Capítulo 18. MINIMOS CUADRADOS GENERALIZADOS

Capítulo 18. MINIMOS CUADRADOS GENERALIZADOS

En este capítulo, el tema desarrollado se articula con el anterior al comprobar el cumplimiento de los supuestos del modelo lineal general sobre el término de perturbación.

Luego de especificar y estimar el modelo de regresión, se contrastan los supuestos sobre la parte sistemática –analizados en el capítulo anterior- y sobre los residuos del modelo; particularmente, las hipótesis de homocedasticidad y no autocorrelación.

La violación de alguno o ambos supuestos da lugar a utilizar modelos en los que se flexibiliza la restricción sobre la matriz de varianzas y covarianzas de las perturbaciones.

18.1. El análisis gráfico de los residuos

El *análisis gráfico de los residuos* va a presentar una primera información sobre estas hipótesis:

- El de los valores de e_t contra los valores de t, si se detecta una tendencia creciente o decreciente en el gráfico, puede existir *autocorrelación*.

Ejemplo 18.1. En la Figura se observan los residuos heterocedásticos y autocorrelacionados del Caso 17.1, para el estudio del consumo en la Argentina en función del PBI y de la tasa de interés.

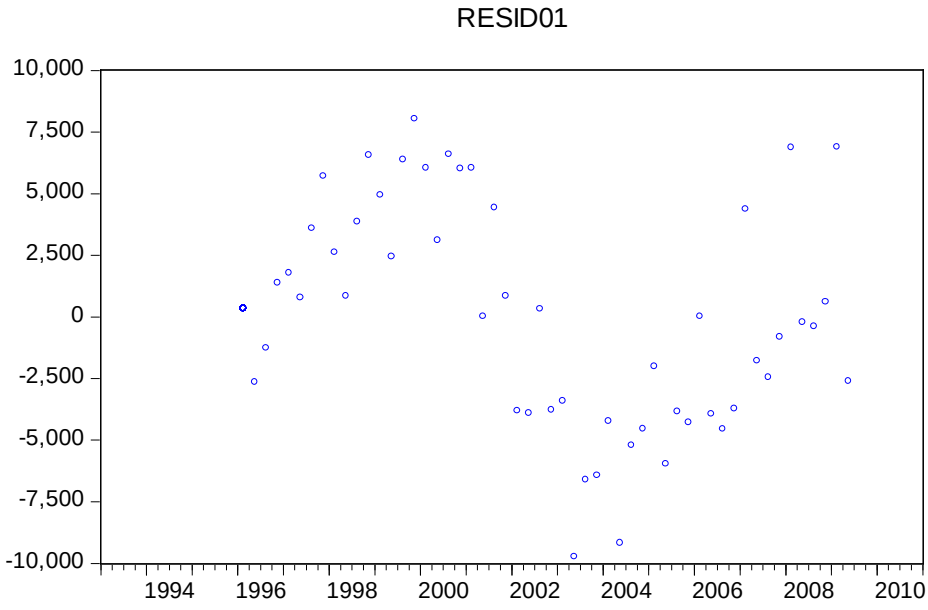


Figura 18.1 Residuos heterocedásticos y autocorrelacionados

- El de los valores de e_t contra los valores de $\hat{Y}_t$, si se comprueba una tendencia de cualquier tipo en el gráfico, puede existir autocorrelación, ya que habrá correlación entre los residuos.
- El de los valores de e_t^2 contra los valores de $\hat{Y}_t$, si se verifica una tendencia de cualquier tipo en el gráfico, puede existir *heterocedasticidad*.
- El de los valores de e_t contra los valores de X_t, si se detecta una tendencia creciente o decreciente en el gráfico, puede existir *autocorrelación*, ya que los residuos no serán ortogonales respecto a las variables explicativas.
- El de los valores de e_t^2 contra los valores de X_t, si se verifica una tendencia de cualquier tipo en el gráfico, puede existir *heterocedasticidad* o *no linealidad* (habrá relación entre la varianza del término del error y las variables explicativas).

18.2. Heterocedasticidad

Si la varianza del término de perturbación del modelo de regresión lineal no es constante para todas las observaciones se dice que es heterocedástica, o que existe heterocedasticidad en las perturbaciones.

La heterocedasticidad puede surgir en numerosas aplicaciones económicas, siendo más común en el análisis de datos de corte transversal.

Entre las causas que originan heterocedasticidad se encuentran:

- Especificación errónea del modelo
- Cambio estructural
- Alta dispersión, absoluta y relativa, a medida que crece el tamaño de la muestra

Ejemplo 18.2. En los estudios que analizan el consumo o gasto familiar, es frecuente encontrar una mayor variabilidad del gasto realizado por familias de renta alta que por familias de renta baja. Esto se debe a que un mayor nivel de renta permite mayor margen para la realización de gastos y, por lo tanto, una mayor varianza. Lo mismo ocurre en estudios sobre beneficios de las empresas, cuya varianza puede depender del tamaño de la empresa, de la diversificación de su producto, de las características del sector industrial al que pertenezca, etc. Y, por lo tanto, puede variar a través de las distintas empresas.

La consecuencia de esto es que la matriz de varianzas y covarianzas de las perturbaciones no es escalar. Suponiendo que no existe autocorrelación en las perturbaciones, la heterocedasticidad implica la siguiente estructura de la matriz de varianzas y covarianzas:

$$E\left(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\right) = \boldsymbol{\Omega} = \begin{bmatrix} \sigma_1^2 & 0 \cdots 0 \\ 0 & \sigma_2^2 \cdots 0 \\ \cdots\cdots\cdots\cdots \\ 0 & 0 \cdots \sigma_T^2 \end{bmatrix}$$

Normalmente, en la práctica, a priori no se sabe si hay o no problemas de heterocedasticidad en las perturbaciones; por esto es que la estimación se realiza bajo el supuesto de homocedasticidad. Luego, para confirmar el cumplimiento del supuesto, se realizan los contrastes.

Se han desarrollado un gran número de métodos para contrastar la hipótesis nula de igualdad de varianzas u homocedasticidad. Esta variedad se debe a que la especificación de la hipótesis alternativa de heterocedasticidad suele no ser conocida, puede ser más o menos general y refleja diferentes comportamientos de las perturbaciones.

A continuación se explican algunos de los contrastes más utilizados en la literatura.

Contraste de Goldfeld y Quandt

En determinados contextos, aunque no se conozca la forma de la heterocedasticidad, se tienen sospechas de que las varianzas, $\sigma_i^2; i = 1,\ldots,T$ mantienen una relación monótona con los valores de alguna variable **Z**.

Se supone que la hipótesis alternativa es $\sigma_1^2 = \sigma_2^2 G(\mathbf{Z}_i, \gamma)$, donde $G(\cdot)$ es una función monótona creciente en $\mathbf{Z}_i$ que puede ser o no uno de los regresores incluidos en el modelo de regresión.

Ejemplo 18.3. En el análisis del gasto familiar, es válido suponer que la varianza del gasto depende del nivel de renta de cada familia; es decir, que $\sigma_i^2 = \sigma^2 G(R_i)$, donde $G(\cdot)$ es una función creciente de la renta familiar y σ^2 es un factor de escala.

Los pasos que se siguen para realizar el contraste, son:

1) Ordenar las observaciones de la tabla de datos, siguiendo el orden creciente de la variable $\mathbf{Z}_i$.

2) Eliminar p observaciones centrales dando lugar a dos bloques de $(T-p)/2$ observaciones, T_1 y T_2 respectivamente; las observaciones centrales que se eliminan permiten mayor independencia entre los dos grupos. El número de observaciones en cada grupo ha de ser mayor que el número de parámetros a estimar. Pulido (1989) indica que si se tienen 30 observaciones en la muestra deben eliminarse 8 observaciones; a partir de un tamaño de muestra de 60 es suficiente eliminar 16 observaciones.

3) Estimar el modelo de regresión en forma separada para cada grupo de observaciones. En el primer grupo se concentran las observaciones con menor valor nominal de **Z**; mientras que, en el segundo grupo, se encuentran las de mayor valor nominal.

Con la estimación del modelo en cada grupo se obtienen los errores de estimación de cada grupo.

4) Construir el estadístico de contraste. Bajo la hipótesis nula de homocedasticidad

$$H_0 : \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_T^2$$

y suponiendo que la perturbación se distribuye como una normal de media cero y no está serialmente correlacionada, el estadístico GQ sigue una distribución *F* de Snedecor:

$$GQ = \frac{\mathbf{e}_2' \mathbf{e}_2 / T_2 - k}{\mathbf{e}_1' \mathbf{e}_1 / T_1 - k} \sim F(T_1 - k, \quad T_2 - k)$$

donde, $\mathbf{e}_2' \mathbf{e}_2$ es la suma de cuadrados de residuos de la regresión de Y sobre X en el segundo grupo de observaciones, y $\mathbf{e}_1' \mathbf{e}_1$ es la suma de cuadrados de residuos de la regresión Y sobre X utilizando el primer grupo de observaciones.

Mientras que, bajo la hipótesis nula, las varianzas deben ser iguales, bajo la hipótesis alternativa, crecerán de un grupo a otro. Cuanto más difieran estas sumas de cuadrados, mayor será el valor del estadístico, y por lo tanto, mayor evidencia habrá en contra de la hipótesis nula.

5) Se rechaza la H_0, a un nivel de significación α, si:

$$GQ > F_\alpha(T_1 - k, \quad T_2 - k)$$

Este contraste se puede utilizar, en principio, para detectar 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 captarla y no ser significativo.

Contraste de Breusch y Pagan

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

$$H_A: \ \sigma_i^2 = \sigma^2 \ G\left(\alpha_0 + \boldsymbol{\alpha}'\mathbf{Z}_i\right)$$

$\mathbf{Z_i}$ es un vector de variables exógenas que pueden ser las explicativas del modelo y la función $G(\cdot)$ no se especifica.

La hipótesis nula del contraste es la de homocedasticidad que, dada la alternativa, implica contrastar:

$$H_0: \ \boldsymbol{\alpha} = \mathbf{0}$$

Una forma operativa de realizar el contraste, es la siguiente:

1) Estimar el modelo $\mathbf{Y_i} = \beta_1 + \beta_2\mathbf{X}_{2i} + \beta_3\mathbf{X}_{3i} + \cdots + \beta_k\mathbf{X}_{ki} + \boldsymbol{\varepsilon}_i$ para obtener los errores

2) Utilizando los residuos $\mathbf{e} = \mathbf{Y} - \mathbf{X}\,\hat{\boldsymbol{\beta}}_{\mathbf{MCO}}$ se construye la siguiente serie

$$r_i = \frac{e_i^2}{\mathbf{e}'\mathbf{e}} \qquad i = 1, \cdots, T$$

3) Especificar y estimar el modelo

$$r_i = \alpha_0 + \boldsymbol{\alpha}' \mathbf{Z}_i + \omega_i \qquad i = 1, \cdots, T$$

De aquí se obtiene la suma de cuadrados explicada (SCE).

4) Se utiliza como estadístico de contraste SCE/2, que bajo hipótesis nula de homocedasticidad se distribuye asintóticamente $\chi^2(S)$, donde S son los grados de libertad igual al número de variables en $\mathbf{Z_i}$. Se rechaza la hipótesis nula a un nivel de significación (α), si el valor muestral del estadístico excede el cuantil $\chi^2_\alpha(S)$.

Contraste de White

Con este método se contrasta la hipótesis nula de homocedasticidad frente a una alternativa general de heterocedasticidad.

White, derivó este contraste comparando dos estimadores de la varianza de los estimadores MCO:

1. $\widehat{\mathbf{V}}(\widehat{\boldsymbol{\beta}}) = \widehat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}$
2. $\mathbf{V}_{\text{WHITE}}(\widehat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{S}\,\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1}$

Donde, **S** es una matriz diagonal cuyos elementos son los residuos mínimo-cuadráticos ordinarios al cuadrado

$$\boldsymbol{S} = diag(\mathrm{e}_1^2, \mathrm{e}_2^2, \ldots, \mathrm{e}_\mathrm{T}^2)$$

El estimador $\mathbf{V}_{\text{WHITE}}(\widehat{\boldsymbol{\beta}})$ es consistente siempre que la matriz $\boldsymbol{\Omega}$ sea diagonal; es decir, en aquellas situaciones en las que el modelo tenga perturbaciones heterocedásticas pero no autocorrelacionadas.

Bajo la hipótesis nula de homocedasticidad, ambos estimadores, 1 y 2, son consistentes; mientras que, bajo la alternativa de heterocedasticidad, el estimador $\hat{\mathbf{V}}(\hat{\boldsymbol{\beta}})$ no lo es.

La forma operativa de realizar el contraste consiste en:

1) Estimar el modelo bajo estudio

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} + \varepsilon_i$$

para obtener los errores, e.

2) White supone que el cuadrado de los errores estimados se comportan de la siguiente manera

$$e_i^2 = \alpha_0 + \sum_{k=1}^{K} \beta_k X_{ki} + \sum_{k=1}^{K} \varphi_k X_{ki}^2 + \sum_{k=1}^{K-1} \sum_{l=k+1}^{K} \delta_{kl} X_{ki} X_{li} + \mu_i \qquad \forall i = 1,2,3 \cdots T$$

Esto significa que se especifica un modelo donde el cuadrado de los errores se explica por las variables independientes del modelo, el cuadrado de ellas y su producto cruzado:

$$e_i^2 = \alpha_0 + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} + \varphi_2 X_{2i}^2 + \varphi_3 X_{3i}^2 + \cdots + \varphi_k X_{ki}^2 + \delta_{23} X_{2i} X_{3i} + \delta_{24} X_{2i} X_{4i} \cdots + \delta_{kl} X_{ki} X_{li} + \mu_i \qquad \forall i = 1,2,3 \cdots T$$

con la estimación de este modelo se obtiene el R2.

Ejemplo 18.4. De este modo, para contrastar un modelo con tres variables explicativas, a través de este test, se realiza la regresión

$$e_i^2 = \alpha_0 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \varphi_2 X_{2i}^2 + \varphi_3 X_{3i}^2 + \varphi_4 X_{4i}^2 + \delta_{23} X_{2i} X_{3i} + \delta_{24} X_{2i} X_{4i} + \delta_{34} X_{3i} X_{4i} + \mu_i \qquad \forall i = 1,2,3 \cdots T$$

---

3) Se construye el estadístico

$$\lambda = TR^2$$

donde R^2 es el coeficiente de determinación que se calcula a partir de la estimación realizada en el punto 2.

Contrastar la hipótesis nula de homocedasticidad, es equivalente a contrastar que todos los coeficientes de la regresión de los errores, exceptuando el intercepto, son conjuntamente cero, es decir:

$$H_0: \ \beta_k, \phi_k, \delta_{kl} = 0 \qquad \forall j, s$$

Se puede demostrar que bajo la hipótesis nula

$$\lambda \overset{a}{\sim} \chi^2(p)$$

donde p es el número de regresores en el modelo especificado en 2 sin incluir el término constante. Se rechaza la hipótesis nula si el valor muestral del estadístico excede el valor crítico de las tablas χ^2, elegido un nivel de significación.

Este contraste tiene la ventaja de ser muy flexible por no tener que especificar la hipótesis alternativa; pero si se rechaza la hipótesis nula de homocedasticidad no indica cual puede ser la dirección a seguir.

El contraste de White puede recoger otro tipo de problemas de mala especificación de la parte sistemática: omisión de variables relevantes, mala forma funcional, etc. Esto es correcto si se identifica

cuál es el problema; en caso contrario, la solución que se tome puede estar equivocada.

18.3. Autocorrelación

En el modelo de regresión, el término de perturbación engloba todos aquellos factores determinantes de la variable endógena que no están recogidos en la parte sistemática del modelo. Estos factores pueden ser innovaciones, errores de medida de la variable endógena, variable omitida, etc.

Si estos factores están correlacionados en el tiempo o en el espacio, entonces no se satisface la hipótesis

$$E(\varepsilon_i \varepsilon_j) = 0, \quad \forall i \neq j$$

Este fenómeno se conoce con el nombre de *autocorrelación* o *correlación serial*, en el caso de datos de series temporales, y de correlación espacial en el caso de datos de corte transversal.

En los modelos que se especifican relaciones en el tiempo entre variables, la propia inercia de las series económicas, donde el impacto de una perturbación en un período de tiempo puede tener efectos en subsiguientes períodos, puede generar autocorrelación en el término de perturbación.

Esta dinámica, aunque no sea relevante en media, refleja un patrón sistemático de comportamiento que hay que considerar a la hora de estimar el modelo. La matriz,

$$\mathrm{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \boldsymbol{\Omega} = \begin{bmatrix} \sigma_{\varepsilon}^2 & \sigma_{12} & \cdots & \sigma_{1T} \\ \sigma_{21} & \sigma_{\varepsilon}^2 & \cdots & \sigma_{2T} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{T1} & \sigma_{T2} & \cdots & \sigma_{\varepsilon}^2 \end{bmatrix}$$

tiene elementos fuera de la diagonal principal, distintos de cero.

En consecuencia, los coeficientes de regresión habrán de ser estimados por métodos de mínimos cuadrados generalizados.

Observación. Las perturbaciones aleatorias que están autocorrelacionadas se pueden modelar utilizando procesos estocásticos. Los más utilizados son los autorregresivos y de medias móviles, denominados ARMA (p,q).

Esta clase de procesos incluye, como casos particulares, los autorregresivos de orden **p** [**AR (p)]** y de medias móviles de orden **q** [**MA (q)].**

La forma general de un proceso **AR (p),** es:

$$\varepsilon_t = \rho_1\, \varepsilon_{t-1} + \rho_2\, \varepsilon_{t-2} + \cdots + \rho_p\, \varepsilon_{t-p} + \mu_t$$

Donde μ_t, se distribuye independientemente en el tiempo con media cero y varianza constante σ_{μ}^2 y $\rho_1, \cdots, \rho_p$, son parámetros constantes en el tiempo. El proceso autorregresivo

más utilizado, dentro del marco del análisis de regresión, es el proceso de orden uno, [**AR (1)]:**

$$\varepsilon_t = \rho\ \varepsilon_{t-1} + \mu_t$$

Donde la perturbación en un período t, depende de la perturbación en el período anterior $t-1$ y un término aleatorio o innovación μ_t que se supone ruido blanco; ser ruido blanco es tener media 0, varianza constante σ_μ^2 y covarianzas nulas. Si se sustituye repetidamente se obtiene que:

$$\varepsilon_t = \sum_{i=0}^{\infty} \rho^i\ \mu_{t-i}$$

La perturbación ε_t es una combinación lineal de las innovaciones pasadas μ_t con ponderaciones $\mathbf{1}, \boldsymbol{\rho}, \boldsymbol{\rho^2}$ ... que decaen geométricamente; si el valor del coeficiente ρ está acotado en el intervalo (-1, 1), las innovaciones μ_{t-i} tienen menor influencia en ε_t cuanto más alejadas están en el tiempo.

Es fácil comprobar que el vector de perturbaciones $\boldsymbol{\varepsilon}$, tiene media cero y matriz de varianzas y covarianzas:

$$E\left(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\right) = \frac{\sigma_\mu^2}{1-\rho^2}\begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{T-1} \\ \rho & 1 & \rho & \cdots & \rho^{T-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{T-3} \\ \vdots & \vdots & \vdots & & \vdots \\ \rho^{T-1} & \rho^{T-2} & \rho^{T-3} & \cdots & 1 \end{bmatrix} = \sigma_\varepsilon^2\ \hat{\Sigma}$$

De esta forma, dado el valor de ρ, la matriz **Ω** queda totalmente determinada, a excepción del factor de escala σ_ε^2 .

Para recoger efectos estacionales en la perturbación con un proceso autorregresivo de datos trimestrales **AR(4)**, se expresa de la siguiente manera

$$\varepsilon_t = \rho \ \varepsilon_{t-4} + \mu_t$$

El proceso de medias móviles general, **MA (q)**, es:

$$\varepsilon_t = \mu_t + \theta_1 \ \mu_{t-1} + \cdots + \theta_q \ \mu_{t-q}$$

Donde se supone que μ_t, es ruido blanco con media cero y varianzas σ_μ^2 y $\theta_1, \cdots \theta_q$ son parámetros constantes.

El proceso de medias móviles más sencillo es el **MA (1):**

$$\varepsilon_t = \mu_t + \theta \ \mu_{t-1}$$

A diferencia de los procesos autorregresivos, en el proceso **MA (1)** la perturbación ε_t es una combinación lineal de solo dos innovaciones μ_t y μ_{t-1} por lo que se dice que es un proceso de memoria corta. En este caso, el vector de perturbaciones tiene media cero y matriz de varianza y covarianza:

$$E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \sigma_\varepsilon^2 \begin{bmatrix} (1+\theta^2) & -\theta & 0 & \cdots & 0 \\ -\theta & (1+\theta^2) & -\theta & \cdots & 0 \\ 0 & -\theta & (1+\theta^2) & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & (1+\theta^2) \end{bmatrix} = \sigma_\varepsilon^2 \boldsymbol{\Sigma}$$

Por último, el modelo más general es el modelo autorregresivo de medias móviles, **ARMA (p, q)**, donde la perturbación ε_t depende de sus valores pasados y de la innovación μ_t y su pasado:

$$\varepsilon_t = \rho_1 \ \varepsilon_{t-1} + \rho_2 \ \varepsilon_{t-1} + \cdots + \rho_\rho \ \varepsilon_{t-\rho} + \mu_t + \theta_1 \ \mu_{t-1} + \cdots + \theta_q \ \mu_{t-q}$$

Cuando modelamos la dependencia en el tiempo de ε_t mediante un proceso **ARMA (p, q)**, estamos especificando la estructura de la matriz de varianza y covarianza **Ω** en términos de los parámetros $\sigma^2, \rho_1, \ldots, \rho_p, \theta_1, \ldots, \theta_q$.

La elección de un proceso **ARMA (p, q)** concreto, depende en cada caso, de las características de los datos y del estudio que se esté realizando. Para simplificar la explicación, a lo largo de este tema se supone que las perturbaciones siguen un proceso **AR (1)**.

Contraste de autocorrelación de Durbin–Watson

En la práctica, no se conoce a priori si existe autocorrelación ni cuál puede ser el proceso más adecuado para modelarla. Al igual que ocurre con con heterocedasticidad, el modelo se especifica y estima bajo el supuesto de perturbaciones no autocorrelacionadas.

Existen varios contrastes de autocorrelación que se construyen utilizando los residuos mínimo–cuadráticos ordinarios. Uno de ellos es

el de Durbin-Watson, que permite detectar la existencia de un proceso **AR(1)** en el término de perturbación.

La hipótesis nula es la no existencia de autocorrelación,

$$H_0: \rho = 0$$

El estadístico de contraste es:

$$DW = \frac{\sum_{t=2}^{T} (e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2}$$

donde e_t, son los residuos mínimo–cuadráticos ordinarios.

Si el número de observaciones es suficientemente grande, este estadístico se puede calcular mediante la aproximación:

$$DW \cong 2\,(1-\hat{\rho})$$

siendo $\hat{\rho}$ el coeficiente estimado por MCO en la regresión:

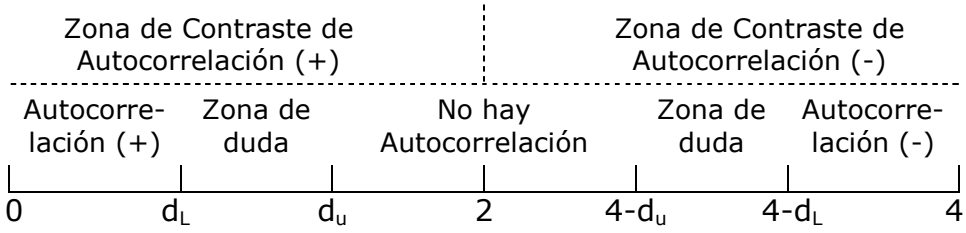
$$e_t = \rho\ e_{t-1} + \nu_t \qquad t = 2, \cdots, T$$

A partir de esta relación se puede establecer el rango de valores que puede tomar el estadístico:

1. $\hat{\rho} = 0$ $\quad DW \cong 2$
2. $-1 < \hat{\rho} < 0$ $\quad DW \in (4,2)$
3. $0 < \hat{\rho} < 1$ $\quad DW \in (2,0)$

Durbin y Watson tabularon los valores críticos, el máximo $\mathbf{d_u}$ y mínimo $\mathbf{d_L}$, que depende de la matriz de datos **X**. Estos valores críticos definen la zona de duda -donde no es posible afirmar o rechazar la existencia de autocorrelación-, las zonas de

autocorrelación positiva y negativa y la zona de no existencia de autocorrelación. La comparación del estadístico empírico **DW** con la escala teórica de variabilidad 0 a 4, donde se explicitan los valores críticos, permite concluir si se acepta o rechaza la hipótesis nula.



Este contraste se puede considerar también, como un contraste de mala especificación del modelo. La omisión de variables relevantes, una forma funcional poco adecuada, cambios estructurales no incluidos en el modelo, entre otros, pueden originar un estadístico **DW** significativo. Esto puede llevar a errores, si se considera que hay evidencia de autocorrelación y se modela con un proceso **AR (1)**. Por otro lado, si ε_t sigue un proceso distinto a un **AR (1)**, puede que la significatividad del estadístico **DW** se vea afectada.

Si los errores están correlacionados positivamente, los valores sucesivos tenderán a estar uno cerca de otro en el cuadrante I y las diferencias tenderán a ser numéricamente más pequeñas que los mismos residuos, como se observa en la Figura 18.2.

Si los errores están correlacionados negativamente, los valores sucesivos tenderán a estar unos en el cuadrante I y otros en el cuadrante IV; de modo que, las primeras diferencias tenderán a ser numéricamente mayores que los mismos residuos, como se observa en la Figura 18.3.

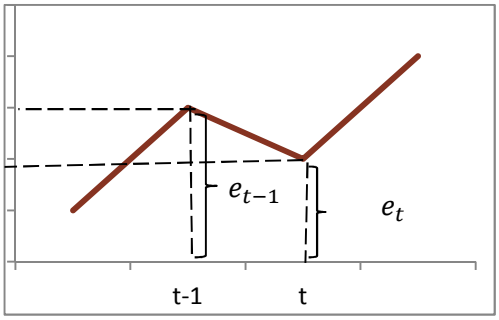


Figura 18.2 Autocorrelación positiva

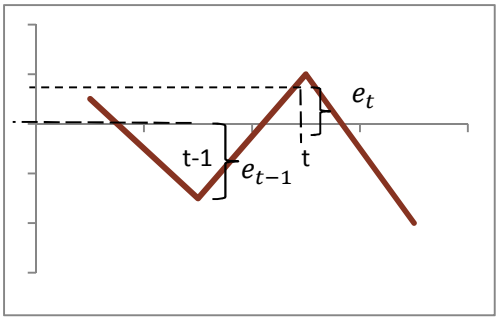


Figura 18.3 Autocorrelación negativa

En resumen, el estadístico de Durbin-Watson es útil porque indica la existencia de problemas en el modelo, pero no ayuda a establecer cuál es el modelo alternativo.

18.4. Estimación por Mínimos Cuadrados Generalizados

En el modelo lineal general, la matriz de varianzas y covarianzas del vector de perturbaciones $\boldsymbol{\varepsilon}$ es un escalar: $V(\boldsymbol{\varepsilon}) = \sigma_{\varepsilon}^2 \mathbf{I}_T$. Con este supuesto se indica la *Homocedasticidad* -varianza es constante para todas las perturbaciones, $E(\varepsilon_t^2) = \sigma_{\boldsymbol{\varepsilon}}^2,\ \forall\, t = 1,2,\ldots.,T$ – y la No *autocorrelación* -no existe *correlación* entre las perturbaciones de diferentes períodos, $E(\varepsilon_t, \varepsilon_s) = 0,\ \forall\, t \neq s$.

Este supuesto describe la información sobre las varianzas y covarianzas entre las perturbaciones que es proporcionada por las variables independientes. Es decir, por sí mismas, las perturbaciones no proporcionan dicha información.

Ejemplo 18.5. En la Figura se observan los residuos homocedásticos y no autocorrelacionados luego de reespecificar el modelo del Caso 17.1, para el estudio del consumo en la Argentina en función del PBI y de la tasa de interés.

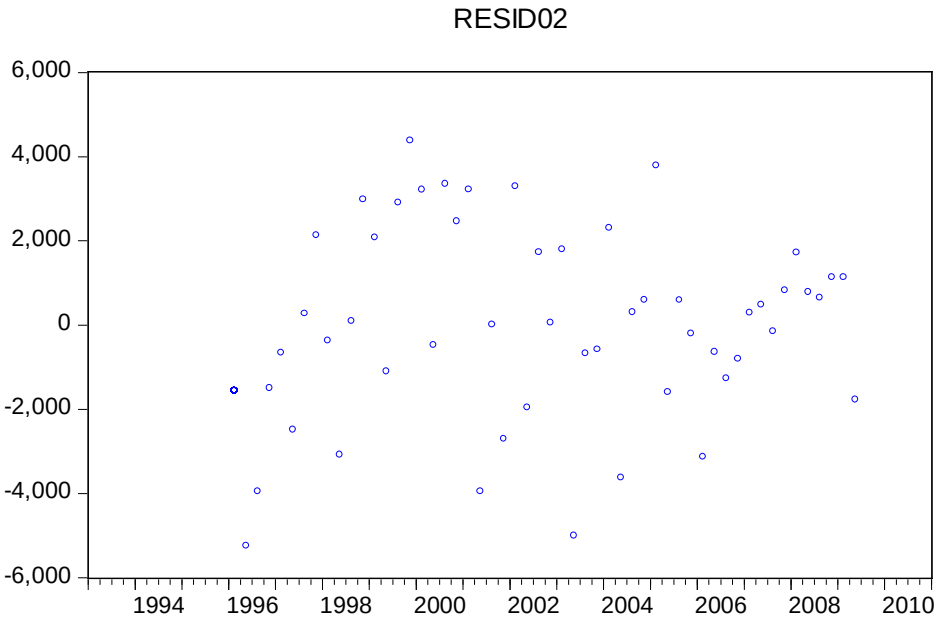


Figura 18.4. Residuos homocedásticos y no autocorrelacionados

---

La restricción de $\mathrm{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \sigma^2 \mathbf{I}_\mathrm{T}$ se puede flexibilizar para recoger situaciones más generales, donde las varianzas de las perturbaciones son distintas y/o las covarianzas no nulas. Si no se imponen

restricciones a priori, la forma general de la matriz de varianzas y covarianzas de las perturbaciones es:

$$E\left(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\right)=\begin{bmatrix}\sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1T}\\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2T}\\ \cdots & \cdots & \cdots & \cdots\\ \sigma_{T1} & \sigma_{T2} & \cdots & \sigma_T^2\end{bmatrix}=\boldsymbol{\Omega}$$

Esto conduce a trabajar dentro del marco más general del modelo de regresión lineal con *matrices de varianzas y covarianzas no escalares,*

$$E\left(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\right)=\boldsymbol{\Omega}$$

que, en la literatura econométrica, se suele denominar *modelo de regresión lineal generalizado*.

En primer lugar, se analiza qué consecuencias tiene sobre los estimadores MCO de los coeficientes de regresión, flexibilizar el supuesto de *perturbaciones esféricas*.

Seguidamente, se introduce un método de estimación alternativo al MCO que tendrá en cuenta la información que recoge la matriz de covarianzas $\boldsymbol{\Omega}$ Este método se conoce con el nombre de *mínimos cuadrados generalizados*, MCG.

En el caso particular de que $\boldsymbol{\Omega}=\sigma^2\mathbf{I}_T$, ambos métodos de estimación coinciden.

Observación. Sea el modelo de regresión lineal generalizado siguiente:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

Donde $E(\boldsymbol{\varepsilon}) = \mathbf{0}$, $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \boldsymbol{\Omega}$ y $\mathbf{X}$, es una matriz no estocástica de rango k.

Bajo los supuestos del modelo, el estimador MCO de $\boldsymbol{\beta}$, es lineal e insesgado con matriz de varianzas y covarianzas dada por,

$$V(\widehat{\boldsymbol{\beta}}_{MCO}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$$

Se puede demostrar que si la matriz de covarianzas de las perturbaciones no es escalar, el estimador habitual de la matriz de varianza $V(\widehat{\boldsymbol{\beta}}_{MCO}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$, es un estimador sesgado de la misma.

Esto tiene graves consecuencias a la hora de realizar contrastes de hipótesis sobre el vector de coeficientes $\boldsymbol{\beta}$, porque los estadísticos habituales, no se distribuyen como una F de Snedecor, ni como una t de Student, de forma que si se compara el valor del estadístico muestral con el correspondiente a esas distribuciones, se puede llegar a una mala elección de la región crítica y a conclusiones erróneas.

Por otro lado, el estimador MCO de $\boldsymbol{\beta}$, es óptimo si se cumplen todos los *supuestos básicos* del modelo de regresión lineal. Al relajar uno de los supuestos, no se puede aplicar el teorema de GAUSS-MARKOV y nada nos garantiza que el estimador MCO de $\boldsymbol{\beta}$, del modelo especificado, sea el de menor varianza dentro de la clase de estimadores lineales e insesgados. Intuitivamente,

es razonable pensar que podemos obtener un estimador más eficiente incorporando la nueva información que tenemos en el modelo a través de la matriz de covarianzas no escalar $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \boldsymbol{\Omega}$ y que no es tenida en cuenta por el método de MCO.

El método de estimación de mínimos cuadrados generalizados se basa en el criterio de estimación mínimo cuadrática, pero la función de distancia a minimizar es distinta a la de este criterio, ya que incorpora la información adicional, en la matriz de varianzas y covarianzas, de las perturbaciones $\boldsymbol{\Omega}$.

La función objetivo a minimizar viene dada por

$$\underbrace{Min}_{\hat{\beta}} \left(\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}\right)' \boldsymbol{\Omega}^{-1} \left(\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}\right)$$

o equivalentemente, si se escribe $\boldsymbol{\Omega} = \sigma^2 \boldsymbol{\Sigma}$, donde $\boldsymbol{\Sigma}$ es conocida y σ^2 es un factor de escala

$$\underbrace{Min}_{\hat{\beta}} \left(\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}\right)' \sigma^2 \boldsymbol{\Sigma}^{-1} \left(\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}\right)$$

A lo largo de este capítulo se trabaja indistintamente con la matriz $\boldsymbol{\Omega}$ o $\sigma^2\boldsymbol{\Sigma}$.

El factor de escala σ^2 no es relevante a la hora de minimizar la suma de cuadrados ponderada con respecto a $\boldsymbol{\beta}$. Lo que si es relevante, es la información incorporada en $\boldsymbol{\Sigma}$.

En el criterio de mínimos cuadrados ordinarios, la función objetivo consta únicamente de la suma de cuadrados de las desviaciones

$(\mathbf{Y}-\mathbf{X}\widehat{\boldsymbol{\beta}})$**.** En la "nueva" función objetivo aparece, como matriz de ponderaciones, la inversa de $\boldsymbol{\Sigma}$; incluyendo, de esta manera, la información existente sobre la dispersión y correlación de las desviaciones $(\mathbf{Y}-\mathbf{X}\widehat{\boldsymbol{\beta}})$.

De las condiciones de primer orden del problema de minimización,

$$\frac{\partial}{\partial\widehat{\beta}}\left\{(\mathbf{Y}-\mathbf{X}\widehat{\boldsymbol{\beta}})'\sigma^2\boldsymbol{\Sigma}^{-1}(\mathbf{Y}-\mathbf{X}\widehat{\boldsymbol{\beta}})\right\}=0$$

$$\sigma^2\frac{\partial}{\partial\widehat{\beta}}\left\{(\mathbf{Y}-\mathbf{X}\widehat{\boldsymbol{\beta}})'(\boldsymbol{\Sigma}^{-1}\mathbf{Y}-\boldsymbol{\Sigma}^{-1}\mathbf{X}\widehat{\boldsymbol{\beta}})\right\}=0$$

$$\sigma^2\frac{\partial}{\partial\widehat{\beta}}\left\{\mathbf{Y}'\boldsymbol{\Sigma}^{-1}\mathbf{Y}\underbrace{-\widehat{\boldsymbol{\beta}}'\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{Y}-\mathbf{Y}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\widehat{\boldsymbol{\beta}}}_{-2\widehat{\boldsymbol{\beta}}'\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{Y}}+\widehat{\boldsymbol{\beta}}'\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\widehat{\boldsymbol{\beta}}\right\}=0$$

$$\left\{-2\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{Y}+\mathbf{2}\left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)\widehat{\boldsymbol{\beta}}\right\}=0$$

se obtiene el sistema de k ecuaciones normales:

$$(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})\widehat{\boldsymbol{\beta}}_{\mathrm{MCG}}=\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{Y}$$

Cuya solución, es el estimador de *mínimos cuadrados generalizados*

$$\widehat{\boldsymbol{\beta}}_{\mathrm{MCG}}=(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-\mathbf{1}}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{Y}$$

Se puede demostrar que el estimador MCG de $\boldsymbol{\beta}$, es lineal, insesgado y óptimo dentro del marco del modelo de regresión lineal generalizado.

Este resultado, se conoce con el nombre de *Teorema de AITKEN* y es una generalización del *Teorema de GAUSS-MARKOV.*

La matriz de varianzas y covarianzas del estimador $\hat{\boldsymbol{\beta}}_{\mathbf{MCG}}$, es

$$\mathbf{V}\left(\widehat{\boldsymbol{\beta}}_{\mathrm{MCG}}\right) = \sigma^2(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-\mathbf{1}}$$

Un estimador insesgado de la matriz de varianzas y covarianzas viene dado por

$$\widehat{\mathbf{V}}\left(\widehat{\boldsymbol{\beta}}_{\mathrm{MCG}}\right) = \widehat{\sigma}^2_{\mathrm{MCG}}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-\mathbf{1}}$$

Donde el estimador insesgado del factor de escala σ^2, es

$$\widehat{\sigma}^2_{\mathrm{MCG}} = \frac{\left(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}_{\mathrm{MCG}}\right)' \boldsymbol{\Sigma}^{-1}\left(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}_{\mathrm{MCG}}\right)}{\mathrm{T} - \mathrm{k}}$$

Si suponemos que la perturbación ε sigue una distribución normal, se puede obtener la siguiente distribución para el estimador MCG:

$$\widehat{\boldsymbol{\beta}}_{\mathrm{MCG}} \sim \mathrm{N}\left(\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-\mathbf{1}}\right)$$

por lo que podemos contrastar restricciones lineales sobre los coeficientes del tipo $H_0 : \mathbf{R}\boldsymbol{\beta}$ con el estadístico:

$$\mathrm{F} = \frac{\left(\mathbf{R}\widehat{\boldsymbol{\beta}}_{\mathrm{MCG}} - \mathbf{r}\right)'(\mathbf{R}(\mathbf{X}'\,\boldsymbol{\Sigma}^{-1}\,\mathbf{X})^{-\mathbf{1}}\mathbf{R}')^{-\mathbf{1}}\left(\mathbf{R}\widehat{\boldsymbol{\beta}}_{\mathrm{MCG}} - \mathbf{r}\right) / \mathrm{q}}{\widehat{\sigma}^2_{\mathrm{MCG}}} \sim \mathrm{F}\left(\mathrm{q}, \mathrm{T} - \mathrm{k}\right)$$

Siguiendo las reglas de decisión habituales.

El estimador $\hat{\boldsymbol{\beta}}_{\mathbf{MCG}}$ es función de $\boldsymbol{\Sigma}$ y, por lo tanto, para obtenerlo es preciso conocer esta matriz de varianza y covarianzas.

Se puede demostrar fácilmente que si $\boldsymbol{\Sigma} = \mathbf{I}_{\mathrm{T}}$, el estimador $\hat{\boldsymbol{\beta}}_{\mathbf{MCG}}$ es igual al estimador $\hat{\boldsymbol{\beta}}_{\mathbf{MCO}}$. ¿Por qué?

18.5. Estimación bajo Heterocedasticidad

Cuando no se conocen los elementos de $\boldsymbol{\Omega}$, no es posible estimar **T** varianzas más **k** coeficientes de regresión con solo **T** observaciones.

Una forma de abordar el problema es "modelar" las varianzas de las perturbaciones en función de un vector **(sx1)** de variables que son observables, $\mathbf{z}_i$ (que pueden ser parte o no del conjunto de regresores), y de un vector de parámetro $\boldsymbol{\theta}$, cuya dimensión es estimable y no crece con el tamaño muestral:

$$\sigma_i^2 = G(\mathbf{z}_i, \boldsymbol{\theta}), \ \forall i$$

de forma que $\boldsymbol{\Omega} = \boldsymbol{\Omega}(\boldsymbol{\theta})$

Una vez obtenido un estimador $\hat{\boldsymbol{\theta}}$, se puede definir un estimador $\hat{\boldsymbol{\Omega}} = \boldsymbol{\Omega}(\hat{\boldsymbol{\theta}})$ y estimar el vector de coeficientes $\boldsymbol{\beta}$ por el método de mínimos cuadrados generalizados factibles.

Bajo ciertas condiciones de regularidad, se sabe que si el estimador $\hat{\boldsymbol{\Omega}}$ es consistente, el estimador $\hat{\boldsymbol{\beta}}_{\mathbf{MCGF}}$ tiene buenas propiedades asintóticas.

Por lo tanto, una primera etapa para obtener el estimador MCGF de $\boldsymbol{\beta}$ se basa en obtener un estimador consistente de $\boldsymbol{\theta}$.

Una forma de conseguirlo, es considerar la siguiente aproximación del residuo mínimo-cuadrático con la perturbación:

$$e_i = Y_i - \mathbf{x}_i'\widehat{\boldsymbol{\beta}}_{MCO}$$

Restando y sumando $\mathbf{x}_i'\boldsymbol{\beta}$

$$e_i = Y_i - \mathbf{x}_i'\boldsymbol{\beta} + \mathbf{x}_i'\boldsymbol{\beta} - \mathbf{x}_i'\widehat{\boldsymbol{\beta}}_{MCO}$$

Teniendo en cuenta que

$$Y_i - \mathbf{x}_i'\boldsymbol{\beta} = \varepsilon_i$$

$$e_i = \varepsilon_i + \mathbf{x}_i'\left(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}_{MCO}\right) = \varepsilon_i + \text{sesgo de estimación}$$

Dado que

$$E\left(\varepsilon_i^2\right) = \sigma_i^2 = G(\mathbf{Z}_i, \boldsymbol{\theta}),$$

Se tiene que,

$$e_i^2 = G(\mathbf{Z}_i, \boldsymbol{\theta}) + \text{sesgo de estimación}$$

Si $G(\mathbf{Z}_i, \boldsymbol{\theta})$ es lineal en $\boldsymbol{\theta}$, por ejemplo $\hat{\boldsymbol{\theta}}'\mathbf{Z}_i$, se puede considerar la siguiente regresión para estimar los parámetros $\boldsymbol{\theta}$:

$$e_i^2 = \boldsymbol{\theta}'\mathbf{Z}_i + \nu_i \qquad i = 1, \cdots, T$$

En esta regresión, el término de perturbación es una combinación de los errores acumulados en las aproximaciones.

Se puede demostrar que, bajo ciertas condiciones, el estimador de $\boldsymbol{\theta}$ así derivado, es consistente.

Una vez obtenido un estimador consistente de $\boldsymbol{\theta}$, se sustituye en la función suma de cuadrados ponderada y se minimiza con respecto a

$\boldsymbol{\beta}$, obteniéndose un estimador mínimo cuadrado generalizado factible (MCGF).

Mínimos cuadrados generalizados ponderados

Existen casos en los que es posible conocer la estructura de la matriz de varianzas y covarianzas $\boldsymbol{\Omega}$.

Ejemplo 18.6. En los casos de agregación de datos de corte transversal o temporal. Si se considera como observaciones en el modelo de regresión las medias de datos agrupados, la varianza de la perturbación en el modelo de regresión dependerá inversamente del número de observaciones en cada grupo T_i esto es $\sigma_i^2 = \sigma^2 (T_i)^{-1}$. Si en lugar de las medias se considera simplemente la suma de las observaciones en cada grupo, la varianza de la perturbación es proporcional al número de observaciones en cada grupo $\sigma_i^2 = \sigma^2 \ T_i$.

El vector de coeficientes $\boldsymbol{\beta}$ se puede estimar por MCG resolviendo el problema de minimización que, para el problema de heterocedasticidad, toma la forma:

$$\underset{\beta}{Min}\ (\mathbf{Y}-\mathbf{X\beta})'\,\mathbf{\Omega^{-1}}\,(\mathbf{Y}-\mathbf{X\beta})=\sum_{i=1}^{T}\frac{\left(Y_i-\mathbf{X_i'\beta}\right)^2}{\sigma_i^2}$$

En la suma de cuadrados, se ponderan más las desviaciones $\left(Y_i-\mathbf{X_i'\beta}\right)$ con menor varianza que las de mayor varianza, por ello, también se conoce este método como de *mínimos cuadrados ponderados.*

En el caso de heterocedasticidad, la matriz $\mathbf{\Omega^{-1}}$ es diagonal

$$\mathbf{\Omega^{-1}}=\ diag\left(\sigma_1^{-2},\ \ \sigma_2^{-2},\ \ \cdots,\ \ \sigma_T^{-2}\right)$$

entonces el estimador MCG se puede obtener también estimando por MCO el modelo transformado

$$\frac{Y_i}{\sigma_i}=\frac{\beta_1}{\sigma_i}+\beta_2\ \frac{X_{2i}}{\sigma_i}+\cdots+\beta_k\ \frac{X_{ki}}{\sigma_i}+\frac{u_i}{\sigma_i}\qquad i=1,\ 2,\ \cdots,\ T$$

$$Y_i^*=\beta_1\ X_{1i}^*+\beta_2\ \ X_{2i}^*+\cdots+\beta_k\ \ X_{ki}^*+u_i^*\qquad i=1,\ 2,\ \cdots,\ T$$

Donde

$$E\left(u_i^*\right)=0$$
$$E\left(u_i^*\right)^2=\frac{E(u_i)^2}{\sigma_i^2}=\frac{\sigma_i^2}{\sigma_i^2}=1,\quad \forall i$$
$$E\left(u_i^*\ u_j^*\right)=0,\qquad \forall i\neq j$$

De esta forma se satisfacen todas las condiciones para que el estimador MCO del vector $\boldsymbol{\beta}$ en el modelo sea un estimador ELIO. Ahora bien, este estimador no es más que el estimador MCG:

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

$$\widehat{\boldsymbol{\beta}}_{\mathrm{MCG}}=(\mathbf{X}^{*\prime}\mathbf{X}^{*})^{-1}(\mathbf{X}^{*\prime}\mathbf{Y}^{*})=(\mathbf{X}'\mathbf{\Omega^{-1}}\mathbf{X})^{-1}(\mathbf{X}'\mathbf{\Omega^{-1}}\mathbf{Y})=(\mathbf{X}'\mathbf{\Sigma^{-1}}\mathbf{X})^{-1}(\mathbf{X}'\mathbf{\Sigma^{-1}}\mathbf{Y})$$

El problema que se presenta ahora es cómo obtener la matriz $\boldsymbol{\Sigma}$. Para ello es necesario premultiplicar por una matriz $\mathbf{P}$, de dimensiones TxT, en el modelo $\mathbf{Y} = \mathbf{X}\beta + \varepsilon$

$$\mathbf{PY} = \mathbf{PX}\boldsymbol{\beta} + \mathbf{P}\varepsilon \qquad (1)$$

Se define $\mathbf{Y}^* = \mathbf{PY}$, $\mathbf{X}^* = \mathbf{PX}$, $\boldsymbol{\varepsilon}^* = \mathbf{P}\varepsilon$

La transformación es, en realidad, un cambio de variable. La varianza del error es

$$\text{Var}(\boldsymbol{\varepsilon}^*) = \text{Var}(\mathbf{P}\boldsymbol{\varepsilon}) = \sigma_\varepsilon^2 \mathbf{P}\boldsymbol{\Sigma}\mathbf{P}' \qquad (2)$$

Ahora bien ¿existe una matriz **P** tal que el término de error del modelo transformado tenga como matriz de covarianzas $\text{Var}(\boldsymbol{\varepsilon}^*) = \sigma_\varepsilon^2 \mathbf{I}_T$?

$\boldsymbol{\Sigma}$ es simétrica definida positiva, cuando esto pasa siempre existe una matriz cuadrada no singular **V** de modo que $\boldsymbol{\Sigma} = \mathbf{V}\mathbf{V}'$ lo cual da lugar a que

$$\mathbf{V}^{-1}\boldsymbol{\Sigma}\left(\mathbf{V}^{-1}\right)' = \mathbf{I}_T \qquad (3)$$

$$\boldsymbol{\Sigma}^{-1} = \mathbf{V}^{-1}\left(\mathbf{V}^{-1}\right)' \qquad (4)$$

De observar (2), (3) y (4) se deduce que si se utiliza la matriz $\mathbf{V}'$ como matriz de transformación **P**, entonces el término de error tiene matriz de covarianza escalar.

La estimación de un modelo transformado

$$\mathbf{Y}^* = \mathbf{X}^*\boldsymbol{\beta} + \boldsymbol{\varepsilon}^* \qquad (5)$$

$$\text{Donde} \quad \begin{aligned} \mathbf{Y}^* &= \mathbf{V}^{-1}\mathbf{Y} \\ \mathbf{X}^* &= \mathbf{V}^{-1}\mathbf{X} \\ \boldsymbol{\varepsilon}^* &= \mathbf{V}^{-1}\boldsymbol{\varepsilon} \end{aligned}$$

Se tiene que $\mathrm{E}(\boldsymbol{\varepsilon}^*) = \mathrm{E}(\mathbf{V}^{-1}\boldsymbol{\varepsilon}) = \mathbf{0}_{\mathrm{T}}$

$$Var(\boldsymbol{\varepsilon}^*) = \sigma_\varepsilon^2 \mathbf{V}^{-1}\boldsymbol{\Sigma}\left(\mathbf{V}^{-1}\right)' = \sigma_\varepsilon^2 \mathbf{I}_T$$

Por lo que el estimador de mínimos cuadrados ordinarios del modelo (5) es

$$\hat{\boldsymbol{\beta}}_{MCG} = \left[\mathbf{X}'(\mathbf{V}^{-1})'(\mathbf{V}^{-1})\mathbf{X}\right]^{-1}\mathbf{X}'(\mathbf{V}^{-1})'(\mathbf{V}^{-1})\mathbf{Y} = \left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{Y}$$

Observación. Otra forma de derivar la función criterio y obtener el estimador MCG, se basa en transformar el modelo de forma que la matriz de varianzas y covarianzas de sus perturbaciones, sea escalar. Dado que $\boldsymbol{\Sigma}^{-1}$ es una matriz simétrica y definida positiva, existe una matriz $\mathbf{P}$ no singular tal que $\boldsymbol{\Sigma}^{-1} = \mathbf{P}'\mathbf{P}$. Por lo tanto $\mathbf{P}'\boldsymbol{\Sigma}^{-1}\mathbf{P} = \mathbf{I}_{\mathrm{T}}$. Este resultado sugiere la siguiente transformación del modelo original:

$$\mathbf{Py} = \mathbf{PX}\boldsymbol{\beta} + \mathbf{P}\boldsymbol{\varepsilon}$$

Donde $E(\mathbf{P}\boldsymbol{\varepsilon}) = \mathbf{0}$ *y* $E\left(\mathbf{P}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\mathbf{P}'\right) = \sigma^2\mathbf{I}_T$. El modelo transformado, satisface todas las hipótesis básicas, por lo que será ELIO. La función objetivo para el modelo es:

$$\underbrace{Min}_{\widehat{\beta}} (\mathbf{Py}-\mathbf{PX\beta})'(\mathbf{Py}-\mathbf{PX\beta}) = (\mathbf{y}-\mathbf{X\beta})'\mathbf{\Sigma^{-1}}(\mathbf{y}-\mathbf{X\beta})$$

La solución de las condiciones de primer orden de este problema de minimización es, de nuevo, el estimador de mínimos cuadrados generalizados

$$\widehat{\boldsymbol{\beta}}_{\mathrm{MCG}} = (\mathbf{X'P'PX})^{-1}(\mathbf{X'P'Py}) = \left(\mathbf{X'\Sigma^{-1}X}\right)^{-1}\left(\mathbf{X'\Sigma^{-1}y}\right)$$

Ahora bien, en la práctica ¿cómo opera la matriz $\mathbf{\Sigma}$? y ¿qué tiene dentro?. Dado el modelo $\mathbf{Y}=\mathbf{X\beta}+\boldsymbol{\varepsilon}$ donde $\boldsymbol{\varepsilon}$ tiene

$$\mathrm{Var}(\boldsymbol{\varepsilon}) = \begin{bmatrix} \sigma_1^2 & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & 0 & \cdots & 0 \\ 0 & 0 & \sigma_3^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \sigma_T^2 \end{bmatrix}$$

Bajo el supuesto de que $\mathrm{E}(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_j)=0$, la varianza de los errores se puede descomponer en el producto

$$\mathrm{Var}(\boldsymbol{\varepsilon}) = \begin{bmatrix} \sigma_1 & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2 & 0 & \cdots & 0 \\ 0 & 0 & \sigma_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \sigma_T \end{bmatrix}\begin{bmatrix} \sigma_1 & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2 & 0 & \cdots & 0 \\ 0 & 0 & \sigma_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \sigma_T \end{bmatrix} = \mathbf{VV'}$$

De donde

$$\mathbf{V}^{-1}=\begin{bmatrix} 1/\sigma_1 & 0 & 0 & \cdots & 0 \\ 0 & 1/\sigma_2 & 0 & \cdots & 0 \\ 0 & 0 & 1/\sigma_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1/\sigma_T \end{bmatrix}$$

Utilizando esta matriz $\mathbf{V}^{-1}$ para transformar las variables se tiene

$$\mathbf{Y^*}=\mathbf{V}^{-1}\mathbf{Y}=\begin{bmatrix} 1/\sigma_1 & 0 & 0 & \cdots & 0 \\ 0 & 1/\sigma_2 & 0 & \cdots & 0 \\ 0 & 0 & 1/\sigma_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1/\sigma_T \end{bmatrix}\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_T \end{bmatrix}=\begin{bmatrix} Y_1/\sigma_1 \\ Y_2/\sigma_2 \\ Y_3/\sigma_3 \\ \vdots \\ Y_T/\sigma_T \end{bmatrix}$$

$$\mathbf{X^*}=\mathbf{V}^{-1}\mathbf{X}=\begin{bmatrix} 1/\sigma_1 & 0 & 0 & \cdots & 0 \\ 0 & 1/\sigma_2 & 0 & \cdots & 0 \\ 0 & 0 & 1/\sigma_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1/\sigma_T \end{bmatrix}\begin{bmatrix} X_{11} & X_{21} & X_{31} & \cdots & X_{k1} \\ X_{12} & X_{22} & X_{32} & \cdots & X_{k2} \\ X_{13} & X_{23} & X_{33} & \cdots & X_{k3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ X_{1T} & X_{2T} & X_{3T} & \cdots & X_{kT} \end{bmatrix}$$

$$=\begin{bmatrix} X_{11}/\sigma_1 & X_{21}/\sigma_1 & X_{31}/\sigma_1 & \cdots & X_{k1}/\sigma_1 \\ X_{12}/\sigma_2 & X_{22}/\sigma_2 & X_{32}/\sigma_2 & \cdots & X_{k2}/\sigma_2 \\ X_{13}/\sigma_3 & X_{23}/\sigma_3 & X_{33}/\sigma_3 & \cdots & X_{k3}/\sigma_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ X_{1T}/\sigma_T & X_{2T}/\sigma_T & X_{3T}/\sigma_T & \cdots & X_{kT}/\sigma_T \end{bmatrix}$$

Lo que se hace es ponderar cada observación con un peso

$1/\sigma_t \quad \forall t=1,2,3\cdots T$

este peso será menor cuanto mayor sea la varianza del término de error en dicha observación. Por esta razón, la aplicación de mínimos cuadrados generalizados con una matriz $\mathbf{\Sigma}$ -del tipo considerado- se lo

conoce como mínimos cuadrados ponderados. El problema es que la matriz $\boldsymbol{\Sigma}$ no se conoce y por lo tanto tampoco es posible conocer los ponderadores $1/\sigma_t$.

En la práctica, lo que se puede realizar es estimar $1/\sigma_t$ a través de los errores mínimo cuadráticos ordinarios siguiendo el procedimiento que se detalla:

1) Dado el modelo

   $$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

   donde $E(\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_j) = \sigma_t^2 \mathbf{I}_T$

   se calcula $\hat{\boldsymbol{\beta}}_{MCO} = [\mathbf{X}'\mathbf{X}]^{-1}\mathbf{X}'\mathbf{Y}$

   y se obtiene $\mathbf{e} = \mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{MCO}$

2) Con el vector de errores se calcula la matriz diagonal de varianzas de las perturbaciones

   $$\hat{\mathbf{V}}\hat{\mathbf{V}}' = \begin{bmatrix} e_1^2 & 0 & \cdots & 0 \\ 0 & e_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & e_T^2 \end{bmatrix}$$

   Obteniéndose la matriz $\hat{\mathbf{V}}^{-1}$

   $$\hat{\mathbf{V}}^{-1} = \begin{bmatrix} 1/e_1^2 & 0 & \cdots & 0 \\ 0 & 1/e_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1/e_T^2 \end{bmatrix}$$

3) Con la matriz $\hat{\mathbf{V}}^{-1}$ se transforman las variables

$$\mathbf{Y}^* = \mathbf{V}^{-1}\mathbf{Y}$$

$$\mathbf{X}^* = \mathbf{V}^{-1}\mathbf{X}$$

4) Realizando la estimación sobre las variables transformadas, se obtiene el estimador mínimo cuadrado generalizado (MCG)

$$\hat{\boldsymbol{\beta}}_{MCG} = \left(\mathbf{X}^{*\prime}\mathbf{X}^*\right)^{-1}\left(\mathbf{X}^{*\prime}\mathbf{Y}^*\right)$$

Estimador de White

Al estimar los coeficientes de regresión $\boldsymbol{\beta}$ por MCO en presencia de heterocedasticidad, da por resultado estimadores insesgados pero no eficientes.

Además, el estimador de la matriz de varianza y covarianza de $\hat{\boldsymbol{\beta}}_{\mathbf{MCO}}$, $\hat{\sigma}^2\left(\mathbf{X}'\mathbf{X}\right)^{-1}$, es inconsistente, por lo que los estadísticos de contraste habituales no son válidos para hacer inferencia sobre $\boldsymbol{\beta}$, ni siquiera para muestras grandes.

En los apartados anteriores se ha analizado cómo, para aplicar métodos de estimación más apropiados, es preciso conocer la matriz $\boldsymbol{\Omega}$, o al menos, cuál es la estructura de la heterocedasticidad para poder especificar $\boldsymbol{\Omega} = \boldsymbol{\Omega}(\boldsymbol{\theta})$.

Dada la dificultad existente para conocer la forma de $\boldsymbol{\Omega}$, es interesante contar con una estimación consistente de $\mathbf{v}\left(\hat{\boldsymbol{\beta}}_{\mathbf{MCO}}\right)$ y de esta forma derivar estadísticos válidos, al menos asintóticamente, para contrastar hipótesis sobre el vector de coeficientes $\boldsymbol{\beta}$.

White (1980) demuestra que es posible obtener un estimador consistente de la matriz de varianzas y covarianzas de $\hat{\boldsymbol{\beta}}_{\mathbf{MCO}}$, sin tener que hacer supuestos sobre $\boldsymbol{\Omega}$, a excepción del que indica que es una matriz diagonal.

Para ello, sólo es necesario obtener un estimador consistente de $\left(\mathbf{X}'\boldsymbol{\Omega}\mathbf{X}\right)$. White demuestra que, bajo ciertas condiciones de regularidad y siendo e_i, $i=1, \cdots, T$ el residuo mínimo-cuadrático ordinario

$$plim\ \frac{\mathbf{X}'\mathbf{S}\mathbf{X}}{\mathrm{T}} = plim\ \frac{\mathbf{X}'\boldsymbol{\Omega}\mathbf{X}}{\mathrm{T}}$$

Donde $\mathbf{S} = diag\ (e_1^2, e_2^2, \ldots, e_T^2)$

Por lo tanto, se puede utilizar:

$$\hat{\mathbf{V}}_{\mathbf{WHITE}} = T\left(\mathbf{X}'\mathbf{X}\right)^{-1}\left(\mathbf{X}'\,\mathbf{S}\,\mathbf{X}\right)\left(\mathbf{X}'\mathbf{X}\right)^{-1}$$

Como un estimador consistente de la matriz de varianzas y covarianzas asintóticas de $\sqrt{T}\ \hat{\boldsymbol{\beta}}_{\mathbf{MCO}}$.

Este resultado es muy importante, ya que si se estima por mínimos cuadrados ordinarios en presencia de heterocedasticidad y se utiliza este estimador de la matriz de covarianzas, es posible realizar inferencia válida sobre los coeficientes $\boldsymbol{\beta}$, al menos para muestras grandes, en base al siguiente resultado:

$$T\left(\mathbf{R}\hat{\boldsymbol{\beta}}_{\mathbf{MCO}} - \mathbf{r}\right)'\left(\mathbf{R}\ \hat{\mathbf{V}}_{\mathbf{WHITE}}\mathbf{R}'\right)^{-1}\left(\mathbf{R}\,\hat{\boldsymbol{\beta}}_{\mathbf{MCO}} - \mathbf{r}\right) \xrightarrow{d} \chi^2(q)$$

Sin tener que especificar a priori la estructura de la heterocedasticidad.

18.6. Estimación bajo Autocorrelación

Dada $\mathbf{\Omega} = \begin{bmatrix} \sigma_{\varepsilon}^2 & \sigma_{12} & \cdots & \sigma_{1T} \\ \sigma_{21} & \sigma_{\varepsilon}^2 & \cdots & \sigma_{2T} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{T1} & \sigma_{T2} & \cdots & \sigma_{\varepsilon}^2 \end{bmatrix}$

Si no se conoce es necesario estimarla; esto significa estimar $T(T-1)/2$ covarianzas distintas con solo T observaciones, lo que no es factible.

Para poder estimar los elementos de $\mathbf{\Omega}$, es necesario especificar la autocorrelación de las perturbaciones en términos de un proceso que depende de un número pequeño y estimable de parámetros.

Bajo el supuesto de que el término de error sigue un proceso de autocorrelación regresivo de primer orden

$\varepsilon_t = \rho_1 \varepsilon_{t-1} + \mu_t$

Donde $|\rho| < 1$ y $E(\mu_t) = 0$ y $E(\mu_i \mu_j) = \sigma^2 \mathbf{I}_T$

Así, ε_2 se puede escribir como

$\varepsilon_2 = \rho \varepsilon_1 + \mu_2$

Al calcular la varianza:

$$V(\varepsilon_2) = \rho^2\ V(\varepsilon_1) + V(\mu_2)$$

$$\sigma_\varepsilon^2 = \rho^2\sigma_\varepsilon^2 + \sigma_\mu^2$$

De modo que $\sigma_\varepsilon^2 = \dfrac{\sigma_\mu^2}{1-\rho^2}$, la cual es constante para todo t.

La covarianza entre ε_1 y ε_2 será:

$$\text{Cov}(1,2) = \sigma_{12} = \text{E}(\varepsilon_1\varepsilon_2) = \text{E}(\varepsilon_1(\rho\varepsilon_1 + \mu_2)) = \text{E}(\rho\varepsilon_1\varepsilon_1 + \varepsilon_1\mu_2) = \text{E}(\rho\varepsilon_1^2 + \varepsilon_1\mu_2)$$
$$= \rho\text{E}(\varepsilon_1^2) + \text{E}(\varepsilon_1\mu_2) = \rho\sigma_\varepsilon^2$$

Siendo $\text{E}(\varepsilon_1^2) = \sigma_\varepsilon^2$ y $\text{E}(\varepsilon_i\mu_j) = 0$

De la misma forma

$$\text{E}(\varepsilon_1\varepsilon_3) = \text{E}(\varepsilon_1(\rho\varepsilon_2 + \mu_3)) = \text{E}(\varepsilon_1(\rho(\rho\varepsilon_1 + \mu_2) + \mu_3)) =$$
$$\text{E}(\rho^2\varepsilon_1\varepsilon_1 + \rho\mu_2\varepsilon_1 + \varepsilon_1\mu_3) = \rho^2\text{E}(\varepsilon_1^2) + \rho\,\text{E}(\mu_2\varepsilon_1) + \text{E}(\varepsilon_1\mu_3) = \rho^2\sigma_\varepsilon^2$$

Extendiendo este razonamiento a la totalidad de observaciones se origina la matriz

$$\Omega = E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \sigma_\varepsilon^2\boldsymbol{\Sigma} = \sigma_\varepsilon^2\begin{bmatrix} 1 & \rho & \cdots & \rho^{(T-1)} \\ \rho & 1 & \cdots & \rho^{(T-2)} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{(T-1)} & \rho^{(T-2)} & \cdots & 1 \end{bmatrix}$$

En esta matriz solo se desconocen ρ y σ_ε^2. Es decir, un problema de estimar $T(T-1)/2$ elementos se ha reducido a la estimación de solo 2

Los elementos de $\boldsymbol{\Sigma}$ se pueden calcular a partir de

1) estimar por mínimos cuadrados ordinarios $\mathbf{y}=\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon}$ para obtener los residuos $\mathbf{e}$

2) estimar por mínimos cuadrados ordinarios $e_t = \rho e_{t-1} + \mu_t$ para obtener $\hat{\rho}$ y $\sum_{t=1}^{T}\hat{\mu}^2$ y calcular $\hat{\sigma}_\mu^2$ y $\hat{\sigma}_\varepsilon^2 = \dfrac{\hat{\sigma}_\mu^2}{1-\hat{\rho}^2}$.

3) Calcular $\sigma_\varepsilon^2\hat{\boldsymbol{\Sigma}}$ para obtener el estimador de mínimos cuadrados generalizados

$$\hat{\boldsymbol{\beta}}_{MCG} = \left(\mathbf{X}'\hat{\Sigma}^{-1}\mathbf{X}\right)^{-1}\left(\mathbf{X}'\hat{\Sigma}^{-1}\mathbf{Y}\right)$$

Es necesario aclarar que el estimador mínimo cuadrado generalizado es de mínima varianza siempre que $\boldsymbol{\Sigma}$ no se sustituya por su estimación. Como esto en la práctica es imposible, será válido el estimador mínimo cuadrado generalizado dependiendo de la *calidad de estimación* de $\hat{\boldsymbol{\Sigma}}$ y esta calidad aumenta asintóticamente, cuanto mayor es el número de observaciones.

Mínimos cuadrados generalizados factibles.

Bajo el supuesto de que las perturbaciones siguen un proceso autorregresivo de orden uno [**AR (1)]**, el modelo de regresión lineal generalizado, es:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \cdots + \beta_k X_{kt} + \varepsilon_t \quad t = 1, \cdots, T$$

$$\varepsilon_t = \rho \varepsilon_{t-1} + \mu_t \qquad \mu_t \sim NID\left(0, \sigma_\mu^2\right)$$

Si el valor de ρ es conocido, el estimador de mínimos cuadrados generalizados de $\boldsymbol{\beta}$ se obtiene minimizando la función criterio. En este caso, como $\boldsymbol{\Sigma}$ es una matriz simétrica y definida positiva, existe una matriz $\mathbf{P}$ tal que $\boldsymbol{\Sigma}^{-1} = \mathbf{P}\mathbf{P}'$, y el estimador de mínimos cuadrados generalizados obtiene estimando el modelo transformado por mínimos cuadrados ordinarios.

En el caso de un modelo AR (1), la matriz $\mathbf{P}$ es la siguiente:

$$P = \begin{bmatrix} \sqrt{1-\rho^2} & -\rho & 0 & \cdots & 0 & 0 \\ -\rho & 1 & -\rho & \cdots & 0 & 0 \\ 0 & -\rho & 1 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & -\rho & 1 \end{bmatrix}$$

y el modelo transformado se puede escribir como:

$$\sqrt{1-\rho^2}\, Y_1 = \beta_1\sqrt{1-\rho^2} + \beta_2\sqrt{1-\rho^2} X_{21} + \cdots + \beta_k\sqrt{1-\rho^2} X_{k1} + \varepsilon_1$$

$$Y_t - \rho Y_{t-1} = \beta_1(1-\rho) + \beta_2 (X_{2t} - \rho X_{2t-1}) + \cdots + \beta_k (X_{kt} - \rho X_{kt-1}) + \varepsilon_t$$

$$t = 2, \cdots, T$$

Es interesante señalar que la primera observación sufre una transformación diferente a todas las demás.

La suma de cuadrados a minimizar con respecto a β, es:

$$S(\beta) = (1-\rho^2)(Y_1 - \beta_1 - \beta_2 X_{2t} - \cdots - \beta_k X_{kt})^2 +$$
$$+ \sum_{t=2}^{T} \left\{ (Y_t - \rho\ Y_{t-1}) - \beta_1\ (1-\rho) - \sum_{j=2}^{k} \beta_j\ (X_{jt} - \rho\ X_{jt-1}) \right\}^2$$

El primer sumando proviene de la primera observación y el segundo, no es sino la suma de cuadrados de residuos del modelo transformado para $t = 2,\ldots,T$.

En el caso de que ρ sea desconocido, no se puede obtener el estimador de β por MCG directamente, sino que hay que estimar conjuntamente ρ y β.

Existen varios métodos que estiman conjuntamente ρ y β, basándose en el modelo transformado; de estos se desarrollan dos: el método **Durbin** y el método de **Cochranne-Orcutt**.

Ambos métodos de estimación se basan en que las perturbaciones siguen un proceso AR (1), por lo que el modelo transformado apropiado es

$$Y_t - \rho Y_{t-1} = \beta_1(1-\rho) + \beta_2\ (X_{2t} - \rho X_{2t-1}) + \cdots + \beta_k (X_{kt} - \rho X_{kt-1}) + \varepsilon_t$$
$$t = 2,\cdots,T$$

pero no tienen en cuenta la transformación de la primera observación.

Método de Durbin

La estimación por el *método de Durbin* (1960), se realiza en dos etapas:

1) Se estima ρ por mínimos cuadrados ordinarios En el modelo:

$$Y_t = \rho\ Y_{t-1} + \gamma_1 + \beta_2\ X_{2t} + \gamma_2\ X_{2,t-1} + \cdots + \beta_k\ X_{kt} + \gamma_k X_{k,t-1} + \varepsilon_t$$

Donde $t = 2, \cdots, T, \gamma_1 = \beta_1(1-\rho), \gamma_i = -\rho\ \beta_i, \quad i = 2, \cdots, k$.

Dadas las propiedades de ε_t el estimador de ρ por mínimos cuadrados ordinarios, $\hat{\rho}$, es consistente.

2) Se utiliza el estimador $\hat{\rho}$, para obtener el modelo transformado:

$$Y_t - \hat{\rho}\ Y_{t-1} = \beta_1(1-\hat{\rho}) + \beta_2(X_{2t} - \hat{\rho}\ X_{2t-1}) + \cdots + \beta_k(X_{kt} - \hat{\rho}\ X_{kt-1}) + V_t$$

y se estima el vector de coeficientes $\boldsymbol{\beta}$ por mínimos cuadrados ordinarios en este modelo; es decir, minimizando la suma de cuadrados con respecto a $\boldsymbol{\beta}$:

$$S^2(\beta) = \sum_{t=2}^{T} \left\{ (Y_t - \hat{\rho}\ Y_{t-1}) - \beta_1\ (1-\hat{\rho}) - \sum_{j=2}^{k} \beta_j\ (X_{jt} - \hat{\rho}\ X_{jt-1}) \right\}^2$$

Método de Cochrane-Orcutt

El *método de Cochrane-Orcutt* (1949) también se realiza en dos etapas:

1) Partiendo del supuesto que $\rho = 0$, se estima por MCO el modelo:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \cdots + \beta_k X_{kt} + u_t \qquad t = 1, \cdots, T$$

El estimador MCO de $\boldsymbol{\beta}$ es consistente.

En segundo lugar, se obtiene un estimador consistente de ρ, esto se logra estimando por MCO la regresión:

$$e_t = \rho\ e_{t-1} + \nu_t \qquad t = 2, \cdots, T$$

2) Se utiliza $\hat{\rho}$ para obtener el modelo transformado:

$$Y_t - \hat{\rho}\ Y_{t-1} = \beta_1(1-\hat{\rho}) + \beta_2(X_{2t} - \hat{\rho}\, X_{2t-1}) + \cdots + \beta_k(X_{kt} - \hat{\rho}\, X_{kt-1}) + \nu_t$$

y se estima $\boldsymbol{\beta}$ por MCO en este modelo minimizando la suma de cuadrados

$$S(\beta) = \sum_{t=2}^{T} \left\{ (Y_t - \hat{\rho}\ Y_{t-1}) - \beta_1\ (1-\hat{\rho}) - \sum_{j=2}^{k} \beta_j\ (X_{jt} - \hat{\rho}\ X_{jt-1}) \right\}^2$$

Este proceso en dos etapas, se suele realizar repitiendo las regresiones hasta que las estimaciones de ρ y $\boldsymbol{\beta}$, no varíen dentro de un margen de valores.

Es preciso tener en cuenta que los dos métodos considerados minimizan la suma de cuadrados, que no tienen en cuenta la primera observación, por lo que solo son aproximaciones al estimador de

mínimos cuadrados generalizados factibles. Asintóticamente, ambos son equivalentes al estimador MCGF, pero para muestras pequeñas puede haber diferencias, a veces importantes.

CASOS DE ESTUDIO, PREGUNTAS Y PROBLEMAS

Problema 18.1: Heterocedasticidad en series de datos de corte transversal

En el modelo estimado a partir de la Tabla 16.1, contraste las hipótesis de homocedasticidad.

Problema 18.2: Contrastes sobre la perturbación aleatoria

En el modelo estimado a partir de la Tabla 16.4, contraste las hipótesis de homocedasticidad, no autocorrelación y normalidad.

Problema 18.3: Especificación y Estimación de modelos lineales

Especifique un modelo para estudiar una temática económica de su interés, construya la tabla de datos, realice la estimación y contraste la validez de los supuestos.

BIBLIOGRAFIA

- Gujarati, Damodar. *Econometría*. México: Mc.Graw Hill, 2004.
- Johnston, J. y Dinardo, J. . *Métodos De Econometría.* Barcelona: Editorial Vicens Vives, 2001.
- Novales, A. *Econometría*. Madrid: McGraw Hill, 1993.
- Pulido San Román, Antonio. *Modelos Econométricos*. Madrid: Editorial Pirámide, 1993.
- Schmidt, S. J. . "Econometría," México: Mc.Graw Hill., 2005.