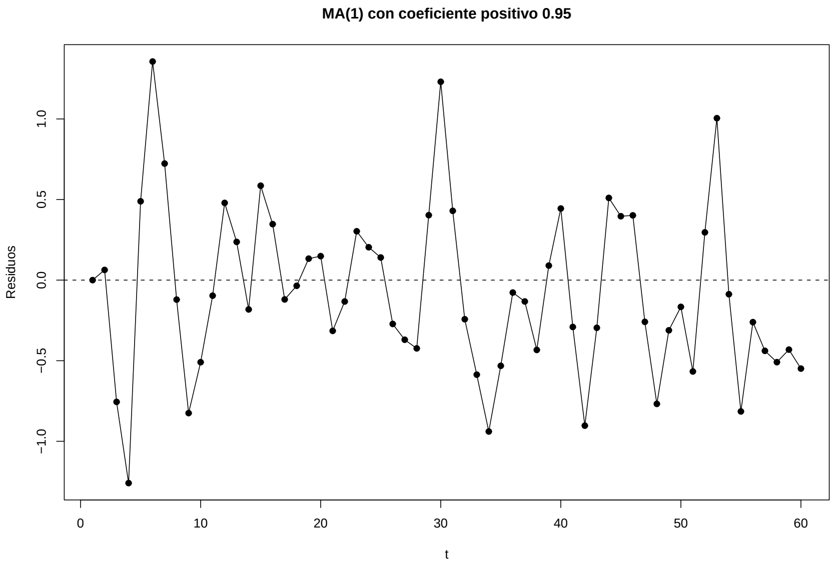
$$E(uu') = \sigma^2_\varepsilon \begin{bmatrix} 1+\theta^2 & \theta & 0 & \cdots & 0 \\ \theta & 1+\theta^2 & \theta & \cdots & 0 \\ 0 & \theta & 1+\theta^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1+\theta^2 \end{bmatrix} \tag{3.4}$$

Ejercicio 3.2 Calcula la matriz de varianzas y covarianzas si consideramos escribir el proceso MA(1) de la forma $u_t = \varepsilon_t - \delta\varepsilon_{t-1}$ donde $\varepsilon_t \stackrel{iid}{\sim} (0, \sigma_\varepsilon^2)$. ¿Si llamamos $\theta = -\delta$, obtienes lo mismo que antes?

Bajo el proceso MA(1) la matriz $E(uu')$ depende también de dos parámetros, σ_ε^2 y θ. A diferencia del proceso AR(1), las covarianzas entre perturbaciones solamente se mantienen distintas de cero entre aquellas separadas un periodo, siendo el resto igual a cero. Esto indica que si θ es mayor que cero, el efecto de un shock positivo en un periodo tiende a mantenerse en el tiempo con el mismo signo solamente por un periodo, siendo luego igual de factible que la realización sea positiva o negativa dependiendo del shock ε de ese periodo. Si el valor de θ es negativo la alternancia será de positivo a negativo en el siguiente periodo o de negativo a positivo.

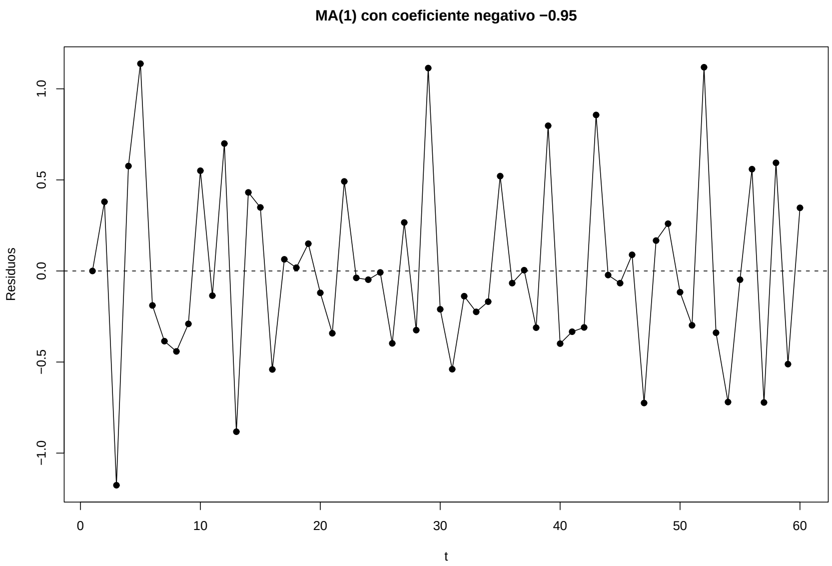
Los siguientes gráficos ilustran las realizaciones en el tiempo de un proceso MA(1) positivo ($u_t = 0,95\varepsilon_{t-1} + \varepsilon_t$) y de un MA(1) negativo ($u_t = -0,95\varepsilon_{t-1} + \varepsilon_t$).

Figura 3.4: *Proceso MA(1) con $\theta = 0,95$*



En los gráficos podemos observar que las series generadas por procesos con autocorrelación de orden uno positiva tienden a ser más suaves que las series temporales generadas por procesos con autocorrelación de orden uno negativa. Los procesos autorregresivos tienden a acentuar esta característica relativamente a los de medias móviles. Procesos de mayor orden o combinaciones de estos dos procesos son más difíciles de detectar con un simple gráfico de la serie temporal. Otro tipo de técnicas de identificación de estos procesos son los llamados gráficos de autocorrelación o correlogramas simple y parcial.

Figura 3.5: *Proceso MA(1) con* $\theta = -0,95$



3.2. Contrastes de autocorrelación y análisis de residuos

Las perturbaciones no son observables, por lo que los residuos serán nuestra aproximación a estas variables tanto en el análisis gráfico como en los contrastes estadísticos. Al igual que en el tema de heterocedasticidad podemos utilizar como preliminar a un contraste estadístico de autocorrelación, el gráfico de la serie temporal de los residuos obtenidos de estimar por MCO la especificación del modelo de interés. Vamos a estudiar dos contrastes diseñados para contrastar la hipótesis nula de ausencia de autocorrelación frente a diversas alternativas de autocorrelación en el error.

3.2.1. Contraste de Durbin y Watson

Este contraste está diseñado para contrastar la hipótesis nula de no autocorrelación

$$H_0 : \rho = 0, \quad u_t = \varepsilon_t \overset{iid}{\sim} (0, \sigma^2_\varepsilon)$$

frente a la alternativa de un proceso AR(1), donde la hipótesis alternativa puede ser

$$H_a : u_t = \rho u_{t-1} + \varepsilon_t \quad 0 < \rho < 1$$

o bien

$$H_a : u_t = \rho u_{t-1} + \varepsilon_t \quad -1 < \rho < 0$$

El procedimiento es el siguiente:

Se estima el modelo de interés por MCO y se calculan los residuos $\hat{u}_t$ para $t = 1, \ldots, T$. El estadístico de Durbin-Watson se calcula como:

$$DW = \frac{\sum_{t=2}^{T}(\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^{T} \hat{u}_t^2}$$

Se puede obtener una relación aproximada entre este estadístico y el coeficiente de autocorrelación de orden uno en los residuos MCO definido como:

$$\hat{\rho} = \frac{\sum_{t=2}^{T} \hat{u}_t \hat{u}_{t-1}}{\sum_{t=2}^{T} \hat{u}_{t-1}^2}$$

siendo esta relación $DW \simeq 2(1 - \hat{\rho})$, por lo tanto, se puede aproximar los valores que tomará el estadístico para ese coeficiente de autocorrelación de orden uno en los residuos.

Así, el valor del estadístico DW está entre $(0, 4)$. Si el valor de DW está cercano a cero indica que el valor de $\hat{\rho}$ está próximo a 1. Si $\hat{\rho} \in (0, 1)$ entonces $DW \simeq\in (2, 0)$, por lo que valores de DW de 2 hacia 0 indicaran evidencia hacia autocorrelación de orden uno positiva. Si $\hat{\rho} \in (-1, 0)$ entonces $DW \simeq\in (4, 2)$, por lo que valores de DW de 2 hacia 4 indicaran evidencia hacia autocorrelación de orden uno negativa.

Formalmente para realizar el contraste necesitamos conocer la distribución del estadístico bajo la hipótesis nula de no autocorrelación, pero la distribución de este estadístico DW no es ninguna de las habituales y se demuestra que depende de la matriz X. Durbin y Watson han tabulado unas cotas, d_i (inferior) y d_s (superior) de valores críticos, que permiten decidir si se rechaza la hipótesis nula a un nivel de significación elegido.

Procedimiento: Dado el tamaño muestral T y el número de regresores exceptuando el término independiente $K' = K - 1$, obtenemos en la tabla de valores críticos tabulados las cotas inferior d_i y superior d_s para un nivel de significación elegido.

- Si $DW \in (d_s, 4 - d_s)$, entonces no se rechaza $H_0 : \rho = 0$ frente a cualquiera de las dos alternativas bien $H_a : 0 < \rho < 1$ o bien $H_a : -1 < \rho < 0$. Por lo tanto, no hay evidencia de autocorrelación de orden uno.
- Si $DW < d_i$ se rechaza $H_0 : \rho = 0$ y hay evidencia de autocorrelación de orden uno positiva $H_a : 0 < \rho < 1$.
- Si $DW > 4 - d_i$ se rechaza $H_0 : \rho = 0$ versus la alternativa de autocorrelación de orden uno negativa $H_a : -1 < \rho < 0$.
- Si $DW \in (d_i, d_s)$ o si $DW \in (4 - d_s, 4 - d_i)$, el estadístico cae en las zonas de indeterminación del contraste, por lo que el contraste no es concluyente.

Comentarios:

- Este contraste puede detectar otros procesos AR o MA de mayor orden siempre y cuando den lugar a residuos con una autocorrelación de orden uno significativa. Por la misma razón puede detectar problemas de mala especificación de la parte sistemática del modelo. Pero el contraste no indica cuál puede ser el caso.

- Si la autocorrelación de orden uno no es significativa puede que este contraste no detecte procesos AR o MA de mayor orden.
- En las zonas de indeterminación el contraste no es concluyente. No obstante, se debería tener en cuenta que tampoco se acepta la hipótesis nula de no autocorrelación siendo cautos con los resultados MCO especialmente en la inferencia utilizando los estadísticos t y F habituales. Esta región de indeterminación es mayor cuanto menor número de grados de libertad se tengan (menor T y mayor K).
- El estadístico de DW no es fiable en modelos que incluyen como variables explicativas la variable dependiente retardada $Y_{t-1}, Y_{t-2}, \ldots$ u otro tipo de regresor estocástico.

3.2.2. Contraste de Breusch y Godfrey

Este es un contraste más general que el de Durbin-Watson. Se contrasta la hipótesis nula:

$$H_0 : u_t \sim iid(0, \sigma_\varepsilon^2)$$

frente a

$$\begin{aligned} H_a : u_t &= \rho_1 u_{t-1} + \ldots + \rho_p u_{t-p} + \varepsilon_t \quad \text{AR(p)} \quad \text{ó} \\ u_t &= \theta_1 \varepsilon_{t-1} + \ldots + \theta_p \varepsilon_{t-p} + \varepsilon_t \quad \text{MA(p)} \end{aligned}$$

donde $\varepsilon_t \sim iid(0, \sigma_\varepsilon^2)$. Luego bajo H_0 el proceso es un ruido blanco, no hay autocorrelación. Bajo la hipótesis alternativa el término de perturbación puede seguir bien un proceso autorregresivo de orden p o bien un proceso de medias móviles del mismo orden, elegido previamente a la realización del contraste. El procedimiento es el siguiente:

- Se estima el modelo de interés por MCO:

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

  Se calculan los residuos $\hat{u}_t$ para $t = 1, \ldots, T$.

- Se obtiene el coeficiente de determinación R^2 de la siguiente regresión auxiliar:

  $$\hat{u}_t = \alpha_1 + \alpha_2 X_{2t} + \ldots + \alpha_K X_{Kt} + \delta_1 \hat{u}_{t-1} + \ldots + \delta_p \hat{u}_{t-p} + \xi_t$$

  donde $t = 1, \ldots, T$ y p es el número de retardos elegido para el orden del proceso AR o MA bajo la alternativa. Al realizar la regresión auxiliar usando todas las observaciones, incluidas las primeras p $(t = 1, \ldots, p)$, se necesita el conocimiento de $\hat{u}_0, \ldots, \hat{u}_{1-p}$. Una forma de solventar el problema de no conocer esos residuos es la de igualar $\hat{u}_0, \ldots, \hat{u}_{1-p}$ a cero.

- El estadístico del contraste y su distribución asintótica bajo la hipótesis nula son:

  $$BG = TR^2 \xrightarrow{d, H_0} \chi_p^2 \qquad \text{bajo } H_0$$

  donde T y R^2 es el tamaño muestral y el coeficiente de determinación, respectivamente, de la regresión auxiliar.

- **Regla de decisión:** Rechazar la hipótesis nula de no autocorrelación a un nivel de significación α, si el valor muestral del estadístico, denotado por BG, es tal que $BG > \chi^2_{p|\alpha}$.

  Otra forma de realizar el contraste consiste en omitir las p observaciones iniciales de la regresión auxiliar

  $$\hat{u}_t = \alpha_1 + \alpha_2 X_{2t} + \ldots + \alpha_K X_{Kt} + \delta_1 \hat{u}_{t-1} + \ldots + \delta_p \hat{u}_{t-p} + \xi_t$$

  donde $t = p+1, \ldots, T$ y considerar el R^2 de esta regresión y el tamaño muestral disponible $(T-p)$. En este caso el estadístico del contraste y su distribución asintótica bajo la hipótesis nula son:

  $$(T-p)R^2 \xrightarrow{d,H_o} \chi^2_p \qquad \text{bajo } H_0$$

  siendo la regla de decisión igual que antes.

Comentarios:

- Este contraste es más general que el de Durbin-Watson. Las alternativas pueden ser $AR(p)$ con $p \geq 1$ y también $MA(p)$.
- A diferencia del contraste de Durbin-Watson, sí se puede utilizar en modelos que incluyen como variables explicativas la variable dependiente retardada $Y_{t-1}, Y_{t-2}, \ldots$ u otro tipo de regresor estocástico.
- Es un contraste asintótico esto es, bajo H_0 se conoce la distribución asintótica del estadístico. Por lo tanto, es válido para muestras grandes siendo cautos con los resultados para muestras pequeñas.
- La elección del número de retardos p es importante. Un p pequeño puede no capturar autocorrelaciones significativas de mayor orden, pero un p grande puede influir negativamente en la potencia del contraste.

3.2.3. Análisis de los residuos y contrastes: ejemplos

En esta sección vamos a presentar varios ejemplos que ilustran la utilización de los gráficos de residuos para detectar la existencia de autocorrelación y en algunos casos detectar una posible mala especificación del modelo. También ilustramos la utilización del contraste de Durbin-Watson no solamente para contrastar la existencia de autocorrelación de orden uno, sino también como indicador de mala especificación.

Ejemplo 3.1 Autocorrelación positiva

Se dispone del fichero de datos ***inv.dat*** del libro *Undergraduate Econometrics* de Hill, Griffiths y Judge disponibles en *http://www.wiley.com/college/econ/hill331848/students.html*. Son datos anuales desde 1974 a 2003 de las siguientes variables:

I: Inversión real en billones de dólares.

GNP: Producto Nacional Bruto real en billones de dólares.

R: Tipo de interés.

Se estima por MCO un modelo que relaciona la Inversión real con el Producto Nacional Bruto real y el tipo de interés incluyendo un término independiente.

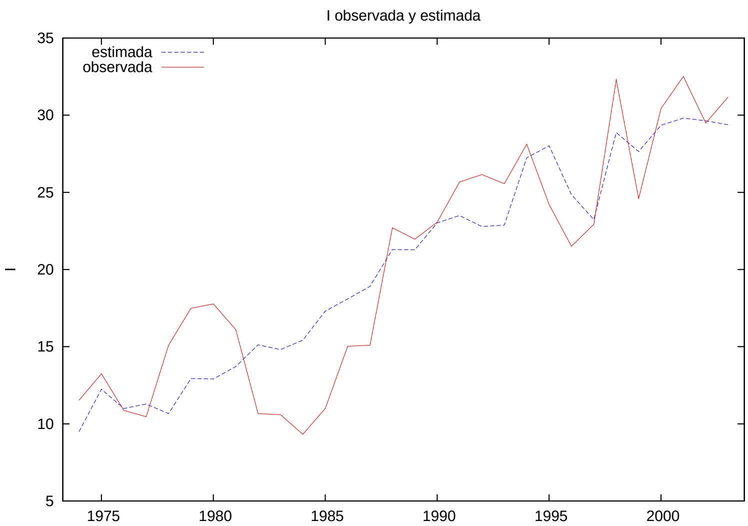
$$I_t = \beta_1 + \beta_2 GNP_t + \beta_3 R_t + u_t \qquad t = 1, \ldots, 30$$

y los resultados de la estimación son:

$$\underset{(\widehat{des}(\hat{\beta}_{MCO}))}{\widehat{I}_t} = \underset{(2{,}51089)}{6,22494} + \underset{(0{,}0717905)}{0,769911}\ GNP_t - \underset{(0{,}126416)}{0,184196}\ R_t$$

$$R^2 = 0,816282 \quad SCR = 299,336 \quad DW = 0,852153 \quad \hat{\rho} = 0,567726$$

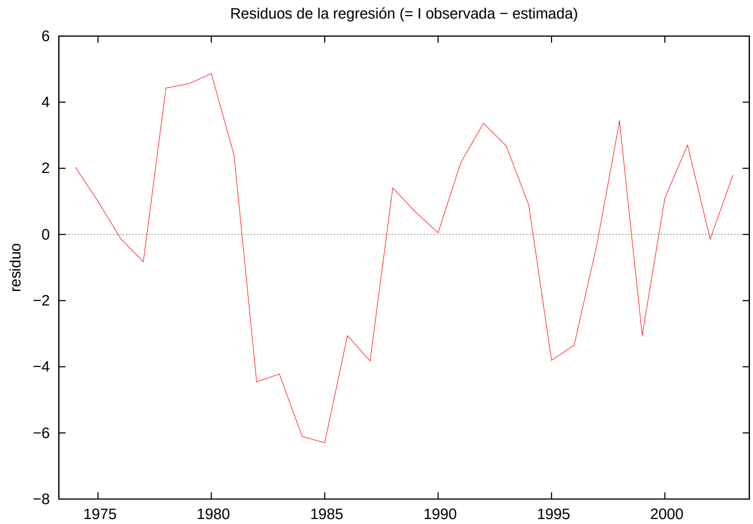
Figura 3.6: *Gráfico de la serie de Inversión observada y estimada*



La Figura 3.6 muestra el gráfico de las series de Inversión observada y estimada, que representa la evolución en el periodo muestral considerado, de los valores observados de la variable Inversión I_t y de la Inversión estimada $\hat{I}_t$ que predice el modelo ajustado dados los valores observados de las variables explicativas. El modelo estimado no recoge bien los ciclos de expansión y recesión observados en la Inversión, especialmente en los años 1977-1982, y 1982-1988, siendo los observados mucho más pronunciados que los generados por el modelo ajustado. Esto explica los ciclos observados en el gráfico de la serie temporal de los residuos. Este hecho sugiere que el término de perturbación del modelo pueda recoger factores correlacionados en el tiempo tal que presente autocorrelación positiva al menos hasta de orden uno, pudiendo ser también de mayor orden.

La Figura 3.7 muestra la evolución de los residuos $\hat{u}_t = I_t - \hat{I}_t$ a lo largo del periodo muestral considerado $t = 1974, \ldots, 2003$. Como es de esperar, ya que hay término constante en la regresión, oscilan alrededor de su media muestral igual a cero. Se puede observar que especialmente de 1973

Figura 3.7: *Gráfico de la serie temporal de los residuos MCO*



a 1995 se van alternando en ciclos, grupos de residuos seguidos del mismo signo, siendo menos evidente al final de la muestra, donde el ajuste es algo mejor.

Contraste de Durbin-Watson:

$$\begin{array}{llll} H_0: \rho = 0 & u_t \sim iid(0, \sigma_u^2) & & \text{No autocorrelación.} \\ H_a: 0 < \rho < 1 & u_t = \rho u_{t-1} + \varepsilon_t & \varepsilon_t \sim iid(0, \sigma_\varepsilon^2) & \text{AR(1) autocorrelación positiva.} \end{array}$$

Para $\alpha = 0,05$, $T = 30$, $K' = 2$ las cotas inferior y superior tabuladas respectivamente son $d_i = 1,2837$ y $d_s = 1,5666$. Como $DW = 0,852153 < d_i = 1,2837$, se rechaza H_0. Hay evidencia de autocorrelación de orden uno positiva en el término de perturbación del modelo .

Ejemplo 3.2 Autocorrelación negativa

Se dispone de observaciones anuales de las variables Consumo (C_t) y Renta (R_t) para un país. Los resultados de la estimación por el método de Mínimos Cuadrados Ordinarios de la función de consumo $C_t = \beta_1 + \beta_2 R_t + u_t$ se muestran a continuación:

Estimación MCO, usando las observaciones 1950-1963 ($T = 14$)
Variable dependiente: C

	Coeficiente	Desv. Típica	Estadístico t	Valor p
const	0,3470	1,1371	0,3052	0,7654
R	0,7001	0,0479	14,6101	0,0000

R^2	0,9467	Suma de cuad. residuos	30,6381
$\hat{\rho}$	−0,7323	Durbin–Watson	3,3158

Figura 3.8: *Gráfico de la serie de Inversión observada y estimada*

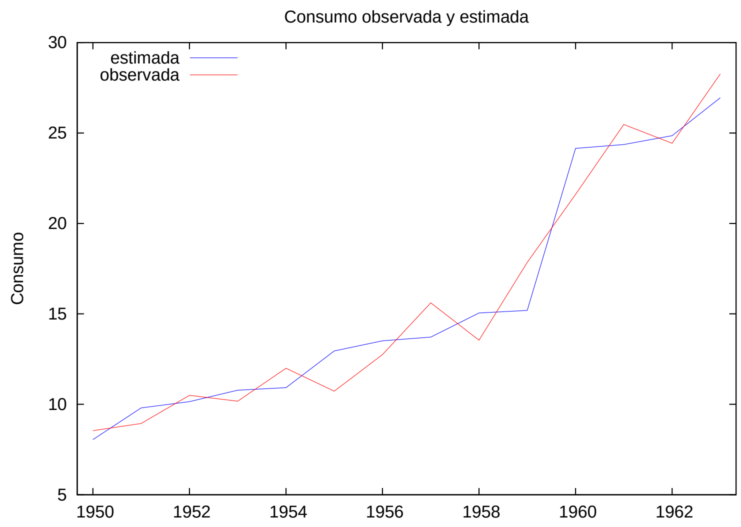
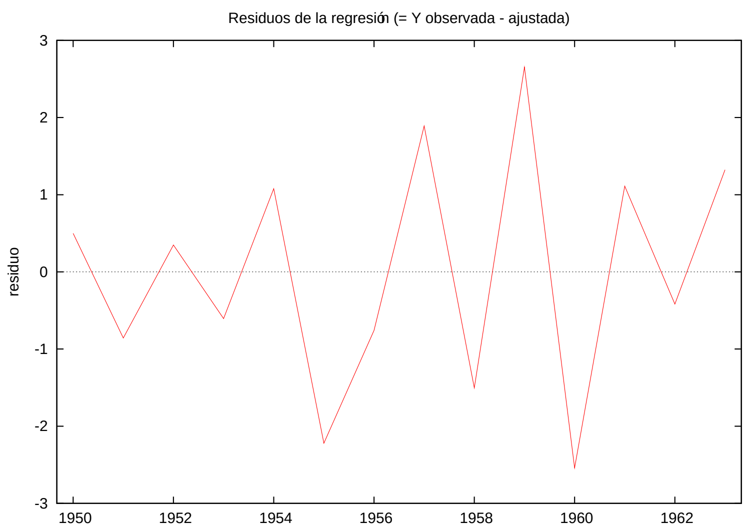


Figura 3.9: *Gráfico de la serie temporal de los residuos MCO*



La Figura 3.8 representa la evolución en el periodo muestral considerado, de los valores observados de la variable Consumo, C_t y del Consumo estimado, $\hat{C}_t$ que predice el modelo ajustado, dados los valores observados de las variables explicativas. El modelo ajustado parece predecir bien la tendencia creciente del consumo, incluso al final de la muestra donde esta tendencia parece ser más acusada. Pero lo que el modelo no predice bien es la dirección del ciclo, ya que cuando se da una expansión el modelo predice una recesión y cuando se observa una recesión el modelo predice una expansión. Esto hace que en la serie temporal de los residuos de la Figura 3.9 se observe una alternacia de signos de positivo a negativo y viceversa a lo largo de todo el periodo. Esto sugiere que puede existir autocorrelación de orden uno negativa.

Contraste de Durbin-Watson:

$$\begin{array}{llll} H_0: \rho = 0 & u_t \sim iid(0, \sigma_u^2) & & \text{No autocorrelación.} \\ H_a: -1 < \rho < 0 & u_t = \rho u_{t-1} + \varepsilon_t & \varepsilon_t \sim iid(0, \sigma_\varepsilon^2) & \text{AR(1) autocorrelación negativa.} \end{array}$$

Para $\alpha = 0,05$, $T = 14$, $K' = 1$ las cotas inferior y superior tabuladas son respectivamente, $d_i = 1,0450$ y $d_s = 1,3503$. Como $DW = 3,315808 > 4 - d_i = 2,955$ se rechaza H_0. Hay evidencia de autocorrelación de orden uno negativa en el término de perturbación del modelo.

Los siguientes ejemplos nos servirán para ilustrar que el contraste de Durbin-Watson puede detectar autocorrelación, pero ser una consecuencia de una mala especificación funcional del modelo, o a cambios estructurales que no se han recogido. En estos casos hay que proceder a reespecificar el modelo.

Ejemplo 3.3 Mala especificación. Especificación lineal versus cuadrática

Consideremos los siguientes datos anuales desde 1990 a 2001 observados para las variables Y y X:

Y	6	3	1	1	1	4	6	16	25	36	49	64
X	-4	-3	-2	-1	1	2	3	4	5	6	7	8

Se propone inicialmente la estimación de una relación lineal entre estas dos variables:

$$Y_t = \beta_1 + \beta_2 X_t + u_t \tag{3.5}$$

Estimación MCO, usando las observaciones 1990-2001 ($T = 12$)
Variable dependiente: Y

	Coeficiente	Desv. Típica	Estadístico t	Valor p
const	8,01220	4,05896	1,9740	0,0766
X	4,45591	0,919174	4,8477	0,0007

Suma de cuad. residuos	1501,071	R^2 corregido	0,671647
$\hat{\rho}$	0,853150	Durbin–Watson	0,321828

El gráfico de la serie observada y ajustada, Figura 3.10, sugiere que el ajuste de una relación lineal no parece ser la más indicada, ya que la serie observada Y presenta una curvatura que no captura el ajuste del modelo. Si analizamos el gráfico de residuos, Figura 3.11, vemos que la serie presenta forma de U, lo que confirma que es mejor una relación cuadrática. La forma que presenta la serie de residuos hace que estos se concentren en grupos de residuos del mismo signo. Esto puede sugerir que el término de error del modelo presenta autocorrelación de orden uno positiva.

El valor del estadístico de Durbin-Watson es $DW = 0,321828$. Las cotas inferior d_i y superior d_s tabuladas para $T = 12$ y $K' = 1$, al nivel de significación del $0,05$ son $d_i = 0,971$ y $d_s = 1,331$. Como el valor del estadístico DW es menor que la cota inferior d_i se rechaza la hipótesis nula

Figura 3.10: *Gráfico de la serie observada y ajustada con especificación lineal*

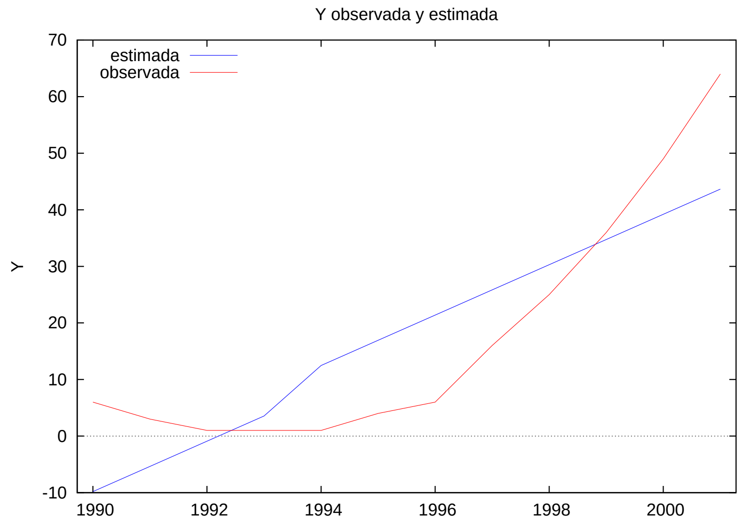
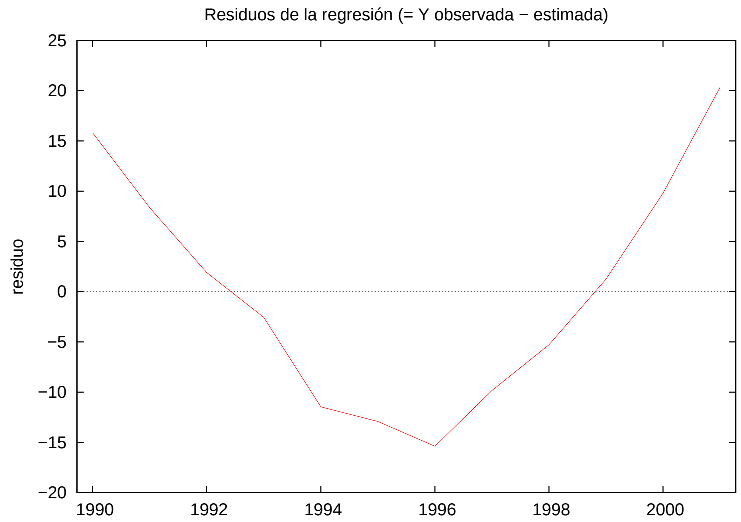


Figura 3.11: *Gráfico de residuos MCO de la especificación lineal*



de no autocorrelación frente a la alternativa de autocorrelación positiva. El contraste parece estar capturando un problema de mala especificación funcional.

Seguidamente se reespecifica el modelo y se estima por MCO una relación cuadrática entre Y_t y X_t introduciendo como nuevo regresor X_t^2.

Estimacion MCO de la especificación cuadrática.

$$Y_t = \beta_1 + \beta_2 X_t + \beta_3 X_t^2 + v_t \tag{3.6}$$

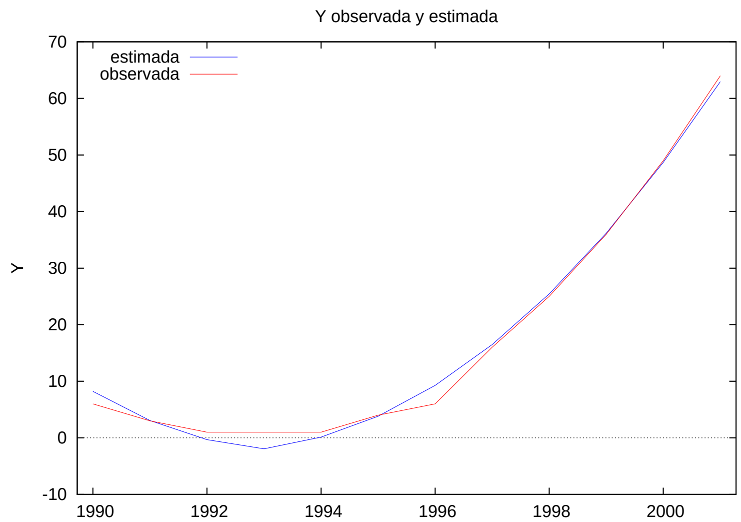
Los resultados, denotando $XSQ = X_t^2$, son los siguientes:

Estimación MCO, usando las observaciones 1990–2001 ($T = 12$)
Variable dependiente: Y

	Coeficiente	Desv. Típica	Estadístico t	Valor p
const	−1,7816	0,7408	−2,4050	0,0396
X	1,0334	0,2064	5,0058	0,0007
XSQ	0,8825	0,0407	21,6641	0,0000

Suma de cuad. residuos	28,2431	R^2 corregido	0,9931
$\hat{\rho}$	0,2984	Durbin–Watson	1,2157

Figura 3.12: *Gráfico de la serie observada y ajustada con especificación cuadrática*

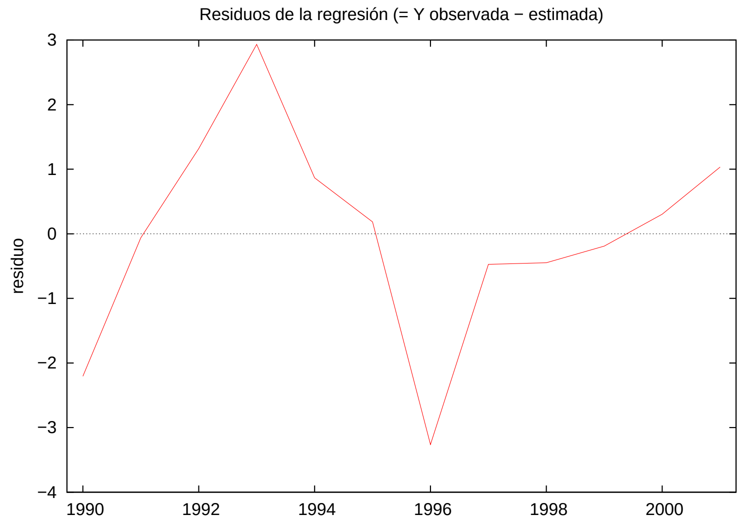


La Figura 3.12 de la relación observada y ajustada muestra claramente que esta especificación se ajusta mucho mejor a los datos. El gráfico de residuos, Figura 3.13, presenta una asociación de residuos positivos y negativos mucho menor que antes. También la magnitud de los residuos, en valor absoluto, es mucho menor.

Por otro lado, el valor del estadístico DW es mucho más cercano a 2 y el valor $\hat{\rho}$ más cercano a cero, por lo que hay una menor evidencia de autocorrelación en el término de perturbación. Mirando a las cotas inferior y superior tabuladas ahora para $T = 12$ y $K' = 2$ son $d_i = 0,812$ y $d_s = 1,579$. El valor muestral del estadístico se encuentra en la zona de indeterminación, $DW = 1,216 \in (0,812; 1,579)$ por lo que el contraste no es concluyente.

Dado que el tamaño muestral es pequeño, la zona de indeterminación del contrastes es bastante amplia y hay que hacer notar que el valor muestral de DW está más cerca de la región de aceptación de la hipótesis nula (hacia cota superior) que de la de rechazo (hacia cota inferior). Aún así, es bastante evidente que el problema de autocorrelación de orden uno positiva detectado anteriormente se debía a una mala especificación funcional o de omisión de la variable X_t^2. De todas formas, pudiera ser que quede aún cierto grado de autocorrelación en el término de error del modelo reespecificado, aunque tampoco podemos concluir si este es el caso. En cuanto a los resultados de la estimación

Figura 3.13: *Gráfico de residuos MCO de la especificación cuadrática*



del modelo (3.6), la variable X_t^2 es significativa al 5 %, ya que el valor p asociado a su coeficiente es menor que $0,05$. Esto confirma que es más adecuada una relación cuadrática que lineal.

Ejemplo 3.4 Mala especificación. Cambio estructural no recogido

Consideramos los siguientes datos sobre salarios (W) y horas trabajadas (H) en un país durante el periodo 2001-2008. La variable ficticia D recoge un cambio en la política de empleo a partir del año 2005.

Año	2001	2002	2003	2004	2005	2006	2007	2008
W	170	180	165	165	105	95	100	90
H	40	50	30	40	50	35	40	35
D	0	0	0	0	1	1	1	1

En primer lugar, se estima el siguiente modelo para los salarios en función de las horas trabajadas, donde no se tiene en cuenta el cambio en la política de empleo.

$$W_t = \beta_1 + \beta_2 H_t + u_t \tag{3.7}$$

Estimación MCO, usando las observaciones 1–8
Variable dependiente: W

	Coeficiente	Desv. Típica	Estadístico t	Valor p
const	102,3210	90,9652	1,1248	0,3036
H	0,7857	2,2436	0,3502	0,7382

Suma de cuad. residuos	10571,43	R^2 corregido	−0,143299
$F(1,6)$	0,122635	Valor p (de F)	0,738160
$\hat{\rho}$	0,752888	Durbin–Watson	0,445789

Figura 3.14: *Gráfico de la serie observada y ajustada con el Modelo (3.7)*

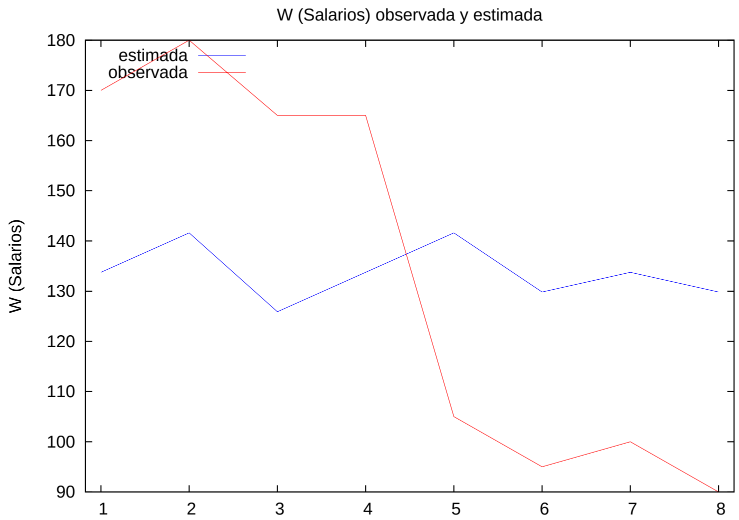
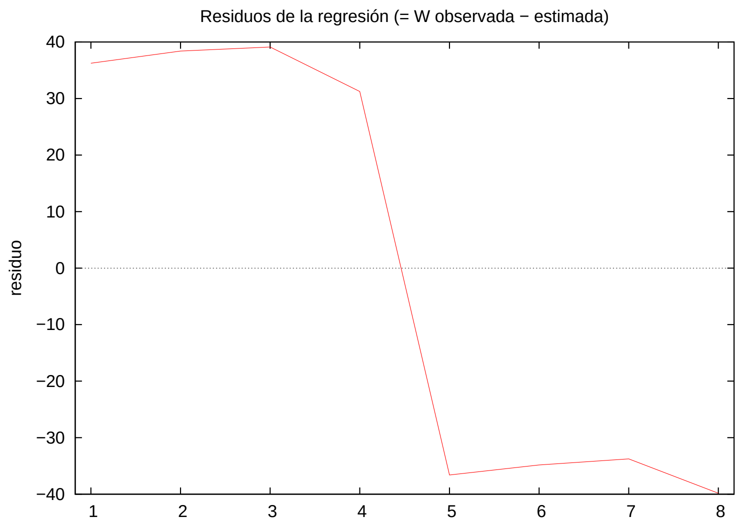


Figura 3.15: *Gráfico de residuos MCO del Modelo (3.7)*



A la vista del gráfico de la serie observada y ajustada, Figura 3.14, se sospecha que a partir del año 2005 con el cambio en la política de empleo, el nivel medio de los salarios bajó de forma significativa. Este cambio estructural en los salarios no se ha recogido en la especificación, por lo que la serie ajustada de salarios se mantiene mucho más estable que la serie observada. En cuanto al gráfico de residuos, Figura 3.15, se observa un primer grupo de valores positivos, correspondientes a los años previos al cambio y posteriormente un segundo grupo de valores negativos. En consecuencia, se detecta autocorrelación de orden uno positiva, ya que el valor del estadístico de Durbin-Watson es menor que la cota inferior d_i tabulada para el tamaño muestral $T = 8$ y $K' = 1$ esto es, $DW = 0,445789 < d_i = 0,7629$. El valor del coeficiente de autocorrelación de orden uno en los residuos $\hat{\rho} = 0,752888$ también indica autocorrelación positiva.

Tenemos que incluir en el modelo una variable ficticia que recoja ese cambio en el nivel de los salarios mínimos a partir del año 2005. Para ello, consideramos estimar el siguiente modelo ampliado donde además de las horas trabajadas incluimos como regresor la variable ficticia D, para recoger ese cambio estructural:

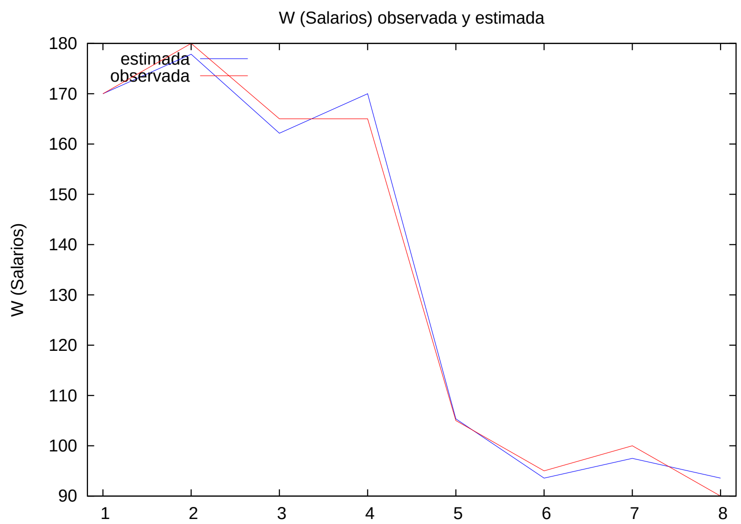
$$W_t = \beta_1 + \beta_2 H_t + \beta_3 D_t + u_t \tag{3.8}$$

Estimación MCO, usando las observaciones 1–8
Variable dependiente: W

	Coeficiente	Desv. Típica	Estadístico t	Valor p
const	138,5710	7,5381	18,3826	0,0000
H	0,7857	0,1835	4,2817	0,0079
D	−72,5000	2,4275	−29,8659	0,0000

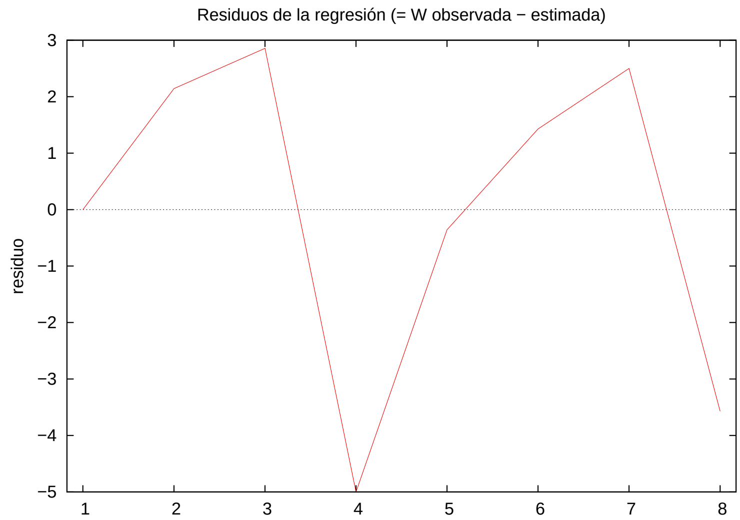
Suma de cuad. residuos	58,92857	R^2 corregido	0,992352
$\hat{\rho}$	−0,265193	Durbin–Watson	2,199134

Figura 3.16: *Gráfico de la serie observada y ajustada con el Modelo (3.8)*



A la luz de los resultados de estimación y de los gráficos de la serie observada y ajustada y de residuos, Figuras 3.16 y 3.17 respectivamente, es claro que la variable ficticia ha recogido el cambio estructural en el nivel medio de los salarios a partir de 2005. El coeficiente que acompaña a la variable ficticia D es significativamente distinto de cero. El ajuste ha mejorado sustancialmente, tanto en términos del R^2 corregido que ha aumentado considerablemente, como gráficamente. El estadístico de Durbin-Watson toma un valor $2,199$ que pertenece a la zona de no rechazo de la hipótesis nula de no autocorrelación ($d_s = 1,7771; 4 - d_s = 2,2229$) dado el nivel de significación $5\,\%$, y $T = 8$, $K' = 2$.Por lo tanto, se confirma que la evidencia de autocorrelación en el modelo (3.7) se debía a una mala especificación. Una vez recogido el cambio estructural en la serie de salarios, incluyendo una variable ficticia como variable explicativa adicional a las horas trabajadas, esta evidencia desaparece.

Figura 3.17: *Gráfico de residuos MCO del Modelo (3.8)*



Ejemplo 3.5 Aplicación de los contrastes de autocorrelación

Utilizamos los datos de la Tabla 6.3 del libro de Greene (1998) *Análisis Econométrico* que están en el Apéndice: Tablas de datos. Son datos de series temporales con frecuencia anual para la economía americana de los años 1968 a 1982 de las siguientes series:

- INVER: Inversión en términos nominales
- PNB: Producto Nacional Bruto en términos nominales
- IPC: Índice de Precios al Consumo
- TI: Tipo de Interés, promedio anual de la tasa de descuento de la Reserva Federal de NY.

Con estos datos se construyen las series de Inversión y Producto Nacional Bruto en términos reales dividiendo las series nominales por el IPC con año base en 1972 y multiplicando por 10^{-1} con lo cual las series están medidas en trillones de dólares. La Tasa de Inflación se ha calculado como el porcentaje de variación del IPC y el Tipo de Interés es el promedio anual de la tasa de descuento del Banco de la Reserva Federal de Nueva York. El Tipo de Interés Real se obtiene como la diferencia entre el Tipo de Interés Nominal y la Tasa de Inflación.

- $INVERR_t = (INVER_t/IPC_t) \times 0,1$
- $PNBR_t = (PNB_t/IPC_t) \times 0,1$
- $INF_t = ((IPC_t - IPC_{t-1})/IPC_{t-1}) \times 100$
- $TIREAL_t = TI_t - INF_t$

Partimos de una primera especificación de la función de inversión incluyendo solamente el tipo de interés real. Los resultados obtenidos son los siguientes:

Modelo sin PNBR como regresor:

$$\underset{\text{(t-estad)}}{\widehat{\text{INVERR}}_t} = \underset{(21{,}44)}{0{,}198273} + \underset{(1{,}42)}{0{,}006442}\,\text{TIREAL}_t$$

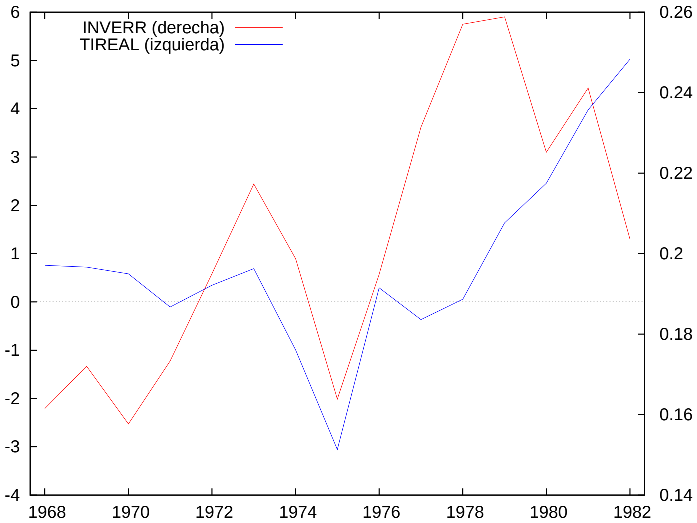
$$T = 15 \quad \bar{R}^2 = 0{,}068 \quad F(1,13) = 2{,}018 \quad DW = 0{,}556216 \quad \hat{\rho} = 0{,}669145$$

La variable Tipo de Interés Real, TIREAL, explica muy poco de la Inversión Real, INVERR. El coeficiente de determinación corregido es muy bajo, la variable TIREAL no es significativa al 5 % ya que el valor muestral del estadístico t, en valor absoluto, es menor que el valor crítico 2, 16037. A la vista de los resultados del ajuste del modelo sin PNBR como regresor, la especificación propuesta no es muy adecuada.

En la Figura 3.18 se muestra la evolución en el tiempo de la Inversión Real y el Tipo de Interés Real, en el periodo muestral considerado. La serie de Tipo de Interés Real es muy estable en los diez primeros años a excepción del periodo 1973-1976 donde hay una bajada de tipos de interés seguida de una recuperación al mismo nivel anterior. Solamente hay una moderada tendencia creciente a partir de 1978.

La Inversión Real en cambio, presenta una tendencia creciente que se mantiene a lo largo de toda la muestra. Alrededor de ella se muestran ciclos cada vez más pronunciados de expansión y recesión siendo la duración del ciclo más pronunciado de expansión aproximadamente tres años y el de las recesiones sobre dos años.

Figura 3.18: *Evolución temporal de las variables INVERR y TIREAL*



En la Figura 3.19 el gráfico de la serie observada y ajustada por el modelo a la serie INVERR indica que el modelo no captura nada bien el comportamiento de la Inversión Real. La serie ajustada no recoge la tendencia creciente en la serie ni tampoco los ciclos.

El gráfico de la serie de residuos, Figura 3.20, es casi igual a la serie de la INVERR observada, y dado que no se ha recogido ni siquiera la tendencia creciente, los residuos asociados a la primera mitad de la muestra son negativos y el resto son todos positivos. Esta agrupación de residuos del mismo signo seguidos hace pensar que se pueda detectar un problema de autocorrelación de orden uno positiva, pero el problema es claramente de una mala especificación del modelo.

Figura 3.19: *Serie INVERR observada y ajustada: Modelo sin PNBR*

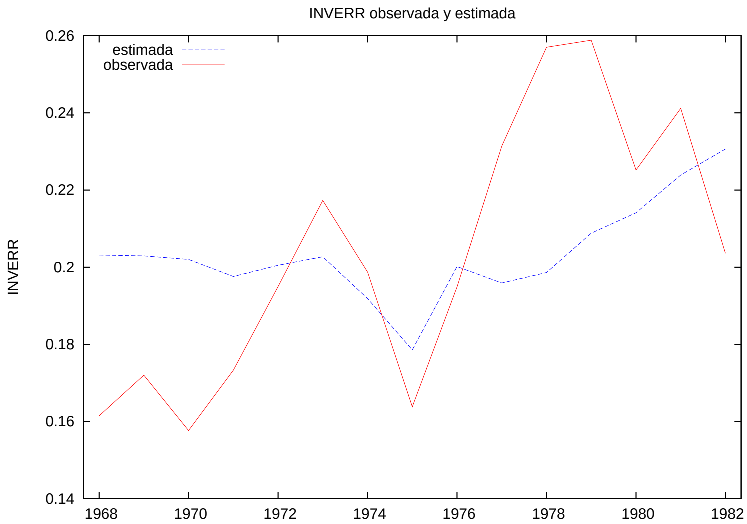
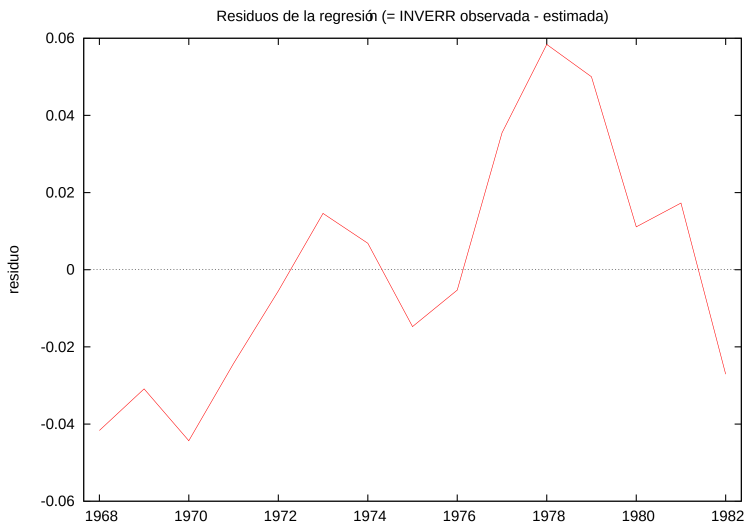


Figura 3.20: *Serie de residuos MCO: Modelo sin PNBR*



Contraste de Durbin-Watson:

$$\begin{array}{llll} H_0: \rho = 0 & u_t \sim iid(0, \sigma_u^2) & & \text{No autocorrelación.} \\ H_a: 0 < \rho < 1 & u_t = \rho u_{t-1} + \varepsilon_t & \varepsilon_t \sim iid(0, \sigma_\varepsilon^2) & \text{AR(1) autocorrelación positiva.} \end{array}$$

Para $\alpha = 0,05$, el tamaño muestral $T = 15$ y el número de regresores exceptuando el término constante $K' = 1$, las cotas tabuladas inferior y superior respectivamente son $d_i = 1,077$ y $d_s = 1,361$. Como $DW = 0,556216 < d_i = 1,077$, se rechaza H_0 al nivel de significación del 5 %. Hay evidencia de autocorrelación de orden uno positiva.

Contraste Breusch-Godfrey:

a) Autocorrelación hasta de orden uno.

$$H_0: \rho = 0 \quad \text{ó} \quad \theta = 0 \qquad u_t \sim iid(0, \sigma_u^2) \quad \text{No autocorrelación.}$$

$$\begin{array}{llll} H_a: \rho \neq 0 & u_t = \rho u_{t-1} + \varepsilon_t & \varepsilon_t \sim iid(0, \sigma_\varepsilon^2) & \text{AR(1)} \quad \text{ó} \\ \quad\;\; \theta \neq 0 & u_t = \theta \varepsilon_{t-1} + \varepsilon_t & \varepsilon_t \sim iid(0, \sigma_\varepsilon^2) & \text{MA(1)} \end{array}$$

Realizamos la regresión auxiliar correspondiente y calculamos el R^2 para obtener el estadístico. Consideramos $t = 2, \ldots, 15$ esto es, 14 observaciones.

$$\begin{array}{rcccc} \widehat{\hat{u}_t} = & 0{,}004 & - \; 0{,}003 \; \text{TIREAL}_t & + \; 0{,}724 \, \hat{u}_{t-1} \\ \text{t-estad.} & (0{,}63) & (-0{,}99) & (3{,}49) \end{array}$$

El coeficiente de determinación de la regresión auxiliar es $R^2 = 0{,}524809$ por lo que el valor muestral del estadístico es $BG = 14(0,524809) = 7,347326$. Dado que el valor muestral del estadístico es mayor que el valor crítico $\chi^2_{1|0,05} = 3,84$, se rechaza H_0 frente a H_a al $0,05$ de significación. Hay evidencia de autocorrelación de orden uno. Veamos que ocurre si consideramos una alternativa más general.

b) Autocorrelación hasta de orden tres.

$$H_0: \rho_1 = \rho_2 = \rho_3 = 0 \quad \text{ó} \quad \theta_1 = \theta_2 = \theta_3 = 0 \quad u_t \sim iid(0, \sigma_u^2) \qquad \text{No autocorrelación.}$$

$$\begin{array}{lll} H_a: \rho_i \neq 0 \text{ para algún } i & u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \rho_3 u_{t-3} + \varepsilon_t & \text{AR(3)} \quad \text{ó} \\ \quad\;\; \theta_i \neq 0 \text{ para algún } i & u_t = \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \theta_3 \varepsilon_{t-3} + \varepsilon_t & \text{MA(3)} \end{array}$$

donde $\varepsilon_t \sim iid(0, \sigma_\varepsilon^2)$. Realizamos la regresión auxiliar correspondiente y calculamos el R^2 para obtener el estadístico. Consideramos $t = 4, \ldots, 15$ esto es, 12 observaciones.

$$\begin{array}{rccccc} \widehat{\hat{u}_t} = & 0{,}005 & - \; 0{,}001 \; \text{TIREAL}_t & + 0{,}838 \, \hat{u}_{t-1} & - \; 0{,}308 \; \hat{u}_{t-2} & - \; 0{,}069 \; \hat{u}_{t-3} \\ \text{t-estad.} & (0{,}58) & (-0{,}22) & (2{,}05) & (-0{,}56) & (-0{,}15) \end{array}$$

El coeficiente de determinación de la regresión auxiliar es $R^2 = 0,468251$ por lo que el valor muestral del estadístico es $BG = 12(0,468251) = 5,619012$. Dado que el valor muestral del estadístico es menor que el valor crítico $\chi^2_{3|0,05} = 7,81$, no se rechaza H_0 frente a H_a al $0,05$ de significación. No hay evidencia de autocorrelación hasta de orden tres. Vemos entonces que el incluir más retardos hace que el contraste pierda potencia si estos no son necesarios. De hecho en este caso no parece detectarse la autocorrelación de orden uno que sí era significativa cuando se considera solamente ese retardo.

Los contrastes de autocorrelación sugieren la existencia de autocorrelación positiva de orden uno. Como hemos observado antes, esta evidencia de autocorrelación de orden uno positiva se debe a un pobre ajuste de este modelo. Esto puede estar capturando la omisión de alguna variable relevante que debería de ser incluida para explicar mejor la inversión real. De hecho se debe de reespecificar el modelo incluyendo al menos una variable relevante para explicar la Inversión Real como es el Producto Nacional Bruto en términos reales PNBR.

Modelo con PNBR como regresor:

$$\begin{array}{cccccc} \widehat{\text{INVERR}}_t = & -0{,}048683 & - & 0{,}002126\,\text{TIREAL}_t & + & 0{,}197170\,\text{PNBR}_t \\ \text{t-estad.} & (-1{,}08) & & (-0{,}72) & & (5{,}51) \end{array}$$

$$T = 15 \qquad \bar{R}^2 = 0{,}714 \qquad F(2,12) = 18{,}496 \qquad DW = 1{,}374911$$

El ajuste ha mejorado sustancialmente en términos del coeficiente de determinación corregido. El valor del estadístico $F(2,12) = 18,496$ para el contraste de significación conjunta es mayor que el valor crítico al $0,05$ de significación $(3,89)$. Luego conjuntamente las variables explicativas TIREAL y PNBR parecen ser relevantes para explicar la Inversión Real, aunque a nivel de significatividad individual, los estadísticos t muestran que solamente es significativa la variable PNBR, no siendo TIREAL individualmente significativa (el valor crítico al 5 % en las tablas de la t-Student con 12 grados de libertad es $2,179$). Por lo tanto, esta especificación del modelo donde además del Tipo de Interés Real se ha incluido el Producto Nacional Bruto en términos reales, parece más adecuada.

La Figura 3.21 muestra el gráfico de la serie INVERR observada y la ajustada por el modelo, indica que esta última especificación captura mejor el comportamiento de la Inversión Real. La serie ajustada recoge la tendencia creciente en la serie, pero no se ajusta del todo bien a los ciclos.

El gráfico de la serie de residuos, Figura 3.22, ya no presenta esa asociación de residuos del mismo signo seguidos, pero muestra los ciclos de expansión y recesión que se daban en la serie de Inversión Real, siendo el ciclo más pronunciado de expansión sobre tres años y las recesiones sobre dos años. Esto sugiere que pueda no ser significativa la autocorrelación de primer orden positiva, pero que pueda existir evidencia de autocorrelación significativa de orden dos o tres.

Contraste de Durbin-Watson:

$$\begin{array}{llll} H_0: \rho = 0 & u_t \sim iid(0, \sigma_u^2) & & \text{No autocorrelación.} \\ H_a: 0 < \rho < 1 & u_t = \rho u_{t-1} + \varepsilon_t & \varepsilon_t \sim iid(0, \sigma_\varepsilon^2) & \text{AR(1) autocorrelación positiva.} \end{array}$$

Para $\alpha = 0{,}05$, $T = 15$, $K' = 2$ las cotas tabuladas son $d_i = 0,946$ y $d_s = 1,543$.

Como el valor muestral del estadístico está entre los valores inferior y superior, $d_i = 0,946 < DW = 1,347911 < d_s = 1,543$, que es la zona de indeterminación del contraste, el contraste no es concluyente. El valor del estadístico esta más cerca de la cota superior o lo que es lo mismo, más cerca de la zona de aceptación de la hipótesis nula.

Figura 3.21: *Serie INVERR observada y ajustada: Modelo con PNBR*

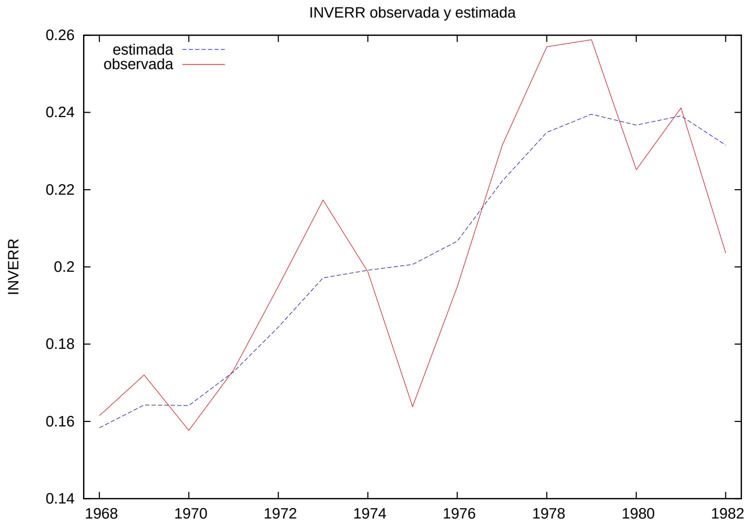
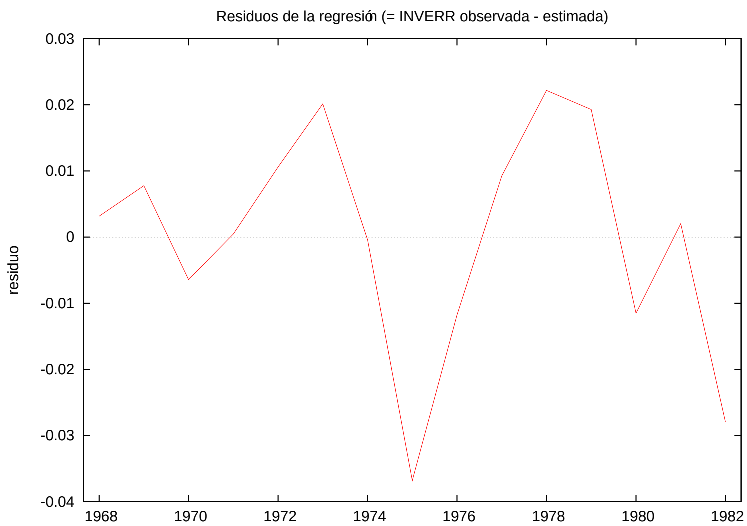


Figura 3.22: *Serie de residuos MCO: Modelo con PNBR*



Contraste Breusch-Godfrey:

a) Autocorrelación hasta de orden uno:

$$H_0: \rho = 0 \quad \text{ó} \quad \theta = 0 \qquad u_t \sim iid(0, \sigma_u^2) \quad \text{No autocorrelación.}$$

$$\begin{array}{llll} H_a: \rho \neq 0 & u_t = \rho u_{t-1} + \varepsilon_t & \varepsilon_t \sim iid(0, \sigma_\varepsilon^2) & \text{AR(1)} \quad \text{ó} \\ \quad\; \theta \neq 0 & u_t = \theta \varepsilon_{t-1} + \varepsilon_t & \varepsilon_t \sim iid(0, \sigma_\varepsilon^2) & \text{MA(1)} \end{array}$$

Realizamos la regresión auxiliar correspondiente a este modelo y esta alternativa. Calculamos el R^2 para obtener el estadístico. Consideramos $t = 2, \ldots, 15$ esto es, 14 observaciones.

$$\widehat{\hat{u}_t} = \underset{(-0{,}029)}{-\,0,001} \; + \underset{(0{,}016)}{0,0007}\,\text{PNBR} - \underset{(-0{,}034)}{0,0001}\,\text{TIREAL} + \underset{(0{,}785)}{0,2712}\,\hat{u}_{t-1}$$

El coeficiente de determinación de la regresión auxiliar es $R^2 = 0,058756$ por lo que el valor muestral del estadístico es $BG = 14\,(0,058756) = 0,822584$. Dado que el valor muestral del estadístico es menor que el valor crítico $\chi^2_{1|0,05} = 3,84$, no se rechaza H_0 frente a H_a al 5 % de significación. No hay evidencia de autocorrelación hasta de orden uno.

Los contrastes de autocorrelación sugieren que no hay problemas de autocorrelación, aunque el de Durbin-Watson no es concluyente. La cuestión es que en estos contrastes las alternativa a contrastar es un proceso con autocorrelación de orden uno. Pudiera ser que la autocorrelación de primer orden no sea muy significativa, pero debido a los ciclos de dos y tres años exista autocorrelación de orden tres. Por ello, hacemos el contraste de Breusch-Godfrey para una alternativa de mayor orden, por ejemplo de orden 3:

b) Autocorrelación hasta de orden 3:

$$H_0: \rho_1 = \rho_2 = \rho_3 = 0 \quad \text{ó} \quad \theta_1 = \theta_2 = \theta_3 = 0 \qquad u_t \sim iid(0, \sigma_u^2) \qquad \text{No autocorrelación.}$$

$$\begin{array}{lll} H_a: \rho_i \neq 0 \text{ para algún } i & u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \rho_3 u_{t-3} + \varepsilon_t & \text{AR(3)} \;\; \text{ó} \\ \quad\; \theta_i \neq 0 \text{ para algún } i & u_t = \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \theta_3 \varepsilon_{t-3} + \varepsilon_t & \text{MA(3)} \end{array}$$

donde $\varepsilon_t \sim iid(0, \sigma_\varepsilon^2)$. Realizamos la regresión auxiliar correspondiente y calculamos el R^2 para obtener el estadístico. Consideramos $t = 4, \ldots, 15$ esto es, 12 observaciones.

$$\begin{aligned} \widehat{\hat{u}_t} = {} & \underset{(0{,}929)}{0,0674} - \underset{(-0{,}918)}{0,051}\;\text{PNBR} + \underset{(1{,}321)}{0,005}\,\text{TIREAL} - \underset{(-0{,}490)}{0,198}\;\hat{u}_{t-1} \\ & - \underset{(-0{,}032)}{0,011}\;\hat{u}_{t-2} - \underset{(-2{,}364)}{1,106}\;\hat{u}_{t-3} \end{aligned}$$

El coeficiente de determinación de la regresión auxiliar es $R^2 = 0,599334$ por lo que el valor muestral del estadístico es $BG = 12\,(0,599334) = 7,192$. Dado que el valor muestral del estadístico es menor que el valor crítico $\chi^2_{3|0,05} = 7,81$, no se rechaza H_0 frente a H_a al 5 % de significación. No hay evidencia de autocorrelación hasta de orden tres.

¿Qué ocurre si consideramos en el contraste todas las observaciones? gretl realiza de esa forma el contraste de Breusch-Godfrey, donde se igualan a cero el valor de los residuos no disponibles

$\hat{u}_0, \hat{u}_{-1}, \hat{u}_{-2}$. En la ventana de estimación por MCO del modelo de interés, eligiendo Contrastes → Autocorrelación se obtiene:

$$\widehat{\hat{u}_t} = \underset{(0,602)}{0,021} - \underset{(-0,579)}{0,016}\ \text{PNBR} + \underset{(1,251)}{0,003}\,\text{TIREAL} - \underset{(-0,591)}{0,209}\ \hat{u}_{t-1}$$
$$- \underset{(-0,099)}{0,031}\ \hat{u}_{t-2} - \underset{(-2,587)}{1,0103}\,\hat{u}_{t-3}$$

El coeficiente de determinación de esta regresión auxiliar es $R^2 = 0,551856$ por lo que el valor muestral del estadístico es $BG = 15\,(0,551856) = 8,277847$. Dado que el valor muestral del estadístico es mayor que el valor crítico $\chi^2_{3|0,05} = 7,81$, en este caso se rechaza H_0 frente a H_a al 5 % de significación. Habría cierta evidencia de autocorrelación de orden tres, aunque no de orden uno.

Por lo tanto, el problema detectado de autocorrelación de orden uno en el modelo donde no se incluye el PNBR parece haberse resuelto incluyendo esa variable que resulta ser relevante en media para explicar la Inversión real. Todavía se puede considerar cierta evidencia marginal de autocorrelación de orden tres, quizá debido a un componente cíclico residual no explicado por el modelo en la serie de Inversión Real. De todas formas, el número de observaciones es algo limitado y el contraste parece verse afectado por este hecho.

3.3. Consecuencias de la detección de autocorrelación

En los ejemplos anteriores hemos visto que a veces se detecta autocorrelación debida a una mala especificación del modelo. En ese caso, el problema es de omisión de variables relevantes más que de pura autocorrelación. Si las variables omitidas están correlacionadas con las incluidas, entonces habrá sesgos en la estimación de los coeficientes de interés por MCO. Así mismo la inferencia no será válida. En ese caso hay que intentar volver a especificar el modelo incluyendo aquellos factores que son relevantes y que hemos omitido en primera instancia.

Por otro lado, si el problema es puramente de autocorrelación es decir, existen factores no relevantes en media que se recogen en el término de perturbación y presentan correlación entre sí a lo largo del tiempo, las consecuencias son otras:

- Al ser factores no relevantes en media esto es, $E(u) = 0$, si los regresores son no estocásticos, el estimador MCO de los coeficientes β será lineal e insesgado.
- La matriz de varianzas y covarianzas del vector de perturbaciones no esféricas, siga el término de perturbación un proceso AR(p) o MA(q) o una combinación de los dos, es tal que $E(uu') \neq \sigma^2 I$. Esto implica que el estimador MCO ya no es el de menor varianza dentro de la clase de estimadores lineales e insesgados. Por lo tanto, no es el estimador eficiente.
- En muestras grandes, en general el estimador MCO es consistente, pero seguirá siendo no eficiente asintóticamente.
- El estimador usual utilizado para estimar $V(\hat{\beta}_{MCO}) = (X'X)^{-1}X'\Sigma X(X'X)^{-1}$ dado por la expresión $\widehat{V}(\hat{\beta}_{MCO}) = \hat{\sigma}^2(X'X)^{-1}$ es un estimador sesgado e inconsistente si $E(uu') \neq \sigma^2 I$.

Por lo tanto, la inferencia utilizando este estimador de $V(\hat{\beta}_{MCO})$ no es válida ni siquiera en muestras grandes.

El estimador de β lineal, insesgado y eficiente en este caso es el obtenido por el método de MCG. Como ya vimos en un tema anterior, el método de estimación de Mínimos Cuadrados Generalizados incorpora en el criterio de estimación la información adicional en la matriz de varianzas y covarianzas de las perturbaciones, $E(uu') = \sigma^2\Omega$. La función objetivo a minimizar viene dada por

$$Min_{\hat{\beta}}\, \hat{u}'\, \Omega^{-1}\, \hat{u} = Min_{\hat{\beta}}\, (Y - X\hat{\beta})'\, \Omega^{-1}\, (Y - X\hat{\beta}) \tag{3.9}$$

donde Ω es conocida y σ^2 es un factor de escala al que es invariante el estimador. Para el caso particular de que el término de perturbación siga un AR(1), si el valor de ρ es conocido, la función criterio a minimizar con respecto a β es la siguiente suma de cuadrados:

$$\begin{aligned} SCR(\hat{\beta}) &= (1-\rho^2)\,(Y_1 - \hat{\beta}_1 - \hat{\beta}_2 X_{21} - \ldots - \hat{\beta}_K X_{K1})^2 \\ &+ \sum_{t=2}^{T}\left[(Y_t - \rho Y_{t-1}) - \hat{\beta}_1(1-\rho) - \sum_{j=2}^{K}\hat{\beta}_j(X_{jt} - \rho X_{j,t-1})\right]^2 \end{aligned} \tag{3.10}$$

Minimizar esta función es equivalente a estimar por MCO el siguiente modelo transformado en el que se verifican las hipótesis básicas. Para la primera observación,

$$\sqrt{1-\rho^2}\, Y_1 = \beta_1 \sqrt{1-\rho^2} + \beta_2 \sqrt{1-\rho^2}\, X_{21} + \ldots + \beta_K \sqrt{1-\rho^2}\, X_{K1} + \varepsilon_1 \tag{3.11}$$

y para el resto de observaciones $t = 2, \ldots, T$ es,

$$Y_t - \rho Y_{t-1} = \beta_1\,(1-\rho) + \beta_2\,(X_{2t} - \rho X_{2t-1}) + \ldots + \beta_K\,(X_{Kt} - \rho X_{Kt-1}) + \varepsilon_t \tag{3.12}$$

donde $\varepsilon_t \sim IID(0,\, \sigma^2_\varepsilon)$. Por tanto, la primera observación sufre una transformación diferente a todas las demás.

Normalmente el valor de los parámetros de los que depende Ω no son conocidos. Por lo general se podrá disponer de un estimador previo de estos parámetros para obtener el estimador MCGF de β. Veamos como obtener este estimador en el caso de que sea un AR(1) el proceso que sigue el término de error del modelo de interés.

3.4. Estimación por MCGF bajo un AR(1)

Suponemos que el modelo de interés es

$$Y_t = \mathbf{X}_t'\beta + u_t \quad u_t = \rho u_{t-1} + \varepsilon_t \quad \varepsilon_t \sim iid(0, \sigma^2_\varepsilon)$$

siendo $\mathbf{X}_t' = [1\ X_{2t}\ \ldots\ X_{Kt}]$ y no se conoce el valor poblacional del parámetro ρ.

El modelo transformado tal que el error es un ruido blanco se puede escribir como:

$$Y_t - \rho Y_{t-1} = (\mathbf{X}_t' - \rho \mathbf{X}_{t-1}')\beta + \varepsilon_t \qquad t = 2, ..., T$$

donde $\mathbf{X}_t' = [1 \; X_{2t} \; \ldots \; X_{Kt}]$.

En general no se considera la primera observación del modelo transformado, porque el estimador así obtenido es asintóticamente equivalente a considerar todas las observaciones y computacionalmente es más sencillo de obtener.

3.4.1. Método de Hildreth y Lu: Red de búsqueda

Dado que $\rho \in (-1, 1)$, se realiza una partición de este intervalo y se estima por MCO el modelo transformado para los diferentes valores del parámetro ρ en esa partición. Se elige aquél valor de ρ y las estimaciones de β por MCO en el modelo transformado correspondientes, que proporcionan una menor suma de cuadrados residual en la estimación.

Los pasos de este método de estimación son:

- Se particiona el intervalo $(-1, 1)$ en una serie de valores equidistantes, $\hat{\rho}_i$, $i = 1, \ldots, S$, por ejemplo, $-0{,}95, -0{,}9, -0{,}85, \ldots, -0{,}05, 0, 0{,}05, \ldots, 0{,}95$. Esta sería la red de búsqueda sobre el parámetro ρ a la hora de minimizar la suma de cuadrados residual al estimar β por MCO en el modelo transformado.
- Para cada ρ_i en la red, se estima β por MCO en el modelo transformado
$$Y_t - \rho_i Y_{t-1} = (\mathbf{X}_t' - \rho_i \mathbf{X}_{t-1}')\beta + w_t \qquad t = 2, \ldots, T$$
y se obtiene su suma de cuadrados residual
$$SCR_i = \sum_{t=2}^{T} (\widehat{w_{it}})^2 = \sum_{t=2}^{T} \Big((Y_t - \rho_i Y_{t-1}) - (\mathbf{X}_t' - \rho_i \mathbf{X}_{t-1}')\hat{\beta}_i \Big)^2$$
donde $\widehat{w_{it}} \quad t = 2, \ldots, T$ son los residuos de estimar β por MCO en el modelo transformado con ese valor ρ_i.
- Los estimadores $\hat{\beta}_{HL}$ y $\hat{\rho}_{HL}$ serán los asociados a la menor suma de cuadrados residual SCR_i del modelo transformado para los valores $\rho_i \quad i = 1, \ldots, S$ en la red. Para mayor precisión, se puede repetir el proceso afinando la red de búsqueda en torno al valor $\hat{\rho}_{HL}$ que ha proporcionado la SCR_i mínima.

Ejemplo 3.6 Ejemplo de estimación por Hildreth-Lu.

Seguimos con el Ejemplo 4.1 sobre el modelo que relaciona la Inversión Real con el Producto Nacional Bruto Real y el Tipo de Interés incluyendo un término independiente:

$$I_t = \beta_1 + \beta_2 GNP_t + \beta_3 R_t + u_t \qquad t = 1, \ldots, 30$$

Consideramos estimar por el método de Hildreth-Lu bajo el supuesto de que u_t sigue un proceso AR(1). El criterio de estimación del método de Hildreth-Lu es elegir el valor de $\hat{\rho}$ en la red tal que la suma de cuadrados residual de estimar el modelo transformado

$$(I_t - \hat{\rho}\, I_{t-1}) = \beta_1(1 - \hat{\rho}) + \beta_2(GNP_t - \hat{\rho}\, GNP_{t-1}) + \beta_3(R_t - \hat{\rho}\, R_{t-1}) + w_t$$

por MCO es la menor.

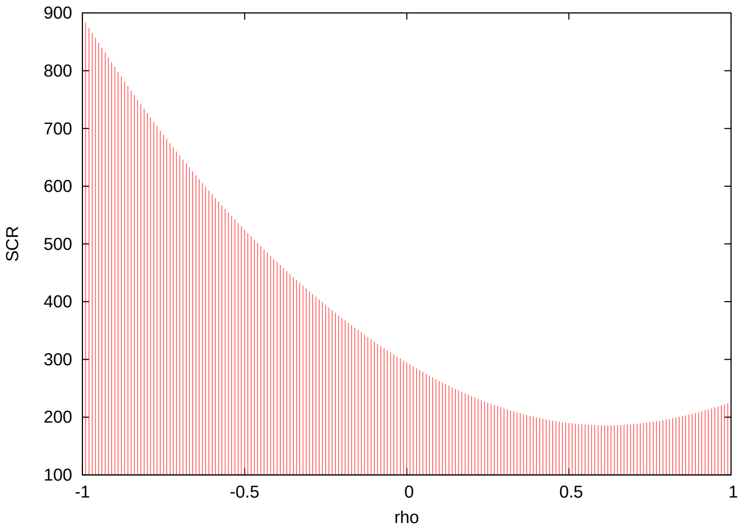
$$Min_{\hat{\beta}}\, SCR(\hat{\beta}) = Min_{\hat{\beta}} \sum_{t=2}^{t=30} \left(Y_t^* - \hat{\beta}_1 X_{1t}^* - \hat{\beta}_2 X_{2t}^* - \hat{\beta}_3 X_{3t}^* \right)^2$$

donde

$$Y_t^* = I_t - \hat{\rho} I_{t-1}; \quad X_{1t}^* = 1 - \hat{\rho}; \quad X_{2t}^* = GNP_t - \hat{\rho} GNP_{t-1}; \quad X_{3t}^* = R_t - \hat{\rho} R_{t-1}$$

En la Figura 3.23 se muestra la Suma de Cuadrados Residual del modelo transformado para cada valor del parámetro ρ en la red de búsqueda.

Figura 3.23: *Función de Suma de Cuadrados Residual del modelo transformado*



El menor valor de la suma de cuadrados residual del modelo transformado es igual a $185,96$ y se alcanza para el valor $\hat{\rho} = 0,61$. Ese valor es la estimación elegida para ρ. Dado ese valor obtenemos las estimaciones de los parámetros de interés β de aplicar MCO al modelo transformado:

$$(I_t - 0,61\, I_{t-1}) = \beta_1(1 - 0,61) + \beta_2(GNP_t - 0,61\, GNP_{t-1}) + \beta_3(R_t - 0,61\, R_{t-1}) + w_t$$

$$\hat{\beta}_{HL} = \begin{bmatrix} \sum X_{1t}^{*2} & \sum X_{1t}^* X_{2t}^* & \sum X_{1t}^* X_{3t}^* \\ \sum X_{1t}^* X_{2t}^* & \sum X_{2t}^{*2} & \sum X_{2t}^* X_{3t}^* \\ \sum X_{1t}^* X_{3t}^* & \sum X_{2t}^* X_{3t}^* & \sum X_{3t}^{*2} \end{bmatrix}^{-1}_{\hat{\rho}=0,61} \begin{bmatrix} \sum Y_t^* X_{1t}^* \\ \sum Y_t^* X_{2t}^* \\ \sum Y_t^* X_{3t}^* \end{bmatrix}_{\hat{\rho}=0,61}$$

Método de Hildreth–Lu, usando las observaciones 1975–2003 ($T = 29$)
Variable dependiente: I
$\hat{\rho}$ = 0,614637

	Coeficiente	Desv. Típica	Estadístico t	Valor p
const	7,3298	3,6585	2,0035	0,0556
GNP	0,7848	0,1442	5,4418	0,0000
R	−0,2957	0,0786	−3,7596	0,0009

Estadísticos basados en los datos rho-diferenciados:

Suma de cuad. residuos 185,9631

La función de regresión muestral obtenida por tanto es la siguiente:

$$\underset{(\widehat{des}(\hat{\beta}_{HL}))}{\hat{I}_t} = \underset{(3{,}6229)}{7,3195} + \underset{(0{,}1428)}{0,7851}\, GNP_t - \underset{(0{,}0788)}{0,2955}\, R_t$$

3.4.2. Método de Cochrane-Orcutt

Este procedimiento, al igual que el de Hildreth-Lu, consiste en minimizar la suma de cuadrados residual del modelo transformado, pero difiere en la forma de estimar el parámetro ρ. Los pasos a seguir son los siguientes:

1. Se estima el modelo original o de interés $Y_t = \mathbf{X}_t'\beta + u_t$ por MCO y se calculan los residuos $\hat{u}_t$.

2. Se obtiene $\hat{\rho}$ de la regresión MCO en:

$$\hat{u}_t = \rho\hat{u}_{t-1} + v_t \qquad t = 2, \ldots, T$$

   El estimador así obtenido es el coeficiente de autocorrelación de orden uno en los residuos:

$$\hat{\rho} = \frac{\sum_{t=2}^{T} \hat{u}_t\hat{u}_{t-1}}{\sum_{t=2}^{T} \hat{u}_{t-1}^2}$$

   Parece razonable utilizar este estimador como aproximación muestral al coeficiente de autocorrelación de orden uno poblacional entre las perturbaciones ρ. Ahora bien, siempre y cuando los residuos sean función de un estimador de β en el modelo de interés que sea consistente.

3. Estimar β por MCO en el modelo transformado:

$$Y_t - \hat{\rho}Y_{t-1} = (\mathbf{X}_t' - \hat{\rho}\mathbf{X}_{t-1}')\beta + w_t \qquad t = 2, \ldots, T$$

Podemos iterar el proceso entre estas dos etapas utilizando en la primera etapa los residuos obtenidos de sustituir en el modelo de interés las estimaciones de los parámetros β, obtenidos en la segunda etapa. El proceso iterativo finalizará cuando se alcance un criterio de convergencia previamente establecido. Por ejemplo, en términos de la diferencia en valor absoluto de l as estimaciones del parámetro ρ, o de la suma de cuadrados residual de la segunda etapa entre dos iteraciones consecutivas.

Ejemplo 3.7 Ejemplo de estimación por Cochrane-Orcutt.

Seguimos con el ejemplo anterior sobre el modelo que relaciona la Inversión Real con el Producto Nacional Bruto real y el tipo de interés incluyendo un término independiente. Estimamos por MCO el modelo de interés

$$I_t = \beta_1 + \beta_2 GNP_t + \beta_3 R_t + u_t \qquad t = 1, \ldots, 30$$

y guardamos los residuos.

$\hat{u}_{1974}$	$\hat{u}_{1975}$	$\hat{u}_{1976}$	$\hat{u}_{1977}$	$\hat{u}_{1978}$	$\hat{u}_{1979}$	$\hat{u}_{1980}$	$\hat{u}_{1981}$
2,036858	1,003142	-0,121917	-0,823617	4,430448	4,556650	4,864468	2,402315
$\hat{u}_{1982}$	$\hat{u}_{1983}$	$\hat{u}_{1984}$	$\hat{u}_{1985}$	$\hat{u}_{1986}$	$\hat{u}_{1987}$	$\hat{u}_{1988}$	$\hat{u}_{1989}$
-4,457423	-4,220092	-6,105541	-6,300801	-3,061876	-3,820975	1,406905	0,682641
$\hat{u}_{1990}$	$\hat{u}_{1991}$	$\hat{u}_{1992}$	$\hat{u}_{1993}$	$\hat{u}_{1994}$	$\hat{u}_{1995}$	$\hat{u}_{1996}$	$\hat{u}_{1997}$
0,052044	2,172243	3,365111	2,684884	0,883149	-3,807123	-3,346817	-0,307411
$\hat{u}_{1998}$	$\hat{u}_{1999}$	$\hat{u}_{2000}$	$\hat{u}_{2001}$	$\hat{u}_{2002}$	$\hat{u}_{2003}$		
3,423955	-3,045531	1,094203	2,702973	-0,142711	1,799846		

Con estos residuos calculamos el valor de $\hat{\rho}$ de la regresión

$$\hat{u}_t = \rho \hat{u}_{t-1} + v_t \qquad t = 2, \ldots, 30$$

MCO, usando las observaciones 1975-2003 ($T = 29$)
Variable dependiente: uhat

	Coeficiente	Desv. Típica	Estadístico t	Valor p
uhat_1	0,567726	0,155221	3,6575	0,0010

El estimador así obtenido es el coeficiente de autocorrelación de orden uno en los residuos:

$$\hat{\rho} = \frac{\sum_{t=1975}^{2003} \hat{u}_t \hat{u}_{t-1}}{\sum_{t=1975}^{2003} \hat{u}_{t-1}^2}$$

El valor muestral inicial para $\hat{\rho}$ es $0,567726$. Con esta primera estimación de ρ se obtiene el modelo transformado:

$$(I_t - 0,56773\, I_{t-1}) = \beta_1(1 - 0,56773) + \beta_2(GNP_t - 0,56773\, GNP_{t-1}) + \beta_3(R_t - 0,56773\, R_{t-1}) + w_t$$

y se estiman por MCO los coeficientes β. De nuevo se calculan los residuos en el modelo de interés utilizando las estimaciones de β en esta etapa y se calcula un nuevo valor de $\hat{\rho}$ en una segunda iteración $\hat{\rho} = 0,61380$. Con esta nueva estimación de ρ se vuelve a obtener $\hat{\beta}$ por MCO en el modelo transformado. Se vuelven a calcular los residuos en el modelo de interés y así sucesivamente se realiza el cálculo iterativo. El proceso se para en la iteración en la que se alcanza el criterio de convergencia en términos de la variación de la suma de cuadrados residual del modelo transformado entre dos iteraciones consecutivas:

Realizando el cálculo iterativo de rho...

ITERACIÓN	RHO	SCR
1	0,56773	186,623
2	0,61380	185,963
final	0,61462	185,963

donde SCR es la suma de cuadrados residual de la estimación MCO de β en el modelo transformado para cada estimación de ρ en cada iteración. Se alcanza el criterio de convergencia en la tercera iteración y las estimaciones finales del proceso iterativo son:

Método de Cochrane–Orcutt , usando las observaciones 1975–2003 ($T = 29$)
Variable dependiente: I
$\hat{\rho} = 0{,}614623$

	Coeficiente	Desv. Típica	Estadístico t	Valor p
const	7,3298	3,6584	2,0036	0,0556
GNP	0,7848	0,1442	5,4420	0,0000
R	−0,2957	0,0786	−3,7596	0,0009

Estadísticos basados en los datos rho-diferenciados:

Suma de cuad. residuos 185,9631

La función de regresión muestral obtenida por el método de Cochrane-Orcutt difiere muy poco de la obtenida por el método de Hildreth-Lu, como era de esperar, y es la siguiente:

$$\underset{(\widehat{des}(\hat{\beta}_{CO}))}{\hat{I}_t} = \underset{(3{,}6584)}{7,3298} + \underset{(0{,}1442)}{0,7848}\ GNP_t - \underset{(0{,}0786)}{0,2957}\ R_t$$

3.5. Inferencia utilizando el estimador MCO con autocorrelación

En el modelo:

$$Y = X\beta + u \qquad\qquad u \sim N(0, \Sigma)$$

donde $\Sigma \neq \sigma^2 I$. Queremos contrastar q restricciones lineales sobre los K parámetros de interés β del tipo $H_0 : R\beta = r$.

En esta situación, el estimador MCO de los coeficientes del modelo, $\widehat{\beta}_{MCO} = (X'X)^{-1}X'Y$ es lineal, insesgado y consistente, pero no es eficiente. Su matriz de varianzas y covarianzas viene dada por la expresión:

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

El estimador de $V(\hat{\beta}_{MCO})$ dado por

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

donde $\hat{\sigma}^2 = \frac{\hat{u}'\hat{u}}{T-k}$, es un estimador sesgado e inconsistente si $\Sigma \neq \sigma^2 I$.

Dado que $\widehat{\beta}_{MCO}$ es consistente, si proponemos un estimador alternativo para $V(\hat{\beta}_{MCO})$ que sea consistente podremos definir estadísticos de contraste válidos al menos asintóticamente.

Estimador de Newey-West de $V(\hat{\beta}_{MCO})$:

El estimador consistente de la matriz de varianzas y covarianzas del estimador MCO, robusto a autocorrelación es:

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

siendo $Q = \sum_{j=1}^{T} \hat{u}_t^2 X_t X_t' + \sum_{\ell=1}^{L} \sum_{t=\ell+1}^{T} w_\ell \hat{u}_t \hat{u}_{t-\ell}[X_t X_{t-\ell}' + X_{t-\ell} X_t']$ donde $w_\ell = 1 - \frac{\ell}{L+1}$ y $X_t' = [1 \quad X_{2t} \; \ldots \; X_{Kt}]$

Para realizar inferencia válida al menos asintóticamente, podemos utilizar los estadísticos basados en el estimador de β por MCO y un estimador consistente de su matriz de varianzas y covarianzas, por ejemplo el anterior $\widehat{V}(\hat{\beta}_{MCO})_{NW}$:

Para $q = 1$, el estadístico de contraste y su distribución asintótica bajo H_0 son:

$$\frac{(R\hat{\beta}_{MCO} - r)}{\sqrt{R\widehat{V}(\hat{\beta}_{MCO})_{NW} R'}} \xrightarrow{d,H_0} N(0,1)$$

Para $q \geq 1$, el estadístico de contraste y su distribución asintótica bajo H_0 son:

$$(R\hat{\beta}_{MCO} - r)'[R\widehat{V}(\hat{\beta}_{MCO})_{NW} R']^{-1}(R\hat{\beta}_{MCO} - r) \xrightarrow{d,H_0} \chi^2_q$$

El utilizar este estimador de $V(\hat{\beta}_{MCO})$ por Newey-West permite seguir utilizando el estimador de los coeficientes MCO para hacer inferencia válida asintóticamente, sin tener que modelizar ni estimar directamente $E(uu') = \Sigma$. Esto permitirá que la inferencia sea válida utilizando los estadísticos anteriores, al menos para muestras grandes. La cuestión es que esto no implica una mejora en la precisión con la que estimamos β por MCO. Para eso tenemos que estimar por MCG y esto implica conocer Σ excepto por un factor de escala, o bien por $MCGF$ lo que requiere modelizar el proceso de autocorrelación y estimar los parámetros de los que es función Σ.

Ejemplo 3.8 Contraste de significatividad individual de la variable tipo de interés. Utilizamos para realizar el contraste los resultados de estimar los coeficientes por MCO junto a las estimaciones de las desviaciones típicas robustas a la posible existencia de autocorrelación.

$$\begin{array}{ccccccc} \widehat{I}_t & = & 6,22494 & + & 0,769911 \; GNP_t - & 0,184196 & R_t \\ (\widehat{des}(\hat{\beta}_{MCO})_{NW}) & & (\text{ 2,04053 }) & & (\text{0,0722822}) & (\text{0,114826}) & \end{array}$$

Contrastamos la significatividad del tipo de interés. Utilizando estos resultados podemos realizar el contraste:

$$H_0 : \beta_3 = 0$$
$$H_a : \beta_3 \neq 0$$

con el estadístico

$$t = \frac{\hat{\beta}_{3,MCO}}{\widehat{des}(\hat{\beta}_{3,MCO})_{NW}} \stackrel{d,H_0}{\rightarrow} N(0,1)$$

Obtenemos el valor muestral del estadístico, que en valor absoluto no supera el valor crítico, $\mid t \mid = 1,6041 < 1,96 = N(0,1)_{|0,025}$, por lo que no rechazamos H_0 para un nivel de significación $\alpha = 5\,\%$ por lo que no habría evidencia para concluir que el tipo de interés es una variable significativa.

3.6. Inferencia con MCGF

Si consideramos modelizar el proceso de autocorrelación en las perturbaciones, estimar Σ con un estimador consistente $\widehat{\Sigma}$ y utilizar el estimador de β por $MCGF$, que será consistente y asintóticamente eficiente, podemos hacer inferencia utilizando los siguientes estadísticos:

Para $q \geq 1$:

$$(R\hat{\beta}_{MCGF} - r)'[R\widehat{V}(\hat{\beta}_{MCGF})R']^{-1}(R\hat{\beta}_{MCGF} - r) \stackrel{d,H_0}{\longrightarrow} \chi^2_q$$

Para $q = 1$:

$$\frac{(R\hat{\beta}_{MCGF} - r)}{\sqrt{R\widehat{V}(\hat{\beta}_{MCGF})R'}} \stackrel{d,H_0}{\longrightarrow} N(0,1)$$

siendo $\widehat{V}(\hat{\beta}_{MCGF})$ un estimador consistente de $V(\hat{\beta}_{MCGF})$. Esto es,

$$\widehat{V}(\hat{\beta}_{MCGF}) = (X'\widehat{\Sigma}^{-1}X)^{-1}$$

Si $\Sigma = \sigma^2\Omega$ entonces un estimador consistente de Σ es $\widehat{\Sigma} = \hat{\sigma}^2\widehat{\Omega}$ donde $\widehat{\Omega}$ es un estimador consistente de Ω utilizado para obtener $\hat{\beta}_{MCGF}$ y

$$\hat{\sigma}^2 = \frac{\hat{u}^{*\prime}\hat{u}^*}{T-K}$$

donde $\hat{u}^{*\prime}\hat{u}^* = (Y - X\hat{\beta}_{MCGF})'\hat{\Omega}^{-1}(Y - X\hat{\beta}_{MCGF}) = (Y^* - X^*\hat{\beta}_{MCGF})'(Y^* - X^*\hat{\beta}_{MCGF})$

Ejemplo 3.9 Contrastamos la significatividad del tipo de interés utilizando los resultados anteriores de la estimación por Hildreth-Lu:

$$\begin{array}{cccccc} \hat{I}_t & = & 7,31954 & + & 0,785198\ GNP_t - & 0,295541\ R_t \\ (\widehat{des}(\hat{\beta}_{HL})) & & (3{,}62292) & & (0{,}142836) & (0{,}0788497) \end{array}$$

$$\begin{array}{c} H_0 : \beta_3 = 0 \\ H_a : \beta_3 \neq 0 \end{array}$$

con el estadístico

$$t = \frac{\hat{\beta}_{3,HL}}{\widehat{des}(\hat{\beta}_{3,HL})} \stackrel{d,H_0}{\rightarrow} N(0,1)$$

Como $\mid t \mid = 3,7482 > 1,96 = N(0,1)_{|0,025}$ es decir, el valor muestral del estadístico en valor absoluto supera el valor crítico, rechazamos H_0 y concluimos que el tipo de interés es una variable significativa, al $5\,\%$ de significación.

¿Cuál es la diferencia entre el contraste de significatividad realizado utilizando el estimador de Hildreth-Lu y el que utiliza el estimador MCO del coeficiente que acompaña al tipo de interés utilizando un estimador robusto a autocorrelación de su desviación típica?

La diferencia entre ambos contrastes está en el estimador utilizado para construir el estadístico. El estimador de Hildreth-Lu es un estimador de MCGF y es asintóticamente más eficiente que el estimador MCO. El resultado del contraste es distinto, ya que en el primer caso no se rechaza la hipótesis nula por lo que la variable tipo de interés es una variable significativa, mientras que en el segundo se rechaza la hipótesis nula, pudiendo ser consecuencia de utilizar en el contraste un estimador más eficiente asintóticamente. Por esa razón, aunque ambos contrastes son válidos asintóticamente, puede ser mejor utilizar el contraste basado en MCGF. Pero para ello es necesario estimar consistentemente Σ, lo que no siempre es sencillo.

3.7. Ejercicios a resolver

A continuación se proponen varios ejercicios para que el alumno vaya haciendo a la vez que se va explicando la materia en las clases magistrales, y cuyas dudas se resolverán a lo largo de estas:

Ejercicio M-A.1

Sea el siguiente modelo:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + u_t \qquad t = 1, 2, \ldots, 52$$

Los resultados de la estimación MCO son:

$$\underset{(\widehat{des}(\hat{\beta}_{MCO}))}{\hat{Y}_t} = \underset{(0{,}12)}{-1,51} - \underset{(0{,}01)}{0,14}\ X_{2t} + \underset{(0{,}015)}{0,998}\ X_{3t} - \underset{(0{,}02)}{0,52}\ X_{4t}$$

$$R^2 = 0,97 \qquad\qquad DW = 0,74$$

Completa las matrices en base a la información disponible. Razona y explica tu respuesta.

$$E(u) = \begin{bmatrix} \\ \\ \\ \\ \end{bmatrix} \qquad E(uu') = \begin{bmatrix} & & & & & & & & & & \\ \\ \\ \\ \end{bmatrix}$$

Ejercicio M-A.2

Se dispone de observaciones anuales de las variables Consumo (C_t) y Renta (R_t) para un país. Los datos se muestran en las primeras columnas de la siguiente tabla:

Obs.	C	R	$\widehat{C}$	$\hat{u}$
1	8,547	11,0	8,0483680	0,498632
2	8,942	13,5	9,7986580	-0,856658
3	10,497	14,0	10,148716	0,348284
4	10,173	14,9	10,778820	-0,605820
5	11,997	15,1	10,918843	1,078157
6	10,729	18,0	12,949180	-2,220180
7	12,750	18,8	13,509273	-0,759273
8	15,611	19,1	13,719307	1,891693
9	13,545	21,0	15,049528	-1,504528
10	17,843	21,2	15,189551	
11	21,610	34,0	24,151036	
12	25,473	34,3	24,361070	
13	24,434	35,0	24,851152	
14	28,274	38,0	26,951500	

Los resultados de la estimación por el método de Mínimos Cuadrados Ordinarios (MCO) de la función de consumo

$$C_t = \beta_1 + \beta_2 R_t + u_t$$

se muestran a continuación:

$$\underset{(t-estad.)}{\hat{C}_t} = \underset{(0{,}31)}{0,347092} + \underset{(14{,}61)}{0,700116}\, R_t \tag{3.13}$$

$$\bar{R}^2 = 0,942 \qquad\qquad SCR = 30,6381$$

1. La última columna de la tabla anterior muestra los residuos de la estimación anterior, complétala y dibuja la serie temporal de los residuos. A la vista del gráfico comenta razonadamente si existe algún problema.

2. Obtén el valor del estadístico de Durbin y Watson y realiza el contraste para el cuál está diseñado. Indica todos los elementos del contraste incluyendo la hipótesis nula y la alternativa.

3. Utilizando la siguiente información realiza el contraste de Breusch y Godfrey. Indica todos los elementos del contraste incluyendo la hipótesis nula y la alternativa.

$$\underset{(t-estad.)}{\hat{u}_t} = \underset{(-0{,}603)}{-0,5679} + \underset{(0{,}0385)}{0,0198}\ R_t - \underset{(-3{,}338)}{0,75}\ \hat{u}_{t-1} + \hat{\omega}_t \quad R^2 = 0,433$$

4. Dada la evidencia obtenida en los apartados anteriores, explica qué consecuencias tiene en:

   a) Las propiedades para muestras finitas del estimador de los coeficientes del modelo. Razona y demuestra tu respuesta.

b) La inferencia utilizando los estadísticos t mostrados en la ecuación (3.13). Razona tu respuesta.

5. ¿Cambiaría tu respuesta al apartado anterior si el problema detectado fuera consecuencia de omitir alguna variable relevante? Razona tu respuesta.

6. Considera la siguiente información y completa lo que falta.

$\hat{\rho}$	-0,99	-0,9	-0,8	-0,7	-0,6	-0,5	-0,4	-0,3	-0,2	-0,1	0,0	0,1
SCR^*	15,9	14,8	14,2	14,1	14,7	15,8	17,5	19,9	22,8	26,2	30,3	34,9

siendo

$$SCR^* = \sum_{t=....}^{t=....} \{(Y_t^* - \hat{\beta}_1 X_{1t}^* - \hat{\beta}_2 X_{2t}^*\}^2 \tag{3.14}$$

$$Y_t^* = C_t - \hat{\rho} C_{t-1}; \qquad X_{1t}^* = ...; \qquad X_{2t}^* = ...$$

$$\begin{bmatrix} \hat{\beta}_1 \\ \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} .... & .... \\ \\ .... & .... \end{bmatrix}^{-1} \begin{bmatrix} .... \\ \\ .... \end{bmatrix}$$

a) ¿Qué método de estimación se está utilizando?

b) ¿Cómo obtendrías las estimaciones finales de β_1 y β_2 por este método? Indica el valor elegido de $\hat{\rho}$ razonando tu elección y la fórmula para obtener el estimador de β_1 y β_2 ¿Qué propiedades tienen los estimadores obtenidos de estos parámetros?

c) ¿Cómo contrastarías $H_0 : \beta_2 = 1$? Indica todos los elementos del estadístico de contraste, así como la regla de decisión.

d) Explica las diferencias entre el método anterior y el de Cochrane-Orcutt.

Ejercicio M-A.3

Consideremos los siguientes datos anuales desde 1990 a 2001 observados para las variables Y y X:

Y	6	3	1	1	1	4	6	16	25	36	49	64
X	-4	-3	-2	-1	1	2	3	4	5	6	7	8

Se propone la estimación de una relación lineal entre estas dos variables donde el término de perturbación sigue un AR(1):

$$Y_t = \beta_1 + \beta_2 X_t + u_t \qquad u_t = \rho u_{t-1} + \varepsilon_t$$

Estimación MCO, usando las observaciones 1990-2001 ($T = 12$)
Variable dependiente: Y

	Coeficiente	Desv. Típica	Estadístico t	Valor p
const	8,01220	4,05896	1,9740	0,0766
X	4,45591	0,91917	4,8477	0,0007

Suma de cuad. residuos	1501,071	R^2 corregido	0,671647
$\hat{\rho}$	0,853150	Durbin–Watson	0,321828

1. Contrasta la hipótesis nula de no autocorrelación utilizando el valor muestral del estadístico de Durbin-Watson.
2. Calcula los residuos de la estimación por MCO y úsalos para obtener una estimación inicial del coeficiente de autocorrelación de orden uno suponiendo que el término de perturbación sigue un AR(1). ¿Coincide con el valor mostrado en los resultados anteriores?
3. Calcula los datos de las variables del modelo transformado:

$$(Y_t - \rho Y_{t-1}) = \beta_1(1-\rho) + \beta_2(X_t - \rho X_{t-1}) + \varepsilon_t$$

4. Obtén las estimaciones de β_1 y β_2 por el método de Cochrane-Orcutt sin realizar ninguna iteración.
5. Calcula la estimación de la matriz de varianzas y covarianzas de los estimadores obtenidos en el apartado anterior.
6. Contrasta la significatividad de la variable X.

3.8. Prácticas de Aula

En el tema de autocorrelación es habitual disponer de una práctica de aula, lo que equivale a una hora de clase presencial y dos horas de trabajo personal. Si el ejercicio ha sido realizado previamente por el alumno, el tiempo es suficiente para su corrección y solución de dudas existentes en clase.

Las competencias a trabajar en esta práctica de aula son:

1. Comprender la importancia de los supuestos empleados en la especificación de un modelo econométrico básico para poder proponer y emplear supuestos más realistas.
2. Diferenciar distintos métodos de estimación y evaluar su uso de acuerdo a las características de las variables económicas de interés para obtener resultados fiables.

Recordatorio: Las prácticas de aula son clases participativas, ello quiere decir que el alumno debe acudir a clase con el ejercicio realizado en su tiempo de trabajo personal. Previamente y con anticipación suficiente se le habrá realizado la debida propuesta. Durante la clase se preguntará aleatoriamente a los alumnos sobre lo realizado.

Práctica de Aula PA-A.1:

Un investigador dispone de una base de datos anuales[1], para el período de 1948 a 1993, de los siguientes índices agrarios de EEUU, todos ellos con base 1982:

output	:	producción agrícola (Rango 51 - 116)
labor	:	mano de obra agrícola (Rango 81 - 278)
land	:	superficie utilizada en la producción agrícola (Rango 89 - 102)
machines	:	maquinaria (duradera) (Rango 38 - 102)

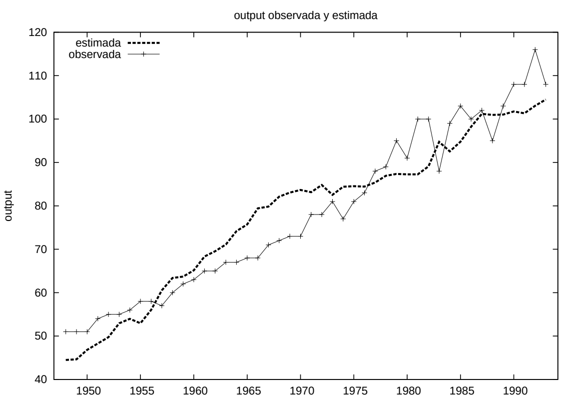
El objetivo del investigador es determinar la función de producción agraria, para ello especifica el siguiente modelo de regresión lineal:

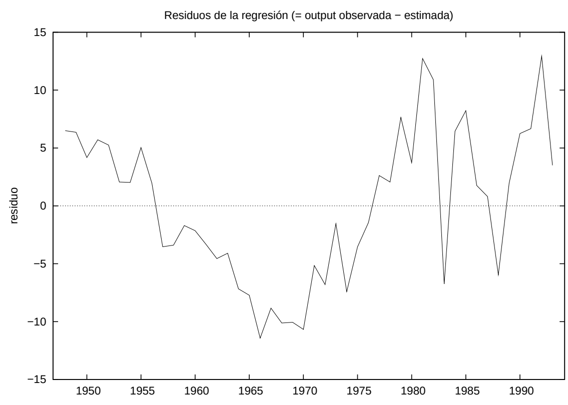
$$output_t = \beta_1 + \beta_2 labor_t + \beta_3 land_t + \beta_4 machines_t + u_t \qquad t = 1, \ldots, T \tag{3.15}$$

en el que se considera que los regresores son no estocásticos. Los resultados obtenidos de la estimación MCO son los que se muestran a continuación:

$$\begin{array}{rcllll} \widehat{output}_t & = & 181,201 & -\ 0,307\ labor_t & -\ 0,517\ land_t & -\ 0,096\ machines_t \\ (\widehat{des}(\hat{\beta}_{MCO})) & & (40{,}194) & (0{,}038) & (0{,}564) & (0{,}169) \end{array}$$

$$R^2 = 0,884 \qquad DW = 0,612 \qquad SCR = 1885,08 \qquad T = 46$$





1. Explica cómo se han calculado los residuos y para qué sirven los gráficos. Interpreta ambos gráficos y señala si existe alguna evidencia de que la perturbación del modelo no cumpla alguna de las hipótesis básicas, justificando tu respuesta.

2. Realiza algún contraste basándote en la información disponible, para cualquier problema detectado en el apartado anterior. Explica detalladamente todos los elementos que intervengan.

[1] Fichero data9-5.gdt, disponible en gretl pestaña Ramanathan. Recogido en Ramanathan, R. (2002), *Introductory Econometrics with Applications*, 5th. edn., South-Western.

3. Con respecto a los contrastes de significatividad individual de las variables explicativas del modelo (3.15):

   a) ¿Es fiable realizarlos utilizando la información disponible? ¿Por qué?

   b) ¿Sería posible llevarlos a cabo si no tuviésemos otra opción que la de estimar los coeficientes del modelo por MCO? Explica cómo lo harías en caso afirmativo.

   Viendo los resultados obtenidos el investigador decide estimar el mismo modelo por el método de Cochrane-Orcutt (CO). A continuación se muestran los resultados obtenidos:

Estimaciones Cochrane–Orcutt utilizando las 45 observaciones 1949–1993
Variable dependiente: output
$\hat{\rho} = 0.791585$

	Coeficiente	Desv. Típica	Estadístico t	Valor p
const	54,3902	30,3065	1,7947	0,0801
labor	−0,4046	0,0649	−6,2284	0,0000
land	1,0727	0,3741	2,8673	0,0065
machines	−0,2874	0,2000	−1,4372	0,1583

Estadísticos basados en los datos rho-diferenciados:

R^2	0,957590	R^2 corregido	0,954487
$F(3, 41)$	12,99460	Valor p (de F)	4,18e–06
$\hat{\rho}$	−0,184791	Durbin–Watson	2,339505

$$\widehat{V}(\hat{\beta}_{CO}) = \begin{bmatrix} 918,4860 & 0,1755 & -9,5945 & -0,0629 \\ 0,1755 & 0,0042 & -0,0096 & 0,0027 \\ -9,5945 & -0,0096 & 0,1399 & -0,0349 \\ -0,0629 & 0,0027 & -0,0349 & 0,0400 \end{bmatrix}$$

4. ¿Cuándo estás dispuesto a aplicar este método de estimación? En particular, ¿consideras adecuado utilizar este método en las circunstancias actuales? Responde razonadamente.

5. Describe detalladamente cómo obtener las estimaciones de los coeficientes del modelo (3.15) utilizando el método del apartado anterior.

6. Con la información disponible, realiza el siguiente contraste $H_0 : \beta_2 = \beta_3$. Escribe la hipótesis nula, la alternativa, el estadístico de contraste junto con su distribución y realiza el contraste. ¿Cómo interpretas el resultado?

3.9. Prácticas de Ordenador

En el tema de autocorrelación es habitual disponer de una práctica de ordenador que equivale a una hora presencial y dos horas de trabajo no presencial para repasar lo visto y repetir la mecánica. A continuación se va a proponer un ejercicio para resolver en el centro de cálculo. El enunciado cubre todo lo aprendido en el tema y se va a marcar la división entre las dos horas de clase que conlleva cada una de las dos prácticas de ordenador.

Las competencias a trabajar en estas prácticas de ordenador son:

2. Diferenciar distintos métodos de estimación y evaluar su uso de acuerdo a las características de las variables económicas de interés para obtener resultados fiables.

3. Utilizar diversas fuentes estadísticas y adquirir destreza en el uso de un software econométrico para analizar relaciones entre variables económicas.

Recordatorio: Es conveniente que una vez acabado el ejercicio y en vuestro tiempo de trabajo personal deis contenido al mismo. Es decir, en clase únicamente os da tiempo a aprender la mecánica, la ejecución sucesiva de instrucciones para obtener los resultados pedidos, aunque el profesor va comentando y explicando los resultados. Vosotros y de forma personal debéis redactar convenientemente las respuestas de cada apartado, en especial los que se recomienda realizar en casa. En el Anexo del tema tenéis un resumen de las instrucciones básicas de gretl para autocorrelación.

Práctica de ordenador PO-A.1

El fichero de datos necesario para la realización de esta práctica se encuentra en los archivos de muestra de gretl y corresponde al fichero Table12-9.gdt de Gujarati[2]. Son datos de series temporales correspondientes a los años 1950 a 1991. De las variables disponibles, se consideran las siguientes:

SALES	Ventas de la industria manufacturera en EE.UU, en millones de dólares
INVENTS	Inventarios de la industria manufacturera en EE.UU, en millones de dólares.

1. Estima por MCO el siguiente modelo

$$INVENTS_t = \beta_1 + \beta_2 SALES_t + u_t \tag{3.16}$$

Completa utilizando los resultados obtenidos con gretl:

$$\begin{array}{lcccc} \widehat{INVENTS}_t = & & + & & SALES_t \\ (\widehat{des}(\hat{\beta}_{MCO})) & (\qquad\qquad) & & (\qquad\qquad) & \end{array}$$

$R^2 =$ $\qquad\qquad$ $SCR =$ $\qquad\qquad$ $T =$

$DW =$ $\qquad\qquad$ $\widehat{Cov}(\hat{\beta}_1, \hat{\beta}_2) =$

Coeficiente de correlación entre $INVENTS$ y $SALES =$

Año t	1950	1951	1952
Residuo $\hat{u}_t$			

[2]Fichero Table12-9.gdt, disponible en gretl pestaña Gujarati. Fuente: Economic Report of the President, 1993, Table B-53, recogida en Ramanathan, R. (2002), *Introductory econometrics with applications*, 5th. Ed., South-Western.

2. Obtén el gráfico de series temporales de las variables dependiente y explicativa, así como el gráfico de la serie observada y ajustada y el de los residuos. Coméntalos. (Los comentarios a realizar en casa)

3. Se considera que u_t puede seguir un proceso AR(p) o MA(p) con p hasta de orden dos.

   Realiza el contraste de Breusch-Godfrey utilizando gretl y completa.

   a) Escribe la hipótesis nula y la alternativa del contraste.
   b) Aplica el contraste y completa:
      Regresión auxiliar obtenida:

$$\ldots\ldots = \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$R^2 =$$

      Estadístico y distribución bajo la hipótesis nula:
      Valor muestral del estadístico =
      Aplica la regla de decisión para un nivel de significación ($\alpha = 5\,\%$)

   c) A la vista de los resultados de los contrastes de autocorrelación DW y Breusch-Godfrey, ¿son las perturbaciones de modelo esféricas? ¿Por qué? (A completar en casa).

4. Estima de nuevo los coeficientes del modelo por MCO, pero obtén desviaciones típicas de los coeficientes estimados robustas a la posible existencia de autocorrelación. Completa.

$$\underset{(\widehat{desv}\ldots\ldots\ldots(\hat{\beta}_{MCO}))}{\widehat{I}_t} = \underset{(\qquad\qquad)}{} \quad \ldots \quad \underset{(\qquad\qquad)}{} \quad SALES_t$$

   ¿Para qué sirven las desviaciones típicas así obtenidas? ¿Cuando son de utilidad? Explica detalladamente. (A completar en casa)

5. Contrasta la hipótesis conjunta de que en media si las ventas son cero no hay inventarios y de que un aumento en el nivel de ventas de un millón de dólares aumentaría los inventarios en 2 millones y medio de dólares. Escribe la hipótesis nula, la alternativa, el estadístico de contraste y su distribución bajo la nula. (A realizar en casa en detalle)

Práctica de ordenador PO-A.2

En esta sesión vamos a continuar trabajando con el fichero de datos que utilizamos en la práctica de ordenador **PO-A.1**. Previamente recordaremos lo realizado en aquella práctica y se preguntará aleatoriamente a los alumnos sobre los apartados que había que realizar en casa.

6. Considera estimar el modelo (3.16) utilizando el método de Cochrane-Orcutt bajo el supuesto de que u_t sigue un proceso AR(1).

   Completa la función de regresión muestral obtenida, la estimación de ρ y el valor mínimo de la función criterio.

$$\underset{(\widehat{des}(\hat{\beta}_{C-O}))}{\ldots\ldots} = \underset{(\qquad\qquad)}{} \quad \underset{(\qquad\qquad)}{\ldots} \quad \underset{(\qquad\qquad)}{\ldots}$$

$\hat{\rho} =$ `valor minimo de` $SCR =$

7. Explica en detalle cómo se han obtenido las estimaciones por este método. (A completar en casa)

8. Considera estimar el modelo (3.16) utilizando el método de Hildreth-Lu bajo el supuesto de que u_t sigue un proceso AR(1).

   Completa la función de regresión muestral obtenida, la estimación de ρ y el valor mínimo alcanzado para la función criterio.

$$\underset{(\widehat{des}(\hat{\beta}_{H-L}))}{\ldots\ldots} = \underset{(\qquad\qquad)}{} \quad \underset{(\qquad\qquad)}{\ldots} \quad \underset{(\qquad\qquad)}{\ldots}$$

$\hat{\rho} =$ `valor minimo de` $SCR =$

   Dibuja el gráfico de la función criterio para cada valor de ρ.

9. Explica en detalle cómo se han obtenido las estimaciones por este método. (A realizar en casa)

10. Utilizando uno de los dos métodos anteriores, contrasta de nuevo la hipótesis conjunta de que en media si las ventas son cero no hay inventarios y de que un aumento en el nivel de ventas de un millón de dólares aumentaría los inventarios en 2 millones y medio de dólares.

11. ¿Difieren las conclusiones del contraste de las obtenidas anteriormente? ¿Es posible que esto ocurra? ¿Cual sería más adecuado? (A realizar en casa)

3.10. Taller sobre todo lo trabajado en el tema

En el tema de autocorrelación en general se lleva a cabo un taller. El objetivo del taller es analizar una serie de resultados en la estimación de varias especificaciones de un modelo e ir practicando en la toma de decisiones y la redacción apropiada de conclusiones. Por ello vamos a proponer un taller utilizando lo ya trabajado en las prácticas de ordenador y cuyo enunciado se habrá repartido previamente. De esta forma el alumno ya está familiarizado con el problema a analizar y se pueden trabajar más en detalle las siguientes competencias:

1. Comprender la importancia de los supuestos empleados en la especificación de un modelo econométrico básico para poder proponer y emplear supuestos más realistas.

2. Diferenciar distintos métodos de estimación y evaluar su uso de acuerdo a las características de las variables económicas de interés para obtener resultados fiables.

4. Elaborar en grupos de trabajo y exponer en público un proyecto empírico, donde se valore adecuadamente los resultados obtenidos del análisis de un modelo econométrico.

El ejercicio debe resolverse en equipo, idealmente en el grupo en que se va a realizar el proyecto. Esto favorece la interrelación y conocimiento entre sus miembros. En los últimos 15 minutos de clase, un representante de cada equipo deberá de exponer al resto de la clase la decisión adoptada por el grupo junto a las razones consideradas.

Enunciado del taller:

Objetivo: Analizar los resultados de la estimación de un modelo por diferentes métodos para los inventarios . Elegir entre ellos utilizando diferentes contrastes de especificación.

Información: Son datos del periodo 1950 a 1991 de las siguientes variables:

SALES	Ventas de la industria manufacturera en EE.UU. en millones de dólares
INVENTS	Inventarios de la industria manufacturera en EE.UU. en millones de dólares

Procedimiento: Debéis analizar la información disponible y decidir cuál es la especificación más adecuada junto con su correcta estimación.

• **ESPECIFICACIÓN A:** Se propone la siguiente relación:

$$INVENTS_t = \beta_1 + \beta_2 SALES_t + u_t \tag{3.17}$$

Estimación 1 de la ESPECIFICACIÓN A: Los resultados de la estimación MCO son los siguientes:

Estimaciones MCO utilizando las 42 observaciones 1950–1991
Variable dependiente: INVENTS

Variable	Coeficiente	Desv. típica	Estadístico t	valor p
const	1668,67	1806,70	0,9236	0,3612
SALES	1,5543	0,0069	222,5282	0,0000

Media de la var. dependiente	311725,00
D.T. de la variable dependiente	259140,00
Suma de cuadrados de los residuos	2,22224e+09
Desviación típica de los residuos ($\hat{\sigma}$)	7453,59
R^2	0,99910
$\bar{R}^2$ corregido	0,999173
Grados de libertad	40
Estadístico de Durbin–Watson	1,37460
Coef. de autocorr. de primer orden.	0,31100

Matriz de covarianzas de los coeficientes

const	SALES	
3,26415e+06	−9,7322	const
	4,87883e-05	SALES

Coeficientes de correlación, usando las observaciones 1950 - 1991
valor crítico al 5 % (a dos colas) = 0,3044 para n = 42

SALES	INVENTS	
1,0000	0,9996	SALES
	1,0000	INVENTS

Además, se dispone de la siguiente regresion auxiliar:

$$\hat{u}_t = 90,048 - 0,000627\, VENTAS_t + 0,287394\, \hat{u}_{t-1} + 0,08407\, \hat{u}_{t-2} \qquad t = 3, \ldots, 42$$

$$R^2 = 0,103545$$

1. Analizar **TODA** la información proporcionada. Debéis decidir si esta especificación junto con su método de estimación es la más adecuada. Debéis razonar en qué medida estos resultados son fiables.

Estimación 2 de la ESPECIFICACIÓN A:

Se estiman de nuevo los coeficientes de la ESPECIFICACIÓN A por MCO, pero se utilizan desviaciones típicas de los coeficientes estimados robustas a la posible existencia de autocorrelación.

Se obtienen los siguientes resultados:

Estimaciones MCO utilizando las 42 observaciones 1950–1991
Variable dependiente: INVENTS
Desviaciones típicas HAC, con ancho de banda 2 (Kernel de Bartlett)

Variable	Coeficiente	Desv. típica	Estadístico t	valor p
const	1668,67	1094,77	1,5242	0,1353
SALES	1,5543	0,0083	187,2301	0,0000

Media de la var. dependiente	311725,00
D.T. de la variable dependiente	259140,00
Suma de cuadrados de los residuos	2,22224e+09
Desviación típica de los residuos ($\hat{\sigma}$)	7453,59
R^2	0,99910
$\bar{R}^2$ corregido	0,99910
Grados de libertad	40
Estadístico de Durbin–Watson	1,374607
Coef. de autocorr. de primer orden.	0,3110

Conjunto de restricciones	b[const] = 0
	b[SALES] = 2,5
Estadístico de contraste	F robusto(2, 40) = 9272,21
	valor p = 4,55217e-054

2. Analizar la información proporcionada. En particular deberías razonar sobre en qué medida y para qué fin estos resultados son fiables. ¿Qué concluyes?

• **ESPECIFICACIÓN B:**

El estudiante considera la inclusión de la variable $time = 1, 2, \ldots, 42$ en el modelo y especifica la siguiente ecuación:

$$INVENTS_t = \beta_1 + \beta_2 SALES_t + \beta_3\, time + u_t \qquad t = 1, \ldots, 42$$

y realiza las siguientes estimaciones:

Estimación 1 de la ESPECIFICACIÓN B:

Estimaciones MCO utilizando las 42 observaciones 1950–1991
Variable dependiente: INVENTS
Desviaciones típicas robustas a autocorrelación

Variable	Coeficiente	Desv. típica	Estadístico t	valor p
const	433,9510	1143,3800	0,3795	0,7064
SALES	1,5434	0,0136	113,3700	0,0000
time	158,8050	169,2250	0,9384	0,3538

Suma de cuadrados de los residuos	2,20257e+09
R^2	0,999200
$\bar{R}^2$ corregido	0,999159
$F(2, 39)$	18497,7
Estadístico de Durbin–Watson	1,375590
Coef. de autocorr. de primer orden.	0,311486

Estimación 2 de la ESPECIFICACIÓN B:

Realizando el cálculo iterativo de rho...

ITERACIÓN	RHO	SCR
1	0,31149	1,9874e+009
2	0,31600	1,98735e+009
final	0,31616	

Estimaciones Cochrane–Orcutt utilizando las 41 observaciones 1951–1991
Variable dependiente: INVENTS
$\hat{\rho} = 0{,}316161$

	Coeficiente	Desv. típica	estadístico t	valor p
const	34,3714	4413,7100	0,0078	0,9938
SALES	1,5375	0,0289	53,1693	0,0000
time	229,7630	410,6420	0,5595	0,5791

Estadísticos basados en los datos rho-diferenciados:

Suma de cuadrados de los residuos	1,98735e+09
R^2	0,999261
$\bar{R}^2$ corregido	0,999222
$F(2, 38)$	12476,9
Estadístico de Durbin–Watson	2,050180
Coef. de autocorr. de primer orden.	−0,027528

3. ¿Por qué crees que el estudiante ha introducido la variable tendencia (*time*) como regresor en el modelo? ¿Es relevante incluirla? ¿Por qué crees que se obtiene ese resultado? Utiliza los contrastes que consideres oportunos. Razona tu respuesta. (A completar en casa)

4. Ayuda al estudiante a decidir sobre la **fiabilidad de los distintos resultados de estimación mostrados** de la especificación B. Razona tu respuesta en base a la información proporcionada.

5. Debéis decidir qué especificación A o B y método de estimación es la más adecuada. Analizar la información proporcionada por los resultados de la estimación. Consejo: empezar por escribir la función de regresión poblacional que se está estimando e indicar cuáles son los supuestos sobre la perturbación que se han asumido. Analizar su coherencia dada **TODA** la información disponible hasta este momento. Finalmente, tomad una decisión.

3.11. Evaluativas - Preguntas cortas

Se recuerda que dado que el curso contempla la evaluación continua es necesario que a lo largo del tema se evalúe a los alumnos, tanto en las clases magistrales, como en las prácticas de aula y ordenador. Se llevan a cabo en clase y de manera individual. A modo de ejemplo se incluyen las siguientes.

- Preguntas cortas en Clases Magistrales:

Al estimar el modelo

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t \qquad t = 1, \ldots, 73$$

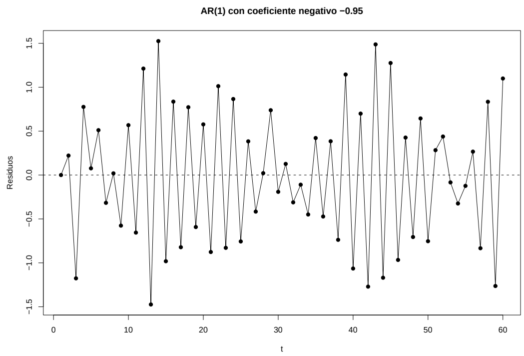
donde la matriz de regresores X es no estocástica, se ha obtenido un valor del estadístico Durbin-Watson igual a 0,468.

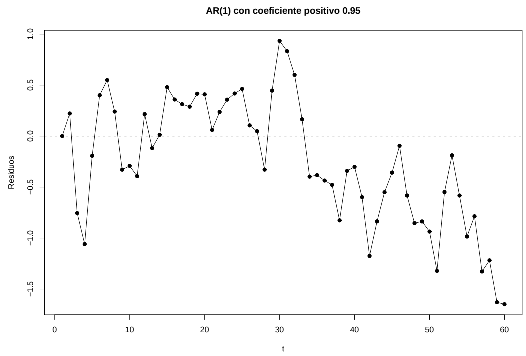
Pc.1 Escribe la fórmula del estadístico de Durbin Watson. ¿Hay evidencia de autocorrelación en el término de perturbación del modelo? (Al 5 %, $d_i = 1,56$ y $d_s = 1,67$)

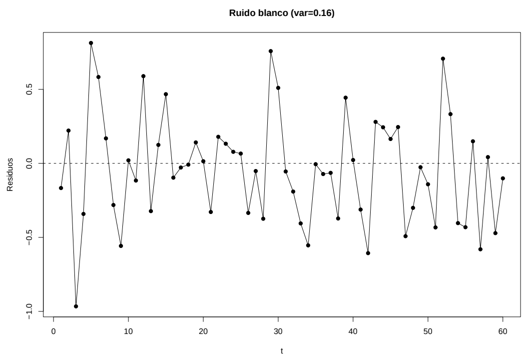
Pc.2 Escribe la regresión auxiliar y el estadístico de Breusch Godfrey para contrastar la existencia de un proceso AR(1) o MA(1) en el término de perturbación.

Pc.3 ¿Es compatible obtener un valor del estadístico de Breusch-Godfrey de 0,2474 dado el valor obtenido del estadístico de Durbin Watson anterior? ¿Por qué? ($\chi^2_{1|0,05} = 3,84$)

Pc.4 ¿Cual de los siguientes gráficos es compatible con la evidencia del valor muestral del estadístico de Durbin-Watson? ¿Y con la del de Breusch-Godfrey?







Pc.5 Escribe el correspondiente modelo con perturbaciones esféricas considerando que el valor del coeficiente de autocorrelación de orden uno de el término de perturbación del modelo de interés es 0,7. Escribe explícitamente la fórmula del estimador eficiente.

• Preguntas cortas en Prácticas de Aula:

Se propone el siguiente modelo para la oferta de café en Colombia:

$$\ln(Q_t) = \alpha + \beta \ln(P_t) + u_t$$

donde Q es el área dedicada a la plantación de café y P es el precio del producto en el mercado. Se dispone de 34 observaciones anuales de Q y P. La estimación MCO es:

$$\underset{(\widehat{des})}{\widehat{\ln(Q_t)}} = \underset{(0{,}20)}{5,1} + \underset{(0{,}23)}{0,85} \ln(P_t) \qquad R^2 = 0,86 \qquad (R1)$$

Se han realizado las siguientes regresiones basadas en los residuos MCO, $\hat{u}$:

$$\begin{array}{rcllll}
\hat{u}_t &=& -0,02 + 0,012\ln(P_t) + 0,34\hat{u}_{t-1} & R^2 = 0,116 & SCR = 2,7 \\
\hat{u}_t &=& -0,38 + 0,01t - 0,18\ln(P_t) + 0,32\hat{u}_{t-1} & R^2 = 0,13 & SCR = 2,61 \\
\hat{e}_t^2 &=& 1,32 - 0,02t & R^2 = 0,023 & SCR = 46,48 \\
\hat{e}_t^2 &=& 5,20 - 0,1t + 1,74\ln(P_t) & R^2 = 0,10 & SCR = 42,76 \\
\hat{e}_t^2 &=& 5,74 - 0,11t + 1,87\ln(P_t) - 0,18v_{t-1} & R^2 = 0,13 & SCR = 41,21 \\
\hat{e}_t &=& -0,22 + 0,01t & R^2 = 0,001 & SCR = 378,62 \\
\hat{e}_t &=& -3,59 + 0,08t - 1,51\ln(P_t) & R^2 = 0,009 & SCR = 375,82 \\
\hat{e}_t &=& 0,51 - 0,009t + 0,17\ln(P_t) - 0,18e_{t-1} & R^2 = 0,13 & SCR = 0,33
\end{array}$$

con $\hat{e}_t = \hat{u}_t/\tilde{\sigma}$ y $\tilde{\sigma}^2 = \sum_t \hat{u}_t^2/34$.

a) Contrasta si existe autocorrelación en el modelo. Indica claramente la hipótesis nula y la alternativa, la regresión auxiliar utilizada, el estadístico de contraste y su distribución bajo la hipótesis nula.

Posteriormente se han obtenido las siguientes estimaciones por MCGF:

$$\underset{(\widehat{des}(\hat{\beta}))}{\widehat{\ln(Q_t)}} = \underset{(0{,}16)}{5,98} + \underset{(0{,}13)}{0,89} \ln(P_t) \qquad SCR = 3,052 \qquad \hat{\sigma}_t = 0,30/\sqrt{t} \qquad (R2)$$

$$\underset{(\widehat{des}(\hat{\beta}))}{\widehat{\ln(Q_t)}} = \underset{(0{,}3)}{5,82} + \underset{(0{,}9)}{1,21} \ln(P_t) \qquad SCR = 5,620 \qquad \hat{\sigma}_t = 5,066 \times t \qquad (R3)$$

$$\underset{(\widehat{des}(\hat{\beta}))}{\widehat{\ln(Q_t)}} = \underset{(0{,}22)}{6,04} + \underset{(0{,}11)}{0,91} \ln(P_t) \qquad SCR = 2,642 \qquad \hat{u}_t = 0,34\hat{u}_{t-1} + e_t \qquad (R4)$$

$$\underset{(\widehat{des}(\hat{\beta}))}{\widehat{\ln(Q_t)}} = \underset{(0{,}28)}{6,21} + \underset{(0{,}15)}{0,96} \ln(P_t) \qquad SCR = 2,532 \quad \hat{u}_t = 0,36\hat{u}_{t-1} + 0,002\hat{u}_{t-2} + e_t \quad (R5)$$

b) Interesa saber si la elasticidad-precio es cero. Explica cómo lo contrastarías indicando claramente el estimador que utilizas y cómo se ha obtenido. Utiliza la información anterior para realizar el contraste.

• Preguntas cortas en Prácticas de Ordenador:

Accede al conjunto de datos **9-7.gdt** del libro de Ramanathan[3] incluido en gretl. Son datos trimestrales desde el primer trimestre de 1975 al cuarto trimestre de 1990. Considera el siguiente subconjunto de variables de este fichero:

QNC	Número de coches nuevos vendidos (en miles de dólares).
PRICE	Índice de precios medios en términos reales de un coche nuevo.
INCOME	Renta personal disponible per capita en miles de dólares.

Para el modelo:

$$QNC_t = \beta_1 + \beta_2 PRICE_t + \beta_3 INCOME_t + u_t \qquad t = 1, \ldots, 64$$

1. Obtén el gráfico de los residuos contra el tiempo. Comenta el gráfico.
2. Dado que son datos trimestrales, contrasta la posibilidad de que u_t siga un proceso autorregresivo o de medias móviles de orden 4, utilizando el estadístico de Breusch-Godfrey. Completa:
   a) H_0 : H_a :
   b) Estadístico de contraste y distribución:
   c) Regresión auxiliar:
   d) Regresión auxiliar estimada:
   e) Computa el estadístico de contraste y lleva a cabo el contraste:

[3] Fichero 9-7.gdt, disponible en gretl pestaña Ramanathan. Recogido en Ramanathan, R. (2002), *Introductory Econometrics with Applications*, 5th. edn., South-Western.

3.12. Bibliografía del tema

Referencias Bibliográficas Básicas:

• Teórica:

[1] Greene, W. (1998), Análisis Econométrico, ed. Prentice Hall, 3ª edition.

[2] Ramanathan, R. (2002), Introductory Econometrics with applications, ed. South-Western, 5th edition.

[3] Wooldridge, J. M. (2003), Introductory Econometrics: A modern Approach, ed. South-Western, 2nd edition.

• Ejercicios:

[1] Alegre, J., Arcarons, J., Bolancé, C. y Díaz, L. (1995), Ejercicios y Problemas de Econometría, Colección Plan Nuevo, ediciones AC.

[2] Fernández, A., González, P., Regúlez, M., Moral, P., Esteban, V. (2005), Ejercicios de Econometría, ed. McGraw-Hill, 2 ª edición.

[3] Recopilación de ejercicios recomendados y exámenes de Econometría. Departamento de Economía Aplicada III (Econometría y Estadística). Mimeo Febrero 2009. Disponible en Lankopi.

• Ejercicios con gretl:

[1] Ramanathan, R. (2002) Instructor's Manual to accompany, del libro Introductory Econometrics with applications, ed. South-Western, 5th edition, Harcourt College Publishers.

[2] Wooldridge, J. M. (2003), Student Solutions Manual, del libro Introductory Econometrics: A modern Approach, ed. South-Western, 2nd edition.

Referencias Bibliográficas Complementarias:

[1] Alonso, A., Fernández, J. y Gallastegui, I. (2005), Econometría, ed. Pearson: Prentice Hall.

[2] Gujarati, D. (2004), Econometría, ed. McGraw-Hill, 4ª edición.

[3] Johnston, J. (1984), Métodos de Econometría, ed. Vicens Vivens, 4ª edición.

[4] Maddala, G. S. (1996), Introducción a la Econometría, ed. Pearson: Prentice Hall.

[5] Novales, A. (1993), Econometría, ed. McGraw-Hill, 2ª edición.

[6] Pindyck, R. S. y Rubinfeld, D. L. (1998), Econometric Models and Economic Forecast, ed. McGraw-Hill, 4ª edición.

3.13. Anexo: Instrucciones básicas de gretl para autocorrelación

En este anexo vamos a indicar las instrucciones básicas para poder realizar la práctica de ordenador del tema. En el anexo del tema anterior ya se han explicado las generales, por lo que nos centraremos solamente en las específicas de autocorrelación.

• **Contrastes de Autocorrelación**: El programa gretl tiene implementados varios contrastes de autocorrelación.

- El estadístico de Durbin-Watson se muestra en los resultados de las estimación por MCO. Para realizar el contraste de Durbin-Watson, hay que obtener los valores de las cotas inferior y superior tabuladas al 5 % para un tamaño muestral dado n, por ejemplo 72 observaciones.

  En la ventana principal hay que elegir *Herramientas*→ *Tablas estadísticas* → *DW* → se completa el tamaño muestral $n = 72$ y el número de regresores del modelo, excluyendo el término constante, por ejemplo 2.

  Se muestran las cotas inferior $d_i \equiv dL$ y superior $d_s \equiv dU$, de la siguiente forma:

  ```
  Valores críticos al 5% del estadístico de Durbin-Watson, n = 72, k = 2

    dL = 1,5611
    dU = 1,6751
  ```

- El contraste de Breusch-Godfrey se puede obtener en la pantalla de resultados de la estimación MCO. Pulsar:

  Contrastes → *Autocorrelación* → *Seleccionar el número de retardos p de la hipótesis alternativa*

  Al realizar la regresión auxiliar gretl usa todas las observaciones, incluidas las primeras p igualando aquellos retardos de los residuos no disponibles $\hat{u}_0, \ldots, \hat{u}_{1-p}$ a cero. Por supuesto puedes obtener el valor del estadístico de contraste siguiendo todos los pasos aprendidos en clase. Veamos como realizar explícitamente y paso a paso el contraste de Breusch-Godfrey:

  1. Se estima el modelo de interés por MCO *Modelo* → *Mínimos Cuadrados Ordinarios*
  2. Se guardan los residuos *Guardar* → *residuos* en la ventana de estimación del modelo.
  3. Se realiza la regresión auxiliar tal y como se ha explicado en la clase magistral. Si no se consideran todas las observaciones, simplemente se realiza la estimación por MCO eligiendo como variable dependiente la que se ha definido para guardar los residuos y como variables explicativas los regresores del modelo de interés. Además, para incluir como regresores los p retardos de los residuos: en la ventana *gretl:especificar modelo* elegir *retardos* y en la que surge elegir los retardos que se deseen, de 1 a p de la variable dependiente.
  4. Se guarda el valor del coeficiente de determinación R^2 de esa regresión auxiliar y se multiplica por el número de observaciones disponibles para el cálculo del estadístico.

• **Estimar de forma consistente la matriz de varianzas y covarianzas de los coeficientes estimados por MCO robusto a autocorrelación**.

La estimación de los coeficientes del modelo es por MCO, pero su matriz de varianzas y covarianzas se estima teniendo en cuenta que hay autocorrelación para que la inferencia usando los estadísticos t y F sea adecuada:

Modelo → Mínimos Cuadrados → elegir la variable dependiente y las variables independientes

Elegir en esa misma ventana **desviaciones típicas robustas**, y en configurar elegir HAC.

De esta forma podemos hacer los contrastes de significatividad individual con los valores de los estadísticos t que nos muestra el output eligiendo el valor crítico en la distribución N(0,1). Cualquier otro contraste de restricciones lineales los podemos realizar eligiendo en la ventana de estimación *Contrastes → Restricciones lineales.* El programa gretl considera para realizar el contraste el estimador robusto a autocorrelación de la matriz de varianzas y covarianzas de los coeficientes estimados por MCO.

• **Estimar la especificación del modelo por el método de Cochrane-Orcutt**.

Modelo → Series Temporales → Cochrane-Orcutt

En la ventana de resultados se muestra el valor de rho y la SCR del modelo transformado para cada iteración, hasta el valor final alcanzado el criterio de convergencia que por defecto usa gretl. A su vez, se muestran los resultados de la estimación del modelo original por este método. También se muestran una serie de estadísticos basados en los datos rho-diferenciados. Hay que ser cautelosos en la interpretación de estos resultados, ya que se refieren al modelo transformado y no al original. Este modelo tiene regresores estocásticos ya que para la transformación se utiliza la estimación del parámetro ρ. Por lo tanto, el valor del estadístico Durbin-Watson no es comparable con las cotas tabuladas bajo el supuesto de regresores no estocásticos.

Cualquier contraste de restricciones lineales utilizando este estimador lo podemos realizar eligiendo en la ventana de estimación *Contrastes → Restricciones lineales*

• **Estimar la especificación del modelo por el método de Hildreth-Lu**

Modelo → Series Temporales → Hildreth-Lu

En la ventana de resultados se muestra el valor de rho en la red para la cual la SCR es mínima junto con los resultados de la estimación de los parámetros del modelo original. Adiccionalmente, se muestra en una nueva ventana la función a minimizar que es la Suma de Cuadrados de Residuos (SCR) del modelo transformado para distintos valores de rho en la red.

- **Tratamiento de los retardos en gretl**.

  - Añadir retardos de una variable en concreto a la ventana principal.

    Añadir $\rightarrow$ *Retardos de las variables seleccionadas*

    En la ventana que surge poner el número de retardos que se quiere añadir. El problema está en que estos retardos no aparecerán en las ventanas de estimación del modelo por MCO o por otro tipo de métodos.

    Otra forma sería *Añadir* $\rightarrow$ *Definir nueva variable*
    En la ventana que surge definir el retardo de la variable deseada, por ejemplo de la variable RD el retardo 4 sería $RD4 = RD(-4)$

    Entre paréntesis el número de retardo deseado se acompaña con el signo menos.

    Vemos los datos de esas variables

Obs	RD	RD4
1960	57,94	
1961	60,59	
1962	64,44	
1963	70,66	
1964	76,83	57,94
1965	80,00	60,59
1966	84,82	64,44
1967	86,84	70,66
1968	88,81	76,83
1969	88,28	80,00

    La primera columna sería RD_t y la segunda RD_{t-4}.

  - Añadir retardos de la variable dependiente o de los regresores y usarlos en la estimación. En este caso se añadirán en la misma ventana de estimación:

    En la ventana *gretl: especificar modelo* elegir *retardos* y en la que surge elegir los retardos que se deseen, bien de forma continua o retardos específicos de forma discontinua, tanto de la variable dependiente como de las variables explicativas que se han definido previamente.

Tema 4

Regresores Estocásticos

Clases Magistrales

En las clases magistrales del tema de Regresores Estocásticos vamos a relajar una de las hipótesis básicas de trabajo en el Modelo de Regresión Lineal General. Vamos a considerar que la matriz de regresores X es estocástica. Bajo este nuevo panorama hemos de revisar las propiedades del estimador de MCO y cómo realizar inferencia válida.

En un contexto de regresores estocásticos la validez del estimador MCO depende de la existencia o no de correlación entre X y la perturbación del modelo. Si esta correlación existe el estimador de MCO no es consistente y nos vemos en la necesidad de proponer un estimador alternativo que sí será consistente, el estimador de Variables Instrumentales, VI. Es claro que es fundamental saber cuándo existe correlación estadísticamente significativa entre la perturbación y el o los regresores estocásticos. Para ello utilizaremos el contraste de Hausman.

Dedicaremos un total de siete clases magistrales a analizar la importancia de que los regresores sean estocásticos y utilizar el estadístico de Hausman; a desarrollar el estimador de VI, probar sus propiedades y obtener en la práctica al estimador así como a mostrar cómo hacer inferencia en este nuevo marco de trabajo. Aproximadamente, tres horas se dedican a resolver ejercicios. Se resolverán las dudas surgidas y se realizaran preguntas al alumno sobre los contenidos vistos, en muchas ocasiones en forma de preguntas cortas que se recogen y evalúan convenientemente.

Competencias a trabajar en estas sesiones:

1. Comprender la importancia de los supuestos empleados en la especificación de un modelo econométrico básico para poder proponer y emplear supuestos más realistas.
2. Diferenciar distintos métodos de estimación y evaluar su uso de acuerdo a las características de las variables económicas de interés para obtener resultados fiables.

Al final de este tema deberíais ser capaces de:

1. Explicar por qué si X es estocástica, el estimador MCO no es lineal en u.
2. Explicar las consecuencias de la no linealidad del estimador de MCO.
3. Explicar qué implicaciones tiene en el estimador de MCO, y sus propiedades, el que los regresores están correlacionados con el término de perturbación del modelo de interés.
4. Describir las propiedades de una variable instrumental. Conocer para qué son necesarias.
5. Estimar utilizando el estimador de Variables Instrumentales bajo los dos casos posibles: cuando el número de instrumentos es igual, o cuando es mayor, al número de variables que lo necesitan.
6. Utilizar el estimador de VI para realizar inferencia.
7. Utilizar el estadístico de Hausman para detectar correlación entre regresores estocásticos y perturbación.

Bibliografía Recomendada:

Al final del tema tenéis recogida la bibliografía correspondiente. En particular os recomendamos leer los capítulos correspondientes a la bibliografía básica detallados a continuación:

- Greene, W. (1998), cap. 4, cap. 6, cap. 9 y cap. 17.
- Ramanathan, R. (2002), cap. 2, cap. 10 y cap 13.
- Wooldridge, J. M. (2003), cap. 5, cap. 9. y cap. 15.

Y para profundizar, podéis leer los capítulos detallados a continuación correspondientes a la bibliografía complementaria:

- Alonso, A. et al. (2005), cap. 8 y cap. 11.
- Gujarati, D. (2004), cap. 13 y cap. 17.
- Johnston, J. (1984), cap. 9 y cap. 10.
- Maddala, G. S. (1996), cap. 9 y cap. 11.
- Novales, A. (1993), cap. 2, cap. 9 y cap. 10.

4.1. Introducción

En los temas anteriores hemos mantenido el supuesto básico sobre los regresores, de que X era una matriz de variables explicativas no estocásticas o fijas. Este supuesto es apropiado para experimentos controlados o de laboratorio, para variables como la tendencia o variables ficticias, o cuando trabajamos condicionando a una muestra dada. En este tema relajaremos dicho supuesto para adecuarnos a situaciones en las que este supuesto no sea razonable.

Objetivo del tema: Relajar la hipótesis básica de que X es una matriz de variables no estocásticas considerando que X es una matriz estocástica. Para ello basta con que uno de los regresores incluidos en X sea estocástico.

En este tema analizaremos si los métodos de estimación e inferencia vistos hasta ahora son válidos cuando X es estocástica. En caso de que no sea así analizaremos qué métodos alternativos están disponibles.

Si X es una matriz de regresores estocásticos, el estimador MCO de β, es una función estocástica no lineal de X y u y por tanto, sus propiedades dependen de la distribución conjunta de ambas variables. Esto dificultará, en general, el conocimiento de las propiedades en muestras finitas de este estimador y sus propiedades asintóticas dependerán de la relación entre X y u. En consecuencia también se verá afectada la inferencia.

Para ilustrar alguna de estas situaciones vamos a considerar los siguientes ejemplos:

Ejemplo 4.1 Sea el siguiente modelo de regresión para explicar la productividad de un trabajador, Y_t, medida como número de piezas realizadas en el día:

$$Y_t = \beta_1 + \beta_2 X_t^\star + u_t^\star \qquad u_t^\star \sim N(0, \sigma^2_{u^\star}) \qquad t = 1, 2, \ldots, T$$

siendo $X_t^\star$ la habilidad de un trabajador, variable no observable ya que es difícil de cuantificar. En su lugar observamos la variable X_t años de experiencia del trabajador en el puesto de trabajo, y vamos a utilizarla para aproximar la habilidad del trabajador[1]. Es probable que cuanto más antigüedad tenga en el puesto el trabajador, más hábil sea y viceversa, así que parece sensato suponer que existe relación entre ambas variables. Sin embargo, la aproximación no será exacta, conllevará un error de medida. Será tal que:

$$X_t = X_t^\star + v_t \qquad v_t \sim N(0, \sigma^2_v) \qquad Cov(u_t^\star, v_t) = 0 \quad \forall t \qquad t = 1, 2, \ldots, T$$

donde v_t es una v.a. que recoge el error de medida en t y se le supone independiente de $u_t^\star$. En esta situación, X_t es una v.a. aunque consideremos a $X_t^\star$ como no estocástica. El modelo estimable en términos de la variable observable sería:

$$\begin{aligned}
Y_t &= \beta_1 + \beta_2 (X_t - v_t) + u_t^\star \\
Y_t &= \beta_1 + \beta_2 X_t + (u_t^\star - \beta_2 v_t) \\
Y_t &= \beta_1 + \beta_2 X_t + u_t
\end{aligned}$$

[1]A una variable que aproxima a otra en la literatura Econométrica se le denomina variable “proxy”.

llamamos $u_t = (u_t^\star - \beta_2 v_t)$, en este caso u_t es una función de $u_t^\star$ y v_t, luego de $X_t^\star$; por ello, no tendría sentido realizar un análisis condicionado a unos valores fijos de X matriz de regresores. En el modelo estimable

$$E(X_t u_t) = E[(X_t^\star + v_t)(u_t^\star - \beta_2\, v_t)] = -\beta_2 \sigma_v^2 \neq 0$$

por lo que, el Teorema de Mann y Wald no se cumple. El estimador MCO será inconsistente y es necesario encontrar un método de estimación consistente alternativo. Este caso lo analizaremos en profundidad más adelante y es conocido en la literatura como errores en variables. En este supuesto la variable medida con error es la variable explicativa.

Ejemplo 4.2 Supongamos que se quiere estimar los coeficientes de la siguiente ecuación de demanda de un bien:

$$Q_t = \beta_1 + \beta_2 P_t + u_t \qquad t = 1, 2, \ldots, T$$

donde Q es la cantidad demandada y P es el precio. Dado que en el momento t observamos la cantidad y precio de equilibrio, ambas variables se determinan simultáneamente en el mercado. Luego, tanto Q como P son variables endógenas. Si en t se produce un shock en la demanda del bien debido, por ejemplo, a un cambio en los gustos de los consumidores, al ser recogido por u_t, se genera un cambio tanto en la cantidad demandada como en el precio. En este contexto dado que las variables se determinan simultáneamente ambas son aleatorias. Este es otro ejemplo donde la matriz de regresores es estocástica y no tiene sentido realizar el análisis condicionado a valores fijos de $P_t \quad t = 1, 2, \ldots, T$, dado que P_t se determina simultáneamente a Q_t. Volveremos a retomar este ejemplo más adelante.

Ejemplo 4.3 Supongamos un modelo con dinámica en la parte sistemática porque incluye como regresor a la variable endógena retardada. Por ejemplo, si estudiamos la función de consumo es muy posible que estemos dispuestos a aceptar que, además de la renta actual, R_t un regresor, posiblemente relevante, para explicar el consumo actual, C_t sea el consumo del período anterior, C_{t-1}:

$$C_t = \beta_1 + \beta_2 C_{t-1} + \beta_3 R_t + u_t \qquad t = 2, \ldots, T$$

En este caso la matriz de regresores del modelo es estocástica ya que incluye como regresor a un retardo de la variable endógena, variable estocástica y no tiene sentido realizar el análisis condicionado a valores fijos.

Es necesario revisar los resultados obtenidos para el MRLG bajo el supuesto de que X es una matriz de regresores estocásticos. Comenzaremos revisando la validez del estimador MCO bajo diferentes supuestos de relación entre X estocástica y u.

4.2. Propiedades del estimador MCO

En presencia de regresores estocásticos el método de estimación y las propiedades de los estimadores dependen de la relación entre u y X. Estudiaremos las propiedades del estimador MCO en las siguientes situaciones:

1. Independencia entre regresor y error.
2. Incorrelación contemporánea entre regresor y error.
3. Correlación entre regresor y error.

4.2.1. Independencia entre regresor y error

Sea $Y = X\beta + u$ donde:

- X es una matriz estocástica, (alguno de sus regresores es una v.a.) con una determinada función de densidad $f(X)$, es decir, X toma diferentes valores con diferentes probabilidades.
- $u \sim N(0, \sigma_u^2 I_T)$
- X y u se distribuyen independientemente, es decir, $E(X'u) = E(X')E(u) = 0$.

Vamos a buscar las propiedades del estimador MCO cuando X y u son independientes.

• Propiedades en muestras finitas para valores de X no condicionados:

- **Linealidad**:
  $\hat{\beta}_{MCO} = \beta + (X'X)^{-1}X'u$ función no lineal de X y u.
  El estimador $\hat{\beta}_{MCO}$ ya no es una combinación lineal de las perturbaciones, sino que es una función estocástica no lineal de X y u, por lo que, sus propiedades dependen de la distribución conjunta de éstas.

- **Insesgadez**:
  Dado que X y u son independientes y $E(u) = 0$ por hipótesis básica, tenemos:

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

  por tanto, $\hat{\beta}_{MCO}$ es insesgado si X y u son independientes y $E[(X'X)^{-1}X']$ existe y es finito[2].

- **Matriz de varianzas y covarianzas**:

  $$\begin{aligned} V(\hat{\beta}_{MCO}) = E(\hat{\beta}_{MCO} - \beta)(\hat{\beta}_{MCO} - \beta)' &= E[\,(X'X)^{-1}X'\,uu'\,X(X'X)^{-1}\,] = \\ &= E_X\{E_{u|X}[\,(X'X)^{-1}X'\,uu'\,X(X'X)^{-1}\,]\} = \\ &= E_X\{(X'X)^{-1}X'\,E_{u|X}(uu')\,X(X'X)^{-1}\} = \\ &= E_X\{(X'X)^{-1}X'\,\sigma_u^2 I_T\,X(X'X)^{-1}\} = \\ &= \sigma_u^2\,E_X\{(X'X)^{-1}\} \end{aligned}$$

  Luego si $\exists\, E_X\{(X'X)^{-1}\}$, $\hat{\beta}_{MCO}$ sigue siendo el estimador insesgado de mínima varianza[3].

[2] Para demostrar esta propiedad hemos utilizado los siguientes resultados estadísticos: $E(a) = E_b[E_{a|b}]$ y $E_{WX} = E_X[E_{W|X}]$ ya que $E(\hat{\beta}_{MCO}) = E_X[E(\hat{\beta}_{MCO}|X)] = E_X[\beta + (X'X)^{-1}X'E(u|X)] = \beta$.

[3] Siendo:

Un estimador insesgado de $V(\hat{\beta}_{MCO})$ donde σ_u^2 y $E_X(X'X)^{-1}$ son los dos elementos desconocidos es:

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

- **Distribución e inferencia en muestras finitas**:
  El estimador $\hat{\beta}_{MCO}$ es una combinación no lineal de X y u y por tanto, no tiene porqué tener una distribución normal, incluso aunque X e u la tengan. Como consecuencia no tenemos garantizado que $\hat{\beta}_{MCO}$ siga una distribución normal y por tanto, los estadísticos t y F no tienen una distribución exacta conocida por ello, la inferencia para tamaños de muestra pequeños no es válida.

- **Conclusión**: La eliminación del supuesto de que X es no estocástica sustituyéndolo por X estocástica, pero independiente de u no altera las propiedades deseables ni la variabilidad de la estimación mínimo cuadrática.

• Propiedades en muestras grandes:
Bajo los supuestos habituales, y si además, se satisface que $\text{plim}\left(\frac{1}{T}X'X\right) = Q$ finita y no singular el estimador de MCO es consistente y es posible derivar sus propiedades asintóticas utilizando el Teorema de Mann y Wald y el Teorema de Cramer. Las condiciones del Teorema de Mann y Wald son:

i) $u_1, u_2, \ldots, u_T$ v. a tal que $u_t \sim iid(0, \sigma_u^2)$

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

ii) $\text{plim}\left(\frac{1}{T}X'X\right) = Q$ finita, simétrica y no singular

Verificamos que se cumplen. La condición iii) se cumple siempre, i) se cumple en este contexto ya que estamos suponiendo perturbaciones esféricas y ii) también se cumple ya X y u son independientes luego $E(X'u) = E(X')E(u) = 0$.

Dado que se cumplen i) + ii) + iii) el Teorema de Mann y Wald produce dos resultados:

a) $\text{plim}\left(\frac{1}{T}X'u\right) = 0$

b) $\frac{1}{\sqrt{T}}X'u \xrightarrow{d} N(0, \sigma_u^2 Q)$

El primer resultado nos garantiza la consistencia del estimador MCO y el segundo nos sirve para encontrar su distribución asintótica. Por lo tanto:

E_X el valor esperado de la distribución marginal de X.

$E_{u|X}$ el valor esperado de la distribución condicional de u dado X.

$E_X(X'X)^{-1}$ la matriz de covarianzas poblacional de los regresores calculada en la distribución marginal de X.

- **Consistencia**: $\hat{\beta}_{MCO}$ es un estimador consistente del parámetro β.

$$\text{plim}\,\hat{\beta}_{MCO} = \text{plim}\,\beta + \text{plim}\left(\frac{1}{T}X'X\right)^{-1} \text{plim}\left(\frac{1}{T}X'u\right) = \beta + Q^{-1}\cdot 0 = \beta$$

- **Distribución asintótica**:
Dado que

$$\text{plim}\,\hat{\beta}_{MCO} = \beta \Longrightarrow \hat{\beta}_{MCO} \xrightarrow{p} \beta \Longrightarrow (\hat{\beta}_{MCO} - \beta) \xrightarrow{p} 0$$

el estimador tiene una distribución degenerada en el límite, por lo que, buscamos la distribución asintótica para $\sqrt{T}(\hat{\beta}_{MCO} - \beta)$ tal que

$$\sqrt{T}(\hat{\beta}_{MCO} - \beta) = \left(\frac{1}{T}X'X\right)^{-1}\left(\frac{1}{\sqrt{T}}X'u\right)$$

Utilizando el Teorema de Mann y Wald y el Teorema de Cramer obtenemos[4]:

$$\sqrt{T}(\hat{\beta}_{MCO} - \beta) = \left(\frac{1}{T}X'X\right)^{-1}\left(\frac{1}{\sqrt{T}}X'u\right) \xrightarrow{d} N(0, \sigma_u^2 Q^{-1})$$

de donde

$$\sqrt{T}(\hat{\beta}_{MCO} - \beta) \xrightarrow{d} N(0, \sigma_u^2 Q^{-1})$$

- **Inferencia asintótica**: Dado que se cumple el Teorema de Mann y Wald las propiedades asintóticas se mantienen y tiene sentido la inferencia asintótica. Bajo $H_0 : R\beta = r$, los estadísticos t y F se distribuyen asintóticamente como $N(0,1)$ y χ^2_q respectivamente, si el tamaño de la muestra es suficientemente grande. Por lo tanto, podemos utilizar estas distribuciones asintóticas para aproximar la distribución exacta de los estadísticos. Así, para contrastar q restricciones lineales de la forma:

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

utilizamos el siguiente estadístico con distribución asintótica:

$$(R\hat{\beta} - r)'\,[R\,\widehat{V}(\hat{\beta})\,R'\,]^{-1}\,(R\hat{\beta} - r) \overset{d,H_0}{\longrightarrow} \chi^2_q$$

luego

$$(R\hat{\beta}_{MCO} - r)'[R\,\hat{\sigma}^2_u\,(X'X)^{-1}R']^{-1}(R\hat{\beta}_{MCO} - r) \overset{d,H_0}{\longrightarrow} \chi^2_q$$

Si el estadístico calculado es menor que el valor en tablas de $\chi^2_{q|\alpha}$ no rechazamos la H_0 para un nivel de significación α.

Si q=1 podemos utilizar el siguiente estadístico y distribución asintótica asociada:

$$\frac{R\hat{\beta}_{MCO} - r}{\sqrt{R\,\hat{\sigma}^2_u\,(X'X)^{-1}R'}} \overset{d,H_0}{\longrightarrow} N(0,1)$$

en cuyo caso no rechazamos H_0 si el valor muestral del estadístico es menor que $N(0,1)_{|\frac{\alpha}{2}}$ para un nivel de significatividad α.

---

[4]El Teorema de Cramer dice: Si $Y_T = A_T \cdot Z_T$, plim$A_T = A$ y $Z_T \overset{a}{\sim} N(\mu, \sum)$ entonces $A_T Z_T \overset{a}{\sim} N(A_\mu,\, A\sum A')$. Del Teorema de Mann y Wald hacemos uso del resultado $\frac{X'u}{\sqrt{T}} \xrightarrow{d} N(0, \sigma^2\, Q)$.

Ejemplo 4.4 Por ejemplo si queremos contrastar la significatividad de una de las variables explicativas del modelo contrastamos: $H_0 : \beta_i = 0$ versus $H_a : \beta_i \neq 0$. En este caso q=1 y podemos escribir el siguiente estadístico a utilizar junto con su distribución asintótica,

$$\frac{\hat{\beta}_{i,MCO}}{\widehat{des}(\hat{\beta}_{i,MCO})} \xrightarrow{d,H_0} N(0,1)$$

Donde $\widehat{des}(\hat{\beta}_{i,MCO})$ es la raíz cuadrada del elemento i-ésimo de la matriz $\hat{\sigma}^2(X'X)^{-1}$. Si el valor muestral del estadístico es menor que $N(0,1)_{|\frac{\alpha}{2}}$ no rechazamos la H_0 para un nivel de significación α.

Ejercicio 4.1 En el modelo lineal simple:

$$Y_t = \beta_1 + \beta_2 X_t + u_t \qquad u_t \sim iid(0, \sigma_u^2) \qquad t = 1, 2, \ldots, T$$

donde X_t es una variable estocástica, pero independiente de la perturbación obtener el estimador de MCO y sus propiedades en muestras finitas.

4.2.2. Incorrelación contemporánea entre regresores y error

Sea $Y = X\beta + u$ con $u \sim N(0, \sigma_u^2 I)$, X estocástica tal que X y u no son independientes, pero son incorreladas contemporáneamente, es decir, mantenemos que $E(X_{it}u_t) = 0 \quad i = 1, 2, \ldots, K$, pero no la independencia entre X_{it} y u_t.

En este caso no podemos derivar analíticamente las propiedades en muestras finitas del estimador, pero se sigue cumpliendo el Teorema de Mann y Wald y por tanto, podemos mantener las propiedades asintóticas.

- $\hat{\beta}_{MCO} = \beta + (X'X)^{-1}X'u$ es una función no lineal de X y u, ambas estocásticas.
- En general es un estimador sesgado, $E(\hat{\beta}_{MCO}) = \beta + E[(X'X)^{-1}X'u]$ y $E[(X'X)^{-1}X'u]$ puede ser distinto de cero ya que X y u no son independientes.
- El cálculo analítico de su matriz de varianzas y covarianzas es complicado dada la no linealidad del estimador en X y u. También lo es el cálculo analítico de $E(\hat{\beta}_{MCO})$.
- Dado que el estimador es no lineal no conocemos su distribución exacta, no sigue una distribución normal aún en el caso de que $X_{it}\ \forall i\ \forall t$ y u la sigan. Como consecuencia los estadísticos t y F no tienen distribución exacta conocida.
- Las propiedades asintóticas de consistencia y distribución asintótica se mantienen ya que podemos aplicar el Teorema de Mann y Wald porque se satisfacen las condiciones de este teorema, $u \sim (0, \sigma_u^2 I)$ y $E(X'u) = 0$, junto con $\text{plim}(\frac{1}{T}X'X) = Q$.

Ejercicio 4.2 En el modelo:

$$Y_t = \beta X_t + u_t \qquad t = 1, 2, \ldots, T \qquad u_t \sim N(0, \sigma_u^2)$$

donde X_t es una v.a. no independiente de u_t, pero tal que $E(X_t u_t) = 0$. Demostrar las propiedades en muestras finitas y asintóticas del estimador MCO del coeficiente β.

4.2.3. Correlación entre regresores y error

En ocasiones la hipótesis $E(X_{it}, u_t) = 0 \ \forall i, \forall t$ no es válida. En este caso los estimadores MCO no son ni siquiera consistentes. Hay cuatro puntos a solucionar en relación a este tema:

1. ¿Cómo aparecen las correlaciones entre X y u? Por ejemplo en las siguientes situaciones:
   a) Cuando la variable explicativa está medida con error.
   b) Si el modelo tiene un problema de omisión de variable relevante y la variable omitida está correlada con los regresores. Esta correlación aparecerá vía la perturbación ya que la perturbación recoge las variables omitidas.
   c) En modelos de ecuaciones simultáneas, por ejemplo en el modelo de oferta y demanda donde P y Q se determinan simultáneamente.
   d) En modelos con variable endógena retardada como regresor y perturbación autocorrelada.
2. ¿Qué importancia puede tener este problema? En realidad depende del caso concreto, pero de forma general podemos decir que el estimador de MCO pierde las propiedades en muestras pequeñas y grandes.
3. ¿Cómo podemos detectar que existe el problema? Usando test de contraste que sean capaces de detectar la correlación entre X y u. Por ejemplo, el contraste de Hausman está diseñado para ello.
4. ¿Cómo podemos solucionar el problema? Si la existencia de correlación entre regresores y perturbación se debe a un problema de omisión de variable relevante debemos intentar especificar correctamente el modelo. En el resto de casos tendremos que buscar un método de estimación alternativo a MCO que tenga buenas propiedades, aunque éstas se logren sólo en muestras grandes.

Es claro que este es el caso relevante de los tres descritos. Comenzaremos viendo algunos ejemplos en los que X y u están correlados.

Ejemplo 4.5 **Variable exógena medida con error.**
Vamos a profundizar en el Ejemplo 5.1 donde analizábamos la productividad de un trabajador, Y, utilizando como variable explicativa su habilidad, X. Sea el modelo de regresión en términos de la variable no observable $X_t^\star \equiv$ productividad:

$$Y_t = \beta_1 + \beta_2 X_t^\star + u_t^\star \qquad t = 1, 2, \ldots, T \qquad u_t^\star \sim iid(0, \sigma_{u^\star}^2)$$

donde la variable $X_t^\star$ es una variable no estocástica y no observable, ya que no se puede cuantificar. Sin embargo, disponemos de X_t los años de antigüedad en la empresa, variable observable, y vamos a utilizarla para aproximar a $X_t^\star$. Definimos $X_t = X_t^\star + v_t$, tal que X_t es una variable observable, pero aleatoria ya que incorpora el efecto de v_t, y lo es, aún en el caso de que $X_t^\star$ sea no estocástica o fija. Además, hacemos las siguientes hipótesis:

$$v_t \sim iid(0, \sigma_v^2) \qquad Cov(u_t^\star, v_t) = Cov(u_t^\star, v_s) = Cov(u_s^\star, v_t) = 0$$

En esta situación el modelo que efectivamente se estima es el siguiente:

$$Y_t = \beta_1 + \beta_2 X_t + u_t \qquad \text{con} \qquad u_t = u_t^\star - \beta_2 v_t$$

- Propiedades de la nueva perturbación $u_t^\star$:

$$\begin{aligned}
E(u_t) &= E(u_t^\star - \beta_2 v_t) = E(u_t^\star) - \beta_2 E(v_t) = 0 \\
Var(u_t) &= E(u_t - E(u_t))^2 = E(u_t)^2 = E(u_t^\star - \beta_2 v_t)^2 \\
&= E(u_t^{\star 2}) + \beta_2^2 E(v_t^2) - 2\beta_2 E(u_t^\star v_t) = \sigma_{u^\star}^2 + \beta_2^2 \sigma_v^2 - 2\beta_2 \cdot 0 \\
&= \sigma_{u^\star}^2 + \beta_2^2 \sigma_v^2 \quad \text{luego homocedástica} \\
Cov(u_t, u_s) &= E[(u_t - E(u_t))(u_s - E(u_s))] = E(u_t^\star - \beta_2 v_t, u_s^\star - \beta_2 v_s) = \\
&= E(u_t^\star u_s^\star) - \beta_2 E(v_t u_s^\star) - \beta_2 E(u_t^\star v_s) + \beta_2^2 E(v_t v_s) = 0, \text{ no autocorrelada}
\end{aligned}$$

- Además, necesitamos conocer la relación entre el regresor y el error, es decir, entre X_t y u_t:

$$\begin{aligned}
Cov(X_t, u_t) &= E(X_t u_t) = E((X_t^\star + v_t)(u_t^\star - \beta_2 v_t)) = \\
&= E(X_t^\star u_t^\star) + E(v_t u_t^\star) - \beta_2 E(X_t^\star v_t) - \beta_2 E(v_t^2) = -\beta_2 \sigma_v^2
\end{aligned}$$

  ya que al ser $X_t^\star$ una variable no estocástica $E(X_t^\star u_t) = E(X_t^\star v_t) = 0$. Al ser la covarianza entre X_t y u_t distinta de cero existe correlación contemporánea entre la variable exógena y la perturbación. Vemos que la correlación depende de la varianza del error de medida, cuanto peor es la proxy o variable que utilizamos para aproximar la variable no observable, mayor es en términos absolutos la correlación. El signo depende del signo del coeficiente β_2. El estimador MCO de β_2 no es consistente, además la diferencia entre $plim\hat{\beta}$ y β depende directamente de esta correlación.

Conclusión: un error de medida en la variable exógena tal que $E(X_t u_t) \neq 0$ implica que los estimadores MCO son sesgados e inconsistentes[5]. El Teorema de Mann y Wald no se satisface. Los coeficientes del modelo deberían ser estimados por un método de estimación que produzca estimadores consistentes. Para conseguirlo propondremos el Método de Variables Instrumentales. Lo veremos en la sección siguiente.

[5]En el Anexo 4.2. Errores de medida en las variables, se han desarrollado con detalle las consecuencias en los estimadores MCO de los coeficientes de un modelo, de que el error de medida se produzca en la variable endógena y simultáneamente, en la variable endógena y en la exógena.

Ejemplo 4.6 Omisión de variable relevante.
Sea el modelo correctamente especificado:

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

donde $v_t \sim iid(0, \sigma_v^2)$, y suponemos $E(X_{2t}v_t) = 0$ y $E(X_{3t}v_t) = 0$. Pero el modelo que se estima es el siguiente:

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

tal que $u_t = \beta_3 X_{3t} + v_t$ y donde:

$$\begin{array}{l} E(X_{2t}u_t) = E[X_{2t}\,(\beta_3 X_{3t} + v_t)] = \beta_3 E(X_{2t}X_{3t}) + E(X_{2t}v_t) = \beta_3 E(X_{2t}X_{3t}) \\ E(u_t) = \beta_3 E(X_{3t}) + E(v_t) = \beta_3 E(X_{3t}) \end{array}$$

dado que X_{3t} es una variable relevante, $\beta_3 \neq 0$ luego:

$$\begin{array}{lll} E(X_{2t}u_t) \neq 0 & \text{si} & E(X_{2t}X_{3t}) \neq 0 \\ E(u_t) \neq 0 & \text{si} & E(X_{3t}) \neq 0 \end{array}$$

Por tanto, X y u son independientes y $E(u|X) = E(u) = 0$ si y solo si X_{2t} y X_{3t} son variables aleatorias independientes y $E(X_{3t}) = 0\ \forall t$. En el resto de los posibles supuestos no lo son. Por tanto, en el caso contemplado, donde se omite la variable X_{3t} correlada con X_{2t}, en el modelo especificado a estimar $E(X_{2t}u_t) \neq 0$ y el estimador MCO es inconsistente.

Por ejemplo, supongamos que estamos estudiando la función de salarios y que proponemos como variable explicativa el nivel de educación del individuo medido en años de educación:

$$WAGE_i = \beta_1 + \beta_2 EDUC_i + u_i$$

Si este es nuestro modelo a estimar es claro que estamos omitiendo factores que determinan el salario aparte del nivel educativo como la experiencia, la habilidad, la motivación. Podemos esperar que los individuos más motivados y/o con más talento tengan a su vez mayor formación medida en más años de educación y por lo tanto, sería lógico pensar que $E(EDUC_i\, u_i) \neq 0$, luego el estimador MCO de β_1 y β_2 sería sesgado e inconsistente.

En términos de la especificación del modelo es importante remarcar que el modelo debe estar correctamente especificado antes de proponer un método de estimación alternativo a MCO. En el ejemplo que nos ocupa medir la experiencia es sencillo, pero variables como la habilidad o la motivación no son sencillas de medir. En general y en el mejor de los casos seremos capaces de aproximar dichas variables con un error de medida, que como ya se ha apuntado anteriormente, y se desarrollará en esta sección, conllevan problemas de correlación entre regresor y perturbación y por tanto, la inconsistencia del estimador de MCO.

Ejemplo 4.7 Simultaneidad.
Sea el modelo formado por las siguientes dos ecuaciones:

$$Y_t = \beta_1 + \beta_2 X_t + u_t \qquad t = 1, 2, \ldots, T \tag{4.1}$$

$$X_t = \gamma_1 + \gamma_2 Y_t + v_t \qquad t = 1, 2, \ldots, T \tag{4.2}$$

tal que:

$$\left[\begin{array}{c} u_t \\ v_t \end{array}\right] \sim N\left(\left[\begin{array}{c} 0 \\ 0 \end{array}\right], \left[\begin{array}{cc} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{array}\right]\right)$$

Es decir, u_t y v_t son v.a. normales. Estamos interesados en estimar β_1 y β_2 en (4.2), para ello queremos saber si X y u son independientes y/o incorreladas. Dado que $E(u_t) = 0 \quad \forall t$:

$$\begin{aligned} E(X_t u_t) &= E[(\gamma_1 + \gamma_2(\beta_1 + \beta_2 X_t + u_t) + v_t)u_t] = \\ &= \gamma_1 E(u_t) + \gamma_2\beta_1 E(u_t) + \gamma_2\beta_2 E(X_t u_t) + \gamma_2 E(u_t^2) + E(v_t u_t) = \\ &= \gamma_2\beta_2 E(X_t u_t) + \gamma_2\sigma_u^2 + \sigma_{uv} \end{aligned}$$

resolviendo:

$$E(X_t u_t) = \frac{1}{1 - \gamma_2\beta_2}(\sigma_{uv} + \gamma_2\sigma_u^2)$$

con lo que, $E(X_t u_t) \neq 0$ si $\gamma_2 \neq 0$ y/o $\sigma_{uv} \neq 0$. En este ejemplo el estimador MCO es inconsistente. Además, el Teorema de Mann y Wald no se satisface. Deberíamos buscar un estimador alternativo a MCO y que sí fuese consistente.

Ejemplo 4.8 Variable endógena retardada como regresor y perturbación autocorrelada.

Supongamos un modelo con dinámica en la parte sistemática porque incluye como regresor a la variable endógena retardada y con dinámica en la perturbación ya que esta sigue un proceso autorregresivo de primer orden.

$$Y_t = \beta_1 + \beta_2 Y_{t-1} + \beta_3 X_t + u_t \qquad u_t = \rho u_{t-1} + \epsilon_t \qquad \epsilon_t \sim iid(0, \sigma_\epsilon^2)$$

En este caso $E(Y_{t-1}u_t) \neq 0$ luego $E(X'u) \neq 0$ y el estimador MCO no es consistente. Este caso se analizará en profundidad en el tema siguiente, Modelos Dinámicos.

En los ejemplos anteriores el estimador MCO es inconsistente ya que la matriz de regresores está correlada con la perturbación. Un estimador alternativo y consistente es el estimador de Variables Instrumentales.

4.3. Estimador de Variables Instrumentales

En modelos donde existe correlación entre regresores y error el estimador MCO es inconsistente y sesgado ya que $E(X'u) \neq 0$. El procedimiento para obtener estimadores consistentes en un modelo de este tipo es el Método de Variables Instrumentales, VI. El método de Variables Instrumentales se basa en encontrar un conjunto de K variables, $Z_{1t}, Z_{2t}, \ldots, Z_{Kt}$, que se llaman instrumentos o variables instrumentales tales que, satisfacen las siguientes condiciones:

1. Cada uno de los instrumentos está incorrelacionado con el término de error de la ecuación de interés, $E(Z_{jt}u_t) = 0 \quad \forall\, t, \forall\, j,\ j = 1, 2, \ldots, K$.

2. Los instrumentos están correlacionados con la variable para la cual hacen de instrumento, $E(Z_{jt}X_{it}) \neq 0$. En cuanto a esta correlación, debe existir, pero no puede ser muy importante pues en este caso, si lo fuera, $E(Z_{jt}u_t) \neq 0$ y Z_{jt} no serviría de instrumento.

Junto con las dos anteriores deben de cumplirse dos condiciones más:

3. $(Z'X)$ es una matriz no singular, es decir, es invertible.
4. $\text{plim}(\frac{1}{T}Z'Z) = Q_{ZZ}$ matriz finita.

El estimador de variables instrumentales, dada una matriz de instrumentos Z se define[6]:

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

Una forma intuitiva de pensar en cómo obtener este estimador podemos encontrarla a partir de las ecuaciones normales del estimador de MCO. Cualquiera de los modelos anteriores podemos escribirlos de la forma matricial habitual $Y = X\beta + u$. Premultiplicando el modelo por X' tenemos:

$$X'Y = X'X\beta + X'u$$

Si $E(X'u) = 0$ entonces $E(X'Y) = E(X'X)\beta + 0$ luego $E(X'Y) - E(X'X)\beta = 0$ que sería la expresión de las ecuaciones normales en términos poblacionales. En términos muestrales podemos escribir $X'Y - X'X\hat{\beta} = 0$ luego $X'(Y - X\hat{\beta}) = 0$ y $X'\hat{u} = 0$.

Si plim $\left(\frac{1}{T}X'u\right) = 0$ los estimadores así obtenidos son consistentes. Por tanto, parece que sugiere que cuando plim $\left(\frac{1}{T}X'u\right) \neq 0$ en vez de premultiplicar por X' lo hagamos por una matriz Z' que satisface plim $\left(\frac{1}{T}Z'u\right) = 0$ tal que, podemos definir el estimador consistente $\hat{\beta}_{VI} = (Z'X)^{-1}Z'Y$ dada la matriz Z que cumple $E(Z'u) = 0$. Llamamos a Z matriz de instrumentos. De las condiciones citadas anteriormente notar que las condiciones 1 y 2 garantizan la consistencia del estimador, mientras que las condiciones 3 y 4 garantizan que el estimador β_{VI} definido como $\hat{\beta}_{VI} = (Z'X)^{-1}Z'Y$ se pueda evaluar en la muestra.

En general en la matriz de regresores X sólo habrá unas variables que no satisfagan la condición $E(X_{it}u_t) = 0$ y son estas variables las que necesitan de variables instrumentales. Para el resto de variables, las no correladas con la perturbación, el mejor instrumento es la propia variable. Es decir, las matrices Z y X tendrán en común aquellas columnas correspondientes a las variables incorrelacionadas con el término de error. Notar que necesariamente Z y X deben tener el mismo número de columnas ya que en otro caso $(Z'X)$ no sería cuadrada.

Antes de ver ejemplos de cómo buscar los instrumentos vamos a ver las propiedades del estimador.

4.3.1. Propiedades del estimador de Variables Instrumentales

- **Linealidad:** Dado que X es una matriz estocástica el estimador de VI no es lineal en u,

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

En la expresión del estimador aparecen X y u ambas variables aleatorias, por tanto, $\hat{\beta}_{VI}$ es una función no lineal de X y u.

[6] En el Anexo 4.3. Estimador de Variables Instrumentales, aparece derivado formalmente este estimador de VI.

- **Insesgadez:** En general el estimador será sesgado ya que $E[(Z'X)^{-1}(Z'u)] \neq 0$. En la expresión $E[(Z'X)^{-1}(Z'u)]$ hemos de notar dos cosas. Por un lado, el requisito impuesto para obtener el estimador de Variables Instrumentales, $\hat{\beta}_{VI}$, es $\text{plim}\left(\frac{1}{T}Z'u\right) = 0$ y no la independencia entre Z y u. Por otro lado, aunque Z y u fuesen independientes la expresión anterior participa de X, matriz estocástica no independiente de u. Luego $E[(Z'X)^{-1}(Z'u)] \neq 0$ y

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

Por tanto, en muestras finitas el estimador de Variables Instrumentales es un estimador no lineal y en general sesgado. Su distribución exacta es desconocida. En muestras grandes el estimador es consistente y tiene distribución asintótica conocida.

- **Consistencia del estimador de Variables Instrumentales y Distribución asintótica:** Vamos a demostrar la consistencia del estimador y a buscar su distribución asintótica utilizando el Teorema de Mann y Wald aplicado a Z y u. Sea X matriz de variables explicativas de orden $(T \times K)$ y Z una matriz de instrumentos de orden $(T \times K)$. Si se cumplen las siguientes condiciones:

  i) $u \sim (0, \sigma_u^2 I)$.
  ii) $E(Z'u) = 0$
  iii) $\text{plim}\left(\frac{1}{T}Z'Z\right) = Q_{ZZ}$ finita, simétrica y definida positiva.
  iv) $\text{plim}\left(\frac{1}{T}Z'X\right) = Q_{ZX}$ finita, no singular

entonces se tiene por el Teorema de Mann y Wald los siguientes resultados:

  a) $\text{plim}\left(\frac{1}{T}Z'u\right) = 0$
  b) $\frac{1}{\sqrt{T}} Z'u \xrightarrow{d} N(0, \sigma_u^2 Q_{ZZ})$

Utilizando el resultado a) podemos demostrar la consistencia del estimador $\hat{\beta}_{VI}$:

$$\text{plim}\, \hat{\beta}_{VI} = \beta + \text{plim}\left(\frac{1}{T}Z'X\right)^{-1} \text{plim}\left(\frac{1}{T}Z'u\right) = \beta + Q_{ZX}^{-1} \cdot 0 = \beta$$

El estimador de Variables Instrumentales es consistente y la condición importante es la ausencia de correlación entre los instrumentos y el término de error del modelo de interés.

Utilizando el resultado b) junto con el Teorema de Cramer obtenemos su distribución asintótica[7]:

$$\sqrt{T}(\hat{\beta}_{VI} - \beta) = \left(\frac{1}{T}Z'X\right)^{-1}\left(\frac{1}{\sqrt{T}}Z'u\right) \xrightarrow{d} N\left(0\,,\, \sigma_u^2 \cdot Q_{ZX}^{-1} \cdot Q_{ZZ} \cdot (Q_{ZX}^{-1})'\right)$$

[7]Este resultado puede generalizarse al caso en el que la perturbación tiene autocorrelación, pero en ese caso la matriz de varianzas y covarianzas del estimador VI no es la derivada anteriormente.

El resultado anterior justifica que en muestras grandes se utilice como matriz de covarianzas asintótica del estimador de variables instrumentales a:

$$\begin{aligned}\widehat{V}(\hat{\beta}_{VI}) &= \frac{\hat{\sigma}_u^2}{T}\left(\frac{Z'X}{T}\right)^{-1}\frac{Z'Z}{T}\left[\left(\frac{Z'X}{T}\right)^{-1}\right]' = \\ &= \hat{\sigma}_{VI}^2\,(Z'X)^{-1}\,Z'Z\,((Z'X)^{-1})'\end{aligned}$$

donde se utilizan las matrices de momentos muestrales $\frac{Z'X}{T}$ y $\frac{Z'Z}{T}$ para aproximar sus límites respectivos Q_{ZX} y Q_{ZZ} y

$$\hat{\sigma}_{VI}^2 = \frac{(Y - X\hat{\beta}_{VI})'\,(Y - X\hat{\beta}_{VI})}{T-K}$$

es un estimador consistente de σ_u^2. Notar que el denominador de $\hat{\sigma}_{VI}^2$ es $T-K$, pero podríamos haber considerado T dado que el estimador es asintótico y por lo tanto, invariante a que el denominador sea T ó $T-K$.

4.3.2. Cómo buscar los instrumentos

Con respecto a los instrumentos sabemos que éstos tienen que cumplir las siguientes condiciones:

1. Estar incorrelacionados con la perturbación del modelo de interés, $E(Z_{jt}u_t) = 0$.
2. Estar correlacionados con la variable para la cuál hacen de instrumento, $E(Z_{jt}X_{it}) \neq 0$.

Además, la matriz $(Z'X)$ debe ser una matriz no singular con $rg(Z) = K < T$, es decir, $\exists (Z'X)^{-1}$. En la práctica existen dos situaciones diferentes en la búsqueda de instrumentos:

1. Que el número de instrumentos disponibles coincida con el número de variables que necesiten instrumento.
2. Que el número de instrumentos disponibles sea mayor que el número de variables que necesiten instrumento.

Número de instrumentos disponibles igual al número de variables explicativas que lo necesitan

Supongamos que el número de instrumentos de que se dispone es igual al número de variables explicativas que necesitan instrumento. En este caso cada instrumento constituye una variable instrumental para su correspondiente variable explicativa correlada con la perturbación. Para el resto de variables, para las que no necesitan instrumento, su mejor instrumento es ella misma. Los instrumentos deben cumplir los requisitos mencionados anteriormente, no deben estar correlados con la perturbación del modelo de interés y deben de estar correlados con la variable explicativa para la que actúan de instrumento. Además, debe existir $(Z'X)^{-1}$. A continuación veremos un ejemplo teórico que nos servirá para mostrar el desarrollo matricial.

Ejemplo 4.9 Sea el modelo:

$$Y_t = \beta_1 + \beta_2 X_t + u_t \quad u_t \sim iid(0, \sigma_u^2) \quad t = 1, 2, \ldots, T$$

$$\text{tal que} \quad \begin{array}{l} X_t = 0,7X_{t-1} + v_t \quad v_t \sim iid(0, \sigma_v^2) \\ E(u_t v_s) = 5 \quad \text{si} \quad t = s \\ E(u_t v_s) = 0 \quad \text{si} \quad t \neq s \end{array}$$

En este ejercicio la matriz $X = [1 \quad \mathbf{X}_t\,]$ es estocástica ya que el regresor X_t es un regresor estocástico[8]. La constante no crea problemas de correlación con u_t ya que es un regresor no estocástico, $E(1\,u_t) = E(u_t) = 0$. Sin embargo, el regresor estocástico X_t está correlado contemporáneamente con la perturbación:

$$E(X_t u_t) = E[(0,7X_{t-1} + v_t)\,u_t] = 0,7E(X_{t-1}u_t) + E(v_t u_t) = 0,7 \cdot 0 + 5 = 5$$

Luego $E(X'u) \neq 0$ y el estimador MCO será no lineal y sesgado. En muestras grandes además, será inconsistente. Deberíamos estimar los coeficientes β_1 y β_2 por el Método de Variables Instrumentales. Para ello buscamos un instrumento para X_t, el único regresor correlado con la perturbación. Podemos pensar en $Z_t = X_{t-1}$ dado que X_{t-1} está incorrelado con la perturbación, $E(X_{t-1}u_t) = 0$ y correlado con X_t ya que ésta se genera por un proceso autorregresivo, $E(X_t X_{t-j}) \neq 0 \quad \forall j > 0$. Además, con $Z_t = X_{t-1} \rightarrow rg(Z) = 2 = K$, luego $\exists (Z'X)^{-1}$.

Aplicamos para $Z_t = X_{t-1}$ el estimador de Variables Instrumentales[9]:

$$\underset{((T-1)\times 1)}{Y} = \begin{bmatrix} Y_2 \\ Y_3 \\ \vdots \\ Y_T \end{bmatrix} \qquad \underset{((T-1)\times 2)}{X} = \begin{bmatrix} 1 & X_2 \\ 1 & X_3 \\ \vdots & \vdots \\ 1 & X_T \end{bmatrix} \qquad \underset{((T-1)\times 2)}{Z} = \begin{bmatrix} 1 & X_1 \\ 1 & X_3 \\ \vdots & \vdots \\ 1 & X_{T-1} \end{bmatrix}$$

El estimador $\hat{\beta}_{VI} = (Z'X)^{-1}(Z'Y)$ es el siguiente[10]:

$$\begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix}_{VI} = \begin{bmatrix} (T-1) & \sum_2^T X_t \\ \sum_2^T X_{t-1} & \sum_2^T X_t X_{t-1} \end{bmatrix}^{-1} \begin{bmatrix} \sum_2^T Y_t \\ \sum_2^T X_{t-1} Y_t \end{bmatrix}$$

[8] En adelante la notación utilizada $X = [1 \quad \mathbf{X}_t\,]$ hace referencia a la siguiente matriz:

$$X = [1 \quad \mathbf{X}_t\,] = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_T \end{bmatrix}$$

[9] Notar en el orden de las matrices que para $Z_t = X_{t-1}$ perdemos una observación al definir el instrumento. Si el instrumento hubiera sido $Z_t = X_{t-2}$ perderíamos dos observaciones y en ese caso las matrices Y, X y Z serían:

$$\underset{((T-2)\times 1)}{Y} = \begin{bmatrix} Y_3 \\ Y_4 \\ \vdots \\ Y_T \end{bmatrix} \qquad \underset{((T-2)\times 2)}{X} = \begin{bmatrix} 1 & X_3 \\ 1 & X_4 \\ \vdots & \vdots \\ 1 & X_T \end{bmatrix} \qquad \underset{((T-2)\times 2)}{Z} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_{T-2} \end{bmatrix}$$

[10] A la hora de trabajar con el estimador de VI es importante notar que la matriz a invertir, $(Z'X)$, no es simétrica.

Su matriz de varianzas y covarianzas estimada se define:

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

$$\hat{\sigma}^2_{VI}\begin{bmatrix} (T-1) & \sum_2^T X_t \\ \sum_2^T X_{t-1} & \sum_2^T X_tX_{t-1} \end{bmatrix}^{-1}\begin{bmatrix} (T-1) & \sum_2^T X_{t-1} \\ \sum_2^T X_{t-1} & \sum_2^T X_{t-1}^2 \end{bmatrix}\left[\begin{bmatrix} (T-1) & \sum_2^T X_t \\ \sum_2^T X_{t-1} & \sum_2^T X_tX_{t-1} \end{bmatrix}^{-1}\right]'$$

siendo:

$$\hat{\sigma}^2_{VI} = \frac{(Y - X\hat{\beta}_{VI})'(Y - X\hat{\beta}_{VI})}{T-K}$$

Ejemplo 4.10 En este ejemplo vamos a volver sobre la estimación de un modelo con variable exógena medida con error y nos va a permitir ilustrar cómo buscar instrumentos de manera adecuada:

Supongamos que queremos estimar el modelo $Y_t = \beta X_t^\star + u_t^\star \quad u_t^\star \sim iid(0, \sigma^2_{u^\star})$ siendo $X_t^\star$ una variable no estocástica e inobservable. Sin embargo, se dispone de la variable observable $X_t = X_t^\star + \epsilon_t$ que nos va a permitir aproximar a la variable inobservable. Además, en cuanto al error de medida, se sabe que $\epsilon_t \sim iid(0, \sigma^2_\epsilon)$ y $Cov(u_t^\star, \epsilon_t) = Cov(u_t^\star, \epsilon_s) = Cov(u_s^\star, \epsilon_t) = 0$. El modelo considerado para estimar en términos de la variable observable X_t es:

$$Y_t = \beta X_t + u_t \qquad t = 1, \ldots, T$$

donde $u_t = (u_t^\star - \beta\epsilon_t)$ y tal que $E(X_t u_t) = -\beta\sigma^2_\epsilon \neq 0$. El estimador de MCO es inconsistente. El estimador consistente de β lo podemos obtener utilizando el estimador de VI, $\hat{\beta}_{VI} = (Z'X)^{-1}Z'Y$ tal que:

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_T \end{bmatrix}; \quad X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_T \end{bmatrix}; \quad Z = \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_T \end{bmatrix};$$

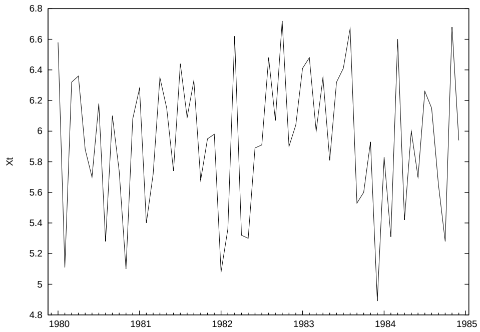
$$\hat{\beta}_{VI} = (Z'X)^{-1}Z'Y = \frac{\sum Z_t Y_t}{\sum Z_t X_t}$$

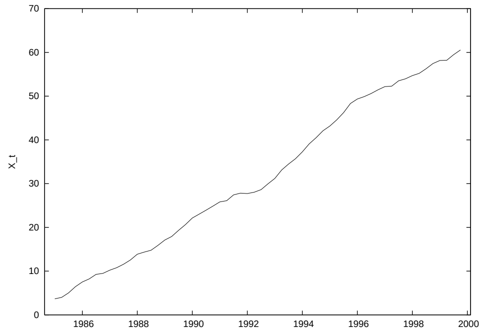
Es fundamental en el estimador de VI la elección del instrumento y ha de cumplir las condiciones explicitadas anteriormente. Vamos a proponer algunos posibles instrumentos y ver sus posibilidades como instrumentos adecuados:

- Supongamos que $Z_t = 1 \; \forall t$. En este caso $E(Z_t u_t) = 1E(u_t) = 0$ y $E(Z_t X_t) = E(X_t) = E(X_t^\star + \epsilon_t) = E(X_t^\star)$. Notar que X_t no tiene que estar en desviaciones a la media ya que en ese caso $\sum X_t = 0$ y $\hat{\beta}_{VI}$ no está definido.
- Otro posible instrumento sería $Z_t = t$, luego la variable instrumental es una tendencia determinista y por lo tanto, variable no estocástica incorrelada con la perturbación u_t, $E(Z_t u_t) =$

$Z_t E(u_t) = tE(u_t) = 0$. Ahora bien, $Z_t = t$ tiene que estar correlacionada con X_t tal que $\sum Z_t X_t \neq 0$. Por lo tanto, X_t tiene que presentar cierta tendencia en el tiempo. Por ejemplo, en el gráfico de la izquierda de la Figura 4.1 la serie no muestra tendencia temporal luego el instrumento $Z_t = t$ no sería adecuado. Sin embargo, en el gráfico de la derecha la serie dibujada si muestra tendencia temporal, en este caso creciente, con lo cual el instrumento $Z_t = t$ tiene sentido.

Figura 4.1: *Serie sin tendencia versus serie con tendencia*





En el caso de $Z_t = t$ la matriz Z es:

$$Z = \begin{bmatrix} 1 \\ 2 \\ \vdots \\ T \end{bmatrix}$$

- Otra variable instrumental puede ser otra variable medida con error, y distinta de X_t, que aproxime a $X_t^\star$, por ejemplo, $Z_t = X_t^\star + \omega_t$, tal que $\omega_t \sim iid(0, \sigma_\omega^2)$ y $E(\omega_t\, u_t^\star) = 0$. Además, los errores de medida tienen que estar incorrelados, $E(\epsilon_t\, \omega_t) = 0$, entonces,

$$E(Z_t u_t) = E((X_t^\star + \omega_t)(u_t^\star - \beta\epsilon_t)) = X_t^\star E(u_t^\star) + E(\omega_t\, u_t^\star) - \beta X_t^\star E(\epsilon_t) - \beta E(\omega_t\, \epsilon_t) = 0$$

  En este caso la única dificultad es encontrar la proxy alternativa. Una posibilidad es que dos instituciones independientes midan la misma variable de forma que los errores de medida se mantengan independientes.

- Actuando como en el ejercicio anterior podemos pensar en $Z_t = X_{t-1}$. En este caso si $E(\epsilon_t\, u_s^\star) = 0$, $E(X_{t-1} u_t) = 0$ ya que no existe correlación no contemporánea:

$$\begin{aligned} E(X_{t-1} u_t) &= E((X_{t-1}^\star + \epsilon_{t-1}) u_t) = E((X_{t-1}^\star + \epsilon_{t-1})(u_t^\star - \beta\epsilon_t)) = \\ &= E(X_{t-1}^\star\, u_t^\star) + E(\epsilon_{t-1}\, u_t^\star) - \beta E(X_{t-1}^\star\, \epsilon_t) - \beta E(\epsilon_t\, \epsilon_{t-1}) = 0 \end{aligned}$$

  y además, $E(X_{t-1} X_t) \neq 0$. Por tanto, también es un instrumento adecuado.

- Si generalizamos el modelo para incluir un término constante:

$$Y_t = \alpha + \beta X_t + u_t \qquad t = 1, 2, \ldots, T$$

entonces:

$$\underset{(T\times 2)}{X} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_T \end{bmatrix} \qquad \underset{(T\times 2)}{Z} = \begin{bmatrix} Z_{11} & Z_{21} \\ Z_{12} & Z_{22} \\ \vdots & \vdots \\ Z_{1T} & Z_{2T} \end{bmatrix}$$

Al menos necesitamos dos instrumentos. Para la constante 1 podemos usar $Z_{1t} = 1 \,\forall\, t$ ya que $E(Z_{1t}u_t) = 1E(u_t) = 0$, es decir, el mejor instrumento es la misma variable. Para X_t hemos de buscar un instrumento Z_{2t} distinto de la constante 1 ya que si repetimos el instrumento $rg(Z) = 1 < 2$ y $\not\exists (Z'X)^{-1}$, con lo que el estimador $\hat{\beta}_{VI} = (Z'X)^{-1}Z'Y$ no se puede obtener. No está definido. Además, tampoco sirven otras constantes como $Z_{2t} = 5$ por el mismo problema.

Ejemplo 4.11 Queremos estimar la relación:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X^{\star}_{3t} + u^{\star}_t \qquad u_t \sim iid(0, \sigma^2_{u^\star}) \qquad t = 1, 2, \ldots, T$$

donde X_{2t} es un regresor no estocástico y $X^{\star}_{3t}$ es una variable no estocástica e inobservable. Pero se dispone de una variable observable X_{3t} tal que $X_{3t} = X^{\star}_{3t} + \epsilon_t$ y

$$\epsilon_t \sim iid(0, \sigma^2_\epsilon) \qquad Cov(u^\star_t, \epsilon_t) = Cov(u^\star_t, \epsilon_s) = Cov(u^\star_s, \epsilon_t) = 0$$

luego el modelo a estimar en términos de las variables observables es:

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

con $u_t = u^\star_t - \beta_3\epsilon_t$, $E(X_{2t}u_t) = 0$ y $E(X_{3t}u_t) = -\beta_3\sigma^2_\epsilon \neq 0$. En este contexto el estimador de MCO es inconsistente y para obtener consistencia los coeficientes β_1, β_2 y β_3 deberían ser estimados por Variables Instrumentales.

Supongamos que disponemos de otra proxy Z_t para aproximar a $X^\star_{3t}$ tal que $Z_t = X^\star_{3t} + \eta_t$ con $\eta_t \sim iid(0, \sigma^2_\eta)$, $Cov(u^\star_t, \eta_t) = Cov(u^\star_t, \eta_s) = 0$ y $E(\epsilon_t\eta_t) = 0$, es decir, los dos errores de medida están incorrelacionados.

Podemos usar esta proxy Z_t como instrumento para X_{3t} ya que:

$$\begin{aligned} E(Z_t u_t) &= E((X^\star_{3t} + \eta_t)(u^\star_t - \beta_3\epsilon_t)) = \\ &= E(X^\star_{3t}u^\star_t) + E(\eta_t u^\star_t) - \beta_3 E(X^\star_{3t}\epsilon_t) - \beta_3 E(\eta_t\epsilon_t) = 0 + 0 - \beta_3\cdot 0 - \beta_3 \cdot 0 = 0 \end{aligned}$$

además, $E(Z_tX_{3t}) \neq 0$ ya que ambas son medidas de la variable $X^\star_{3t}$ no observable. Para la constante 1 y X_{2t} el mejor instrumento son ellas mismas. En este caso las matrices involucradas en el estimador de Variables Instrumentales son:

$$\underset{(T\times 1)}{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_T \end{bmatrix} \qquad \underset{(T\times 3)}{X} = \begin{bmatrix} 1 & X_{21} & X_{31} \\ 1 & X_{22} & X_{32} \\ \vdots & \vdots & \vdots \\ 1 & X_{2T} & X_{3T} \end{bmatrix} \qquad \underset{(T\times 3)}{Z} = \begin{bmatrix} 1 & X_{21} & Z_1 \\ 1 & X_{22} & Z_2 \\ \vdots & \vdots & \vdots \\ 1 & X_{2T} & Z_T \end{bmatrix}$$

con $rg(Z) = 3$, luego $\exists(Z'X)^{-1}$ y el estimador $\hat{\beta}_{VI} = (Z'X)^{-1}(Z'Y)$ está bien definido.

En este caso no podemos contemplar dos posibilidades para definir el instrumento Z_t, este no puede ser ni la constante 1 ni el regresor X_{2t} ya que la matriz no sería de rango completo y $\nexists(Z'X)^{-1}$.

Número de instrumentos disponibles mayor al número de variables explicativas que lo necesitan. Método de Mínimos Cuadrados en dos Etapas

Generalmente, se dispondrá de un número de instrumentos mayor que de variables explicativas que lo necesiten, en este caso habría muchas formas de construir las variables instrumentales que precisamos para obtener un estimador consistente. Pero dado que la matriz de covarianzas del estimador de VI depende de los valores de éstas, el modo en que se combinan los instrumentos para generar variables instrumentales influye sobre la eficiencia del estimador de VI respecto a otro estimador de VI de su misma clase. De ahí que en ocasiones se hable de la eficiencia relativa de los estimadores de Variables Instrumentales. A continuación vamos a proponer algunos ejemplos:

Ejemplo 4.12 En el modelo:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + u_t \qquad t = 1, 2, \ldots, T$$

donde:

$$\begin{array}{ll} u_t \sim iid(0, \sigma_u^2) & \\ X_{3t}, X_{4t} \quad \text{no estocásticas} & \\ X_{2t} = 0,5 X_{2,t-1} + u_t & v_t \sim iid(0, \sigma_v^2) \\ E(u_t v_t) = 5 \quad \forall t = s & E(u_t v_s) = 0 \quad \forall t \neq s \end{array}$$

En este ejercicio la matriz de regresores $X = [\,1 \quad \mathbf{X}_{2t}\ \mathbf{X}_{3t}\ \mathbf{X}_{4t}\,]$ es estocástica ya que incluye al regresor estocástico X_{2t}. Necesitamos ver como es $E(X'u)$. Dado que X_{3t} y X_{4t} son no estocásticos $E(X_{3t}u_t) = 0$ y $E(X_{4t}u_t) = 0$, pero la variable explicativa X_{2t} y la perturbación están contemporáneamente correladas,

$$E(X_{2t}u_t) = E[(0,5X_{2,t-1} + v_t)\,u_t] = 0,5E(X_{2,t-1}u_t) + E(v_t u_t) = 0,5 \cdot 0 + 5 = 5$$

Por tanto, $E(X'u) \neq 0$ y el estimador MCO será no lineal y sesgado. En muestras grandes además, será inconsistente. Deberíamos estimar por el Método de Variables Instrumentales, pero sólo la variable X_{2t} necesita instrumento, para X_{3t} y X_{4t} el mejor instrumento es la propia variable. En este caso hay varios instrumentos que cumplen los requisitos. Los retardos de los dos regresores no estocásticos X_{3t} y X_{4t}, así como el retardo del regresor estocástico X_{2t} están incorrelados con la perturbación:

$$E(X_{2,t-1}\,u_t) = 0; \quad E(X_{3,t-1}\,u_t) = 0; \quad E(X_{4,t-1}\,u_t) = 0$$

Además, las variables $X_{2,t-1}$, $X_{3,t-1}$ y $X_{4,t-1}$ están correladas con X_{2t} por la propia relación del modelo. Por tanto, $X_{2,t-1}$, $X_{3,t-1}$ y $X_{4,t-1}$ serían buenos instrumentos. También lo serían combinaciones lineales de los mismos. La cuestión es que si utilizamos todos los posibles instrumentos, supongamos que su número es por ejemplo l, y los introducimos en $Z_{(T\times l)}$, la matriz de instrumentos tiene más columnas que $X_{(T\times K)}$. En esta situación en la que hay más hay más instrumentos que los que se necesitan, $l > K$, la matriz $(Z'X)$ no sería cuadrada y no existiría su inversa, $\nexists(Z'X)^{-1}$, con lo que no podemos definir el estimador de VI como $\hat{\beta}_{VI} = (Z'X)^{-1}(Z'Y)$.

Dado que en esta situación el número de instrumentos supera al de variables explicativas que lo precisan, se trata de buscar cuál de todos los posibles instrumentos minimiza la varianza del estimador resultante. Una posibilidad consiste en generar la variable instrumental con mayor correlación con X_{2t}. Para ello se estima por MCO una regresión auxiliar de esta variable sobre todos los posibles instrumentos. Para el ejemplo que nos ocupa, si suponemos que los únicos instrumentos de que disponemos son los primeros retardos de los regresores originales la regresión auxiliar sería[11]:

$$X_{2t} = \gamma_1 + \gamma_2 X_{2,t-1} + \gamma_3 X_{3,t-1} + \gamma_4 X_{4,t-1} + \gamma_5 X_{3t} + \gamma_6 X_{4t} + \eta_t \qquad t = 2, 3, \ldots, T$$

así $\widehat{X}_{2t} = \hat{\gamma}_1 + \hat{\gamma}_2 X_{2,t-1} + \hat{\gamma}_3 X_{3,t-1} + \hat{\gamma}_4 X_{4,t-1} + \hat{\gamma}_5 X_{3t} + \hat{\gamma}_6 X_{4t}$ es una combinación lineal de todos los posibles instrumentos que se incluyen en la matriz de instrumentos Z, cada uno ponderado por su correlación con X_{2t}, la variable a instrumentalizar, siendo Z:

$$Z = \begin{bmatrix} 1 & X_{21} & X_{31} & X_{41} & X_{32} & X_{42} \\ 1 & X_{23} & X_{32} & X_{42} & X_{33} & X_{43} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & X_{2T-1} & X_{3T-1} & X_{4T-1} & X_{3T} & X_{4T} \end{bmatrix}$$

A continuación se estima el modelo por VI con:

$$\widehat{X} = [\,1 \quad \widehat{\mathbf{X}}_{2t} \quad \mathbf{X}_{3t} \quad \mathbf{X}_{4t}\,]$$

$$X = [\,1 \quad \mathbf{X}_{2t} \quad \mathbf{X}_{3t} \quad \mathbf{X}_{4t}\,]$$

Al estimador de VI así obtenido se le llama estimador de **Mínimos Cuadrados en dos Etapas**, MC2E, y se define[12]:

$$\hat{\beta}_{MC2E} = (\widehat{X}'X)^{-1}\widehat{X}'Y$$

donde

$$\widehat{X} = \begin{bmatrix} 1 & \widehat{X}_{22} & X_{32} & X_{42} \\ 1 & \widehat{X}_{23} & X_{33} & X_{43} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \widehat{X}_{2T} & X_{3T} & X_{4T} \end{bmatrix} \qquad X = \begin{bmatrix} 1 & X_{22} & X_{32} & X_{42} \\ 1 & X_{23} & X_{33} & X_{43} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & X_{2T} & X_{3T} & X_{4T} \end{bmatrix}$$

El estimador de MC2E es un estimador de VI que utiliza todos los instrumentos de forma óptima en el sentido de que es un estimador consistente y es el más eficiente asintóticamente dentro de la clase de estimadores de VI que toman solamente un subconjunto de todos los posibles instrumentos.

[11]Notar que en este ejemplo perdemos una observación al construir $\widehat{X}_{2t} \quad t = 2, 3, \ldots, T$.

[12]En el Anexo 4.4. Estimador de Mínimos Cuadrados en dos Etapas se desarrolla formalmente el estimador.

Ejemplo 4.13 Sea el modelo:

$$Y_t = \beta_1 + \beta_2 X_t + \beta_3 W_{1t} + u_t \qquad t = 1, 2, \ldots, T$$

donde

$$X_t = \gamma_1 + \gamma_2 Y_t + \gamma_3 W_{2t} + \gamma_4 W_{3t} + v_t$$
$$u_t \sim NID(0, \sigma_u^2) \qquad v_t \sim NID(0, \sigma_v^2) \qquad Cov(u_t, v_t) = \sigma_{uv}$$
$$W_{1t}, W_{2t} \text{ y } W_{3t} \text{ son no estocásticas}$$

Dado que $\gamma_2 \neq 0$ y/o $\sigma_{uv} \neq 0$ existe correlación entre X_t y u_t, $E(X_t u_t) \neq 0$. El estimador adecuado para obtener al menos la propiedad de consistencia es el estimador de Variables Instrumentales. Necesitamos buscar un instrumento para la variable X_t. Tenemos dos posibles instrumentos disponibles W_{2t} y W_{3t}. Para combinarlos de forma óptima podemos realizar la siguiente regresión:

$$X_t = \alpha_0 + \alpha_1 W_{1t} + \alpha_2 W_{2t} + \alpha_3 W_{3t} + \eta_t \qquad t = 1, 2, \ldots, T$$

Así el instrumento adecuado para X_t es $\widehat{X}_t$ obtenido de la estimación por MCO de la regresión anterior tal que:

$$Z = \begin{bmatrix} 1 & W_{11} & W_{21} & W_{31} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & W_{1T} & W_{2T} & W_{3T} \end{bmatrix}$$

Utilizamos el instrumento para fijar la matriz de instrumentos $\widehat{X}$. Aplicamos el estimador de Variables Instrumentales para el cuál:

$$\widehat{X} = \begin{bmatrix} 1 & \widehat{X}_1 & W_{11} \\ 1 & \widehat{X}_2 & W_{12} \\ \vdots & \vdots & \vdots \\ 1 & \widehat{X}_T & W_{1T} \end{bmatrix} \qquad X = \begin{bmatrix} 1 & X_1 & W_{11} \\ 1 & X_2 & W_{12} \\ \vdots & \vdots & \vdots \\ 1 & X_T & W_{1T} \end{bmatrix} \qquad Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_T \end{bmatrix}$$

Sobre el estimador de VI podemos hacer dos observaciones:

- Hay que notar que en el estimador de MC2E se regresan **todas** las variables explicativas sobre los posibles instrumentos. Para aquellas variables incorreladas con la perturbación el mejor instrumento son ellas mismas.

- La inclusión de retardos de las variables exógenas como instrumentos aumentaría el conjunto de información utilizado en la construcción del estimador de MC2E. Así, hay un estimador de MC2E para cada conjunto de instrumentos que se considere. Al utilizar más información, el estimador MC2E resultante sería más eficiente asintóticamente relativamente a otro estimador de VI que utilizase un subconjunto de los posibles instrumentos. Sin embargo, el uso de retardos obliga a prescindir de algunas observaciones muestrales, lo que disminuye el número de grados de libertad y puede influir en la eficiencia.

4.3.3. Contraste de hipótesis con el estimador de VI

Con el estimador de VI podemos utilizar el siguiente estadístico para hacer contrastes de restricciones lineales de la forma:

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

$$(R\hat{\beta}_{VI} - r)'[R\,(\widehat{V}(\hat{\beta}_{VI})\,R']^{-1}(R\hat{\beta}_{VI} - r) \xrightarrow{d,H_0} \chi^2_q$$

donde q es el número de restricciones que se contrastan.

Podemos escribir el estadístico anterior como:

$$(R\hat{\beta}_{VI} - r)'[\,R\,\hat{\sigma}^2_{VI}(Z'X)^{-1}\,Z'Z\,((Z'X)^{-1})'\,R'\,]^{-1}(R\hat{\beta}_{VI} - r) \xrightarrow{d,H_0} \chi^2_q$$

Las reglas de aceptación y rechazo son las usuales. No rechazaremos la hipótesis nula si el valor muestral del estadístico es menor que el valor crítico $\chi^2_{q|\alpha}$ para un valor de significación elegido α.

Ejemplo 4.14 Para contrastar la significatividad individual de un regresor, $H_0 : \beta_i = 0$ versus $H_a : \beta_i \neq 0$ dado que $q = 1$ podemos utilizar:

$$\frac{\hat{\beta}_{i,VI}}{\widehat{des}(\hat{\beta}_{i,VI})} \xrightarrow{d,H_0} N(0,1)$$

Donde $\widehat{des}(\hat{\beta}_{i,VI})$ es la raíz cuadrada del elemento i-ésimo de la matriz $\hat{\sigma}^2(Z'X)^{-1}\,Z'Z\,((Z'X)^{-1})'$. Si el valor muestral del estadístico es mayor que $N(0,1)_{|\frac{\alpha}{2}}$ rechazamos la hipótesis nula para un nivel de significatividad α.

4.4. Contraste de Hausman

Necesitamos derivar un test de contraste que sea capaz de juzgar la incorrelación entre X y u. Este test es el contraste de Hausman. Sirve para contrastar si los regresores están o no correlacionados con la perturbación. Las hipótesis de contraste son:

$$\begin{array}{lll} H_0\text{: } X \text{ y } u \text{ no están correlacionadas} & \Leftrightarrow & E(X'u) = 0 \\ H_a\text{: } X \text{ y } u \text{ están correlacionadas} & \Leftrightarrow & E(X'u) \neq 0 \end{array}$$

Consideremos el modelo de regresión lineal general en forma matricial:

$$Y = X\beta + u \Longleftrightarrow Y = \underset{(T\times K_1)}{X_1}\ \beta_1 + \underset{(T\times K_2)}{X_2}\ \beta_2 + u \qquad u \sim (0, \sigma^2 I)$$

donde K_1 variables explicativas en la matriz X no están correlacionadas con u, $E(X_1'u) = 0$, pero se cree que K_2 variables pueden estarlo[13]. El contraste de Hausman considera, en este contexto,

[13] Un caso particular sería que en X_1 se incluyera únicamente a la columna de unos y en X_2 el resto de variables explicativas.

contrastar:

$$H_0 : E(X_2'u) = 0$$
$$H_a : E(X_2'u) \neq 0$$

El estadístico de contraste considera la diferencia de dos estimadores de los coeficientes β, el estimador de MCO y un estimador de VI tal que:

$\hat{\beta}_{MCO}$ bajo H_0 es consistente y eficiente asintóticamente, pero inconsistente bajo H_a.
$\hat{\beta}_{VI}$ consistente bajo H_0 y H_a, pero menos eficiente asintóticamente que MCO bajo H_0.

El estadístico de contraste podemos escribirlo como:

$$(\hat{\beta}_{2,VI} - \hat{\beta}_{2,MCO})' \, [\widehat{V}(\hat{\beta}_{2,IV}) - \widehat{V}(\hat{\beta}_{2,MCO})]^{-1} \, (\hat{\beta}_{2,VI} - \hat{\beta}_{2,MCO}) \xrightarrow{d,H_0} \chi^2_p$$

con las reglas de decisión habituales y $p = K_2$.

Supongamos que queremos contrastar una única restricción, $H_0 : E(X_{it}, u_t) = 0$ versus $H_a : E(X_{it}u_t) \neq 0$, y $p = K_2 = 1$. En este caso podemos escribir el estadístico de contraste como:

$$\frac{(\hat{\beta}_{i,VI} - \hat{\beta}_{i,MCO})^2}{\widehat{V}(\hat{\beta}_{i,IV}) - \widehat{V}(\hat{\beta}_{i,MCO})} \xrightarrow{d,H_0} \chi^2_1$$

Si el estadístico calculado es menor que el valor $\chi^2_{1|\alpha}$ no rechazamos $H_0 : E(X_{it}u_t) = 0$ para un nivel de significatividad α y concluiríamos que no existe correlación estadísticamente significativa entre el regresor X_{it} y u_t. En este caso si las perturbaciones del modelo son esféricas el estimador adecuado sería MCO. Este estimador a pesar de ser no lineal y sesgado sería consistente.

Si el estadístico calculado es mayor que el valor $\chi^2_{1|\alpha}$ rechazamos $H_0 : E(X_{it}u_t) = 0$ para un nivel de significatividad α, y concluimos que el regresor X_{it} y u_t están correlacionados. El estimador MCO es inconsistente. En este caso si las perturbaciones del modelo son esféricas el estimador adecuado sería VI, estimador no lineal y sesgado en general, pero consistente y asintóticamente eficiente si se buscan correctamente el instrumento o instrumentos necesarios.

Al computar el estadístico de contraste es necesario calcular $\widehat{V}(\hat{\beta}_{VI})$ y $\widehat{V}(\hat{\beta}_{MCO})$, ambas expresiones participan de un estimador consistente de σ^2. Suponiendo bajo H_0 que $u \sim (0, \sigma^2_u I)$, podemos utilizar:

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

con:

$$\hat{\sigma}^2_{VI} = \frac{(Y - X\hat{\beta}_{VI})'(Y - X\hat{\beta}_{VI})}{T-K}$$

$$\hat{\sigma}^2_{MCO} = \frac{(Y - X\hat{\beta}_{MCO})'(Y - X\hat{\beta}_{MCO})}{T-K}$$

Asintóticamente ambos coinciden y el test es asintótico. Sin embargo, se puede demostrar que la potencia del contraste aumenta si utilizamos como estimador consistente de σ^2_u al estimador de VI, $\hat{\sigma}^2_{VI}$. Por otro lado, y con respecto al denominador de ambos estimadores, dado que son asintóticos es indiferente que éste sea $T-K$ ó T.

Ejemplo 4.15 Se propone la siguiente especificación para la función de demanda de vino de un país:

$$Q_t = \beta_1 + \beta_2 P_t + u_t \qquad t = 1, 2, \ldots, T$$

donde $u_t \sim (0\,,\,0{,}0921)$. Dado que el precio se determina simultáneamente con la cantidad Q_t, se sospecha que P_t pueda estar correlacionada con u_t. Se dispone de datos de un índice de costes de almacenamiento, S_t que se determina exógenamente, por lo que se considera independiente de u_t. Dados los siguientes datos para los años de 1955-1975:

$$\begin{array}{ll}
\sum p_t q_t = \sum (P_t - \bar{P})(Q_t - \bar{Q}) = 1,78037 & \sum s_t^2 = \sum (S_t - \bar{S})^2 = 2,1417 \\
\sum p_t s_t = \sum (P_t - \bar{P})(S_t - \bar{S}) = 0,500484 & \sum p_t^2 = \sum (P_t - \bar{P})^2 = 0,507434 \\
\sum s_t q_t = \sum (S_t - \bar{S})(Q_t - \bar{Q}) = 2,75474 &
\end{array}$$

Se pide proponer un estimador consistente de los coeficientes β_1 y β_2 del modelo.

Para proponer un estimador consistente de los coeficientes de la relación es necesario saber si existe o no correlación entre P_t y u_t. Lo contrastamos:

$$H_0 : E(P_t u_t) = 0$$

$$H_a : E(P_t u_t) \neq 0$$

con

$$\frac{(\hat{\beta}_{2,VI} - \hat{\beta}_{2,MCO})^2}{\widehat{V}(\hat{\beta}_{2,IV}) - \widehat{V}(\hat{\beta}_{2,MCO})} \xrightarrow{d,H_0} \chi^2_1$$

Calculamos los estimadores MCO y VI utilizando en este último como instrumento para P_t al el índice de coste de existencias de almacén S_t:

$$\hat{\beta}_{2,MCO} = \frac{\sum p_t q_t}{\sum p_t^2} = \frac{1,78037}{0,507434} = 3,5085$$

$$\hat{\beta}_{2,VI} = \frac{\sum s_t q_t}{\sum s_t p_t} = \frac{2,75474}{0,500484} = 5,4862$$

A continuación calculamos las correspondientes varianzas estimadas utilizando el valor $\sigma^2 = 0,09217$, conocido por el enunciado,

$$\widehat{V}(\hat{\beta}_{2,MCO}) = \frac{\sigma_u^2}{\sum p_t^2} = \frac{0,09217}{0,507434} = 0,18164$$

$$\widehat{V}(\hat{\beta}_{2,VI}) = \sigma_u^2 \frac{\sum s_t^2}{(\sum s_t p_t)^2} = \frac{0,0921 \cdot 2,1417}{(0,500484)^2} = 0,78747$$

y finalmente, computamos el estadístico:

$$\mathrm{H} = \frac{(\hat{\beta}_{2,VI} - \hat{\beta}_{2,MCO})^2}{\widehat{V}(\hat{\beta}_{2,VI}) - \widehat{V}(\hat{\beta}_{2,MCO})} = \frac{(5,4862 - 3,5085)^2}{0,78747 - 0,18164} = 6,5721$$

$\mathrm{H} = 6,5721 > 3,841 = \chi^2_{1|0,05}$ por tanto, rechazo la hipótesis nula para $\alpha = 5\,\%$ y $E(P_t u_t) \neq 0$, P_t y u_t no son incorreladas.

Este ejercicio recoge un ejemplo práctico de la dificultad de estimar relaciones simultáneas. En un contexto donde la variable endógena y y el regresor se determinan simultáneamente como es éste, donde, lo que se especifica es la función de demanda del vino, la matriz de regresores es estocástica, como ya se apuntó en el Ejemplo 5.2. Es una de las situaciones en que existe correlación muestral significativa entre regresor y perturbación, como se ha detectado utilizando el estadístico de Hausman. El estimador MCO es inconsistente, los coeficientes del modelo deben ser estimados por VI, estimador consistente y asintóticamente eficiente. Además, existe un único instrumento, S_t y una única variable que lo necesita, P_t. La FRM podemos escribirla:

$$\widehat{Q}_t = \hat{\beta}_{1,VI} + 5,4862\, P_t$$

Por otra parte, el estimador de VI bajo perturbaciones esféricas permite hacer inferencia asintótica válida. Como podemos comprobar, la variable precio es significativa individualmente. Para ello contrastamos $H_0 : \beta_2 = 0$ versus $H_a : \beta_2 \neq 0$ con el estadístico y distribución asintótica:

$$\frac{\hat{\beta}_{2,VI}}{\widehat{des}(\hat{\beta}_{2,VI})} \xrightarrow{d,H_0} N(0,1)$$

Dado que el valor muestral del estadístico es $t = \frac{5,4862}{\sqrt{0,78747}} = 6,1823 > 1,96 = N(0,1)_{|0,025}$ rechazamos H_0 para un nivel de significatividad del 5 % y el precio es una variable individualmente significativa.

A continuación vamos a ver un ejemplo, utilizando el programa gretl, de cómo trabajar con el estimador de Variables Instrumentales. En el ejemplo se resume todo el tema. Las instrucciones de gretl necesarias para conseguir los resultados que se muestran las encontraréis en el Anexo 4.1: Instrucciones de gretl para el estimador de Variables Instrumentales. Estas instrucciones serán mostradas en la Práctica de Ordenador propuesta en este tema. También serán de aplicabilidad en el tema siguiente.

Ejemplo 4.16 En este ejemplo vamos a analizar los determinantes de la oferta laboral de las mujeres casadas[14]. Para ello vamos a utilizar el archivo *mroz87.gdt* incluido en el programa gretl, en la carpeta de archivos de muestra "Gretl". En él se dispone de observaciones de las siguientes variables, entre otras[15]:

- LFP: Variable ficticia que toma valor 1 si la mujer ha trabajado en 1975 y cero en otro caso
- $HOURS$: Número de horas trabajadas por la esposa en 1975, ($WHRS$).
- $KL6$: Número de hijos menores de seis años, en la familia.
- $K618$: Número de hijos entre seis y dieciocho años, en la familia.

[14] Fichero mrox87.gdt, disponible en gretl pestaña Gretl. Fuente: "The Sensitivity of an Empirical Model of Married Women's Hours of Work to Economic and Statistical Assumptions", *Econometrica* 55, 765-799.

[15] Algunas de las variables han sido renombradas para que el ejercicio resulte más cómodo. Entre paréntesis aparece el nombre original utilizado en el fichero para que las podáis reconocer fácilmente. A la hora de reproducir el ejercicio en vuestro trabajo personal es aconsejable hacer el cambio de nombre sugerido.

- AGE: Edad de la esposa, (WA).
- $EDUC$: Años de educación recibidos por la esposa, (WE).
- $WAGE$: Salario de la esposa en el momento de la encuesta, 1976, ($RPWG$).
- $FAMINC$: Renta familiar en dólares de 1975.
- $EXPER$: Años de experiencia en la actualidad, (AX).

Se trata de una muestra de sección cruzada con 428 observaciones de mujeres trabajadoras donde el término trabajadoras implica que tienen un salario monetario[16].

Con la muestra anterior creamos las siguientes variables:

$$\begin{aligned}
&l_WAGE_i = ln(WAGE)_i,\\
&sq_EXPER_i = EXPER_i^2,\\
&NWIFEINC_i = FAMINC_i - (WAGE_i \times HOURS_i)
\end{aligned}$$

Dado nuestro objetivo de estudiar los determinantes de la oferta laboral de la población femenina casada, la variable a estudiar es $HOURS$. Como determinantes de la misma podemos pensar en incluir en el modelo el salario en logaritmos, l_WAGE, los años de educación recibidos, $EDUC$, la edad, AGE, el número de hijos de la familia, $KL6$ y $K618$, y la variable $NWIFEINC$, que de alguna manera mide la importancia de las rentas familiares que no dependen de los ingresos de la esposa. Así, el modelo a estimar sería:

$$\begin{aligned}
HOURS_i = \beta_1 + \beta_2\, l_WAGE_i + \\
\beta_3 EDUC_i + \beta_4 AGE_i + \beta_5 KL6_i + \beta_6 K618_i + \beta_7\, NWIFEINC_i + u_i
\end{aligned} \tag{4.3}$$

A priori esperaríamos que los coeficientes de las variables l_WAGE, $EDUC$ y $K618$ fuesen positivos, ceteris paribus. Es de esperar que si se tiene un sueldo alto la estimulación a seguir trabajando sea mayor. También es lógico pensar que cuando la preparación (estudios) es mayor, se tenga un empleo mejor y mejor remunerado, luego es más probable que la mujer trabaje fuera de casa. Familias con mayor número de hijos necesitan de una mayor renta, por lo que es probable que la esposa decida trabajar. Por otro lado, cuando hay niños pequeños podemos pensar que la esposa se retire del mercado de trabajo para cuidarlos, por lo que esperaríamos un signo negativo en el coeficiente que acompaña a $KL6$ manteniendo el resto de factoes constante. A mayor edad, también es posible que no se quiera trabajar fuera de casa, por lo que, esperamos signo negativo para la variable AGE ceteris paribus. Finalmente, manteniendo el resto de factores constantes también esperaríamos signo negativo para el coeficiente de la variable $NWIFEINC$ ya que, si la renta disponible que no depende de la remuneración de la mujer es alta, es probable que eso desincentive a trabajar fuera de casa a la esposa. Es importante tener en cuenta la fecha de la recogida de datos, 1975, cuando todavía la participación en el mercado de trabajo de la mujer no estaba consolidada.

[16]Este archivo es en realidad una submuestra de otro archivo original con un total de 753 observaciones de las cuales solo trabajan 428 mujeres. El archivo original es *mroz.gdt*, está incluido en la carpeta Wooldridge de gretl. Sin embargo, como a lo largo del ejercicio de verá, para todos los individuos de la muestra todas las observaciones no están completas por lo que finalmente, el número utilizado de observaciones completas en el ejercicio es de 326.

Los resultados de la estimación MCO de la ecuación (4.3) son:

Estimaciones MCO utilizando 326 observaciones desde 1–428
Se han quitado las observaciones ausentes o incompletas: 102
Variable dependiente: HOURS

	Coeficiente	Desv. típica	estadístico t	valor p
const	1909,20	344,1980	5,5468	0,0000
l_WAGE	311,7460	95,1025	3,2780	0,0012
EDUC	−18,6620	18,6519	−1,0006	0,3178
AGE	−8,2063	5,5617	−1,4755	0,1411
KL6	−296,62	107,6080	−2,7565	0,0062
K618	−80,759	31,8324	−2,5370	0,0117
NWIFEINC	−0,0098	0,0037	−2,6651	0,0081

Media de la var. dep.	1424,990	R^2	0,09628
D.T. de la var. dep.	687,015	$\bar{R}^2$ corregido	0,07928
SCR	1,38627e+08	$F(6, 319)$	5,66458
Desv. típica regresión ($\hat{\sigma}$)	659,217	valor p para $F()$	1,31042e-05

Lo primero que llama la atención es que las variables $EDUC$ y $K618$ no tienen los signos esperados. Por otro lado $EDUC$ y AGE no son significativas. ¿Es posible que nuestro modelo no esté correctamente estimado? ¿Hay razones para sospechar que l_WAGE esté correlada con la perturbación del modelo? $HOURS$ y l_WAGE se determinan conjuntamente en equilibrio entre la oferta y demanda. Luego efectos recogidos en u_i que afectan al nivel de horas trabajadas también afectaran al nivel de salarios de equilibrio. Se hace indispensable saber si el nivel de salarios l_WAGE_i está o no correlada con la perturbación. Es necesario contrastar:

$$H_0 : E(l\_WAGE_i \, u_i) = 0$$
$$H_a : E(l\_WAGE_i \, u_i) \neq 0$$

El contraste de Hausman implica estimar el modelo además de por MCO, por VI.

Vamos a estimar por VI utilizando como instrumento para la variable l_WAGE a la variable $EXPER$. En principio la variable $EXPER$ no tiene porque estar correlada con la perturbación del modelo, así $E(EXPER_i \, u_i) = 0$. Por otro lado, la correlación muestral entre estas dos variables es baja $\mathrm{r}_{l\_WAGE,EXPER} = 0,216$. Sin embargo, si regresamos l_WAGE sobre $EXPER$ incluyendo una constante la variable es significativa, el valor muestral estadístico es 3,985. Por lo que podemos concluir que el instrumento está significativamente correlado con la variable para la cual hace de instrumento, $E(l\_WAGE_i \; EXPER_i) \neq 0$. La variable cumple las características de una variable instrumental.

Los resultados de la estimación por VI utilizando como instrumento para l_WAGE a la variable $EXPER$ son los siguientes:

Estimaciones MC2E utilizando 326 observaciones desde 1–428
Se han quitado las observaciones ausentes o incompletas: 102
Variable dependiente: HOURS
Instrumented: l_WAGE
Instrumentos: const EXPER EDUC AGE KL6 K618 NWIFEINC

	Coeficiente	Desv. típica	z-stat	valor p
const	1334,86	527,6410	2,5299	0,0114
l_WAGE	2099,23	545,3280	3,8495	0,0001
EDUC	−162,5070	50,3543	−3,2273	0,0012
AGE	−9,4465	8,0821	−1,1688	0,2425
KL6	−372,3520	157,8050	−2,3596	0,0183
K618	−3,1880	51,5713	−0,0618	0,9507
NWIFEINC	−0,0124	0,0054	−2,2833	0,0224

Media de la var. dependiente	1424,988	Desv. tip. var. dependiente.	687,0155
Suma de cuadrados de los residuos	2,92e+08	Desv. tip. regresión	956,9807
R^2	0,050719	Adjusted R^2	0,032864
$F(6,319)$	4,307881	P-value(F)	0,000340

Si nos fijamos en el output que muestra gretl este nos indica cual es la variable que estamos instrumentalizando, l_WAGE. Si observamos la lista de instrumentos veremos que la matriz de instrumentos Z consta de siete columnas, las cuales coinciden con las columnas de la matriz de regresores X salvo la segunda donde en lugar de la variable l_WAGE aparecerá el instrumento $EXPER$. La referencia a MC2E de gretl se corresponde con el estimador de VI, ya que en este caso hay un instrumento y una única variable que lo necesita.

Ya podemos computar el contraste de Hausman. Recordemos que el estadístico de contraste y distribución son:

$$\frac{(\hat{\beta}_{2,VI} - \hat{\beta}_{2,MCO})^2}{\widehat{V}(\hat{\beta}_{2,IV}) - \widehat{V}(\hat{\beta}_{2,MCO})} \xrightarrow{d,H_0} \chi^2_1$$

Aplicado a la muestra obtenemos:

$$H = \frac{(\hat{\beta}_{2,VI} - \hat{\beta}_{2,MCO})^2}{\widehat{V}(\hat{\beta}_{2,IV}) - \widehat{V}(\hat{\beta}_{2,MCO})} = \frac{(2099,23 - 311,746)^2}{(545,328)^2 - (95,1025)^2} = \frac{3195099,05}{288338,1421} = 11,081$$

dado que el valor muestral del estadístico es mayor que $3,84 = \chi^2_{1|0,05}$ rechazamos la hipótesis nula y concluimos que $E(l\_WAGE_i \; u_i) \neq 0$. Luego el estimador apropiado es VI ya que MCO es inconsistente.

Los signos obtenidos de la estimación VI se mantienen con respecto a MCO. Con respecto a la significatividad de las variables en este caso la variable $EDUC$ es individualmente significativa, al contrario que con la estimación MCO. La no significatividad de AGE se mantiene, pero ahora obtenemos que la variable $K618$ no es significativa, cosa que no ocurría con el estimador MCO.

Podemos pensar en mejorar nuestro instrumento. ¿Que pasaría si utilizáramos como instrumento el cuadrado de la experiencia, sq_EXPER? La correlación muestral entre estas dos variables es baja $r_{l\_WAGE,sq\_EXPER} = 0,1853$, algo más baja que en el caso anterior. Si regresamos l_WAGE sobre sq_EXPER incluyendo una constante la variable es significativa, el valor muestral del estadístico es 3,395. Concluimos que el instrumento está significativamente correlado con la variable para la cual hace de instrumento, $E(l\_WAGE_i \; sq\_EXPER_i) \neq 0$. La variable es un instrumento adecuado.

Los resultados de la estimación por VI utilizando como instrumento para l_WAGE a la variable sq_EXPER son los siguientes:

Estimaciones MC2E utilizando 326 observaciones desde 1–428
Se han quitado las observaciones ausentes o incompletas: 102
Variable dependiente: HOURS
Instrumented: l_WAGE
Instrumentos: const sq_EXPER EDUC AGE KL6 K618 NWIFEINC

	Coeficiente	Desv. típica	z-stat	valor p
const	1356,310	524,6080	2,5854	0,0097
l_WAGE	2032,460	599,0830	3,3926	0,0007
EDUC	−157,1340	53,9486	−2,9127	0,0036
AGE	−9,4002	7,9271	−1,1858	0,2357
KL6	−369,5240	155,1580	−2,3816	0,0172
K618	−6,0855	51,9093	−0,1172	0,9067
NWIFEINC	−0,0123	0,0053	−2,3050	0,0212

Media de la var. dependiente	1424,988	Desv. tip. var. dependiente.	687,0155
Suma de cuadrados de los residuos	2,81e+08	Desv. tip. regresión	938,3666
R^2	0,051376	Adjusted R^2	0,033534
$F(6, 319)$	3,830092	P-value(F)	0,001056

Los resultados de la estimación por VI utilizando como instrumento a sq_EXPER no varían con respecto a utilizar como instrumento a la variable $EXPER$. Ni en cuanto a los signos obtenidos, ni en cuanto a significatividad[17]. Sin embargo, hemos de notar que si ambas son instrumentos adecuados podemos ampliar el conjunto de instrumentos incluyendo a ambas variables y considerar el estimador de MC2E utilizando todos los instrumentos para mejorar en eficiencia asintótica.

Vamos a mostrar ahora los resultados de la estimación por MC2E utilizando ambas variables como instrumento junto con los regresores originales no correlados con la perturbación:

[17] Ya hemos determinado que l_WAGE es una variable correlada con la perturbación y no es necesario computar de nuevo el contraste de Hausman con el nuevo instrumento propuesto. Aún así, si lo calculamos, el estadístico de Hausman aplicado a la muestra es:

$$H = \frac{(\hat{\beta}_{2,VI} - \hat{\beta}_{2,MCO})^2}{\widehat{V}(\hat{\beta}_{2,IV}) - \widehat{V}(\hat{\beta}_{2,MCO})} = \frac{(2032,23 - 311,746)^2}{(599,083)^2 - (95,1025)^2} = \frac{2960065,194}{349855,9554} = 8,46$$

dado que el valor muestral del estadístico es mayor que $3,84 = \chi^2_{1|0,05}$ rechazamos la hipótesis nula y concluimos que $H_a : E(l\_WAGE_i \; u_i) \neq 0$. Luego el estimador apropiado es VI ya que MCO es inconsistente.

Estimaciones MC2E utilizando 326 observaciones desde 1–428
Se han quitado las observaciones ausentes o incompletas: 102
Variable dependiente: HOURS
Instrumented: l_WAGE
Instrumentos: const EDUC AGE KL6 K618 NWIFEINC EXPER sq_EXPER

	Coeficiente	Desv. típica	z-stat	valor p
const	1328,65	530,2510	2,5057	0,0122
l_WAGE	2118,57	544,7060	3,8894	0,0001
EDUC	−164,0630	50,3799	−3,2565	0,0011
AGE	−9,4599	8,1281	−1,1639	0,2445
KL6	−373,1720	158,6810	−2,3517	0,0187
K618	−2,3488	51,7912	−0,0454	0,9638
NWIFEINC	−0,0124	0,0054	−2,2757	0,0229

Media de la var. dependiente	1424,988	Desv. tip. de la var. dependiente.	687,0155
Suma de cuadrados de los residuos	2,95e+08	Desv. tip. regresión	962,4351
R^2	0,050536	Adjusted R^2	0,032678
$F(6, 319)$	4,338567	P-value(F)	0,000316

Si analizamos el output que muestra gretl la variable que estamos instrumentalizando coincide con los casos anteriores, pero ahora en la lista de instrumentos hay ocho variables. Se incluyen $EXPER$ y sq_EXPER y los regresores originales. En este caso, el término MC2E se corresponde totalmente con el estimador que hemos visto en la teoría. Hay más de un instrumento y una única variable que lo necesita, gretl ha computado el instrumento utilizando todas las variables indicadas para obtener el estimador de VI.

Los resultados de la estimación por MC2E no cambian frente a las anteriores regresiones de VI, ni en cuanto a los signos obtenidos, ni en cuanto a significatividad de las variables. Los coeficientes de determinación no deben ser utilizados como medida de bondad del ajuste ya que en este caso no tienen la interpretación habitual.

Ejercicio 4.3 Para afianzar la comprensión del ejemplo anterior.

1) Repetir todas las estimaciones y contrastes.

2) Calcular el estimador de MC2E paso a paso. Es decir:

    i) Estimar por MCO la regresión:

$$\begin{aligned} l\_WAGE_i = \alpha_1 + \alpha_2 EXPER_i + \alpha_3 sq\_EXPER_i + \\ \alpha_4 EDUC_i + \alpha_5 AGE_i + \alpha_6 KL6_i + \alpha_7 K618_i + \alpha_8 NWIFEINC_i + w_i \end{aligned}$$

    y guardar las estimaciones obtenidas, $l\_\widehat{WAGE}_i$.

    ii) Computar el estimador de VI utilizando como instrumento para l_WAGE_i a la variable $Z_i = l\_\widehat{WAGE}_i$.

iii) Comparar las estimaciones obtenidas con las del ejemplo anterior.

3) Contrastar que las variables $KL6$ y $K618$ no son conjuntamente significativas. Razona la estimación elegida para realizar el contraste.

4.5. Ejercicios a resolver

A continuación se proponen varios ejercicios para que el alumno vaya resolviendo a la vez que se va explicando la materia en las clases magistrales y cuyas dudas se resolverán a lo largo de estas:

Ejercicio M-RE.1
En el modelo:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t \qquad u_t \sim iid(0, \sigma^2)$$

donde X_{2t} es una variable no estocástica y X_{3t} es una variable aleatoria. Denotamos por β al vector de coeficientes desconocidos de orden (3×1).

1. ¿Por qué el estimador de β MCO no es lineal?
2. ¿Qué supuesto te garantiza que el estimador de β por el método de Mínimos Cuadrados Ordinarios sea insesgado? Demuéstralo.
3. Si X_{3t} es estocástica y no independiente de u_t, pero $E(X_{3t}u_t) = 0 \quad \forall t$, ¿es el estimador de β por MCO consistente? Demuéstralo e indica los supuestos adicionales que te sean necesarios.
4. Si X_{3t} es estocástica, pero se satisface el Teorema de Mann y Wald, ¿podemos hacer inferencia sobre β a pesar de no conocer la distribución de u_t? Razona tu respuesta.

Ejercicio M-RE.2
Se dispone de 62 observaciones sobre las siguientes características de los terremotos registrados en Alaska durante el periodo 1969-1978[18]:

Y_t : El logaritmo de la amplitud de onda en metros por segundo, (m/sg).

X_t : El logaritmo de la amplitud del cuerpo longitudinal de la onda en m/sg.

W_t : El logaritmo de la traza máxima de amplitud de onda a corta distancia en m/sg.

Se quiere estimar cuál es el efecto sobre Y_t de la velocidad de amplitud del cuerpo de la onda de un terremoto, X_t, mediante el modelo:

$$Y_t = \beta_1 + \beta_2 X_t^* + v_t \qquad v_t \sim NID(0, \sigma_v^2) \tag{4.4}$$

[18] *Fuente:* Fuller, W.A. (1987), *Measurement Error Models*. Wiley, New York.

La tecnología existente no permite obtener directamente el valor de la variable no estocástica X_t^* por lo que se aproxima mediante $X_t = X_t^* + e_t$, donde X_t es la variable observada y el error de medida es $e_t \sim \text{NID}(0, \sigma_e^2)$. Además, la perturbación del modelo, v, y el error de medida, e, son independientes. Se han obtenido los siguientes resultados a partir del estimador de Mínimos Cuadrados Ordinarios:

$$\underset{(\widehat{des}(\hat{\beta}_{i,MCO}))}{\widehat{Y_t}} = \underset{(0{,}780)}{-1{,}491} + \underset{(0{,}149)}{1{,}261}\ X_t + \hat{u}_t\,, \qquad SCR = 17{,}242.$$

$$(X'X)^{-1} = \begin{bmatrix} 2,118 & -0,403 \\ -0,403 & 0,077 \end{bmatrix}^{-1}$$

1. Obtén paso a paso cada uno de los siguientes valores:

$u_t =$

$E(u_t) =$

$Var(u_t) =$

$Cov(u_t, u_s) =$

$E(X_t u_t) =$

2. Razona las propiedades en muestras finitas y asintóticas del estimador MCO.

El modelo anterior ha sido reestimado por Variables Instrumentales. Para ello se ha utilizado como instrumento para el regresor X_t a la variable W_t, cuya medición se puede realizar con exactitud. Se han obtenido los siguientes resultados:

$$\underset{(\widehat{des}(\hat{\beta}_{i,VI}))}{\widehat{Y_t}} = \underset{(1{,}114)}{-4{,}287} + \underset{(0{,}213)}{1{,}797}\ X_t + \hat{\omega}_t\,, \qquad SCR = 20{,}961.$$

3. Escribe explícitamente la fórmula del estimador de VI y su expresión en términos de sumatorios.

4. Escribe explícitamente las condiciones necesarias y suficientes para que el estimador de VI sea consistente.

5. Lleva a cabo el contraste de Hausman para analizar si es o no importante el problema de error de medida. Escribe la hipótesis nula, la alternativa y todos los elementos del contraste, así como su conclusión.

6. Contrasta la hipótesis de que, en media, la amplitud del cuerpo longitudinal de la onda recogida en un sismógrafo no es relevante sobre la amplitud de la onda.

Ejercicio M-RE.3
Se quiere estimar el modelo:

$$Y_t = \beta X_{1t} + u_t \qquad u_t \sim iid(0, \sigma^2) \tag{4.5}$$

y se sabe que X_{1t} se determina con Y_t ya que $X_{1t} = Y_t + X_{2t}$ donde $E(X_{2t}u_t) = 0 \; \forall t$.

1. Demuestra que $E(X_{1t}u_t) = (1-\beta)^{-1}\sigma^2$. Se supone que $\beta \neq 1$.
2. ¿Qué implicaciones tiene este hecho en el estimador de β aplicando el método de Mínimos Cuadrados Ordinarios (MCO) a (4.5)? Razona la respuesta.
3. Escribe explícitamente la fórmula de un estimador de β alternativo para este modelo concreto razonando por qué lo escogerías.

Si se dispone de una muestra de 60 observaciones donde se han obtenido los siguientes productos cruzados:

	Y_t	X_{1t}	X_{2t}
Y_t	100	40	-60
X_{1t}		80	40
X_{2t}			100

por ejemplo $\sum Y_t X_{2t} = -60$.

4. Obtén la estimación de β por el método propuesto en el apartado 3 y por el método de MCO.
5. Contrasta al nivel de significación del 5 % la $H_0 : \beta = 0$. (Suponer que $\sigma^2 = 1$.)
6. Si el investigador ignorara que $X_{1t} = Y_t + X_{2t}$, ¿cómo podría darse cuenta de que $E(X_{1t}u_t) \neq 0$? Explica y realiza el contraste. (Suponer que $\sigma^2 = 1$.)

Ejercicio M-RE.4
Se quiere evaluar el rendimiento de la educación en términos del siguiente modelo

$$Y_i = \beta_1 + \beta_2 EDU_i + w_i \qquad i = 1, \ldots, N$$

donde Y_i y EDU_i son las ganancias salariales anuales (en decenas de miles de euros) y el nivel de educación de un individuo respectivamente. Además, $E(EDU_i w_i) = 0$ para todo i y w_i es un ruido blanco.

Se dispone de una muestra de 1000 individuos. Sin embargo, se mide el nivel de educación a través de la variable observada, años de estudio, S_i, que está medida con error, tal que $S_i = EDU_i + \epsilon_i$ donde ϵ_i es un ruido blanco independiente de EDU_i y de w_i.

Utilizando el método de Mínimos Cuadrados Ordinarios en base a la información muestral disponible, se han obtenido los siguientes resultados:

$$\begin{array}{ccccc} \hat{Y}_i & = & 2,431 & + & 0,03332\ S_i \\ (\widehat{des}(\hat{\beta}_{MCO})) & & (0{,}078) & & (0{,}0046) \end{array}$$

1. Interpreta qué indica la estimación obtenida para el parámetro β_2.

2. Explica en detalle qué propiedades tendrá el estimador MCO de β_1 y β_2 si se ha utilizado la medida de educación disponible S_i en lugar de EDU_i en el modelo. Razona tu respuesta.

Disponemos de una variable adicional P_i, que mide los años de educación del padre de ese individuo i. Para la muestra de 1000 individuos se tiene la siguiente información:

$$\begin{array}{llll} \sum_i Y_i = 2988,232 & \sum_i S_i = 16707 & \sum_i Y_i S_i = 50071,6 & \sum_i S_i^2 = 283539 \\ \sum_i P_i = 14343 & \sum_i Y_i P_i = 42914,7 & \sum_i P_i S_i = 240466 & \sum_i P_i^2 = 206469 \\ \sum_i Y_i^2 = 9028,9 & & & \end{array}$$

3. Propón un estimador consistente, alternativo al de MCO, razonando bajo qué condiciones sería consistente y cuál será su distribución asintótica.

4. Calcula la estimación de β_1 y β_2 en base al estimador propuesto en el apartado anterior.

5. Si se ha utilizado un estimador consistente, ¿cómo se ha obtenido la siguiente estimación de la matriz de varianzas y covarianzas asintótica del estimador propuesto en el apartado 3? Indica todos los pasos que se han realizado hasta llegar a este resultado.

$$\hat{V}(\hat{\beta}) = \frac{98,88}{998}\begin{bmatrix} 0,2984084 & -0,017800 \\ -0,017800 & 0,001065 \end{bmatrix}$$

6. Utilizando el estimador propuesto en el apartado 3, contrasta la hipótesis de que un año adicional de educación supone un incremento medio en las ganancias salariales anuales de 720 euros. Escribe la hipótesis nula, la alternativa y todos los elementos del contraste.

7. Lleva a cabo el contraste de Hausman para analizar si es o no importante el problema de error de medida. Escribe la hipótesis nula, la alternativa y todos los elementos del contraste.

8. Indica de manera razonada cuál de los dos estimadores elegirías teniendo en cuenta el resultado del contraste de Hausman.

Ejercicio M-RE.5

Considera el siguiente modelo de regresión:

$$Y_i = \beta_1 + \beta_2 X_i + u_i \qquad i = 1, \ldots, N \tag{4.6}$$

donde X_i es estocástica, $u_i \sim N(0, \sigma^2)$, $E(u_i u_j) = 0$ para $i \neq j$ y donde $E(X_i u_i) = 0,9$.

1. ¿Qué problema existe en el modelo anterior? ¿Cómo podría detectarse? Explica en detalle el contraste que propones y las consecuencias de rechazar o no la nula.

2. ¿Qué consecuencias tiene en los contrastes de hipótesis sobre β_1 y β_2 la utilización del estimador MCO y el de VI? Razona tu respuesta en cada caso.

   Se dispone de una muestra de 500 observaciones que da lugar a las siguientes sumas de cuadrados y de productos cruzados[19]:

	1	Y_i	X_i	Z_{1i}
1	500	1530,17	14,48	-0,23
Y_i		7163,54	1551,83	448,79
X_i			1037,57	451,24
Z_{1i}				509,40

   donde por ejemplo $\sum Y_i X_i = 1551,83$ y $\sum Y_i = 1530,17$.

3. Utilizando la información muestral dada, completa todos los elementos dentro de las matrices para la obtención de las estimaciones de β_1 y β_2 por VI, considerando como único instrumento a Z_1:

$$\widehat{\beta}_{VI} = \begin{bmatrix} \qquad & \qquad \\ \qquad & \qquad \end{bmatrix}^{-1} \begin{bmatrix} \qquad \\ \qquad \end{bmatrix} = \begin{bmatrix} 3,03 \\ 0,996 \end{bmatrix}$$

[19]Fuente: Libro *Undergraduate Econometrics,* de Hill, Griffiths y Judge (2001).

Con los datos de la tabla anterior se ha obtenido la siguiente estimación de la matriz de varianzas y covarianzas del anterior estimador VI de β_1 y β_2:

$$\widehat{V}(\hat{\beta}_{VI}) = \begin{bmatrix} 0{,}00203608 & -0{,}000074 \\ & 0{,}002544 \end{bmatrix}$$

Completa la ecuación del modelo estimado:

$$\begin{array}{ccccc} \hat{Y}_t & = & & \ldots & \\ (\widehat{des}(\hat{\beta}_{VI})) & & (\qquad\qquad) & & (\qquad\qquad) \end{array}$$

4. ¿Bajo qué condiciones es consistente el estimador VI del apartado anterior? ¿Es un estimador asintóticamente eficiente? Razona tu respuesta.

5. Para realizar el contraste $H_0 : \beta_1 = 3 \quad \beta_2 = 1$:

   Escribe el estadístico de contraste y su distribución:

   Detalla cada uno de los elementos del estadístico anterior:

   Utilizando la siguiente información y el estimador del apartado 3) realiza el contraste:

```
Conjunto de restricciones
 1: b[const] = 3
 2: b[X] = 1
Valor muestral del estadístico de contraste: chi^2(2) = 0,490224,
con valor p = 0,782617.
```

   .

Se ha considerado un estimador alternativo al utilizado en el apartado 3) obteniéndose los siguientes resultados en gretl.

Estimaciones MC2E utilizando las 500 observaciones 1–500
Variable dependiente: Y
Instrumentos: const Z1 Z2

Variable	Coeficiente	Desv. típica	Estadístico t	valor p
const	3,03113	0,0445796	67,9936	0,0000
X	1,00899	0,0448997	22,4721	0,0000

6. Explica paso a paso la obtención de este estimador ¿Es mejor que el anterior? ¿En qué sentido? Razona tu respuesta.

4.6. Prácticas de Aula

El tema de Regresores Estocásticos tiene una dificultad considerable ya que los conceptos teóricos relacionados con teoría asintótica no son sencillos de comprender. Por ello las prácticas de aula son de una gran importancia. Más si cabe cuando el tema siguiente seguirá profundizando en los conceptos manejados en este tema. La resolución individual por parte del alumno de los ejercicios propuestos en las prácticas de aula, en su tiempo de estudio no presencial, le permitirá participar en clase contestando a las preguntas que se le realicen o poniendo en común dudas y comentarios surgidos. Previamente y con anticipación suficiente se le habrá realizado la debida propuesta.

En el tema de Regresores Estocásticos es habitual disponer de dos prácticas de aula, lo que equivale a dos horas de clase presencial y cuatro horas de trabajo personal. Si el ejercicio ha sido realizado previamente por el alumno el tiempo es suficiente para su corrección y solución de dudas existentes. A continuación se van a proponer los enunciados correspondientes a las dos prácticas citadas.

Competencias a trabajar en estas sesiones.

1. Comprender la importancia de los supuestos empleados en la especificación de un modelo econométrico básico para poder proponer y emplear supuestos más realistas.
2. Diferenciar distintos métodos de estimación y evaluar su uso de acuerdo a las características de las variables económicas de interés para obtener resultados fiables.

Práctica de Aula PA-RE.1:

Un estudiante está realizando su proyecto de fin de carrera sobre la demanda de pescado en el Fulton Fish Market, un mercado localizado en Nueva York y que opera desde hace 150 años. Para ello dispone de una muestra de 111 observaciones de datos diarios, desde el 2 de diciembre de 1991 al 8 de mayo de 1992, sobre las siguientes variables:

- *lquan*: Cantidad de merluza vendida en libras (en logaritmos).
- *lprice*: Precio de merluza por libra (en logaritmos).
- *mon*: Variable ficticia que toma valor 1 en lunes y 0 en otro caso.
- *tue*: Variable ficticia que toma valor 1 en martes y 0 en otro caso.
- *wed*: Variable ficticia que toma valor 1 en miércoles y 0 en otro caso.
- *thu*: Variable ficticia que toma valor 1 en jueves y 0 en otro caso.
- *stormy*: Variable ficticia que toma valor 1 si ese día hizo mucho viento y oleaje, 0 en otro caso.

La especificación para la ecuación de demanda es la siguiente:

$$lquan_t = \beta_1 + \beta_2 lprice_t + \beta_3 mon_t + \beta_4 tue_t + \beta_5 wed_t + \beta_6 thu_t + u_t$$

y los resultados de la estimación por MCO se muestran a continuación:

Ecuación de demanda: estimaciones MCO
Variable dependiente: lquan

Variable	Coeficiente	Desv. típica	Estadístico t	valor p
const	8,606890	0,143043	60,1698	0,0000
lprice	−0,562550	0,168213	−3,3443	0,0011
mon	0,014316	0,202647	0,0706	0,9438
tue	−0,516242	0,197690	−2,6114	0,0103
wed	−0,555373	0,202319	−2,7450	0,0071
thu	0,081621	0,197817	0,4126	0,6807

Suma de cuadrados de los residuos	47,1672
R^2	0,220486

El estudiante en su proyecto se cuestiona si, al ser un modelo en el que el precio y la cantidad se determinan conjuntamente en equilibrio entre oferta y demanda, la variable *lprice* pueda ser endógena, y estar correlacionada con el error de la ecuación. Por ello realiza la siguiente estimación:

Ecuación de demanda: estimaciones MC2E
Variable dependiente: lquan
Instrumented: stormy
Instrumentos: const stormy mon tue wed thu

Variable	Coeficiente	Desv. típica	Estadístico t	valor p
const	8,505910	0,166167	51,1890	0,0000
lprice	−1,119400	0,428645	−2,6115	0,0090
mon	−0,025402	0,214774	−0,1183	0,9059
tue	−0,530769	0,208000	−2,5518	0,0107
wed	−0,566351	0,212755	−2,6620	0,0078
thu	0,109267	0,208787	0,5233	0,6007

1. Explica en detalle como se han obtenido las estimaciones MC2E mostradas. Escribe de forma explícita cada una de las matrices que intervienen en la expresión del estimador.

2. Escribe y explica las condiciones tanto para poder obtener el estimador MC2E como para que éste sea consistente.

3. Contrasta la sospecha del estudiante. Escribe la hipótesis nula, la alternativa, el estadístico de contraste y su distribución bajo la hipótesis nula.

4. A la luz del resultado del contraste, ¿qué estimador elegirías? Razona tu respuesta en términos de las propiedades de los estimadores.

5. Contrasta la hipótesis nula de que una variación porcentual unitaria en el precio de la merluza se traduce en una variación porcentual unitaria en la cantidad demandada de merluza en ese mercado.

Práctica de Aula PA-RE.2:

Sea el siguiente modelo:

$$Y_t = \beta_1 X_{1t} + \beta_2 X_{2t} + u_t \qquad u_t \sim iid(0, \sigma_u^2) \qquad t = 1, 2, \ldots, T \tag{4.7}$$

donde $X_{1t} = \gamma Z_t + \eta_t \quad \eta_t \sim iid(0, \sigma_\eta^2)$ y X_{2t} y Z_t son variables no estocásticas.

1. ¿Cuándo estimarías el modelo por el método de Variables Instrumentales utilizando la variable Z_t como instrumento para la variable X_{1t}? ¿Por qué? ¿Crea problemas la variable X_{2t}? ¿Por qué?

A partir de una muestra de 52 observaciones se han obtenido los siguientes productos cruzados:

	Y_t	X_{1t}	X_{2t}	Z_t
Y_t	100	80	-60	60
X_{1t}		100	-40	-10
X_{2t}			80	50
Z_t				40

por ejemplo $\sum X_{1t} X_{2t} = -40$

2. Siendo Z_t el instrumento para X_{1t}, estima los coeficientes β_1 y β_2 del modelo utilizando el método de variables instrumentales.

Los resultados de estimar por MCO el modelo han sido:

$$\underset{(\widehat{des}(\hat{\beta}_{i,MCO}))}{\widehat{Y}_t} = \underset{(0{,}077)}{0,625}\ X_{1t} - \underset{(0{,}086)}{0,4375}\ X_{2t} \tag{4.8}$$

3. Contrasta la $H_0 : E(X_{1t}u_t) = 0$ sabiendo que:

$$\widehat{V}(\hat{\beta}_{VI}) = \begin{bmatrix} 2,1166 & 1,0583 \\ 1,0583 & 1,2254 \end{bmatrix}$$

Como conclusión del resultado del contraste ¿cuál es el método adecuado para estimar el modelo (4.8)? ¿Qué propiedades tienen dichos estimadores?

4.7. Prácticas de Ordenador

En el tema de Regresores Estocásticos es habitual disponer de una práctica de ordenador que equivale a una hora presencial y dos horas de trabajo no presencial. A continuación se va a proponer un ejercicio para resolver en el centro de cálculo. El enunciado cubre todo lo aprendido en el tema.

Enfatizamos que es conveniente que una vez acabado el ejercicio y en vuestro tiempo de trabajo personal deis contenido al mismo. Es decir, en clase únicamente nos da tiempo a aprender como ir obteniendo los resultados con gretl. Aunque el profesor va comentando y explicando los resultados. Vosotros y de forma personal debéis redactar convenientemente las respuestas de cada apartado. Notar que en algunos de ellos se incluye la coletilla "A realizar en casa" por esa razón. En el Anexo 4.1. aparecen las instrucciones de gretl para regresores estocásticos.

Competencias a trabajar en estas sesiones.

2. Diferenciar distintos métodos de estimación y evaluar su uso de acuerdo a las características de las variables económicas de interés para obtener resultados fiables.
3. Utilizar diversas fuentes estadísticas y adquirir destreza en el uso de un software econométrico para analizar relaciones entre variables económicas.

Práctica de Ordenador PO-RE.1

Para la realización de este ejercicio utilizamos el fichero de datos *smoke* del libro de Wooldridge que tenéis como archivo de muestra en gretl. Son datos para 807 individuos varones residentes en distintos estados americanos en el año 1979. Las variables que están en este fichero son:

- *income*: renta familiar anual en miles de dólares.
- *lincome*: logaritmo de la renta familiar anual en miles de dólares.
- *cigs*: el número medio de cigarrillos fumados por día.
- *educ*: el número de años escolarizado.
- *age*: edad del individuo en años.

- *agesq*: edad del individuo en años elevado al cuadrado.
- *cigpric*: el precio de un paquete de cigarrillos en centavos.
- *lcigpric*: logaritmo del precio de un paquete de cigarrillos en centavos.
- *restaurn*: variable ficticia que es igual a la unidad si una persona reside en un estado donde hay restricciones al tabaquismo en los restaurantes, cero en otro caso.
- *white*: variable ficticia que es igual a la unidad si el individuo es blanco, cero en otro caso.

Puedes acceder a estos datos ejecutando gretl → En Archivo → Abrir datos → Archivo de muestra → Elige Wooldridge, fichero **smoke**[20]

Considera la especificación del Modelo (4.9):

$$lincome_i = \beta_1 + \beta_2 cigs_i + \beta_3 educ_i + \beta_4 age_i + \beta_5 agesq_i + u_i \tag{4.9}$$

1. Muestra los resultados de la estimación por MCO del Modelo (4.9).
2. Comenta los resultados obtenidos sobre la bondad de ajuste, los coeficientes estimados y su significatividad. (“A realizar en casa”).
3. ¿Hay evidencia de que la relación entre la variable lincome y age sea cuadrática, manteniendo constante el resto de las variables explicativas? Muestra los resultados del contraste utilizado para tus conclusiones. (“A realizar en casa”).
4. Se cree que el consumo de cigarrillos puede estar determinado conjuntamente con la renta, tal que la variable **cigs** es un regresor estocástico correlacionado con el término de perturbación del Modelo (4.9). Explica las consecuencias que esto tiene sobre los resultados obtenidos en los apartados anteriores. (“A realizar en casa”).
5. Muestra los resultados de estimar el Modelo (4.9) por el método de Variables Instrumentales utilizando la variable **restaurn** como instrumento para **cigs**. ¿Son los resultados muy diferentes a los obtenidos por MCO? Comenta estos resultados.
6. Escribe la matriz de instrumentos Z y la matriz de variables explicativas del modelo, X. No pongas los valores muestrales, simplemente utiliza los nombres de las variables en las columnas. Escribe su dimensión. (“A realizar en casa”).
7. Escribe la expresión del estimador de VI utilizado. Escribe los elementos de $Z'X$ y $Z'Y$ utilizando sumatorios y su dimensión. ¿Qué características tiene que tener Z para que el estimador se pueda obtener? ¿Qué condiciones tiene que satisfacer Z para que el estimador tenga propiedades deseables y los contrastes sean válidos? (“A realizar en casa”).
8. Se dispone además de otro instrumento adicional para **cigs**, la variable **lcigpric**. Estima el Modelo (4.9) por Mínimos Cuadrados en 2 Etapas utilizando todos los instrumentos. Muestra el resultado obtenido en gretl. Compara estos resultados con los obtenidos en el apartado 5).

[20]Fichero smoke, disponible en gretl pestaña Wooldridge. Fuente: Wooldridge, J. M. (2003): *Introductory Econometrics*, fichero smoke.

9. Calcula las correlaciones entre los instrumentos y la variable **cigs**. ¿Qué indican estas correlaciones sobre la bondad de estos instrumentos?

10. Realiza la regresión de la variable **cigs** sobre **todos los posibles instrumentos** incluida la constante:

$$cigs = \alpha_1 + \alpha_2 educ + \alpha_3 age + \alpha_4 age^2 + \alpha_5 lcigpric + \alpha_6 restaurn + u_i \tag{4.10}$$

    Guarda la serie ajustada de la regresión $\widehat{cigs}_i \quad i = 1, \ldots, 879$ y utiliza esta variable como instrumento para cigs. ¿Obtienes los mismos resultados que en el apartado 8)? ¿Por qué obtienes esos resultados? ¿Son las variables *lcigpric* y *restaurn* significativas?

11. Realiza el contraste de Hausman para los resultados del apartado 5). A la vista del resultado ¿Cómo estimarías los coeficientes del Modelo (4.9)?

12. Contrasta si la variable edad es significativa. ¿En cuánto se estima el cambio en la renta cuando el individuo tiene un año adicional y el resto de las características se mantienen constantes? ¿Es esta variación la misma para todos los individuos en la muestra?

4.8. Evaluativas - Preguntas cortas

Se recuerda que dado que el curso contempla la evaluación continua es necesario que a lo largo del tema se evalúe a los alumnos, tanto en las clases magistrales, como en las prácticas de aula y ordenador. Se llevan a cabo en clase y de manera individual. A modo de ejemplo se incluyen las siguientes.

- Preguntas cortas en Clases Magistrales:

Pc1. Marca lo que sea **cierto**:

a) Si $X_t = 5 + v_t \qquad v_t \sim iid(0, \sigma_v^2)$
a) X_t es una variable constante b) X_t es una variable aleatoria

b) Sea $u_t \sim iid(0, \sigma_u^2)$ y X_t una variable no estocástica:
a) $E(X_t u_t) = 0$ b) $E(X_t u_t) \neq 0$

c) Sea $u_t \sim iid(0, \sigma_u^2)$ y X_t una variable estocástica independiente de $u_t \quad \forall t$:
a) $E(X_t u_t) = 0$ b) $E(X_t u_t) \neq 0$

d) Sean $u_t \sim iid(0, \sigma_u^2)$, X_t una variable estocástica y además, la correlación entre X_t y u_t no es cero.
a) $E(X_t u_t) = 0$ b) $E(X_t u_t) \neq 0$

Pc2. En el MRLG $Y = X\beta + u$ con $u_t \sim iid(0, \sigma_u^2)$ y X matriz de regresores estocástica independiente de u. Para cada apartado, marca lo que sea **cierto**.
El estimador $\hat{\beta}_{MCO} = \beta + (X'X)^{-1}(X'u)$:

1. a) Es lineal en u.

   b) Es no lineal en u.

2. a) Es insesgado.

   b) Es sesgado.

3. a) Tiene distribución conocida en muestras finitas.

   b) No tiene distribución conocida en muestras finitas.

• Preguntas cortas en Práctica de Aula:

Se dispone de 585 observaciones sobre las siguientes variables para el año 1970:

SAVE: Ahorro anual de la familia i.
INCOME: Renta agregada anual de la familia i.
SIZE: Tamaño de la familia i.

Se considera la especificación del Modelo (4.11):

$$SAVE_i = \beta_1 + \beta_2 INCOME_i + u_i \tag{4.11}$$

tal que $E(INCOME_i\, u_i) \neq 0$, $E(SIZE_i\, u_i) = 0$, $E(u_i) = 0$, $E(u_i^2) = \sigma_u^2$.

Pc1. ¿Por qué se debería estimar por VI y no por MCO los coeficientes de la relación anterior?

Pc2. Escribe la matriz de instrumentos Z y la matriz de variables explicativas del modelo, X. Escribe su dimensión.

$$Z = \begin{bmatrix} \quad \\ \quad \\ \quad \end{bmatrix} \qquad X = \begin{bmatrix} \quad \\ \quad \\ \quad \end{bmatrix}$$

Pc3. Escribe la expresión del estimador de VI utilizado. Escribe los elementos de $Z'X$ y $Z'Y$ utilizando sumatorios.

$$\widehat{\beta}_{VI} = \begin{bmatrix} \quad & \quad & \quad \\ \quad & \quad & \quad \\ \quad & \quad & \quad \end{bmatrix}^{-1} \begin{bmatrix} \quad \\ \quad \\ \quad \end{bmatrix} =$$

Pc4. ¿Qué características tiene que tener Z para que el estimador se pueda obtener? ¿Qué condiciones tiene que satisfacer Z para que el estimador tenga propiedades deseables y los contrastes sean válidos?

• Preguntas cortas en Práctica de Ordenador:

Se dispone de una muestra de 111 observaciones de datos diarios, desde el 2 de diciembre de 1991 al 8 de mayo de 1992, sobre las siguientes variables:

- *lquan*: Cantidad de merluza vendida en libras (en logaritmos).
- *lprice*: Precio de merluza por libra (en logaritmos).
- *mon*: Variable ficticia que toma valor 1 en lunes y 0 en otro caso.
- *tue*: Variable ficticia que toma valor 1 en martes y 0 en otro caso.
- *wed*: Variable ficticia que toma valor 1 en miércoles y 0 en otro caso.
- *thu*: Variable ficticia que toma valor 1 en jueves y 0 en otro caso.
- *stormy*: Variable ficticia que toma valor 1 si ese día hizo mucho viento y oleaje, 0 en otro caso.

El fichero de datos está disponible en la dirección: *http://people.brandeis.edu/∼kgraddy/data.html*

La especificación para la ecuación de demanda es la siguiente:

$$lquan_t = \beta_1 + \beta_2 lprice_t + \beta_3 mon_t + \beta_4 tue_t + \beta_5 wed_t + \beta_6 thu_t + u_t$$

Pc1. Estima el modelo anterior por VI utilizando como instrumento para *lprice* a la variable *stormy* y completa:

$$\underset{(\widehat{desv}(\hat{\beta}_{i,VI}))}{\widehat{lquan}_t} = \underset{(.............)}{..............} \quad \underset{(.............)}{..............}\, lprice_t \quad \underset{(.............)}{..............}\, mon_t$$

$$\underset{(.............)}{..............}\, tue_t \quad \underset{(.............)}{..............}\, wed_t \quad \underset{(.............)}{..............}\, thu_t$$

Pc2. Contrasta la posibilidad de que la variable estocástica *lprice* esté correlada con la perturbación.

a) H_0 : H_a :

b) Estadístico de contraste y distribución:

c) Computa el estadístico de contraste y lleva a cabo el contraste:

d) A la vista de los resultados del contraste propón un método de estimación adecuado para los coeficientes de la relación y enumera sus propiedades.

4.9. Bibliografía del tema

Referencias Bibliográficas Básicas:

• Teórica:

[1] Greene, W. (1998), Análisis Econométrico, ed. Prentice Hall, 3ª edition.

[2] Ramanathan, R. (2002), Introductory Econometrics with applications, ed. South-Western, 5th edition.

[3] Wooldridge, J. M. (2003), Introductory Econometrics: A modern Approach, ed. South-Western, 2nd edition.

• Ejercicios:

[1] Alegre, J., Arcarons, J., Bolancé, C. y Díaz, L. (1995), Ejercicios y Problemas de Econometría, Colección Plan Nuevo, ediciones AC.

[2] Fernández, A., González, P., Regúlez, M., Moral, P., Esteban, V. (2005), Ejercicios de Econometría, ed. McGraw-Hill, 2 ª edición.

[3] Recopilación de ejercicios recomendados y exámenes de Econometría. Departamento de Economía Aplicada III (Econometría y Estadística). Mimeo Febrero 2009.

• Ejercicios con gretl:

[1] Ramanathan, R. (2002) Instructor's Manual to accompany, del libro Introductory Econometrics with applications, ed. South-Western, 5th edition, Harcourt College Publishers.

[2] Wooldridge, J.M. (2003), Student Solutions Manual, del libro Introductory Econometrics: A modern Approach, ed. South-Western, 2nd edition.

Referencias Bibliográficas Complementarias:

[1] Alonso, A., Fernández, J. y Gallastegui, I. (2005), Econometría, ed. Pearson: Prentice Hall.

[2] Gujarati, D. (2004), Econometría, ed. McGraw-Hill, 4ª edición.

[3] Johnston, J. (1984), Métodos de Econometría, ed. Vicens Vivens, 4ª edición.

[4] Maddala, G. S. (1996), Introducción a la Econometría, ed. Pearson: Prentice Hall.

[5] Novales, A. (1993), Econometría, ed. McGraw-Hill, 2ª edición.

[6] Pindyck, R. S. y Rubinfeld, D. L. (1998), Econometric Models and Economic Forecast, ed. McGraw-Hill, 4ª edición.

4.10. Anexo 4.1: Instrucciones básicas de gretl para regresores estocásticos

En este anexo se van a recoger las instrucciones básicas que incluye gretl para poder computar el estimador de VI con el fin de que podáis llevar a cabo la práctica de ordenador con cierta soltura. Instrucciones de carácter básico ya han sido recogidas en temas anteriores.

• **Para computar el estimador de Variables Instrumentales:**

Pulsar sucesivamente en

Modelo $\longrightarrow$ *Otros modelos lineales* $\longrightarrow$ *Mínimos cuadrados en dos etapas*

Seleccionar la variable endógena y las variables independientes de la manera habitual.

A continuación seleccionar los instrumentos.

Hay que tener en cuenta que le estamos indicando a gretl como es la matriz de instrumentos Z y que debe tener tantas columnas como la matriz X que recoge a las variables independientes. Luego debemos incluir como mínimo tantos instrumentos como regresores tenga el modelo. Para aquellas variables no estocásticas o que no estén correlacionadas con la perturbación el mejor instrumento son ellas mismas. Para aquellas variables correlacionadas con la perturbación se selecciona el instrumento-(s) adecuado-(s).

No es necesario respetar el mismo orden que al incluir las variables independientes.

En ocasiones los instrumentos a incluir son retardos de los regresores. En este caso en la ventana gretl: *especificar modelo* tenemos que elegir la variable dependiente, las variables explicativas en t o corrientes y los instrumentos que estarán en la lista en t.

A continuación se pulsa en retardos y aparece una ventana donde se tienen que seleccionar los retardos de las variables que aparecen en el modelo como regresores, tanto de variables exógenas como de la dependiente y también las variables que vamos a usar como instrumentos.

4.11. Anexo 4.2. Errores de medida en las variables

Hasta ahora hemos supuesto que las variables utilizadas en el proceso de estimación se medían sin error. En la práctica es muy posible que existan errores de medida en las variables o que simplemente las variables a utilizar no sean, sino estimaciones de conceptos teóricos que no se observan en la realidad, por ejemplo el stock de capital, el PIB, o las variables de Contabilidad Nacional. Estas situaciones alterarán las propiedades de los estimadores de los coeficientes en concreto introduciendo sesgos en las estimaciones y generando estimadores de MCO inconsistentes. A continuación vamos a repasar esas propiedades. Podemos distinguir tres situaciones:

1. Variable endógena medida con error.

2. Variable exógena medida con error. Analizada en el tema.

3. Variable exógena y endógena medidas con error.

4.11.1. Variable endógena medida con error

Sea el verdadero modelo:

$$Y_t^\star = \alpha + \beta X_t + u_t^\star \qquad u_t^\star \sim iid(0, \sigma^2_{u^\star}) \qquad t = 1, 2, \ldots, T$$

donde X_t es no estocástica. Por alguna razón la variable endógena disponible no es $Y_t^\star$, sino $Y_t = Y_t^\star + \epsilon_t$ con $\epsilon_t \sim iid(0, \sigma^2_\epsilon)$, $Cov(X_t, \epsilon_t) = 0$, $Cov(u_t^\star, \epsilon_t) = 0$.

El modelo estimable en términos de las variables observables es:

$$\begin{aligned} Y_t - \epsilon_t = \alpha + \beta X_t + u_t^\star \quad &\Rightarrow Y_t = \alpha + \beta X_t + (u_t^\star + \epsilon_t) \\ &\Rightarrow Y_t = \alpha + \beta X_t + u_t \end{aligned}$$

El error de medida en la variable endógena se acumula en la perturbación original. Las propiedades de la nueva perturbación $u_t^\star$ son:

$$\begin{array}{lll} E(u_t) & = E(u_t^\star + \epsilon_t) = E(u_t^\star) + E(\epsilon_t) = 0 & \\ Var(u_t) & = E(u_t - E(u_t))^2 = E(u_t)^2 = E(u_t^\star + \epsilon_t)^2 & \\ & = E(u_t^{\star 2}) + E(\epsilon_t^2) + 2E(u_t^\star \epsilon_t) = \sigma^2_{u^\star} + \sigma^2_\epsilon & \text{homocedástica} \\ Cov(u_t, u_s) & = E[(u_t - E(u_t))(u_s - E(u_s))] = & \\ & = E(u_t u_s) = E(u_t^\star + \epsilon_t, u_s^\star + \epsilon_s) = 0 & \text{no autocorrelada} \end{array}$$

$$u_t \sim iid(0, \sigma^2_{u^\star} + \sigma^2_\epsilon)$$

En este caso, en presencia de errores de medida en la variable endógena y cuando el error de medida es incorrelado con la perturbación del modelo original el error de medida se traslada a la perturbación del modelo estimable, incrementándola, pero se mantienen sus propiedades esféricas. Por tanto, los MCO son apropiados y tienen buenas propiedades.

4.11.2. Variable exógena y variable endógena medidas con error

Sea el verdadero modelo:

$$Y_t^\star = \alpha + \beta X_t^\star + u_t^\star \qquad t = 1, 2, \ldots, T$$

$$\begin{array}{ll} \text{donde}: & Y_t = Y_t^\star + \epsilon_t \quad \text{es la variable endógena disponible.} \\ & X_t = X_t^\star + v_t \quad \text{es la variable exógena disponible, y } X_t^\star \text{ es no estocástica} \\ & u_t^\star \sim iid(0, \sigma^2_{u^\star}) \quad v_t \sim iid(0, \sigma^2_v) \quad \epsilon_t \sim N(0, \sigma^2_\epsilon) \\ & Cov(u_t^\star, \epsilon_t) = Cov(u_t^\star, v_t) = Cov(\epsilon_t, v_t) = 0 \end{array}$$

El modelo estimable en términos de las variables observables es:

$$Y_t - \epsilon_t = \alpha + \beta(X_t - v_t) + u_t^\star \qquad t = 1, 2, \ldots, T$$

$$Y_t = \alpha + \beta X_t + (u_t^\star + \epsilon_t - \beta v_t) \qquad t = 1, 2, \ldots, T$$

$$Y_t = \alpha + \beta X_t + u_t \qquad t = 1, 2, \ldots, T$$

La perturbación del modelo estimable es homocedástica y no autocorrelada

$$u_t \sim iid(0, \sigma^2_{u^\star} + \sigma^2_\epsilon + \beta^2 \sigma^2_v)$$

Existe correlación contemporánea ya que:

$$\begin{aligned} E(X_t u_t) &= E(X_t^\star + v_t)(u_t^\star + \epsilon_t - \beta v_t) = \\ &= E(X_t^\star u_t^\star) + E(v_t u_t^\star) + E(X_t^\star \epsilon_t) + E(v_t \epsilon_t) - \beta E(X_t^\star v_t) - \beta E(v_t^2) = \\ &= -\beta \sigma^2_v \end{aligned}$$

No existe correlación no contemporánea ya que $E(X_s u_t) = 0$

En este caso existen dos fuentes de error, el error de medida en la variable endógena y el error de medida en la variable exógena. El error de medida en la variable endógena implica un incremento en la varianza de la perturbación del modelo estimable mientras que el error de medida en la variable exógena implica que los estimadores MCO de α y β serán sesgados e inconsistentes. El modelo estimable, bajo todos los supuestos realizados anteriormente, debería ser estimado por el Método de Variables Instrumentales.

4.12. Anexo 4.3. Estimador de Variables Instrumentales

A continuación se va a mostrar, a partir de la función objetivo, cómo obtener el estimador de Variables Instrumentales. Demostración:

$$\begin{aligned} &Y = X\beta + u \\ &Z'Y = Z'X\beta + Z'u \end{aligned}$$

Denotamos por $u_\star = Z'u$

$$\begin{aligned} &u_\star = Z'u = Z'(Y - X\beta) \\ &\hat{u}_\star = Z'\hat{u} = Z'(Y - X\hat{\beta}) \\ &\hat{u}_\star'\hat{u}_\star = \left[Z'(Y - X\hat{\beta})\right]' \left[Z'(Y - X\hat{\beta})\right] = (Y - X\hat{\beta})'ZZ'(Y - X\hat{\beta}) \end{aligned}$$

La función objetivo podemos escribirla como:

$$\underset{\hat{\beta}}{Min}\ \hat{u}_\star'\hat{u}_\star = \underset{\hat{\beta}}{Min}\ \left[Y'ZZ'Y - 2\hat{\beta}'X'ZZ'Y + \hat{\beta}'X'ZZ'X\hat{\beta}\right]$$

por las condiciones de primer orden obtenemos:

$$\begin{aligned} \frac{\partial(\hat{u}_\star'\hat{u}_\star)}{\partial\hat{\beta}} &= -2X'ZZ'Y + 2X'ZZ'X\hat{\beta} = 0 \\ &= X'ZZ'(Y - X\hat{\beta}) = 0 \end{aligned}$$

De donde las ecuaciones normales del modelo serían:

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

tal que:

$$\begin{aligned} \hat{\beta}_{VI} &= (X'ZZ'X)^{-1}(X'ZZ'Y) \\ \hat{\beta}_{VI} &= (Z'X)^{-1}(X'Z)^{-1}(X'Z)(Z'Y) \\ \hat{\beta}_{VI} &= (Z'X)^{-1}(Z'Y) \end{aligned}$$

4.13. Anexo 4.4. Estimador de Mínimos Cuadrados en dos etapas

A continuación vamos a mostrar como derivar el estimador de MC2E. Sea el modelo en forma matricial:

$$\underset{(T\times 1)}{Y} = \underset{(T\times K)}{X}\ \underset{(K\times 1)}{\beta} + \underset{(T\times 1)}{u} \iff \underset{(T\times 1)}{Y} = \underset{(T\times K_1)}{X_1}\ \underset{(K_1\times 1)}{\beta_1} + \underset{(T\times K_2)}{X_2}\ \underset{(K_2\times 1)}{\beta_2} + \underset{(T\times 1)}{u} \qquad u \sim (0,\sigma^2 I)$$

La matriz $X = [\,\mathbf{X}_1 \vdots \mathbf{X}_2\,]$ es estocástica y tal que $E(\mathbf{X}_1 u) = 0$ y $E(\mathbf{X}_2 u) \neq 0$. Para la matriz de instrumentos Z necesitaremos al menos K_2 instrumentos para las variables X_2 además, de poder utilizar a X_1 como instrumentos para ellas mismas.

$$\underset{(T\times l)}{Z} = [\mathbf{X}_1 \vdots \mathbf{X}_2] \quad \text{donde} \quad l \geq K$$

- Si $l = K$ entonces tenemos exactamente el mismo número de instrumentos que necesitamos. $(Z'X)$ es una matriz cuadrada de orden $(K \times K)$ y no singular tal que existe $(Z'X)^{-1}$ y podemos definir el estimador de Variables Instrumentales de la forma $\hat{\beta}_{VI} = (Z'X)^{-1}Z'Y$.
- Si $l > K$ entonces tenemos más instrumentos que los que necesitamos $(Z'X)$ es de orden $(l \times K)$ y no es cuadrada.

El proceso de estimación que comprende el estimador de Mínimos Cuadrados en dos Etapas es el siguiente:

En la primera etapa se lleva a cabo la regresión de cada una de las columnas de la matriz X sobre Z para obtener una matriz de valores ajustados $\widehat{X}$.

$$\underset{(T\times K)}{\widehat{X}} = \underset{(T\times l)}{Z} \underbrace{\underset{(l\times l)}{(Z'Z)^{-1}}\ \underset{(l\times T)}{Z'}\ \underset{(T\times K)}{X}}_{\text{matriz de coeficientes de las K-regresiones}} = P_z X \quad \text{donde} \quad P_z = Z(Z'Z)^{-1}Z'$$

P_z es idempotente y simétrica, $P_z' = P_z$ y $P_z P_z = P_z$.

En la segunda etapa se realiza la regresión de Y sobre $\widehat{X}$ para obtener el estimador de los coeficientes β por Mínimos Cuadrados en 2 Etapas.

$$\begin{aligned}
\hat{\beta}_{MC2E} &= (\widehat{X}'\widehat{X})^{-1}\widehat{X}'Y = \\
&= (X'\underbrace{P_z P_z}_{P_z} X)^{-1}X'P_zY = (X'P_zX)^{-1}X'P_zY = \\
&= (\widehat{X}'X)^{-1}\widehat{X}'Y = (X'Z(Z'Z)^{-1}Z'X)^{-1}X'Z(Z'Z)^{-1}Z'Y
\end{aligned}$$

luego el estimador de MC2E es un estimador de Variables Instrumentales que utiliza como matriz de instrumentos a $\widehat{X} = P_z X$ donde combina todos los instrumentos de forma óptima en el sentido de que es un estimador consistente y es el más eficiente asintóticamente dentro de la clase de estimadores de VI que toman solamente un subconjunto de todos los posibles instrumentos Z.

Hay que notar que, si $l = K$, entonces $(X'Z)$ y $(Z'X)$ son cuadradas y $(X'Z(Z'Z)^{-1}Z'X)^{-1} = (Z'X)^{-1}Z'Z(X'Z)^{-1}$ por lo que en ese caso:

$$\hat{\beta}_{MC2E} = (Z'X)^{-1}Z'Z(X'Z)^{-1}X'Z(Z'Z)^{-1}Z'Y = (Z'X)^{-1}Z'Y$$

Para estimar la matriz de varianzas y covarianzas asintótica de $\hat{\beta}_{MC2E}$ se utiliza el estimador consistente:

$$\widehat{V}(\hat{\beta}_{MC2E}) = \frac{(Y - X\hat{\beta}_{MC2E})'(Y - X\hat{\beta}_{MC2E})}{T}(\widehat{X}'\widehat{X})^{-1}$$

Tema 5

Modelos Dinámicos

Clases Magistrales

Habitualmente se dedican a este tema tres clases magistrales. El alumno tiene los instrumentos necesarios para abordar este tema sin dificultad, aunque es un tema en el que se debe de tener una buena visión de conjunto, tanto en términos de modelización con datos en el tiempo, como de estimación por diversos métodos. En una primera clase magistral se introduce el tema y el profesor ilustra al alumno lo que se entiende por un modelo dinámico distinguiendo entre diversas especificaciones. Se considera la especificación dinámica solamente en la parte sistemática teniendo como variables explicativas retardos de variables exógenas o incorreladas con el término de error, para ampliar seguidamente con retardos de la variable endógena. Más adelante se combina con dinámica a través del término de error. Las dos clases siguientes se dedican a ilustrar los contenidos del tema con un ejemplo sobre la influencia de la exención fiscal en la tasa de fertilidad usando datos históricos para EE.UU. Se parte de una especificación estática y se va reespecificando el modelo hacia distintas formas de introducir dinámica. De esta forma, el alumno recorre todos los casos propuestos en la primera clase y se da una visión de conjunto revisando la utilización del análisis gráfico de las series utilizadas en el modelo y de los residuos, los contrastes de autocorrelación vistos en el tema de autocorrelación, así como diversos métodos de estimación MCO, VI y MCGF haciendo hincapié en los problemas específicos que surgen en estos modelos. Se analizan cuestiones de multicolinealidad, sesgo de estimación por MCO en modelos con retardos de la variable endógena como regresor y autocorrelación y las posibles alternativas para solventar en su caso estos problemas.

A lo largo de las clases magistrales se pedirá al alumno, además de la lectura de las notas sobre el tema, la realización de diversos ejercicios que se resolverán en clase. Para conocer el grado de seguimiento del alumno, se podrán realizar en clase preguntas cortas a resolver.

Competencias a trabajar en estas sesiones.

1. Comprender la importancia de los supuestos empleados en la especificación de un modelo econométrico básico para poder proponer y emplear supuestos más realistas.

2. Diferenciar distintos métodos de estimación y evaluar su uso de acuerdo a las características

de las variables económicas de interés para obtener resultados fiables.

Al final de este tema deberíais ser capaces de

1. Explicar que se entiende por dinámica en un modelo econométrico.
2. Saber analizar gráficamente series en el tiempo, tanto la serie observada y ajustada como la de los residuos MCO.
3. Utilizar el contraste de Breusch-Godfrey para contrastar la posible existencia de autocorrelación y saber cuando no es adecuado el contraste de Durbin-Watson.
4. Describir y comparar las propiedades de los estimadores MCO, VI y MCGF en diversas especificaciones dinámicas.
5. Estimar y hacer inferencia.
6. Saber valorar los resultados obtenidos y elegir los más adecuados.

Bibliografía Recomendada. Al final del tema tenéis recogida la bibliografía correspondiente. En particular os recomendamos leer los capítulos correspondientes a la bibliografía básica detallados a continuación:

- Greene, W. (1998), cap. 17
- Ramanathan, R. (2002), cap. 6 y 10
- Wooldridge, J.M. (2003), cap. 15 y 18

Y para profundizar, podéis leer los capítulos detallados a continuación correspondientes a la bibliografía complementaria:

- Alonso, A. y otros (2005), cap. 11
- Gujarati, D. (2004), cap. 17
- Johnston, J. (1984), cap. 9 y 10
- Novales, A. (1993), cap. 9
- Pindyck R. S. y Rubinfeld, D.L. (1998), cap. 14

5.1. Introducción

Este tema vuelve a ocuparse de la modelización de relaciones entre variables dentro de un contexto de datos en el tiempo o series temporales. Como ya comentamos en el tema de autocorrelación, estos datos presentan un ordenamiento natural, ya que las observaciones están ordenadas de acuerdo al momento del tiempo en que han sido observadas.

A su vez, será probable que a la hora de explicar el comportamiento de una variable observada en el tiempo, se tenga en cuenta que bien su propio pasado pueda ser relevante en periodos sucesivos, o bien que el efecto de una variable explicativa no sea sólo contemporáneo sino que puede tener efectos distribuidos en años sucesivos. Finalmente, además se pueden tener factores no observables recogidos en el término de perturbación del modelo cuya influencia se mantiene a lo largo del tiempo. De esa última parte ya nos ocupamos en el tema de autocorrelación. En este nuevo tema trataremos de estudiar la manera de recoger, en un modelo de regresión lineal, estos efectos temporales o dinámicos en la parte que llamamos sistemática, o de factores observables, y su posible interacción con el tema de autocorrelación, que ya de por sí introduce dinámica en el error, donde se recogen efectos o variables no observables. Primeramente nos ocuparemos de ver distintas especificaciones y en cada caso se comentarán los problemas que pueden surgir en la estimación por el método de Mínimos Cuadrados Ordinarios. Si hay métodos alternativos a MCO también se discutirá su utilización y propiedades.

Por último, nos ocuparemos de un ejemplo empírico que nos llevará por diferentes especificaciones: desde un modelo estático donde se especifica una relación contemporánea entre la tasa de fertilidad anual en EE.UU. en el siglo XX y las exenciones fiscales en ese año, hacia diferentes especificaciones más generales donde se introduce explícitamente la modelización dinámica en todas sus fuentes. Este ejemplo también será la base para las prácticas de ordenador. De esta forma, lo estudiado en las clases magistrales se reforzarán a la hora de obtener los resultados en el centro de cálculo.

5.2. Especificación y estimación de modelos dinámicos

En el contexto del modelo de regresión lineal general, las relaciones dinámicas entre la variable dependiente y las variables explicativas se pueden introducir de diversas maneras:

5.2.1. Dinámica solamente en la parte sistemática

Entendemos por parte sistemática del modelo a aquella que contiene las variables explicativas observables o regresores del modelo. Se pueden considerar dos tipos de regresores: retardos de la variable dependiente o endógena que son regresores estocásticos, y por otro lado retardos de las variables consideradas fijas o exógenas, esto es, que en ningún caso están correlacionadas con el término de error del modelo e incluso se pueden considerar independientes de éste. En el modelo podemos tener solamente un tipo de cada uno o también una mezcla, pero sus características diferentes hacen que tengamos que ser cautelosos a la hora de estudiar el tipo de problemas que pueden introducir a la hora de la estimación e inferencia.

Modelos con variables no estocásticas o exógenas retardadas

La formulación general del modelo, para el caso de una sola variable exógena X_t o regresor fijo, es:

$$Y_t = \beta_1 + \beta_2 X_t + \beta_3 X_{t-1} + \beta_4 X_{t-2} + \ldots + \beta_{m+1} X_{t-m} + u_t \tag{5.1}$$

Este tipo de modelos recibe el nombre de modelo de retardos distribuidos finitos.

Si $u_t \sim iid(0, \sigma^2)$ y los regresores son variables no estocásticas, el estimador MCO es lineal, insesgado y eficiente. El problema que puede surgir es la poca precisión en la estimación de los efectos individuales de la variable X_t y sus retardos, ya que el grado de multicolinealidad entre estos regresores puede ser muy alto. Esto es debido a que las series temporales económicas suelen presentar cambios lentos de un período a otro. Por otro lado, cuantos más retardos introduzcamos más coeficientes tendremos que estimar a la vez que dispondremos de menos observaciones, por lo que existe una disminución por ambos lados de los grados de libertad.

Aquí surge otro problema ya que el número de retardos a introducir tal que el error sea ruido blanco puede ser elevado. Por lo tanto, la inferencia para poder determinar qué retardos son relevantes o cuantos introducir en el modelo desde el inicio de la especificación puede verse afectada por este problema de multicolinealidad, agravado por tener pocos grados de libertad. Este es un problema de difícil solución, a no ser que se imponga cierta estructura *ad-hoc* o restricciones en los coeficientes reduciendo el número de parámetros a estimar.

Modelos con variable endógena retardada

Si el modelo incluye entre sus variables explicativas retardos de la variable endógena, se deja de cumplir el supuesto de regresores no estocásticos. La matriz de datos X es estocástica, ya que los regresores que son retardos de la variable dependiente $Y_{t-1}, Y_{t-2}, \cdots, Y_{t-s}$, son variables aleatorias que vienen determinados por los términos de perturbación $u_{t-1}, u_{t-2}, \cdots, u_{t-s}, \ldots$. Nos encontramos, por lo tanto, en el caso de regresores estocásticos y es preciso hacer supuestos sobre la distribución conjunta de X y u.

Ejemplo 5.11 En el modelo:

$$Y_t = \beta_1 + \beta_2 X_t + \beta_3 Y_{t-1} + u_t \tag{5.2}$$

El supuesto de que la matriz X sea independiente del vector u significa que tanto X_t como Y_{t-1} han de ser independientes de todos los valores pasados, presentes y futuros de la perturbación. El modelo (5.2) implica que el regresor Y_{t-1} está relacionado con u_{t-1}, $u_{t-2}, \ldots$, por lo que este supuesto no se cumple. Luego implica que el estimador MCO es un estimador sesgado y no se conoce su distribución exacta para un tamaño de muestra dado.

Sin embargo, si el término de perturbación es un ruido blanco $u_t \sim iid(0, \sigma^2)$, entonces Y_{t-1} no está correlacionado con u_t. Es decir, se cumple que $E(Y_{t-1}\, u_t) = 0$, lo que significa que nos encontramos en el caso de un modelo con regresores estocásticos que no están correlacionados contemporáneamente con la perturbación, por lo que, como vimos en el tema de Regresores Estocásticos, los estimadores MCO son consistentes y tienen distribución asintótica Normal.

Por lo tanto, los modelos dinámicos con variable endógena retardada y perturbaciones bien comportadas se pueden estimar por MCO, conservando estos estimadores sus buenas propiedades asintóticas y siendo válidos asintóticamente los resultados habituales de inferencia estadística.

5.2.2. Dinámica en la parte sistemática y en la perturbación

Consideremos los modelos anteriores pero añadamos que tenemos evidencia de autocorrelación en el término de perturbación. En general para contrastar la existencia de autocorrelación en el término de perturbación en modelos con retardos de variables explicativas, se pueden utilizar tanto el contraste de Durbin-Watson como el de Breusch-Godfrey. Ahora bien, el test de Durbin-Watson no es aplicable cuando en el modelo la variable endógena retardada aparece como regresor. En ese caso el estadístico está sesgado hacia valores que indican ausencia de autocorrelación. Además las cotas de valores críticos tampoco son válidas. Luego, en este caso utilizaremos el contraste de Breusch-Godfrey que sigue siendo válido.

El hecho de tener autocorrelación en el error introduce diferentes implicaciones en las propiedades del estimador MCO al tener solamente retardos de variables exógenas o regresores no estocásticos o tener también retardos de la variable endógena. Por eso los trataremos por separado.

Ejemplo 5.12 Modelos con variables fijas o exógenas retardadas y autocorrelación

Por ejemplo, consideremos el modelo:

$$Y_t = \beta_1 + \beta_2 X_t + \beta_3 X_{t-1} + \beta_4 X_{t-2} + u_t \qquad u_t = \rho u_{t-1} + \varepsilon_t \quad \varepsilon_t \sim iid(0, \sigma^2_\varepsilon) \tag{5.3}$$

En este caso se combinan los problemas comentados anteriormente debido a la posible multicolinealidad entre los regresores y el problema de autocorrelación, teniendo ambos como consecuencia falta de precisión en la estimación por MCO. Además, como ya hemos comentado en temas anteriores, al existir autocorrelación los estadísticos t y F calculados con el estimador habitual de la matriz de varianzas y covarianzas $\widehat{V}(\hat{\beta}_{MCO}) = \hat{\sigma}^2(X'X)^{-1}$ no son fiables.

La estimación por MCG, o por MCGF si ρ no es conocido, será más eficiente que MCO, al menos asintóticamente, y la inferencia sobre los coeficientes basados en estos métodos será válida. Pero el problema de la multicolinealidad y sus efectos sobre la precisión en las estimaciones de los coeficientes individuales por estos métodos persiste.

Ejemplo 5.13 Modelos con variable endógena retardada y autocorrelación

Supongamos por ejemplo el modelo:

$$Y_t = \beta_1 + \beta_2 X_t + \beta_3 Y_{t-1} + u_t \tag{5.4}$$

donde X_t es un regresor no estocástico y u_t no es un ruido blanco, sino que presenta autocorrelación, bien un proceso AR(p) o MA(q).

La existencia de autocorrelación en las perturbaciones hace que Y_{t-1} esté relacionada con u_t mediante la función de regresión y u_t está correlacionado con u_{t-1} debido a la autocorrelación de al menos orden uno en las perturbaciones. Por lo que, por ejemplo en el caso del modelo (5.4) si u_t sigue un AR(1):

$$E(Y_{t-1}u_t) = E\left(\beta_1 + \beta_2 X_{t-1} + \beta_3 Y_{t-2} + u_{t-1}\right)\left(\rho u_{t-1} + \varepsilon_t\right) \neq 0$$

Estamos en el caso de regresores estocásticos correlacionados con el término de perturbación del modelo que tratamos en el tema de Regresores Estocásticos. En esta situación se demostró que los estimadores MCO de los parámetros del modelo no son consistentes, perdiendo así tanto las buenas propiedades para muestra finitas como asintóticas. Para obtener estimadores consistentes de los parámetros del modelo (5.4) podemos aplicar los siguientes métodos:

- **Estimador de variables instrumentales.** En el modelo (5.4) sólo

  hemos de encontrar un instrumento para Y_{t-1} porque la variable X_t no plantea problemas al ser una variable exógena o regresor no estocástico para todo t. Por esa razón, un instrumento podría ser la variable X_{t-1}, que cumple las condiciones de no estar correlacionada con u_t y sí con Y_{t-1} a través del propio modelo. La matriz Z de instrumentos, en este caso, puede ser la siguiente:

$$\underset{((T-1)\times 3)}{Z} = \begin{bmatrix} 1 & X_2 & X_1 \\ 1 & X_3 & X_2 \\ \vdots & \vdots & \vdots \\ 1 & X_T & X_{T-1} \end{bmatrix}$$

  siendo

$$\underset{((T-1)\times 3)}{X} = \begin{bmatrix} 1 & X_2 & Y_1 \\ 1 & X_3 & Y_2 \\ \vdots & \vdots & \vdots \\ 1 & X_T & Y_{T-1} \end{bmatrix}; \quad \underset{((T-1)\times 1)}{Y} = \begin{bmatrix} Y_2 \\ Y_3 \\ \vdots \\ Y_T \end{bmatrix}$$

  Al no estar X_{t-1} como regresor en el modelo, el rango de la matriz $(Z'X)$ es completo e igual a 3, por lo que al existir su inversa, el estimador de VI está bien definido como:

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

$$\hat{\beta}_{VI} = \begin{bmatrix} T-1 & \sum_{t=2}^{T} X_t & \sum_{t=2}^{T} Y_{t-1} \\ \sum_{t=2}^{T} X_t & \sum_{t=2}^{T} X_t^2 & \sum_{t=2}^{T} X_t Y_{t-1} \\ \sum_{t=2}^{T} X_{t-1} & \sum_{t=2}^{T} X_{t-1}X_t & \sum_{t=2}^{T} X_{t-1}Y_t \end{bmatrix} \begin{bmatrix} \sum_{t=2}^{T} Y_t \\ \sum_{t=2}^{T} X_t Y_t \\ \sum_{t=2}^{T} X_{t-1}Y_t \end{bmatrix}$$

  Ahora bien, este estimador, aunque consistente, no es asintóticamente eficiente. Además, al existir autocorrelación en el término de error, la distribución asintótica derivada en el tema de regresores estocásticos utilizando el teorema de Mann-Wald no es la adecuada en este caso.

Por lo tanto, hay que ser cuidadosos a la hora de computar adecuadamente las desviaciones típicas para realizar los contrastes utilizando este estimador.

Una opción alternativa, pero que puede requerir de este estimador como una primera etapa, es utilizar la estimación obtenida por el método de Mínimos Cuadrados Generalizados Factibles.

- **Estimador de Mínimos Cuadrados Generalizados Factibles.** Si la perturbación del modelo (5.4) sigue un $AR(1)$, el modelo transformado que cumple las hipótesis básicas conocido el valor de ρ, para $t = 3, \ldots, T$ es:

$$Y_t - \rho Y_{t-1} \;=\; \beta_1 (1-\rho) + \beta_2 (X_t - \rho X_{t-1}) + \beta_3 (Y_{t-1} - \rho\, Y_{t-2}) + \varepsilon_t \tag{5.5}$$

El término de error de este modelo, ε_t, es un ruido blanco, por lo que aunque los regresores $(Y_{t-1} - \rho\, Y_{t-2})$ son estocásticos y no son independientes de ε_t, están contemporáneamente incorrelacionadas, $E[(Y_{t-1} - \rho\, Y_{t-2}),\, \varepsilon_t] = 0,\ \forall t$.

Así, tanto el método de Hildreth-Lu como el de Cochrane-Orcutt, son procesos iterativos que se basan en, dado un valor de ρ, minimizar respecto a β la función:

$$SCR(\hat{\beta}) = \sum_{t=3}^{T} \left[(Y_t - \hat{\rho} Y_{t-1}) - \hat{\beta}_1 (1-\hat{\rho}) - \hat{\beta}_2 (X_t - \hat{\rho} X_{t-1}) - \hat{\beta}_3 (Y_{t-1} - \hat{\rho}\, Y_{t-2}) \right]^2 \tag{5.6}$$

La cuestión relevante es cómo fijar el valor de $\hat{\rho}$ para que el estimador de β así obtenido sea consistente y asintóticamente eficiente. En el caso de el método de Hildreth-Lu no hay problema porque la estimación de ρ se hace en función de una red de valores para este parámetro en el intervalo $(-1, 1)$ tal que se elige aquél asociado con el menor valor en la red calculada como $SCR(\hat{\beta})$ en (5.6).

En el caso del método de Cochrane-Orcutt no es adecuado obtener un estimador de ρ basado en los residuos de estimar los coeficientes β del modelo (5.4) por MCO, ya que éste no es consistente. El proceso se puede iniciar, por ejemplo, calculando los residuos partiendo del estimador por Variables Instrumentales, $\hat{\beta}_{VI}$, de los coeficientes del modelo (5.4). Así, calcularíamos el estimador inicial de ρ como:

$$\hat{\rho} = \frac{\sum_{t=2}^{T} \hat{u}_{VI,t} \hat{u}_{VI,t-1}}{\sum_{t=2}^{T} \hat{u}^2_{VI,t-1}}$$

donde $\hat{u}_{VI,t} = Y_t - \hat{\beta}_{1,VI} + \hat{\beta}_{2,VI} X_t + \hat{\beta}_{3,VI} Y_{t-1} \qquad t = 2, \ldots, T$. Posteriormente procederíamos de igual forma, sustituyendo ese valor en la función (5.6) y obtendríamos el valor de $\hat{\beta}$ que la minimiza, que sería el obtenido de la regresión en el modelo transformado con ese valor de $\hat{\rho}$. El estimador así obtenido en esta segunda etapa es consistente y asintóticamente eficiente. El procedimiento puede iterarse comenzando con este estimador de β para volver a obtener los residuos y así sucesivamente. Asintóticamente, no se mejora en eficiencia si iteramos.

5.3. Ejemplo magistral: hacia una modelización dinámica

En un estudio sobre la política de natalidad del gobierno de EE.UU. en el siglo XX, los autores disponen de datos anuales sobre las siguientes variables[1], para el período 1913-1984:

gfr:	tasa general de fertilidad[2].
pe :	valor real en dólares de la exención en el pago de impuestos personales.
year :	tendencia temporal t=1,...,72 (de 1913 a 1984).
pill :	variable ficticia igual a 1 en el año de introducción de la píldora 1963.
ww2 :	variable ficticia igual a 1 si el año está entre 1941 y 1945 (2ª guerra mundial)

La cuestión principal del estudio es analizar si, en el agregado, la decisión de tener hijos está relacionada con el valor impositivo de tener un hijo. Inicialmente se especifica un modelo en el que para explicar la tasa de fertilidad en un año concreto solamente se considera como variable explicativa las exenciones fiscales en ese año.

$$\text{gfr}_t = \beta_1 + \beta_2 \text{pe}_t + u_t \tag{5.7}$$

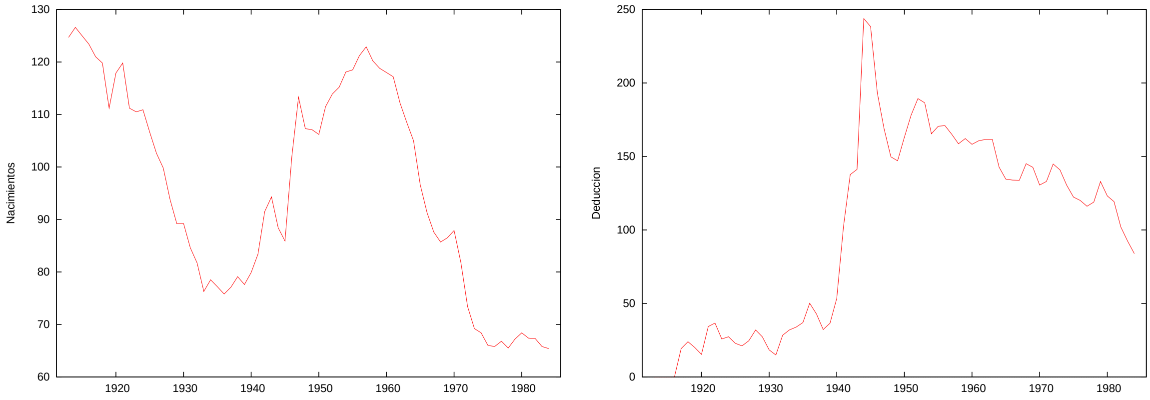
Por lo tanto, dado que las variables dependiente y explicativa están en el mismo periodo de tiempo, el efecto de una variación en la reducción fiscal sobre el número de nacimientos se supone que, de tener relevancia, es en el mismo año o de forma contemporánea. Si se supone que el término de perturbación no está correlacionado en el tiempo, el modelo de partida no es un modelo dinámico sino estático. Esta especificación de partida puede no ser la más adecuada, bien porque para analizar el efecto de las exenciones fiscales sobre la tasa de fertilidad se ha de controlar por otros factores que pueden afectar bien en el mismo periodo, de forma contemporánea, o bien en periodos consecutivos. Además, la propia tasa de fertilidad puede presentar cierta dependencia o correlación en el tiempo, tal que pueda ser relevante su propio pasado como variable explicativa. A su vez, pudiera ocurrir que hubiera factores no observables recogidos en el término de perturbación que presentaran correlación en el tiempo. De todo ello nos iremos ocupando de forma escalonada, tratando de encontrar una especificación adecuada.

Comencemos pues con esta primera especificación, para la cual consideramos el siguiente análisis:

Del estudio de los gráficos de la evolución temporal de ambas variables podemos concluir que: la serie temporal de la tasa de fertilidad presenta un ciclo con una caída continuada en el periodo de 1913 a 1936 y una recuperación progresiva a partir de esa fecha hasta 1957, año donde se alcanza casi la misma tasa que al comienzo de la muestra. A partir de 1957 vuelve a darse una caída en la tasa de fertilidad, no mostrando señales de recuperación en lo años posteriores hasta el final de la muestra.
Por otro lado, podemos apreciar que las exenciones fiscales experimentaron un crecimiento muy importante entre los años 1940-1945, período de la 2ª guerra mundial, para luego descender, primero de forma abrupta hasta el año 1950 y posteriormente, de forma progresiva, a partir de ese año hasta el final de la muestra.

[1] Fichero fertil3.gdt, disponible en gretl pestaña Wooldridge. Fuente: Wooldridge, J.M. (2001), *Introducción a la Econometría.*

Figura 5.1: *Gráficos de las series de fertilidad y de las exenciones fiscales*



- Analicemos los primeros resultados obtenidos de la estimación del modelo (5.7)

Estimaciones MCO utilizando las 72 observaciones 1913–1984
Variable dependiente: gfr

Variable	Coeficiente	Desv. típica	Estadístico t	valor p
const	96,344300	4,30473	22,3810	0,0000
pe	−0,007095	0,03592	−0,1975	0,8440

Media de la var. dependiente	95,63190
D.T. de la variable dependiente	19,80460
Suma de cuadrados de los residuos	27832,4
Desviación típica de los residuos ($\hat{\sigma}$)	19,94000
R^2	0,00055
$\bar{R}^2$ corregido	−0,01372
Grados de libertad	70
Estadístico de Durbin–Watson	0,04691
Coef. de autocorr. de primer orden.	0,97791

El valor del coeficiente de determinación $R^2 = 0,000556$ indica que el ajuste es muy pobre. Las exenciones fiscales no explican por si solas casi nada de la tasa de fertilidad en ese mismo año. Esto también se ve reflejado en la no significatividad del coeficiente de regresión, ya que el valor p asociado al valor muestral del estadístico t, 0,844, es mucho mayor que el nivel de significación $\alpha = 0,05$. El contraste de autocorrelación con el estadístico de Durbin-Watson indica que en el modelo (5.7) la perturbación está autocorrelada ya que $DW = 0,046 < 1,58 = d_L$, la cota inferior obtenida para $T = 72$ y $K' = 3$.

Los gráficos de residuos, Figura 5.2, y de la serie observada y ajustada por el modelo, Figura 5.3, indican que el ajuste ha dejado sin explicar toda la evolución temporal de la tasa de fertilidad, y simplemente ha recogido un nivel aproximadamente igual a la media de esta tasa en el periodo muestral. Todos estos elementos llaman a incluir otras variables explicativas en el modelo.

Figura 5.2: *Gráfico de residuos MCO*

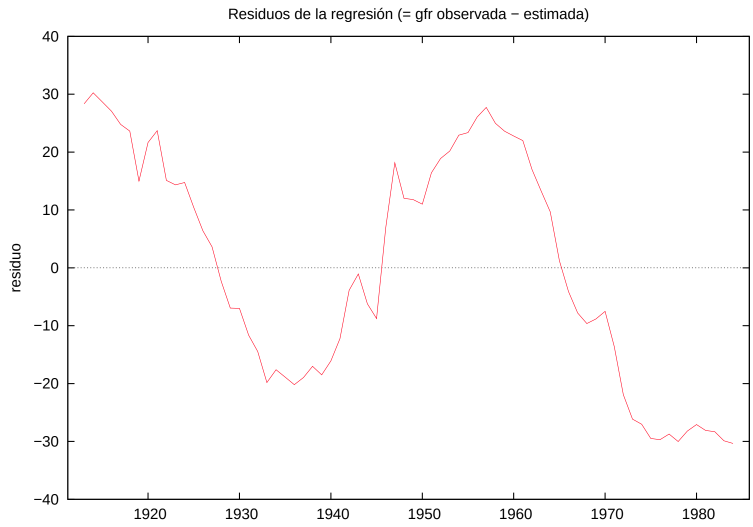
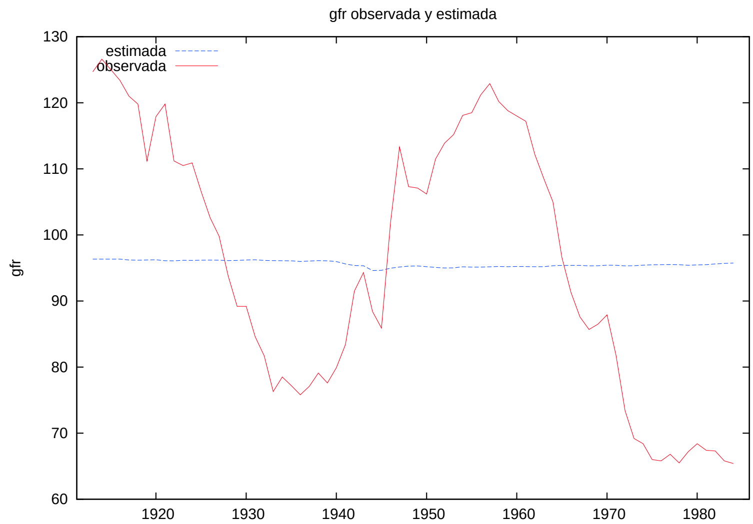


Figura 5.3: *Gráfico de la serie gfr observada y ajustada*



• Consideramos incluir dos variables ficticias, $ww2$ y $pill$, que tienen en cuenta un posible cambio en el valor medio de la tasa de fertilidad durante el periodo de la segunda guerra mundial y a partir de la introducción de la píldora como método anticonceptivo, respectivamente. La especificación sigue siendo la de un modelo estático, ya que no introducimos valores retardados de ninguna variable. El modelo a estimar entonces es el siguiente:

$$\text{gfr}_t = \beta_1 + \beta_2 \text{pe}_t + \beta_3 \text{pill}_t + \beta_4 \text{ww2}_t + u_t \tag{5.8}$$

y los resultados de la estimación se muestran a continuación:

Estimaciones MCO utilizando las 72 observaciones 1913–1984
Variable dependiente: gfr

Variable	Coeficiente	Desv. típica	Estadístico t	valor p
const	98,6818	3,20813	30,7599	0,0000
pe	0,0825	0,02964	2,7842	0,0069
pill	−31,5940	4,08107	−7,7416	0,0000
ww2	−24,2380	7,45825	−3,2499	0,0018

Media de la var. dependiente	95,6319
D.T. de la variable dependiente	19,8046
Suma de cuadrados de los residuos	14664,3
Desviación típica de los residuos ($\hat{\sigma}$)	14,6851
R^2	0,4734
$\bar{R}^2$ corregido	0,4501
$F(3, 68)$	20,3780
Estadístico de Durbin–Watson	0,1768
Coef. de autocorr. de primer orden.	0,8904

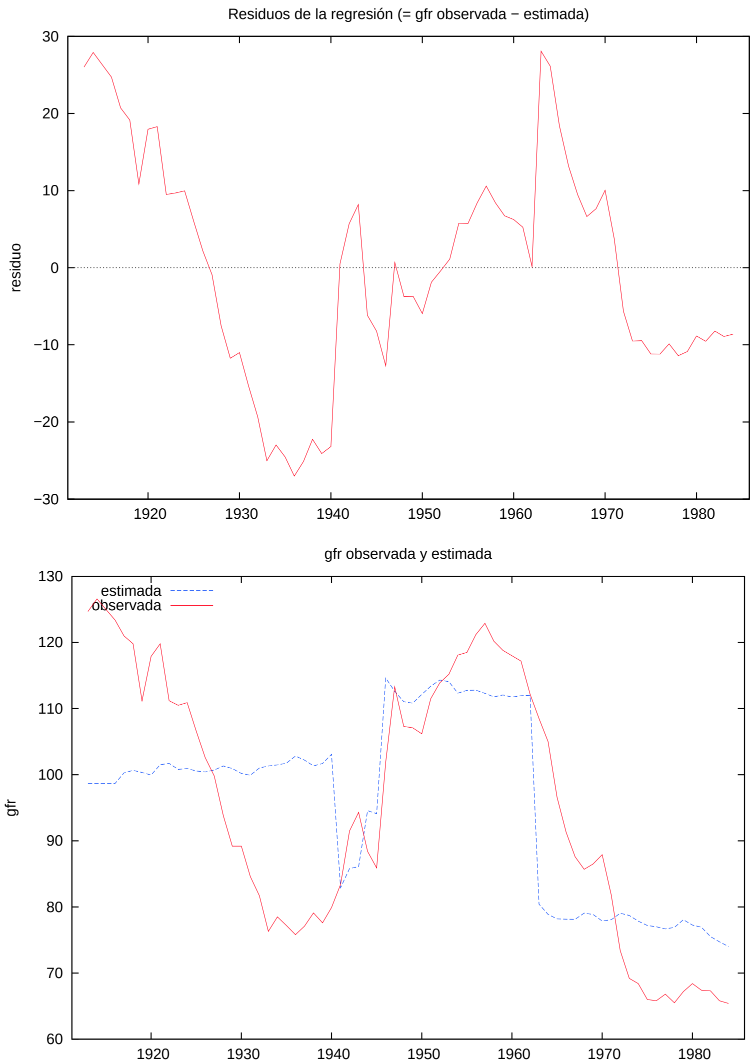
Las dos variables ficticias presentan coeficientes estimados negativos y son significativas individualmente, indicando que parece existir una caída significativa en media en la tasa de fertilidad, tanto durante el periodo de la 2ª guerra mundial, como después de la introducción de la píldora. A su vez, las exenciones fiscales presentan un coeficiente estimado positivo siendo su efecto sobre la tasa de fertilidad significativo al menos de forma contemporánea, ya que el valor p asociado a su coeficiente igual a $0,0069$ es menor que el nivel de significación tanto del $0,05$ como del $0,01$. En términos de bondad de ajuste se ha mejorado sustancialmente tanto en términos del coeficiente de determinación R^2 como del corregido. El ajuste explica un 47 % de la variabilidad de la tasa de fertilidad en la muestra.

Viendo los gráficos de la serie observada y ajustada junto con el de los residuos, Figura 5.4, se observa que el ajuste no es bueno, especialmente en el periodo inicial de 1913 a 1940. En este periodo la tasa de fertilidad observada disminuye progresivamente mientras que la ajustada se mantiene constante. Por este motivo, el gráfico de los residuos muestra una agrupación de residuos consecutivos del mismo signo que sugiere la posible existencia, también en esta segunda especificación del modelo, de autocorrelación al menos de orden uno positiva. El valor del estadístico de Durbin y Watson igual a $0,176$ es menor que el valor de la cota inferior $d_L = 1,52$ para $T = 72$ y $K' = 3$. Esto confirma la sospecha, ya que se rechaza la hipótesis nula de no autocorrelación frente a un proceso AR(1) con coeficiente positivo en la perturbación del modelo.

La cuestión está ahora en si la dinámica en la tasa de fertilidad es algo a recoger únicamente por el término de error o bien existe un problema de mala especificación tal que se necesita introducir retardos de alguna variable, bien pe, bien gfr o ambas.

Dado que hay autocorrelación en el error, para que la inferencia utilizando el estimador MCO de los coeficientes sea válida, tenemos que estimar de forma consistente la matriz de varianzas y covarianzas de los coeficientes estimados por MCO. Estos serían los resultados utilizando un estimador de la matriz de varianzas y covarianzas de los coeficientes estimados robusto a la existencia

Figura 5.4: *Gráfico de residuos MCO y de la serie gfr observada y ajustada*



de autocorrelación:

Estimaciones MCO utilizando las 72 observaciones 1913–1984
Variable dependiente: gfr
Desviaciones típicas robustas a autocorrelación

Variable	Coeficiente	Desv. típica	Estadístico t	valor p
const	98,6818	7,69679	12,8212	0,0000
pe	0,0825	0,04772	1,7294	0,0883
pill	−31,5940	5,43135	−5,8170	0,0000
ww2	−24,2380	3,55876	−6,8109	0,0000

De esta forma podemos hacer los contrastes de significatividad individual con los valores de los estadísticos t que nos muestra el output eligiendo el valor crítico en la distribución N(0,1). El valor crítico de la N(0,1) para un contraste a dos colas al nivel de significación del 5 % es 1,96 luego las variables *pill* y *ww*2 son individualmente significativas. En cambio las ayudas fiscales dejan de serlo al 5 %, aunque aún lo serían al 10 % de significación.

• Si el problema de autocorrelación detectado es debido a la existencia de un proceso AR(1) en la perturbación, $u_t = \rho u_{t-1} + \epsilon_t$, el modelo debería ser estimado por MCGF, bien Cochrane-Orcutt o Hildreth-Lu. Utilizamos el método de Cochrane-Orcutt y obtenemos los siguientes resultados:

Estimaciones por Cochrane–Orcutt, usando las observaciones 1914–1984 ($T = 71$)
Variable dependiente: gfr
$\hat{\rho} = 0{,}976137$

	Coeficiente	Desv. Típica	Estadístico t	Valor p
const	66,8660	21,5987	3,0958	0,0029
pe	−0,0159	0,0318	−0,5028	0,6168
pill	−3,4131	4,2129	−0,8101	0,4207
ww2	−5,4514	3,3389	−1,6327	0,1072

La estimación del coeficiente ρ es muy cercana a la unidad, $\hat{\rho} = 0{,}976137$, lo que indica una alta persistencia en la serie de los residuos, tardan en cambiar el signo. Los estadísticos de significatividad individual muestran que ninguna de las variables explicativas es significativa. Este resultado es bastante llamativo. Además el coeficiente que acompaña a la variable *pe* es de signo contrario al esperado. Parece que esa alta persistencia en la serie de los residuos puede estar afectando a estos resultados.

La evidencia existente de autocorrelación en el modelo anterior llama a considerar otras especificaciones alternativas donde se tenga en cuenta características en el tiempo de la serie a explicar, como pueden ser tendencias, dependencia en el tiempo de la propia serie o efectos distribuidos en el tiempo de la exención fiscal sobre la tasa de fertilidad.

• Especificaciones alternativas:

A. Modelización de tendencias añadiendo al modelo una tendencia temporal lineal, t, o cuadrática, t^2.

   En la Figura 5.1 se observa que la serie gfr o tasa de fertilidad presenta una caída continuada en todo el periodo muestral a excepción de los años 1936-1957. Viendo el gráfico de la serie gfr observada y ajustada, Figura 5.4, esta tendencia decreciente parece que no está bien recogida por la especificación (5.8). Esta puede ser una de las causas de detectar autocorrelación positiva en el error. Una forma de recoger esta tendencia es incluir como regresor la variable t, tiempo o tendencia temporal; luego $t = 1, 2, \ldots, 72$. Esta variable puede recoger parte de la influencia de factores no observables que están creciendo o decreciendo sistemáticamente en el tiempo.

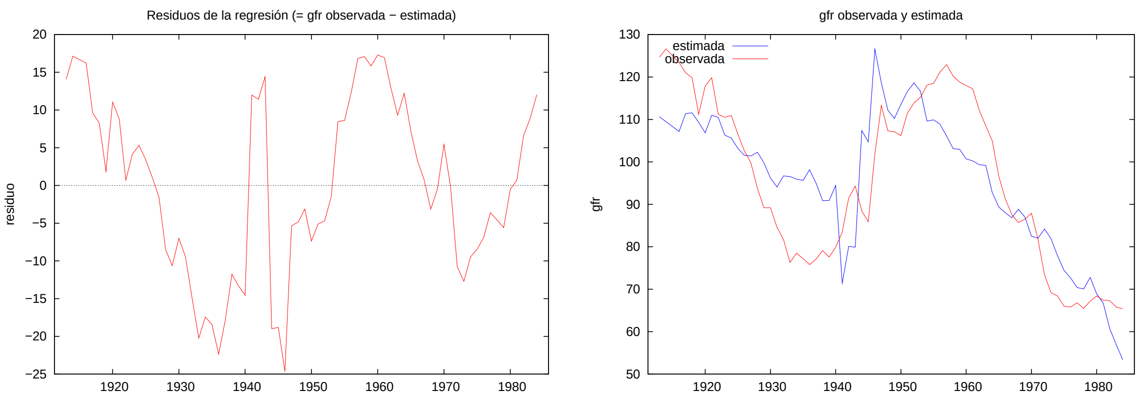
$$\mathrm{gfr}_t = \beta_1 + \beta_2 t + \beta_3 \mathrm{pe}_t + \beta_4 \mathrm{pill}_t + \beta_5 \mathrm{ww2}_t + u_t \tag{5.9}$$

Estimaciones MCO, usando las observaciones 1913–1984 ($T = 72$)
Variable dependiente: gfr

	Coeficiente	Desv. Típica	Estadístico t	Valor p
const	111,7690	3,35777	33,2868	0,0000
t	−1,1498	0,18790	−6,1195	0,0000
pe	0,2788	0,04001	6,9685	0,0000
pill	0,9974	6,26163	0,1593	0,8739
ww2	−35,5923	6,29738	−5,6519	0,0000

R^2	0,662213	R^2 corregido	0,642047
$F(4, 67)$	32,83745	Valor p (de F)	3,76e–15
$\hat{\rho}$	0,819265	Durbin–Watson	0,350212

Figura 5.5: *Gráficos de residuos y de la serie gfr observada y ajustada*



Vemos en los resultados de la estimación por MCO de esta especificación (5.9) que el ajuste ha mejorado, tanto en términos del coeficiente de determinación corregido como gráficamente, en el gráfico de la serie gfr observada y ajustada, Figura (5.5.) La tendencia t es una variable significativa y el coeficiente estimado es negativo, recogiendo que la tasa de fertilidad gfr está descendiendo, en media, a lo largo del período muestral, manteniendo el resto de variables fijo.

Por otro lado, el coeficiente de la variable *pe* se ha triplicado en relación al estimado en el modelo (5.8) y es significativo. En cambio la variable que recoge el efecto de la introducción de la píldora anticonceptiva *pill* no es significativa. De todas formas, hay que ser cautos con estos resultados de significatividad dado que el estadístico de Durbin-Watson toma un valor menor que el valor de la cota inferior obtenida para $T = 72$ y $K' = 4$ $DW = 0,350212 < d_L = 1,5029$ indicando evidencia de autocorrelación positiva en el término de perturbación de esta especificación. Dada la existencia de autocorrelación, los contrastes realizados en base a los resultados de la estimación MCO no son válidos para realizar inferencia.

Si tenemos en cuenta una estimación robusta a autocorrelación de las desviaciones típicas para realizar los contrastes de significatividad con las estimaciones por MCO de los coeficientes, las conclusiones no cambian. No mejoramos en cuanto a significatividad ni en cuanto a los signos de los coeficientes estimados.

Estimaciones MCO, usando las observaciones 1913–1984 ($T = 72$)
Variable dependiente: gfr
Desviaciones típicas robustas a autocorrelación

	Coeficiente	Desv. Típica	Estadístico t	Valor p
const	111,7690	5,8204	19,2027	0,0000
t	−1,1498	0,3226	−3,5643	0,0007
pe	0,2788	0,0713	3,9077	0,0002
pill	0,9974	10,7323	0,0929	0,9262
ww2	−35,5923	7,5391	−4,7210	0,0000

Así y todo, vemos que la tasa de fertilidad también presenta un periodo, de 1936 a 1947, de tendencia creciente, por lo que podemos considerar ver qué ocurre si incluimos también una tendencia cuadrática, $tsq \equiv t^2$. El modelo a estimar en este caso sería entonces:

$$\text{gfr}_t = \beta_1 + \beta_2 t + \beta_3 t^2 + \beta_4 \text{pe}_t + \beta_5 \text{pill}_t + \beta_6 \text{ww2}_t + u_t \tag{5.10}$$

y los resultados de la estimación son los siguientes:

Estimaciones MCO, usando las observaciones 1913–1984 ($T = 72$)
Variable dependiente: gfr

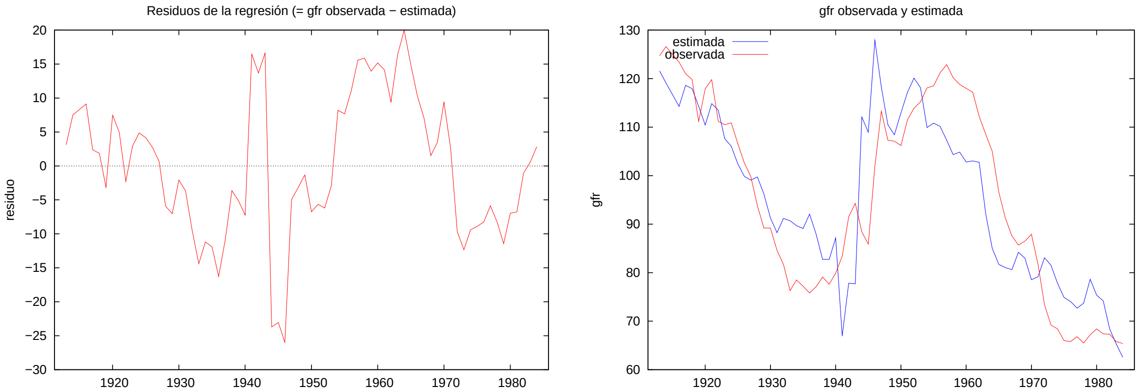
	Coeficiente	Desv. Típica	Estadístico t	Valor p
const	124,0920	4,36074	28,4566	0,0000
t	−2,5314	0,38938	−6,5011	0,0000
tsq	0,0196	0,00497	3,9454	0,0002
pe	0,3478	0,04025	8,6392	0,0000
ww2	−35,8803	5,70792	−6,2861	0,0000
pill	−10,1197	6,33600	−1,5972	0,1150

R^2	0,726676	R^2 corregido	0,705970
$F(5, 66)$	35,09434	Valor p (de F)	2,44e–17
$\hat{\rho}$	0,742433	Durbin–Watson	0,514369

Los dos coeficientes de tendencia son significativos y, como era de esperar, de signo contrario; el coeficiente que acompaña a t es negativo, indicando la tendencia decreciente en promedio de la tasa de fertilidad, y el que acompaña a t^2 es positiva, recogiendo el hecho de que hay periodos en los que esa tendencia es creciente. El coeficiente de *pe* es incluso mayor cuantitativamente que antes, y estadísticamente significativo. Ahora el coeficiente estimado que acompaña a la variable *pill* tiene el signo esperado negativo pero sigue sin ser significativo. De todas formas, sigue existiendo evidencia de autocorrelación positiva[3], ya que $DW = 0,514369$ es menor que la cota inferior $d_L = 1,4732$ para $T = 72$ y $K' = 5$.

[3]Una posibilidad es incluir más elementos de tendencia $t^3, t^4 \ldots$ tal que se llegaría a ajustar perfectamente la serie. Pero el interés en el análisis de regresión es capturar movimientos tendenciales generales de la serie a explicar con el objetivo de descubrir cuáles son las variables explicativas que afectan a gfr.

Figura 5.6: *Gráfico de residuos MCO y de la serie gfr observada y ajustada*



Teniendo en cuenta la autocorrelación a la hora de estimar la matriz de varianzas y covarianzas del estimador MCO, los resultados de significatividad no cambian sustancialmente.

Estimaciones MCO, usando las observaciones 1913–1984 ($T = 72$)
Variable dependiente: gfr
Desviaciones típicas robustas a autocorrelación

	Coeficiente	Desv. Típica	Estadístico t	Valor p
const	124,0920	4,88820	25,3860	0,0000
t	−2,5314	0,62030	−4,0809	0,0001
tsq	0,0196	0,00744	2,6330	0,0105
pe	0,3478	0,08289	4,1956	0,0001
pill	−10,1197	11,49250	−0,8805	0,3818
ww2	−35,8803	9,02103	−3,9774	0,0002

Las especificaciones anteriores consideran un modelo con autocorrelación en la perturbación que no parece ser adecuadamente recogida en media, incluso cuando incluimos una tendencia lineal y/o cuadrática. Por ello, parece sensato intentar una especificación del modelo donde incorporemos explícitamente dinámica en la parte sistemática.

B. Modelizaciones dinámicas: Incorporación de retardos de la variable endógena gfr o/y de la variable explicativa *pe*.

• Primeramente especificamos un modelo de retardos distribuidos incluyendo como regresores cuatro retardos consecutivos de la variable pe_t, es decir, añadimos pe_{t-1}, ..., pe_{t-4} a la lista de regresores del modelo 5.8. La especificación sería la siguiente:

$$\text{gfr}_t = \beta_1 + \beta_2 \text{pe}_t + \beta_3 \text{pe}_{t-1} + \beta_4 \text{pe}_{t-2} + \beta_5 \text{pe}_{t-3} + \beta_6 \text{pe}_{t-4} + \beta_7 \text{pill}_t + \beta_8 \text{ww2}_t + u_t \quad (5.11)$$

Los resultados de la estimación por MCO se muestran a continuación:

Estimaciones MCO, usando las observaciones 1917–1984 ($T = 68$)
Variable dependiente: gfr

	Coeficiente	Desv. Típica	Estadístico t	Valor p
const	92,5016	3,32548	27,8160	0,0000
pe	0,0887	0,12618	0,7033	0,4846
pe_1	−0,0039	0,15311	−0,0260	0,9794
pe_2	0,0073	0,16510	0,0448	0,9644
pe_3	0,0180	0,15358	0,1177	0,9067
pe_4	0,0139	0,10502	0,1327	0,8948
pill	−31,0816	3,89687	−7,9761	0,0000
ww2	−21,3435	11,54080	−1,8494	0,0693

R^2	0,536811	R^2 corregido	0,482772
$F(7, 60)$	9,933828	Valor p (de F)	3,63e–08
$\hat{\rho}$	0,862026	Durbin–Watson	0,215806

De partida, hemos perdido cuatro observaciones porque necesitamos ese número de condiciones iniciales para definir los retardos, luego nuestro tamaño de muestra disponible es ahora $T = 68$. En esta regresión solamente tenemos retardos de la variable exógena *pe* que consideramos regresores no estocásticos. Por lo tanto, si el resto de hipótesis básicas sobre el término de perturbación se satisfacen, podremos hacer inferencia válida en muestras finitas y el estimador MCO tendría buenas propiedades, tanto en muestras finitas como asintóticas.

La cuestión es que el valor del estadístico de Durbin-Watson indica con claridad que hay evidencia de autocorrelación positiva en el término de error, ya que $DW = 0,215806 < 1,3893 = d_L$ obtenida para $T = 68$ y $K' = 7$. Tenemos que contrastar la significatividad, individual y conjunta, a través de estadísticos válidos. Para ello, si seguimos utilizando el estimador de los coeficientes por MCO, debemos estimar consistentemente su matriz de varianzas y covarianzas. Los resultados en este caso son los siguientes:

Estimaciones MCO, usando las observaciones 1917–1984 ($T = 68$)
Variable dependiente: gfr
Desviaciones típicas robustas a autocorrelación

	Coeficiente	Desv. Típica	Estadístico t	Valor p
const	92,5016	7,35339	12,5794	0,0000
pe	0,0887	0,08820	1,0062	0,3184
pe_1	−0,0039	0,06700	−0,0594	0,9529
pe_2	0,0073	0,08428	0,0877	0,9304
pe_3	0,0180	0,06640	0,2723	0,7863
pe_4	0,0139	0,07249	0,1923	0,8482
pill	−31,0816	5,07610	−6,1231	0,0000
ww2	−21,3435	10,41060	−2,0502	0,0447

R^2	0,536811	R^2 corregido	0,482772
$F(7, 60)$	12,70587	Valor p (de F)	7,09e–10
$\hat{\rho}$	0,862026	Durbin–Watson	0,215806

Contraste de omisión de variables –
Hipótesis nula: los parámetros son cero para $pe_{t-1}, pe_{t-2}, pe_{t-3}, pe_{t-4}$
Estadístico de contraste asintótico:
valor muestral del estadístico $\chi^2(4)$ = 0,392538 con valor p = 0,982663

Contraste de omisión de variables –
Hipótesis nula: los parámetros son cero para $pe_t, pe_{t-1}, pe_{t-2}, pe_{t-3}, pe_{t-4}$
Estadístico de contraste asintótico: $\chi^2(5)$ = 12,1584
con valor p = 0,045022

Contraste de omisión de variables por eliminación secuencial –
(nivel de significación 0, 05 a dos colas)
Quitando pe_{t-1} (valor p 0,953) no se rechaza que ese coeficiente sea cero
Quitando pe_{t-2} (valor p 0,960) no se rechaza que ese coeficiente sea cero
Quitando pe_{t-4} (valor p 0,846) no se rechaza que ese coeficiente sea cero
Quitando pe_{t-3} (valor p 0,605) no se rechaza que ese coeficiente sea cero

¿Introduce esta especificación algún otro problema conocido? ¿Hay un problema de multicolinealidad entre pe_t y sus retardos $pe_{t-1}, \ldots, pe_{t-4}$?

Vemos que los coeficientes de las variables pe_t y sus retardos $pe_{t-1}, \cdots, pe_{t-4}$ no son individualmente significativos. Por otro lado, al realizar el contraste de omisión de variables bajo la hipótesis nula de que los parámetros son cero para $pe_t, pe_{t-1}, pe_{t-2}, pe_{t-3}, pe_{t-4}$ conjuntamente se rechaza esa hipótesis. Por lo tanto, puede existir un problema de multicolinealidad entre pe_t y sus retardos. Si miramos a los coeficientes de correlación entre estas variables, vemos que son muchos cercanos a uno, indicando esa alta correlación entre estos regresores.

Coeficientes de correlación, usando las observaciones 1913 - 1984
(se ignoraron los valores perdidos)
valor crítico al 5 % (a dos colas) = 0,2319 para n = 72

pe	pe_1	pe_2	pe_3	pe_4	
1,0000	0,9636	0,9090	0,8554	0,7996	pe
	1,0000	0,9637	0,9097	0,8579	pe_1
		1,0000	0,9639	0,9111	pe_2
			1,0000	0,9645	pe_3
				1,0000	pe_4

Por otro lado, al realizar tanto el contraste secuencial, como el contraste conjunto de que solamente los coeficientes que acompañan a $pe_{t-1}, pe_{t-2}, pe_{t-3}, pe_{t-4}$ son igual a cero, no se rechaza la hipótesis nula en cada caso. Por lo tanto se puede decir que el problema no es tanto de multicolinealidad entre los propios retardos sino entre pe_t y ellos.

Si comenzamos con un número de retardos menor, por ejemplo dos, tenemos los mismos resultados:

Estimaciones MCO, usando las observaciones 1915–1984 ($T = 70$)
Variable dependiente: gfr
Desviaciones típicas robustas a autocorrelación

	Coeficiente	Desv. Típica	Estadístico t	Valor p
const	95,8705	7,55497	12,6897	0,0000
pe	0,0726	0,07856	0,9250	0,3584
pe_1	−0,0057	0,07128	−0,0811	0,9356
pe_2	0,0338	0,07622	0,4438	0,6587
pill	−31,3050	5,29729	−5,9096	0,0000
ww2	−22,1265	7,58264	−2,9180	0,0049

R^2	0,498599	R^2 corregido	0,459427
$F(5, 64)$	15,66942	Valor p (de F)	4,75e–10
$\hat{\rho}$	0,875014	Durbin–Watson	0,188715

Contraste de omisión de variables –
Hipótesis nula: los parámetros son cero para las variables pe_t, pe_{t-1}, pe_{t-2}
Estadístico de contraste asintótico: valor muestral del estadístico $\chi^2(3) = 5{,}62982$ con valor p = 0,14243 no se rechaza la hipótesis nula

Contraste de omisión de variables –
Hipótesis nula: los parámetros son cero para las variables pe_{t-1}, pe_{t-2}
Estadístico de contraste: valor muestral del estadístico $F(2, 64) = 0{,}0995195$
con valor p = 0,905412 no se rechaza la hipótesis nula

Por ello, parece que incluir retardos de pe_t no mejora la especificación del modelo, sigue existiendo autocorrelación en el error, y la multicolinealidad entre pe_t y sus retardos hace que se estime de forma imprecisa el efecto contemporáneo de la exención de impuestos sobre la tasa de fertilidad.

Veamos otra alternativa de introducir dinámica en el modelo.

• Seguidamente especificamos un modelo dinámico incluyendo como regresores cuatro retardos consecutivos de la variable endógena gfr_t, es decir, añade $gfr_{t-1}, \ldots, gfr_{t-4}$ a la lista de regresores del modelo (5.8).

$$\mathrm{gfr}_t = \beta_1 + \beta_2 \mathrm{pe}_t + \beta_3 \mathrm{pill}_t + \beta_4 \mathrm{ww2}_t + \beta_5 gfr_{t-1} + \beta_6 gfr_{t-2} + \beta_7 gfr_{t-3} + \beta_8 gfr_{t-4} + u_t \quad (5.12)$$

En este caso lo regresores incluidos son estocásticos, ya que aunque determinadas en el pasado, o predeterminadas en t, han sido generadas por las propias perturbaciones pasadas del modelo, de ahí que sean variables aleatorias gfr_{t-1}, gfr_{t-2}, gfr_{t-3}, gfr_{t-4}. El estimador MCO es sesgado y no se conoce su distribución exacta o en muestras finitas pero si las perturbaciones del modelo son un ruido blanco, el estimador será consistente y los estadísticos t y F válidos asintóticamente.

Estimaciones MCO, usando las observaciones 1917–1984 (T = 68)
Variable dependiente: gfr

	Coeficiente	Desv. Típica	Estadístico t	Valor p
const	8,43045	3,15685	2,6705	0,0097
pe	0,02864	0,00967	2,9599	0,0044
pill	−5,34771	1,49830	−3,5692	0,0007
ww2	−2,52153	2,04522	−1,2329	0,2224
gfr_1	1,14486	0,12777	8,9601	0,0000
gfr_2	−0,58858	0,18519	−3,1782	0,0023
gfr_3	0,46819	0,18511	2,5292	0,0141
gfr_4	−0,13390	0,12220	−1,0958	0,2776

R^2	0,966703	R^2 corregido	0,962818
$F(7,60)$	248,8517	Valor p (de F)	7,40e–42
$\hat{\rho}$	−0,048412	Durbin–Watson	2,088677

Por el contrario, si se detecta autocorrelación en el término de error, entonces el estimador MCO no será consistente y la inferencia tampoco será válida. Por esa razón es importante contrastar si hay o no evidencia de autocorrelación. Para ello, el estadístico de Durbin-Watson no es fiable, dado que como regresores tenemos retardos de la variable endógena. Utilizaremos el contraste de Breusch-Godfrey. Contrastamos la hipótesis nula de que el término de perturbación es un ruido blanco, frente a la alternativa de que sigue un proceso AR(1) o MA(1).

El resultado del contraste se muestra a continuación:

Contraste Breusch-Godfrey de autocorrelación de primer orden MCO, usando las observaciones 1917-1984 (T = 68)
Variable dependiente: uhat

	Coeficiente	Desv. Típica	Estad. t	Valor p
const	-3,78284	4,20750	-0,8991	0,3723
pe	-0,01652	0,01557	-1,0610	0,2930
$pill$	2,83789	2,57773	1,1010	0,2754
$ww2$	1,53037	2,32700	0,6577	0,5133
gfr_1	0,58349	0,45098	1,2940	0,2008
gfr_2	-0,65901	0,52223	-1,2620	0,2119
gfr_3	0,29550	0,28607	1,0330	0,3058
gfr_4	-0,16737	0,17361	-0,9641	0,3390
$uhat_1$	-0,61992	0,45977	-1,3480	0,1827

El coeficiente de determinación de esta regresión auxiliar es $R^2 = 0,029892$, y el valor muestral del estadístico: $TR^2 = 2,032657$ con un valor p de 0,154. Dado el valor p asociado al valor muestral del estadístico del contraste de Breusch-Godfrey no se rechaza la hipótesis nula de que el término de perturbación sea un ruido blanco frente a la alternativa de que siga un AR(1) o MA(1). Por lo tanto, no parece haber evidencia de autocorrelación hasta de orden uno.

Podemos considerar como fiables los resultados para hacer inferencia, al menos válida asintóticamente, sobre la significatividad de las variables incluidas. Todas las variables son individualmente significativas a excepción de la variable ww_2 que recoge el periodo de la segunda guerra mundial y el cuarto retardo de la variable endógena gfr_{t-4}.

• Consideramos incluir todos los retardos de pe y gfr de las especificaciones anteriores en el modelo (5.8). El modelo más general sería:

$$\begin{aligned} \mathrm{gfr}_t = \beta_1 + \beta_2 \mathrm{pe}_t + \beta_3 \mathrm{pe}_{t-1} + \beta_4 \mathrm{pe}_{t-2} + \beta_5 \mathrm{pe}_{t-3} + \beta_6 \mathrm{pe}_{t-4} + \beta_7 \mathrm{pill}_t + \\ \beta_8 \mathrm{ww2}_t + \beta_9 gfr_{t-1} + \beta_{10} gfr_{t-2} + \beta_{11} gfr_{t-3} + \beta_{12} gfr_{t-4} + u_t \end{aligned} \tag{5.13}$$

Los resultados de la estimación de este modelo y del contraste de Breusch-Godfrey para contrastar si el término de perturbación es un ruido blanco frente a que siga un AR(1) o MA(1) se muestran a continuación:

Estimaciones MCO, usando las observaciones 1917–1984 ($T = 68$)
Variable dependiente: gfr

	Coeficiente	Desv. Típica	Estadístico t	Valor p
const	7,21746	2,96855	2,4313	0,0183
pe	−0,03757	0,03196	−1,1754	0,2448
pe_1	0,03203	0,03940	0,8128	0,4198
pe_2	0,09221	0,04277	2,1556	0,0354
pe_3	−0,06771	0,04170	−1,6237	0,1101
pe_4	0,00404	0,02925	0,1382	0,8906
pill	−4,70508	1,51670	−3,1022	0,0030
ww2	0,28547	3,22531	0,0885	0,9298
gfr_1	1,12341	0,13766	8,1604	0,0000
gfr_2	−0,44953	0,18893	−2,3793	0,0208
gfr_3	0,33774	0,17747	1,9030	0,0622
gfr_4	−0,10568	0,12009	−0,8800	0,3826

R^2	0,972925	R^2 corregido	0,967606
$F(11, 56)$	182,9370	Valor p (de F)	1,08e–39
$\hat{\rho}$	−0,042547	Durbin–Watson	2,068701

El coeficiente de determinación de esta regresión auxiliar es $R^2 = 0,028959$, y el valor muestral del estadístico: $TR^2 = 1,969205$ con un valor p de 0,161. Dado el valor p asociado al valor muestral del estadístico del contraste de Breusch-Godfrey no se rechaza la hipótesis nula de que el término de perturbación sea un ruido blanco frente a la alternativa de que siga un AR(1) o MA(1). Por lo tanto, no parece haber evidencia de autocorrelación hasta de orden uno. Podemos considerar como fiables los resultados para hacer inferencia al menos válida asintóticamente sobre la significatividad de las variables incluidas.

Contraste Breusch-Godfrey de autocorrelación de primer orden MCO, usando las observaciones 1917-1984 (T = 68) Variable dependiente: uhat

	Coeficiente	Desv. Típica	Estadístico t	Valor p
const	-3,711730	4,136670	-0,8973	0,3735
pe	-0,007805	0,032362	-0,2412	0,8103
pe_1	0,030136	0,045708	0,6593	0,5124
pe_2	-0,024036	0,046493	-0,5170	0,6072
pe_3	-0,059447	0,062240	-0,9551	0,3437
pe_4	0,045326	0,045809	0,9895	0,3268
$pill$	2,892880	2,715980	1,0650	0,2915
$ww2$	0,238492	3,212430	0,0742	0,9411
gfr_1	0,662843	0,535351	1,2380	0,2209
gfr_2	-0,735456	0,604202	-1,2170	0,2287
gfr_3	0,248948	0,262536	0,9482	0,3472
gfr_4	-0,124160	0,153815	-0,8072	0,4230
$uhat_1$	-0,685005	0,534860	-1,2810	0,2057

- A continuación omitimos de forma secuencial todas las variables que encontramos no significativas al nivel de significación del 5 %, incluyendo variables retardadas y no retardadas.

  Eliminación secuencial utilizando alfa= 0,05 a dos colas

Quitando ww2	(valor p 0,930)	no se rechaza que el coeficiente sea cero
Quitando pe_{t-4}	(valor p 0,907)	no se rechaza que el coeficiente sea cero
Quitando pe_{t-1}	(valor p 0,403)	no se rechaza que el coeficiente sea cero
Quitando gfr_{t-4}	(valor p 0,327)	no se rechaza que el coeficiente sea cero
Quitando pe_t	(valor p 0,270)	no se rechaza que el coeficiente sea cero

  Hipótesis nula: los parámetros de regresión asociados a $pe_t, pe_{t-1}, pe_{t-4}, ww2, gfr_{t-4}$ son cero.

  Estadístico de contraste: F(5, 56) = 0,567449, con valor p = 0,724516 no se rechaza.

  Por lo tanto, la especificación final sería:

$$\mathrm{gfr}_t = \beta_1 + \beta_2 \mathrm{pe}_{t-2} + \beta_3 \mathrm{pe}_{t-3} + \beta_4 \mathrm{pill}_t + \beta_5 gfr_{t-1} + \beta_6 gfr_{t-2} + \beta_7 gfr_{t-3} + u_t \tag{5.14}$$

Estimaciones MCO, usando las observaciones 1917–1984 ($T = 68$)
Variable dependiente: gfr

	Coeficiente	Desv. Típica	Estadístico t	Valor p
const	6,77128	2,79399	2,4235	0,0184
pe_2	0,09150	0,02395	3,8199	0,0003
pe_3	−0,06207	0,02510	−2,4726	0,0162
pill	−5,13607	1,38690	−3,7033	0,0005
gfr_1	1,06757	0,11527	9,2609	0,0000
gfr_2	−0,40369	0,16115	−2,5051	0,0149
gfr_3	0,23931	0,10765	2,2230	0,0299

R^2	0,971553	R^2 corregido	0,968755
$F(6,61)$	347,2224	Valor p (de F)	3,41e–45
$\hat{\rho}$	0,015698	h de Durbin	0,388041

Contraste Breusch-Godfrey de autocorrelación de primer orden MCO, usando las observaciones 1916-1984 (T = 69)
Variable dependiente: uhat

	Coeficiente	Desv. Típica	Estad. t	Valor p
const	0,398790	3,1035	0,1285	0,8982
pe_2	0,000711	0,02436	0,02918	0,9768
pe_3	0,000818	0,02542	0,03219	0,9744
$pill$	-0,262979	1,65241	-0,1591	0,8741
gfr_1	-0,060770	0,240580	-0,2526	0,8014
gfr_2	0,064050	0,275406	0,2326	0,8169
gfr_3	-0,008623	0,111933	-0,07704	0,9388
$uhat_1$	0,077286	0,268101	0,2883	0,7741

El coeficiente de determinación de esta regresión auxiliar es $R^2 = 0,001360$, y el valor muestral del estadístico: $TR^2 = 0,093872$ con un valor p de 0,759. No hay evidencia de autocorrelación hasta de orden uno.

Esta especificación dinámica parece adecuada, las variables son individual y conjuntamente significativas y no hay evidencia de que el término de perturbación no sea un ruido blanco.

Finalmente podemos decir que la existencia de autocorrelación en las especificaciones anteriores se debía a una mala especificación del modelo y que incluyendo dinámica en la parte sistemática mediante las variables gfr_{t-1}, gfr_{t-2}, gfr_{t-3}, pe_{t-2}, pe_{t-3} la hemos controlado.

5.4. Ejercicios a resolver

Además de los ejercicios que se resuelven en clase tenéis a vuestra disposición una colección de ejercicios de examen que se actualiza cada año. No deberíais dar por acabado el trabajo de cada tema hasta que están hechos los ejercicios recomendados en clase.

Ejercicio M-MD.1:
Se quiere estudiar la dependencia de la bolsa de Madrid de las bolsas de Japón y París. Para ello se dispone de 30 observaciones diarias de los índices IBEX, NIKKEI y CAC 40. Suponiendo que la bolsa madrileña es un seguidor de estas bolsas se propone el siguiente modelo:

$$IBEX_t = \beta_1 + \beta_2 NIKKEI_{t-1} + \beta_3 CAC_{t-1} + u_t \qquad \text{con } t: 2, \ldots, 50.$$

Su estimación por MCO proporciona el siguiente resultado:

$$\begin{array}{lcllll} \widehat{IBEX}_t & = & 0,0096 & + \ 0,500 \ NIKKEI_{t-1} & - \ 0,1799 \ CAC_{t-1} & \\ (\widehat{des}(\hat{\beta}_{MCO})) & & (0,0022) & \ \ (0,1199) & \ \ (0,1899) & \end{array} \tag{5.15}$$

$$R^2 = 0,68 \quad DW = 0,82$$

1. Contrasta la significatividad individual de las variables explicativas; ¿es la bolsa madrileña seguidor de la bolsa de Japón? ¿y de la de Paris?
2. ¿Es el modelo dinámico?
3. Contrasta la existencia de autocorrelación de tipo AR(1) en las perturbaciones. Especifica claramente la hipótesis nula, la alternativa, el estadístico de contraste y la regla de decisión.
4. Comenta la siguiente afirmación: “Al existir como regresores variables retardadas y autocorrelación, el estimador MCO es inconsistente”.

Posteriormente se añade la variable explicativa $IBEX_{t-1}$ al modelo, ya que se considera relevante el valor del índice al cierre del día anterior en la evolución del mismo al día siguiente. Se estima por MCO, con la misma muestra, obteniendo:

$$\underset{(\widehat{des}(\hat{\beta}_{MCO}))}{\widehat{IBEX}_t} = \underset{(0{,}0011)}{0,003} + \underset{(0{,}0800)}{0,1910}\ IBEX_{t-1} + \underset{(0{,}2460)}{0,8400}\ NIKKEI_{t-1} - \underset{(0{,}0120)}{0,0600}\ CAC_{t-1}$$

$$DW = 1,9 \tag{5.16}$$

$$\underset{(\widehat{des})}{\widehat{v}_t} = \underset{(0{,}002)}{0,0001} + \underset{(0{,}009)}{0,03}\ \hat{v}_{t-1} + \underset{(0{,}3)}{0,009} IBEX_{t-1} + \underset{(0{,}1)}{0,04}\ NIKKEI_{t-1} - \underset{(0{,}03)}{0,006}\ CAC_{t-1}$$

$$R^2 = 0,09$$

5. Contrasta la hipótesis de existencia de autocorrelación de tipo AR(1) en v_t.
6. A la vista de los resultados del contraste anterior razona las propiedades del estimador MCO.
7. Explica un método de estimación alternativo a MCO para el modelo (5.16) que tenga mejores propiedades.

Ejercicio M-MD.2:

Una empresa considera que sus beneficios anuales (Y) dependen de sus gastos en publicidad (X) y de sus beneficios del año anterior. Para saber cuál ha sido la relación entre estas variables durante los últimos 5 años, decide estimar el modelo:

$$Y_t = \beta_1 + \beta_2 X_t + \beta_3 Y_{t-1} + u_t, \qquad t = 1, 2, \ldots, 5 \tag{5.17}$$

con los siguientes datos:

t	1	2	3	4	5
Y_t	-10	16	2	4	-4
X_t	4	16	6	6	0

1. Realiza la estimación de los coeficientes de (5.17) por MCO.
2. ¿Es este estimador lineal? ¿Es insesgado? ¿Por qué?
3. Realiza la estimación de los coeficientes de (5.17) por el método de variables instrumentales, utilizando X_{t-1} como instrumento para Y_{t-1}. Escribe cuáles son las matrices hasta llegar al resultado final.
4. ¿Es este estimador lineal? ¿Es insesgado? ¿Por qué?
5. ¿Te parece X_{t-1} un instrumento adecuado para Y_{t-1}? ¿Por qué?
6. Si se considera como instrumento adicional a X_{t-2}:
   i) Explica el procedimiento de MC2E para estimar los coeficientes del modelo.
   ii) Completa:

$$Z = \left[\quad\quad\quad\quad \right] \qquad \hat{\beta}_{MC2E} = \left[\quad\quad\quad\quad\quad\quad \right]^{-1} \left[\quad\quad\quad \right]$$

$$\hat{Y}_{t-1} = \ldots\ldots\ldots\ldots\ldots\ldots\ldots \qquad t = \ldots\ldots$$

Ejercicio M-MD.3:
Sea el modelo

$$Y_t = \beta_1 + \beta_2 X_t + \beta_3 X_{t-1} + \beta_4 Y_{t-1} + u_t \tag{5.18}$$

con X fija y $u_t = \rho u_{t-1} + \varepsilon_t$, donde $\varepsilon_t \sim iid(0, \sigma^2_\varepsilon)$.

1. ¿Es el estimador MCO consistente? Explica razonadamente por qué.
2. ¿Cambiaría tu respuesta si en lugar de Y_{t-1} tuviésemos Y_{t-2} como variable explicativa en el modelo? ¿Por qué?
3. En el modelo (5.18), propón un estimador del vector de parámetros β que por lo menos sea **consistente**. Razona la respuesta.

5.5. Prácticas de Aula

Las prácticas de aula son clases participativas, ello quiere decir que el alumno debe acudir a clase con el ejercicio realizado en su tiempo de trabajo personal. Previamente y con anticipación suficiente se le habrá realizado la debida propuesta. Durante la clase se preguntará aleatoriamente a los alumnos sobre lo realizado.

En el tema de modelos dinámicos es habitual disponer de una práctica de aula, lo que equivale a una hora de clase presencial y dos horas de trabajo personal. Si el ejercicio ha sido realizado previamente por el alumno el tiempo es suficiente para su corrección y solución de dudas existentes. A continuación se va a proponer un enunciado correspondiente a las práctica citada.

Competencias a trabajar en estas sesiones.

1. Comprender la importancia de los supuestos empleados en la especificación de un modelo econométrico básico para poder proponer y emplear supuestos más realistas.
2. Diferenciar distintos métodos de estimación y evaluar su uso de acuerdo a las características de las variables económicas de interés para obtener resultados fiables.
3. Utilizar diversas fuentes estadísticas y adquirir destreza en el uso de un software econométrico para analizar relaciones entre variables económicas.

Práctica de Aula PA-MD.1:
Un investigador dispone de una base de datos anuales, para el período de 1948 a 1993, de los siguientes índices agrarios de EEUU, todos ellos con base 1982[4]:

output : producción agrícola (Rango 51 - 116).

labor : mano de obra agrícola (Rango 81 - 278).

land : superficie utilizada en la producción agrícola (Rango 89 - 102)

machines : maquinaria (duradera) (Rango 38 - 102)

El objetivo del investigador es determinar la función de producción agraria, para ello especifica el siguiente modelo de regresión lineal:

$$output_t = \beta_1 + \beta_2 labor_t + \beta_3 land_t + \beta_4 machines_t + u_t \qquad t = 1, \ldots, T \tag{5.19}$$

en el que se considera que los regresores son no estocásticos. Los resultados obtenidos de la estimación MCO son los que se muestran a continuación:

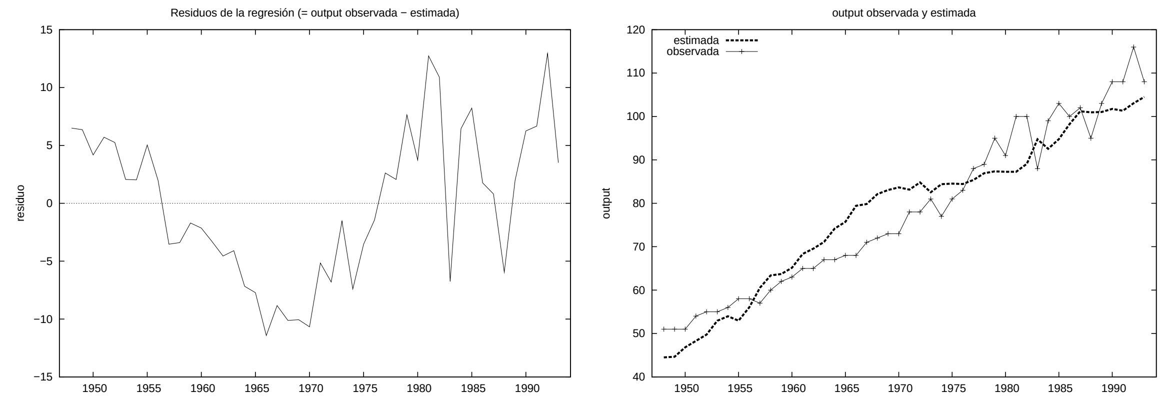
$$\begin{array}{lcccc}
\widehat{output}_t = & 181,201 & -\ 0,307\ labor_t & -\ 0,517\ land_t & -\ 0,096\ machines_t \\
(\widehat{des}(\hat{\beta}_{MCO})) & (40{,}194) & (0{,}038) & (0{,}564) & (0{,}169)
\end{array}$$

$$R^2 = 0,884 \qquad DW = 0,612 \qquad SCR = 1885,08 \qquad T = 46$$

1. Explica cómo se han calculado los residuos y para qué sirven los gráficos. Interpreta ambos gráficos y señala si existe alguna evidencia de que la perturbación del modelo no cumpla alguna de las hipótesis básicas, justificando tu respuesta.

[4] Fichero data9-5.gdt disponible en gretl pestaña Ramanathan. Fuente: Ramanathan, Ramu (2002): *Introductory Econometrics with Applications*.

Figura 5.7: *Gráfico de residuos y de la serie gfr observada y ajustada*



2. Realiza algún contraste basándote en la información disponible, para cualquier problema detectado en el apartado anterior. Explica detalladamente todos los elementos que intervengan.

3. Con respecto a los contrastes de significatividad individual de las variables explicativas del modelo (5.19):

   a) ¿Es fiable realizarlos utilizando la información disponible? ¿Por qué?

   b) ¿Sería posible llevarlos a cabo si no tuviésemos otra opción que la de estimar los coeficientes del modelo por MCO? Explica cómo lo harías en caso afirmativo.

Viendo los resultados obtenidos el investigador decide estimar el mismo modelo por el método de Cochrane-Orcutt (CO). A continuación se muestran los resultados obtenidos:

Estimaciones Cochrane–Orcutt utilizando las 45 observaciones 1949–1993
Variable dependiente: output
$\hat{\rho} = 0.791585$

	Coeficiente	Desv. Típica	Estadístico t	Valor p
const	54,3902	30,3065	1,7947	0,0801
labor	−0,4046	0,0649	−6,2284	0,0000
land	1,0727	0,3741	2,8673	0,0065
machines	−0,2874	0,2000	−1,4372	0,1583

Estadísticos basados en los datos rho-diferenciados:

R^2	0.957590	R^2 corregido	0.954487
$F(3,41)$	12.99460	Valor p (de F)	4.18e–06
$\hat{\rho}$	−0.184791	Durbin–Watson	2.339505

$$\widehat{V}(\hat{\beta}_{CO}) = \begin{bmatrix} 918,486 & 0,175515 & -9,59451 & -0,06293 \\ 0,175515 & 0,00422165 & -0,00963229 & 0,00270274 \\ -9,59451 & -0,00963229 & 0,139972 & -0,034905 \\ -0,06293 & 0,00270274 & -0,034905 & 0,0400013 \end{bmatrix}$$

4. ¿Cuándo estás dispuesto a aplicar este método de estimación? En particular, ¿consideras adecuado utilizar este método en las circunstancias actuales? Responde razonadamente.
5. Describe detalladamente cómo obtener las estimaciones de los coeficientes del modelo (5.19) utilizando el método del apartado anterior.
6. Con la información disponible, realiza el siguiente contraste $H_0 : \beta_2 = \beta_3$. Escribe la hipótesis nula, la alternativa, el estadístico de contraste junto con su distribución y realiza el contraste. ¿Cómo interpretas el resultado?

A continuación el investigador introduce un retardo de la variable endógena como variable explicativa en el modelo inicial, con la pretensión de recoger la influencia de la producción agrícola del año anterior:

$$output_t = \beta_1 + \beta_2 labor_t + \beta_3 land_t + \beta_4 machines_t + \beta_5 output_{t-1} + v_t \qquad t = 2, \ldots, T \tag{5.20}$$

Estimado el modelo por MCO se obtienen los siguientes resultados:

Estimaciones MCO utilizando las 45 observaciones 1949–1993
Variable dependiente: output

	Coeficiente	Desv. Típica	Estadístico t	Valor p
const	−26,6869	33,4166	−0,7986	0,4292
labor	−0,0868	0,0356	−2,4339	0,0195
land	0,6694	0,3649	1,8342	0,0741
machines	−0,1715	0,1013	−1,6929	0,0983
output_1	0,8535	0,0979	8,7126	0,0000

R^2	0.959224	R^2 corregido	0.955146
$F(4, 40)$	235.2394	Valor p (de F)	3.26e–27
$\hat{\rho}$	−0.299970	h de Durbin	−2.617899
$BG(1)$	6,199	Valor p (de $BG(1)$)	0.0128

7. Utilizando esta información, explica detalladamente la validez del siguiente estadístico de contraste
$$\frac{\hat{\beta}_{5,MCO} - 0}{\widehat{desv}(\hat{\beta}_{5,MCO})} \xrightarrow{H_0,d} N(0,1)$$
para argumentar a favor de incluir en el modelo el retardo de la variable endógena como variable explicativa.

Alternativamente se ha obtenido la siguiente estimación consistente y asintóticamente eficiente de los coeficientes del modelo (5.20):

$$\begin{aligned}
\underbrace{output_t - \hat{\rho}\; output_{t-1}}_{Q_t^*} &= -27,47\underbrace{(1-\hat{\rho})}_{X_t^*} - \underset{(0{,}028)}{0,058}\;\underbrace{(labor_t - \hat{\rho}\; labor_{t-1})}_{LB_t^*} \\
&+ \underset{(0{,}304)}{0,546}\;\underbrace{(land_t - \hat{\rho}\; land_{t-1})}_{LN_t^*} - \underset{(0{,}081)}{0,130}\;\underbrace{(machines_t - \hat{\rho}\; machines_{t-1})}_{MA_t^*} \\
&+ \underset{(0{,}076)}{0,925}\;\underbrace{(output_{t-1} - \hat{\rho}\; output_{t-2})}_{Q_{t-1}^*} + \hat{\epsilon}_t \qquad (5.21)\\
&R^2 = 0,976 \qquad DW = 2,30
\end{aligned}$$

siendo ϵ_t es un ruido blanco tal que $\epsilon_t = v_t - \rho v_{t-1}$ y v_t son las perturbaciones del modelo (5.20).

8. Completa y/o realiza lo siguiente:

   a) $\epsilon_t \sim \ldots\ldots(\quad , \quad)$

   b) ¿Cuál es el método de estimación que se ha utilizado?

   c) ¿Cuál es el estimador consistente de ρ empleado? Describe todos los elementos y las condiciones que te garantizan la consistencia del parámetro ρ estimado.

   d) Escribe la expresión matricial del estimador utilizado:

$$\begin{bmatrix} -27,47 \\ -0,058 \\ 0,546 \\ -0,130 \\ 0,925 \end{bmatrix} = \begin{bmatrix} \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \end{bmatrix}^{-1} \begin{bmatrix} \dots \\ \dots \\ \dots \\ \dots \\ \dots \end{bmatrix}$$

9. Utilizando la información contenida en la ecuación (5.21), explica detalladamente la validez del siguiente estadístico de contraste

$$\frac{\hat{\beta}_5 - 0}{\widehat{des}(\hat{\beta}_5)} \xrightarrow{H_0,d} N(0,1)$$

   para argumentar a favor de incluir en el modelo el retardo de la variable endógena como variable explicativa. Finalmente, ¿incluirías dicha variable en el modelo?

10. ¿Cuál es el estimador óptimo, dada toda la información de que dispones, para los parámetros de la ecuación (5.20)? Razona tu respuesta detalladamente en relación a todas las alternativas posibles.

5.6. Prácticas de Ordenador

En el tema de modelos dinámicos es habitual disponer de dos prácticas de ordenador que equivalen a dos horas presenciales y cuatro horas de trabajo no presencial para repasar lo visto y repetir la mecánica. A continuación se va a proponer un ejercicio para resolver en el centro de cálculo. El enunciado sigue el ejemplo empírico de modelización dinámica que se ha presentado anteriormente en las clases magistrales. Cubre todo lo aprendido en la asignatura sobre modelización en especial utilizando datos de series temporales.

Las competencias a trabajar en estas prácticas de ordenador son:

2. Diferenciar distintos métodos de estimación y evaluar su uso de acuerdo a las características de las variables económicas de interés para obtener resultados fiables.
3. Utilizar diversas fuentes estadísticas y adquirir destreza en el uso de un software econométrico para analizar relaciones entre variables económicas.

Recordatorio: Es conveniente que una vez acabado el ejercicio y en vuestro tiempo de trabajo personal relacionéis lo aprendido en la práctica de ordenador con el contenido del ejemplo empírico de este tema, ya que se van explicando los resultados de esta práctica. Es decir, en clase únicamente os da tiempo a aprender la mecánica, la ejecución sucesiva de instrucciones para obtener los resultados pedidos, aunque el profesor va comentando y explicando los resultados. Debéis entender convenientemente los resultados obtenidos, y a la vez vosotros mismos experimentar con otras posibles especificaciones dinámicas.

Práctica de ordenador PO-MD.1

En un estudio sobre la política de natalidad del gobierno de E.E.U.U. en el siglo XX se tienen datos anuales sobre las siguientes variables[5], para el período 1913-1984:

gfr:	tasa general de fertilidad[6].
pe :	valor real en dólares de la exención en el pago de impuestos personales.
year :	tendencia temporal t=1,...,72 (de 1913 a 1984).
pill :	variable ficticia igual a 1 en el año de introducción de la píldora 1963.
ww2 :	variable ficticia igual a 1 si el año está entre 1941 y 1945 (2ª guerra mundial)

Inicialmente se especifica el modelo

$$\text{gfr}_t = \beta_1 + \beta_2 \text{pe}_t + u_t \tag{5.22}$$

1. Da estructura de series temporales al conjunto de datos disponible pulsando en la pantalla principal en
Datos → Estructura del conjunto de datos → ...

[5]Fichero data 9-5.gdt disponible en gretl pestaña Wooldridge. Fuente: Wooldridge, J.M. (2001), *Introducción a la Econometría.*

2. Estima por MCO el modelo propuesto en (5.22). Interpreta los resultados. (A resolver en casa)

3. Obtén el gráfico de series temporales de la variable gfr_t y de los residuos. Coméntalos a la vista del R^2 obtenido en el apartado anterior. (A resolver en casa)

4. Reestima el modelo incluyendo los regresores $pill_t$ y $ww2_t$, ¿Qué pretenden recoger cada uno de ellos? ¿Tiene su inclusión algún efecto sobre los gráficos anteriores?

5. Contrasta la existencia de autocorrelación de primer orden con el estadístico de Durbin Watson.

6. Teniendo en cuenta toda la información disponible hasta ahora, contrasta la significatividad individual de la variable *pe*.

7. Estima el modelo por el método de Cochrane-Orcutt. Comenta los resultados obtenidos, realizando los contrastes que fuesen necesarios.

8. Añade como regresor la variable retardada un periodo gfr_{t-1} al modelo y estímalo por MCO. Obtén el gráfico de los residuos y coméntalo. Realiza un contraste de autocorrelación de orden 1 y, a partir de ese resultado, comenta los resultados del análisis. ¿Crees necesario utilizar algún otro estimador? ¿Por qué?

9. Añade al modelo una tendencia temporal t. Dado el gráfico de residuos que se obtiene, prueba a incluir también una tendencia cuadrática, t^2. (A resolver en casa)

10. Realiza el contraste de Durbin-Watson en este modelo. Comenta los resultados. (A resolver en casa)

11. ¿Cómo contrastarías la significatividad individual de las variables explicativas incluidas en el modelo? Utiliza siempre un estimador adecuado para los errores estándar de los coeficientes con la información de la que dispones hasta ahora.

Práctica de ordenador PO-MD.2

En un estudio sobre la política de natalidad en los E.E.U.U. en el siglo XX se tienen los datos siguientes de frecuencia anual sobre las variables siguientes[7], para el periodo 1913-1984:

gfr:	tasa general de fertilidad[8].
pe :	valor real en dólares de la exención en el pago de impuestos personales.
year :	tendencia temporal t=1,...,72 (de 1913 a 1984).
pill :	variable ficticia igual a 1 en el año de introducción de la píldora 1963.
ww2 :	variable ficticia igual a 1 si el año está entre 1941 y 1945 (2ª guerra mundial)

Se parte de la siguiente especificación:

$$\text{gfr}_t = \beta_1 + \beta_2 \text{pe}_t + \beta_3 \text{pill}_t + \beta_4 \text{ww2}_t + u_t \tag{5.23}$$

[7]Fichero data 9-5.gdt disponible en gretl pestaña Wooldridge. Fuente: Wooldridge, J.M. (2001), *Introducción a la Econometría*.

1. Dale al conjunto de datos estructura de series temporales pulsando en la ventana principal de Gretl,
   Datos → Estructura de datos → ...

2. Estimar por MCO el modelo propuesto en (5.23). Comprueba la existencia de autocorrelación en este modelo.

3. Especifica un modelo dinámico incluyendo como regresores cuatro retardos consecutivos de la variable endógena gfr_t, es decir, añade $gfr_{t-1}, \ldots, gfr_{t-4}$ a la lista de regresores. Comprueba su significatividad, individual y conjunta, a través de estadísticos válidos.

4. Especifica un modelo dinámico diferente incluyendo como regresores cuatro retardos consecutivos de la variable pe_t, es decir, añade $pe_{t-1}, \ldots, pe_{t-4}$ a la lista de regresores. Comprueba su significatividad, individual y conjunta, a través de estadísticos válidos. ¿Introduce esta especificación algún otro problema conocido?

5. Incluye todas las variables consideradas en las preguntas 3 y 4 anteriores. Después, basándote en contrastes de hipótesis, y de forma secuencial:

   a) Omite la variable gfr_{t-4}.

   b) Omite todas las variables que encuentres no significativas al nivel de significación del 5 %, incluyendo variables retardadas y no retardadas. Puedes tener que reestimar el modelo varias veces.

   c) Guarda a sesión como icono el modelo final que hayas considerado como el mejor y escribe su Función de Regresión Muestral.

(OPCIONAL) Ahora, considera el modelo siguiente:

$$\text{gfr}_t = \beta_1 + \beta_2 \text{pe}_{t-2} + \beta_3 \text{pill}_t + \beta_4 \text{ww2}_t + u_t \tag{5.24}$$

6. Contrasta la existencia de autocorrelación en el modelo (5.24). En vez de añadir variables retardadas, obtén un estimador eficiente (asintóticamente) de sus parámetros. Escribe el modelo transformado relacionado y los valores para todos los parámetros estimados $\hat{\beta}_i$ y $\hat{\rho}$.

7. Intenta obtener la mejor especificación, de entre las de las preguntas 5*c*) y 6. Ayuda:

   a) Contrasta las restricciones pertinentes en el modelo final de la pregunta 5*c*).

   b) Examina cuidadosamente los gráficos de residuos de los dos modelos estimados.

5.7. Taller sobre todo lo trabajado en el tema

Los talleres son clases participativas, ello quiere decir que el alumno debe acudir a clase con el ejercicio realizado en su tiempo de trabajo personal. Previamente y con anticipación suficiente se le habrá realizado la debida propuesta. El ejercicio debe resolverse en grupo, idealmente en el grupo en que se va a realizar el proyecto. Esto favorece la interrelación y conocimiento entre los miembros. En los últimos 15 minutos de clase, un representante de cada grupo deberá exponer al resto de la clase la decisión adoptada por el grupo junto a las razones consideradas. Durante la clase se preguntará aleatoriamente y de forma individual a los alumnos sobre lo realizado.

En el tema de modelos dinámicos en general se lleva a cabo un taller. El objetivo del taller es analizar una serie de resultados de la estimación de varias especificaciones de un modelo o diferentes estimaciones de la misma especificación y que el alumno pueda ir practicando la toma de decisiones y la redacción apropiada de conclusiones.

Una hora de trabajo en un taller en clase equivale a dos horas de trabajo personal sobre el mismo. Además es necesario que en este tiempo se reflexione y redacten los argumentos y conclusiones a las que se haya llegado en clase.

Competencias a trabajar en esta sesión.

1. Comprender la importancia de los supuestos empleados en la especificación de un modelo econométrico básico para poder proponer y emplear supuestos más realistas.
2. Diferenciar distintos métodos de estimación y evaluar su uso de acuerdo a las características de las variables económicas de interés para obtener resultados fiables.
3. Utilizar diversas fuentes estadísticas y adquirir destreza en el uso de un software econométrico para analizar relaciones entre variables económicas.
4. Elaborar en grupos de trabajo y exponer en público, un proyecto empírico donde se valore adecuadamente los resultados obtenidos del análisis de un modelo econométrico.

Enunciado del taller:

Objetivo: Analizar las distintas especificaciones dinámicas para proponer un modelo adecuado para los inventarios.

Información: La información se basa en datos datos anuales[9] para el periodo 1950 a 1991 de las siguientes variables:

VENTAS : Ventas de la industria manufacturera en EE.UU. en millones de dólares.
INVENTARIOS : Inventarios de la industria manufacturera en EE.UU. en millones de dólares.

Se considera a la variable $VENTAS$ no estocástica.

Procedimiento: Debéis analizar la información disponible y decidir cuál es la especificación más adecuada junto con su correcta estimación.

[9] Fichero Table12.9.gdt disponible en gretl pestaña Gujarati. Fuente: Ramanathan, Ramu (2002): *Introductory Econometrics with Applications.*

• **ESPECIFICACIÓN A**. Se propone la siguiente relación:

$$INVENTARIOS_t = \beta_1 + \beta_2 VENTAS_t + \beta_3\, time + u_t \qquad t = 1, \ldots, 42 \tag{5.25}$$

Estimación de la ESPECIFICACIÓN A:

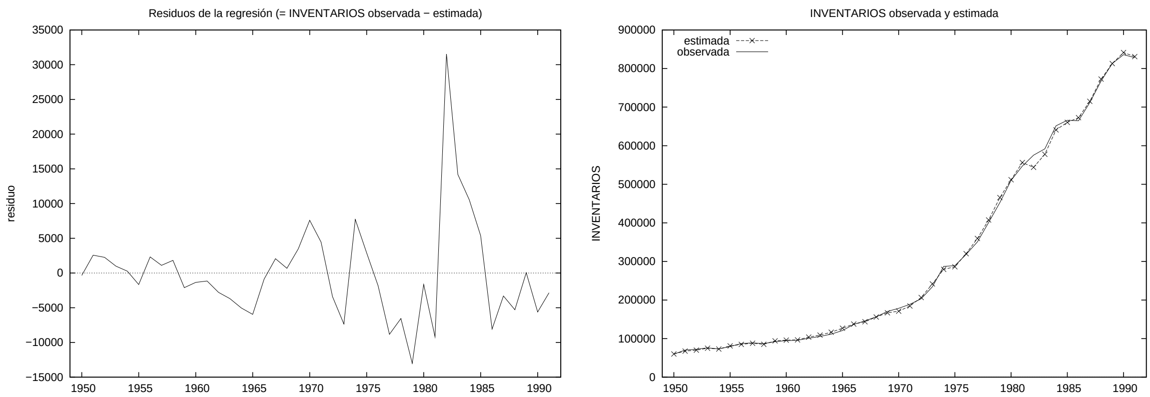
Los resultados de la estimación MCO son los siguientes:

$$\begin{array}{ccccccc} \widehat{INVENTARIOS}_t = & 433,951 & + & 1,543 & VENTAS_t + & 158,805 & time \\ \hat{des}(\hat{\beta}_i) & (2774,17) & & (0,019) & & (269,107) & \\ \hat{des}(\hat{\beta}_i)_{robusto} & (1143,38) & & (0,01361) & & (169,225) & \end{array}$$

$$R^2 = 0,9992 \qquad SCR = 2,202257+09 \qquad DW = 1,3755 \qquad BG(1) = 4,061$$

Junto con los siguientes gráficos:

Figura 5.8: *Gráfico de residuos y de la serie INVENTARIOS observada y ajustada*



1. Analiza la información mostrada: Interpreta los **dos** gráficos y si la perturbación del modelo presenta alguna característica que puede presentar algún problema en la estimación e inferencia de los coeficientes del modelo. Realiza el contraste o contrastes que sean pertinentes. Discute por qué crees que el estudiante ha introducido la variable tendencia (*time*) como regresor en el modelo y si es relevante incluirla. Utiliza los contrastes que consideres oportunos. Razona tu respuesta.

• **ESPECIFICACIÓN B:**
El estudiante considera ahora la inclusión de la variable $INVENTARIOS_{t-1}$ en el modelo y estima la siguiente especificación:

$$INVENTARIOS_t = \beta_1 + \beta_2 VENTAS_t + \beta_3\, time + \beta_4 INVENTARIOS_{t-1} + u_t \tag{5.26}$$

donde u_t es una variable aleatoria con distribución normal. Los resultados de la estimación de la ecuación (5.26) son los siguientes:

Estimaciones MCO utilizando las 41 observaciones 1951–1991
Variable dependiente: INVENTARIOS

	Coeficiente	Desv. típica	estadístico t	valor p
const	−156,9500	2750,0100	−0,0571	0,9548
VENTAS	1,2438	0,0950	13,0803	0,0000
time	320,9310	265,4460	1,2090	0,2343
INVENTARIOS_1	0,1937	0,0602	3,2154	0,0027

Suma de cuadrados de los residuos	1,72144e+09
Desviación típica de la regresión ($\hat{\sigma}$)	6820,94
$\bar{R}^2$ corregido	0,999308
$F(3, 37)$	19248,1
Estadístico de Durbin–Watson	1,59811
Coef. de autocorr. de primer orden.	0,172087
BG(1)	1,285206

Ayuda al estudiante a decidir sobre la **fiabilidad de los distintos resultados de estimación mostrados** de la especificación B. Razona tu respuesta en base a la información proporcionada.

- **Estimación de la ESPECIFICACIÓN C:**

Modelo C: estimaciones MCO utilizando las 41 observaciones 1951–1991
Variable dependiente: INVENTARIOS

	Coeficiente	Desv. típica	estadístico t	valor p
const	−56,1070	2881,5400	−0,0195	0,9846
VENTAS	1,2743	0,1104	11,5356	0,0000
VENTAS_1	0,2688	0,1085	2,4775	0,0179
time	331,1920	281,9270	1,1747	0,2476

DW = 1,20853 BG(1)=5,70 $\hat{u}_{MCO,t} = 0,383549\ \hat{u}_{MCO,t-1}$

Debéis decidir qué especificación A, B o C y qué método de estimación os parece es el más adecuado dada la información proporcionada o si pensáis que puede haber un método alternativo.

5.8. Evaluativas - Preguntas cortas

Se recuerda que dado que el curso contempla la evaluación continua es necesario que a lo largo del tema se evalúe a los alumnos, tanto en las clases magistrales, como en las prácticas de aula y ordenador. Se llevan a cabo en clase y de manera individual. A modo de ejemplo se incluyen las siguientes.

• <u>Preguntas cortas en Clases Magistrales:</u>

Pc.1 Dado el modelo:

$$output_t = \beta_1 + \beta_2 labor_t + \beta_3 land_t + \beta_4 output_{t-1} + v_t \qquad t = 2, \ldots, T \tag{5.27}$$

Marca lo que sea **cierto**:

1. $output_{t-1}$ es un regresor fijo porque está predeterminado en t.
   a) Cierto   b) Falso
2. $output_{t-1}$ siempre es una variable aleatoria independiente de v_t
   a) Cierto   b) Falso
3. Sea $v_t \sim iid(0, \sigma_v^2)$
   a) $E(output_{t-1} v_t) = 0$ b) $E(Output_{t-1} v_t) \neq 0$
4. Sea $v_t = \rho v_{t-1} + \varepsilon_t$ donde $\varepsilon_t \sim iid(0, \sigma_\varepsilon^2)$
   a) $E(output_{t-1} v_t) = 0$ b) $E(output_{t-1} v_t) \neq 0$

Pc.2 utilizando los siguientes resultados,

$$\begin{array}{ccccccc} \widehat{Y}_t & = & 4,84 & - & 0,66 \; X_t - & 0,88 \; Y_{t-1} & t = 2, \ldots, 100 \\ (\widehat{des}(\hat\beta_{MCO})) & & (0{,}36) & & (0{,}038) & (0{,}03) & \\ & & & & & R^2 = 0,91 & DW = 1,7 \end{array}$$

y las siguientes regresiones auxiliares:

$$\begin{array}{llcll} i) & \hat u_t & = & 0,09 + 0,16\hat u_{t-1} + 0,004X_t - 0,01Y_{t-1} + \hat v_{1t} & R^2 = 0,024, \; SCT = 738,3 \\ ii) & \hat u_t & = & 0,35\hat u_{t-1} + 0,1X_t + 0,06Y_{t-1} + \hat v_{2t} & R^2 = 0,018, \; SCT = 738,3 \\ iii) & \hat u_t & = & 0,3 + 0,24\hat u_{t-1} + \hat v_{3t} & R^2 = 0,005, \; SCT = 738,3 \\ iv) & \frac{\hat u_t}{\hat\sigma^2} & = & 0,13 + 0,2\frac{\hat u_{t-1}}{\hat\sigma^2} + 0,19X_t + 0,02Y_{t-1} + \hat v_{4t} & R^2 = 0,354, \; SCT = 98,7 \end{array}$$

Marca lo que sea **cierto**:

1. $\hat\beta_{MCO} = \beta + (X'X)^{-1}(X'u)$
   a) Es lineal en u b) Es no lineal en u
2. $\hat\beta_{MCO} = \beta + (X'X)^{-1}(X'u)$
   a) Es insesgado   b) Es sesgado
3. $\hat\beta_{MCO} = \beta + (X'X)^{-1}(X'u)$
   a) Es consistente   b) No es consistente
4. $\hat\beta_{MCO} = \beta + (X'X)^{-1}(X'u)$
   a) Tiene distribución conocida en muestras finitas
   b) No tiene distribución conocida en muestras finitas

- Preguntas cortas en Prácticas de Aula:

Pc.1 Sea el siguiente modelo $Y_t = \beta_1 + \beta_2 X_t + \beta_3 Y_{t-1} + u_t$, donde $u_t = \rho u_{t-1} + \epsilon_t \quad |\rho| < 1$ $\epsilon_t \sim iid(0, \sigma^2_\epsilon)$ y X_t es no estocástica.

1. ¿Por qué es no lineal el estimador de $\hat{\beta}_{MCO}$?
2. Razona la relación entre las variables Y_{t-1} y u_t y completa $E(Y_{t-1}u_t) = \ldots\ldots\ldots$
3. Propón un estimador consistente de los coeficientes $\beta_i \quad i = 1, 2, 3$.
   a) Nombre:
   b) Expresión matemática:
   c) Comenta razonadamente cuáles son sus propiedades:

Pc.2 Considera el siguiente modelo:

$$Y_t = \beta_1 \; Y_{t-1} + \beta_2 \; X_t + u_t \tag{5.28}$$

Considera las siguientes alternativas de estimación:

Método 1: Usando $t = 2, \ldots, 101$ observaciones

$$\begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} \sum Y^2_{t-1} & \sum Y_{t-1}X_t \\ \sum Y_{t-1}X_t & \sum X^2_t \end{bmatrix}^{-1} \begin{bmatrix} \sum Y_{t-1}Y_t \\ \sum X_t Y_t \end{bmatrix} \tag{5.29}$$

1. ¿Qué método de estimación está utilizando?
2. Si se tiene evidencia de autocorrelación en el término de perturbación, ¿qué propiedades tiene el estimador?

Método 2: Usando $t = 2, \ldots, 101$ observaciones

$$\begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} \sum X_{t-1}Y_{t-1} & \sum X_{t-1}X_t \\ \sum X_t Y_{t-1} & \sum X^2_t \end{bmatrix}^{-1} \begin{bmatrix} \sum X_{t-1}Y_t \\ \sum X_t Y_t \end{bmatrix} \tag{5.30}$$

3. ¿Qué método de estimación está utilizando?
4. ¿Qué propiedades tiene el estimador?

Método 3: Siendo $Y^*_t = (Y_t - \rho^* Y_{t-1})$, $X^*_t = (X_t - \rho^* X_{t-1})$, utilizando $t = 3, \ldots, 101$ observaciones

$$\begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} \sum Y^{*2}_{t-1} & \sum Y^*_{t-1}X^*_t \\ \sum Y^*_{t-1}X^*_t & \sum X^{*2}_t \end{bmatrix}^{-1} \begin{bmatrix} \sum Y^*_{t-1}Y^*_t \\ \sum X^*_t Y^*_t \end{bmatrix} \tag{5.31}$$

$$\rho^* = \frac{\sum \hat{u}_t \hat{u}_{t-1}}{\sum \hat{u}^2_{t-1}} \tag{5.32}$$

y $\hat{u}_t = Y_t - \hat{\beta}_{1,VI} \; Y_{t-1} + \hat{\beta}_{2,VI} \; X_t$

5. ¿Qué método de estimación está utilizando?
6. ¿Qué propiedades tiene el estimador?
7. A la vista de lo comentado en los apartados anteriores, ¿qué investigador ha utilizado el mejor estimador? Razona tu respuesta.

• Preguntas cortas en Prácticas de Ordenador:

Un estudiante pretende estudiar los determinantes de consumo de gasolina en U.S. Para ello dispone de observaciones anuales en el período de 1960 a 1995 sobre las siguientes variables[10]:

G: Consumo de gasolina total, en U.S., gasto total dividido por su índice de precios.
Pg: Índice de precios de la gasolina.
R: Renta disponible, per cápita.
Ps: Índice de precios agregado del consumo de servicios.

Se considera a las variables Pg, R y Ps no estocásticas. El estudiante propone la siguiente especificación:

$$G_t = \beta_1 + \beta_2 Pg_t + \beta_3 R_t + \beta_4 Ps_t + \beta_5 G_{t-1} + u_t \qquad t = 1, \ldots, 42 \tag{5.33}$$

Pc.1 Estima el modelo por MCO y completa:

$$\begin{array}{ccccccccccc} \widehat{G}_t & = & & + & & Pg_t + & & R_t + & & Ps_t + & G_{t-1} \\ (\widehat{des}(\hat{\beta}_{MCO})) & & (\quad\quad) & & (\quad\quad) & & (\quad\quad) & & (\quad\quad) & & (\quad\quad) \end{array}$$

$$R^2 = \qquad\qquad\qquad DW =$$

Se considera que u_t puede seguir un proceso AR(p) o MA(p) con p hasta de orden 2. Realiza el contraste oportuno.

i) Escribe la hipótesis nula, la alternativa y el estadístico de contraste que vas a utilizar junto con su distribución bajo la hipótesis nula. Indica claramente de dónde salen cada uno de los elementos de este estadístico.

ii) Aplícalo a los datos del archivo y completa:
Regresión auxiliar obtenida:

...... = ...

$R^2 =$

Valor muestral del estadístico =
Valor crítico para un nivel de significación ($\alpha = 5\,\%$)=
Aplica la regla de decisión:

iii) ¿Qué puedes decir de la consistencia del estimador empleado? ¿cómo es plim $\hat{\beta}_{MCO}$?

[10]Fichero greene7-8.gdt disponible en gretl pestaña Greene. Fuente: Ramanathan, Ramu (2002): *Introductory Econometrics with Applications*.

Pc.2 Estima el modelo por MC2E usando como instrumentos const, Pg_t, Pg_{t-1}, R_t, R_{t-1}, Ps_t y Ps_{t-1} y completa:

$$\underset{(\widehat{des}(\hat{\beta}_{VI}))}{\widehat{G}_t} = \underset{(\qquad)}{\qquad} + \underset{(\qquad)}{\qquad} Pg_t + \underset{(\qquad)}{\qquad} R_t + \underset{(\qquad)}{\qquad} Ps_t + \underset{(\qquad)}{\qquad} G_{t-1}$$

i) Razona si los instrumentos son adecuados.

ii) ¿Qué propiedades tiene el estimador obtenido? ¿Son fiables las desviaciones típicas obtenidas?

Pc.3 Estima el modelo por el método de Hildreth-Lu. Aplícalo a los datos del archivo y completa:

$$\underset{(\widehat{des}(\hat{\beta}_{HL}))}{\widehat{G}_t} = \underset{(\qquad)}{\qquad} + \underset{(\qquad)}{\qquad} Pg_t + \underset{(\qquad)}{\qquad} R_t + \underset{(\qquad)}{\qquad} Ps_t + \underset{(\qquad)}{\qquad} G_{t-1}$$

$\hat{\rho} =$ `valor mínimo de` $SCR =$

¿Qué propiedades tiene el estimador obtenido? ¿Son fiables las desviaciones típicas obtenidas?

5.9. Bibliografía del tema

Referencias Bibliográficas Básicas:

• Teórica:

[1] Greene, W. (1998), Análisis Econométrico, ed. Prentice Hall, 3ª edition.

[2] Ramanathan, R. (2002), Introductory Econometrics with applications, ed. South-Western, 5th edition.

[3] Wooldridge, J. M. (2003), Introductory Econometrics: A modern Approach, ed. South-Western, 2nd edition.

• Ejercicios:

[1] Alegre, J., Arcarons, J., Bolancé, C. y Díaz, L. (1995), Ejercicios y Problemas de Econometría, Colección Plan Nuevo, ediciones AC.

[2] Fernández, A., González, P., Regúlez, M., Moral, P., Esteban, V. (2005), Ejercicios de Econometría, ed. McGraw-Hill, 2 ª edición.

[3] Recopilación de ejercicios recomendados y exámenes de Econometría. Departamento de Economía Aplicada III (Econometría y Estadística). Mimeo Febrero 2009.

• Ejercicios con gretl:

[1] Ramanathan, R. (2002) Instructor's Manual to accompany, del libro Introductory Econometrics with applications, ed. South-Western, 5th edition, Harcourt College Publishers.

[2] Wooldridge, J.M. (2003), Student Solutions Manual, del libro Introductory Econometrics: A modern Approach, ed. South-Western, 2nd edition.

Referencias Bibliográficas Complementarias:

[1] Alonso, A., Fernández, J. y Gallastegui, I. (2005), Econometría, ed. Pearson: Prentice Hall.

[2] Gujarati, D. (2004), Econometría, ed. McGraw-Hill, 4ª edición.

[3] Johnston, J. (1984), Métodos de Econometría, ed. Vicens Vivens, 4ª edición.

[4] Maddala, G. S. (1996), Introducción a la Econometría, ed. Pearson: Prentice Hall.

[5] Novales, A. (1993), Econometría, ed. McGraw-Hill, 2ª edición.

[6] Pindyck, R. S. y Rubinfeld, D. L. (1998), Econometric Models and Economic Forecast, ed. McGraw-Hill, 4ª edición.

5.10. Anexo: Instrucciones básicas de gretl para modelos dinámicos

• Añadir retardos de la variable dependiente y/o de las variables exógenas como regresores al estimar por MCO
Modelo → Mínimos Cuadrados Ordinarios
En la ventana de gretl: especificar modelo elegir la variable dependiente y añadir las variables exógenas corrientes, en el mismo t. Entonces se ve remarcada la opción de *Retardos*. Al elegir con el botón izquierdo aparece una ventana orden de retardos.

Si se quieren incluir retardos, por ejemplo de RD del 0 al 4 se especifica en la parte izquierda de la caja al lado de la variable correspondiente.

En este caso el modelo especificado será

$$PATENTS_t = b1 + b2RD_t + b3RD_{t-1} + b4RD_{t-2} + b5RD_{t-3} + b6RD_{t-4} + u_t$$

Donde $t = 5, \ldots, T$, se pierden 4 observaciones del total T disponible al definir RD_{t-4}.

Si solamente se quiere incluir el cuarto retardo además de RD_t, esto es, el modelo:

$$PATENTS_t = b1 + b2RD_t + b3RD_{t-4} + u_t$$

Entonces se elige la casilla de *retardos específicos* y se escribe en la caja correspondiente para ello los valores $0, 4$.

Si se quiere añadir al modelo retardos de la variable endógena como regresores en la ventana de orden de retardos se marca la opción *Retardos de la variable dependiente*. De igual forma si se quieren introducir un continuo de retardos del 1 al 2 por ejemplo se seleccionan en la parte izquierda el intervalo de valores.

$$PATENTS_t = b1 + b2RD_t + b3RD_{t-4} + b4PATENTS_{t-1} + b5PATENTS_{t-2} + u_t$$

Si son retardos de la variable dependiente alternos, se selecciona la ventana retardos específicos y se indica cuales en concreto, por ejemplo 1,3 introduce como regresores $PATENTS_{t-1}$ y $PATENTS_{t-3}$ pero no introduce el segundo retardo.

$$PATENTS_t = b1 + b2RD_t + b3RD_{t-4} + b4PATENTS_{t-1} + b5PATENTS_{t-3} + u_t$$

• Añadir retardos al estimar por Variables Instrumentales tanto en los regresores como en la lista de instrumentos.

Modelo → Otros modelos lineales → Mínimos Cuadrados en dos etapas

En la ventana gretl: *especificar modelo* tenemos que elegir la variable dependiente, las variables explicativas en t o corrientes y los instrumentos que estarán en la lista en t.

Se pulsa en retardos y aparece una ventana donde se tienen que seleccionar los retardos de las variables que aparecen en el modelo como regresores, tanto de variables exógenas como de la dependiente y también las variables que vamos a usar como instrumentos.

Tema 6

Guía para el desarrollo de un proyecto empírico

6.1. Características básicas del proyecto

Esta sección desarrolla una guía básica para estructurar la realización de un proyecto. Es del todo recomendable la lectura del Capítulo 19 del libro *Introdución a la Econometría*, de Wooldridge, J. M. (2003), que aparece en las referencias bibliográficas. Los seminarios y tutorías personalizadas de los equipos individuales servirán para marcar el ritmo en la evolución del proyecto y para que el profesor tutorice tanto al equipo como individualmente a cada uno de los integrantes en su aprendizaje. No hay que perder de vista que el proyecto debe de contribuir a obtener las competencias específicas de la asignatura en su totalidad aunque se incida en la cuarta en cuanto a su evaluación.

- **Estructura básica del trabajo:**

  1. **Portada:** Título del trabajo y nombres y apellidos del autor/autores.
  2. **Introducción:** Presenta la motivación principal del trabajo, el problema a analizar y posibles referentes en la literatura económica. Por ejemplo, puede ayudar plantearse una pregunta, muy concreta, sobre un fenómeno económico. Por ejemplo, ¿como afecta a la oferta laboral femenina y casada el nivel salarial, la educación y/o el número de hijos?
  3. **Revisión de la bibliografía:** En muchas ocasiones existe literatura en el tema elegido. Si es así debe de ser incluida, en un resumen breve.
  4. **Datos:** Se describen los datos que componen la muestra a utilizar y las fuentes de donde han sido obtenidos. Se describen las variables a utilizar en el modelo junto con las unidades de medida. Finalmente se realiza un análisis descriptivo básico.
  5. **Modelo:** Se introduce el modelo de partida propuesto, las variables que entran en el modelo, el signo esperado de los coeficientes. También si hay diferentes alternativas a la especificación propuesta que se quieran contrastar o analizar.
     Es claro que el modelo inicial evolucionará a lo largo del trabajo. Por lo tanto en el trabajo y la exposición se debe mostrar los puntos más importantes de esta evolución, sin ser exhaustivos en la repetición de interpretación de signos y coeficientes.

6. **Resultados empíricos:** Se muestran y explican razonadamente los resultados obtenidos aplicando las técnicas vistas en clase que se crean oportunas: análisis de residuos, contrastes sobre las hipótesis básicas, elección de métodos alternativos de estimación y realización de contrastes sobre los parámetros de interés del modelo. Al final del trabajo se crea un apéndice en el que incluyen todos los outputs de gretl y gráficos que se hayan obtenido a lo largo de todo el trabajo.
7. **Conclusiones:** Por último, al final del trabajo, se redactan las conclusiones del trabajo, teniendo en cuenta su objetivo y la especificación y estimación del modelo después del análisis econométrico. En esta sección se explica lo realizado en conjunto y se razona si es el caso sobre la especificación final elegida. Este apartado es imprescindible.
8. **Bibliografía:** Recoge las referencias completas citadas a lo largo del texto. Fuentes de consultas bibliográficas y de datos, así como referencias a páginas web si es que se utilizan.

- **Cronología:** El proyecto debe evolucionar a lo largo del curso, no debe ser un trabajo realizado en los últimos diez días antes de su exposición pública por ello es importante que los equipos se marquen una cronología de acuerdo a las directrices marcadas en lo seminarios que se lleven a cabo en el aula. Como ya se ha indicado los seminarios y tutorías personalizadas de los equipos servirán para marcar el ritmo en la evolución del proyecto y para que el profesor tutorice tanto al equipo como individualmente a cada uno de los integrantes en su aprendizaje. No hay que perder de vista que el proyecto debe de contribuir a obtener las competencias específicas de la asignatura en su totalidad aunque se incida en la cuarta en cuanto a su evaluación.

- **Exposición:** En general el formato de la exposición puede ser elegido libremente, transparencias, power point etc. Habrá de tenerse en cuenta que todos los alumnos intervienen en la exposición y que el profesor realizará preguntas a los miembros del grupo sobre el trabajo realizado.

- **Datos a utilizar:** Los que se quieran, siempre y cuando se refiera la referencia en el trabajo de la fuente utilizada.

1. Se pueden utilizar datos de los ficheros disponibles en Gretl siempre y cuando el trabajo no sea replicar un ejercicio que ya esté en estos libros.
2. Bases de datos disponibles en Gretl sobre servidor
   Archivo ⇒ Bases de datos ⇒ sobre servidor
3. Bases de datos disponibles en Biblioteca.
4. Otras direcciones de interés y bases de datos:
   Eurostat: http://epp.eurostat.cec.eu.int/
   Banco Mundial: http://devdata.worldbank.org/
   Fondo Monetario Internacional: http://www.imf.org/
   OCDE: http://www.oecd.org/
   Banco Central Europeo: http://www.ecb.int/
   Economic and Social Data Services: Guía a Recursos internacionales de datos de libre acceso, http://www.esds.ac.uk/international/access/access.asp

Banco de España: http://www.bde.es/
Instituto Nacional de Estadística: http://www.ine.es/
Eustat: http://www.eustat.es/

- **Algunos consejos para la preparación del trabajo:**

  1. Primero es importante pensar en el modelo o relación básica que se quiere analizar junto con los datos a emplear. Es aconsejable estudiar previamente los datos de las variables que entran en el modelo, sus estadísticos descriptivos, gráficos etc.
  2. También es importante la forma de presentar los resultados. Una vez realizados los diversos análisis, estimaciones etc., se pone en común todos los resultados y se selecciona lo importante. Una vez elegido lo más relevante de todo lo realizado esta información se presenta de forma que sea fácil comparar y valorar los resultados obtenidos.
  3. No mezclar diferentes fuentes y tipos de letra en el texto.
  4. No utilizar abreviaciones coloquiales, tipo SMS.
  5. Todas las figuras y tablas deben estar numeradas. Añadir una pequeña leyenda tanto a los gráficos como a las tablas explicando que recogen (por ejemplo, Figura 2: Gráfico de los residuos contra el tiempo). Se puede poner arriba o abajo de la figura pero siempre hacerlo de la misma forma.
  6. Las figuras y tablas también pueden ir en un apéndice y hacer referencia en el texto con la numeración utilizada.
  7. Numerar todas las páginas.
  8. Es recomendable que se revisen las distintas versiones del trabajo antes de su presentación, cuidando que no haya errores gramaticales.

- **Cómo presentar y comparar resultados:**

  A continuación se va a mostrar la presentación de resultados de tres especificaciones alternativas y tres métodos de estimación como guía de presentación de resultados.

  Vamos a estimar las siguientes especificaciones o modelos alternativos para explicar el precio de la vivienda:

$$\begin{array}{ll} \textbf{Modelo A} & PRICE_i = \beta_1 + \beta_2\, SQFT_i + u_i \\ \textbf{Modelo B} & PRICE_i = \beta_1 + \beta_2\, SQFT_i + \beta_3\, BEDRMS_i + u_i \\ \textbf{Modelo C} & PRICE_i = \beta_1 + \beta_3\, BEDRMS_i + \beta_4\, BATHS_i + u_i \end{array}$$

  Estos tres modelos difieren en las variables explicativas incluidas.

  La Tabla 6.1 muestra los coeficientes estimados por MCO y distintos estadísticos asociados a la cuatro especificaciones o modelos alternativos anteriores. Esta forma de presentar los resultados puede estar más indicada cuando, como ahora, se presentan distintas especificaciones con la misma variable dependiente. En la parte de abajo de la tabla, también para cada una de las especificaciones, podéis incluir los valores muestrales de los estadísticos de diversos

Tabla 6.1: Modelos estimados para el precio de la vivienda $PRICE$

Variable dependiente: $PRICE$

Variable	**Modelo A**	**Modelo B**	**Modelo C**
CONSTANT	52,351	121,179	27,2633
	(1,404)	(1,511)	(0,182)
SQFT	0,13875	0,14831	
	(7,407)	(6,993)	
BEDRMS		-23,911	-10,1374
		(-0,970)	(-0,216)
BATHS			138,795
			(2,652)
Suma de cuadrados residual	18273,6	16832,8	55926,4
R^2	0,821	0,835	0,450706
$\bar{R}^2$	0,806	0,805	0,350834
F de significación conjunta	54,861	27,767	4,51285
Grados de libertad	12	11	11

Los valores entre paréntesis son los correspondientes estadísticos t de significatividad individual sin tener en cuenta la posible heterocedasticidad y o autocorrelación.

contrastes bien de heterocedasticidad, autocorrelación etc. si los habéis realizado para cada modelo estimado.

Si se considera estimar por distintos métodos la misma especificación, puede ayudar presentar los resultados como en la Tabla 6.2:

Tabla 6.2: Función de Salarios

Variable dependiente: Salarios Privados W^p

Explicativas	**MCO**	**Cochrane-Orcutt**	**Hildreth-Lu**
const	$1,497$	$1,5$	$1,53^*$
	$(1,27)$	$(1,27)$	$(1,32)$
X_t	$0,439^*$	$0,439^*$	$0,434^*$
	$(0,032)$	$(0,039)$	$(0,075)$
X_{t-1}	$0,146^*$	$0,147^*$	$0,151^*$
	$(0,037)$	$(0,043)$	$(0,074)$
A_t	$0,13^*$	$0,13^*$	$0,132^*$
	$(0,032)$	$(0,032)$	$(0,035)$

Entre paréntesis se muestran las desviaciones típicas estimadas. En el caso de MCO son robustas a autocorrelación. El símbolo $*$ denota que es significativo al 5 %. El tamaño muestral es $T = 80$ datos trimestrales.

De esta forma es más fácil comparar los resultados a la vez que mostrarlos en una transparencia a la hora de la presentación.

Bibliografía

Referencias Bibliográficas Básicas:

• Teórica:

[1] Greene, W. (1998), Análisis Econométrico, ed. Prentice Hall, 3ª edition.

[2] Ramanathan, R. (2002), Introductory Econometrics with applications, ed. South-Western, 5th edition.

[3] Wooldridge, J. M. (2003), Introductory Econometrics: A modern Approach, ed. South-Western, 2nd edition.

• Ejercicios:

[1] Alegre, J., Arcarons, J., Bolancé, C. y Díaz, L. (1995), Ejercicios y Problemas de Econometría, Colección Plan Nuevo, ediciones AC.

[2] Fernández, A., González, P., Regúlez, M., Moral, P., Esteban, V. (2005), Ejercicios de Econometría, ed. McGraw-Hill, 2 ª edición.

[3] Recopilación de ejercicios recomendados y exámenes de Econometría. Departamento de Economía Aplicada III (Econometría y Estadística). Mimeo Febrero 2010.

• Ejercicios con gretl:

[1] Ramanathan, R. (2002), Instructor's Manual to accompany, del libro Introductory Econometrics with applications, ed. South-Western, 5th edition, Harcourt College Publishers.

[2] Wooldridge, J. M. (2003), Student Solutions Manual, del libro Introductory Econometrics: A modern Approach, ed. South-Western, 2nd edition.

Referencias Bibliográficas Complementarias:

[1] Alonso, A., Fernández, J. y Gallastegui, I. (2005), Econometría, ed. Pearson: Prentice Hall.

[2] Breusch, T. S. y Pagan, A. R. (1979), A Simple Test for Heteroscedasticity and Random Coefficient Variation, *Econometrica*, 47, pp. 1287-1294.

[3] Dougherty, Ch. (2006), Introduction to Econometrics, 3rd. Ed., Oxford University Press.

[4] Goldfeld, S. M. y Quandt, R. E. (1965), Some Test for Homoscedasticity, *Journal of the American Statistical Association*, 60, pp. 539-547.

[5] Gujarati, D. (2004), Econometría, ed. McGraw-Hill, 4ª edición.

[6] Harvey, A., y Phillips, G. (1974), A Comparison of the Power of Some Test for Heteroscedasticity in the General Linear Model, *Journal of Econometrics*, 2, pp. 307-316.

[7] Johnston, J. (1984), Métodos de Econometría, ed. Vicens Vivens, 4ª edición.

[8] Maddala, G. S. (1996), Introducción a la Econometría, ed. Pearson: Prentice Hall.

[9] Novales, A. (1993), Econometría, ed. McGraw-Hill, 2ª edición.

[10] Pindyck, R.S. y Rubinfeld, D.L. (1998), Econometric Models and Economic Forecast, ed. McGraw-Hill, 4ª edición.

[11] White, H. (1980), A heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity, *Econometrica*, 48, pp. 817-838.

Apéndice: Tablas de datos

- Datos sobre Consumo y Renta. Ejemplo magistral del Tema 2: Heterocedasticidad.

Tabla A.3: Observaciones de Consumo y Renta

Observación	Consumo	Renta	Observación	Consumo	Renta
i= 1	52,25	258,3	i= 21	98,14	719,8
i= 2	58,32	343,1	i= 22	123,94	720,0
i= 3	81,79	425,0	i= 23	126,31	722,3
i= 4	119,90	467,5	i= 24	146,47	722,3
i= 5	125,80	482,9	i= 25	115,98	734,4
i= 6	100,46	487,7	i= 26	207,23	742,5
i= 7	121,51	496,5	i= 27	119,80	747,7
i= 8	100,08	519,4	i= 28	151,33	763,3
i= 9	127,75	543,3	i= 29	169,51	810,2
i= 10	104,94	548,7	i= 30	108,03	818,5
i= 11	107,48	564,6	i= 31	168,90	825,6
i= 12	98,48	588,3	i= 32	227,11	833,3
i= 13	181,21	591,3	i= 33	84,94	834,0
i= 14	122,23	607,3	i= 34	98,70	918,1
i= 15	129,57	611,2	i= 35	141,06	918,1
i= 16	92,84	631	i= 36	215,40	929,6
i= 17	117,92	659,6	i= 37	112,89	951,7
i= 18	82,13	664,0	i= 38	166,25	1014,0
i= 19	182,28	704,2	i= 39	115,43	1141,3
i= 20	139,13	704,8	i= 40	269,03	1154,6

Tabla A.4: Datos de la empresa Lydia Pinkham (1907-1960)

Año	1907	1908	1909	1910	1911	1912	1913	1914	1915	1916	1917	1918	1919
V_t	1016	921	934	976	930	1052	1184	1089	1087	1154	1330	1980	2223
G_t	6,08	4,51	5,29	5,43	5,25	5,49	5,25	5,78	6,09	5,04	7,52	6,13	8,62
Año	1920	1921	1922	1923	1924	1925	1926	1927	1928	1929	1930	1931	1932
V_t	2203	2514	2726	3185	3351	3438	2917	2359	2240	2196	2111	1806	1644
G_t	8,66	10,16	13,60	14,48	16,08	18,00	19,41	12,29	13,73	16,11	15,68	9,83	10,46
Año	1933	1934	1935	1936	1937	1938	1939	1940	1941	1942	1943	1944	1945
V_t	1518	1103	1266	1473	1423	1767	2161	2336	2602	2518	2637	2177	1920
G_t	8,07	3,39	5,62	7,45	7,49	8,62	10,34	10,54	11,64	11,02	11,45	10,12	8,36
Año	1946	1947	1948	1949	1950	1951	1952	1953	1954	1955	1956	1957	1958
V_t	1910	1814	1770	1984	1787	1689	1866	1896	1684	1633	1657	1569	1390
G_t	9,41	14,53	15,04	9,81	9,74	7,66	9,20	9,64	8,11	7,89	8,02	7,70	6,39
Año	1959	1960											
V_t	1387	1289											
G_t	6,44	5,64											

V_t: ventas de la empresa medidos en miles de dólares. G_t: gastos en publicidad medidos en cientos de miles de dólares.
Fuente: Palda, K. S. (1964), *The measurement of cumulative adversing effects*, Englewood Cliffs, N. J. Prentice-Hall.