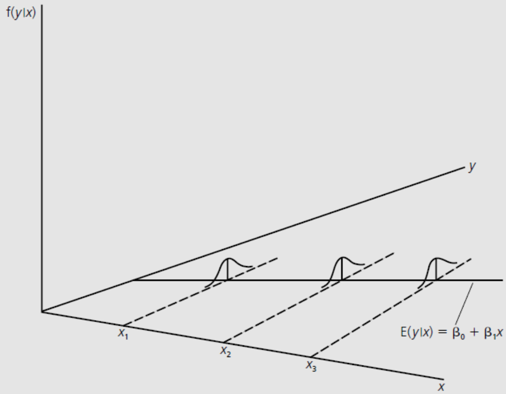
Estadística para la Economía y la Gestión IN 3401 Clase 7

Perturbaciones no Esféricas

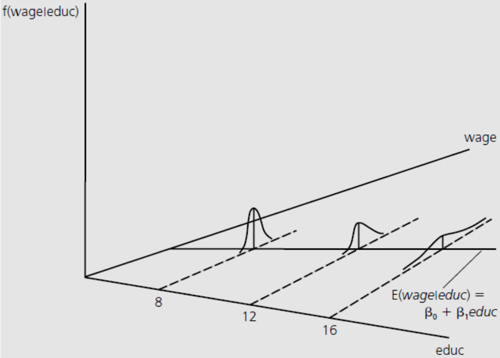
1 de noviembre de 2009

Un supuesto importante en el modelo clásico de regresión lineal es que los errores ϵ_i son homocedásticos, es decir la varianza es constante para todo valor de X_i:

$$V(\epsilon_i) = V(\epsilon_j) \quad \text{para} \quad i \neq j$$



Cuando el supuesto de homocedasticidad no se cumple, los errores son heterocedásticos...



Además se suponía que los términos de error no estaban correlacionados entre si:

$$Cov(\epsilon_i, \epsilon_j) = 0 \quad \text{para} \quad i \neq j$$

Es decir, teníamos que $E(\epsilon \cdot \epsilon') = \sigma^2 I_n$, ahora si el término de error no cumple con los supuestos del modelo de regresión lineal tenemos que $E(\epsilon \cdot \epsilon') = \sigma^2 \Omega$. Donde Ω es una matriz definida positiva.

Recordemos que el estimador MCO es:

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

Como el supuesto de que $E[\epsilon|X] = 0$ se mantiene, tenemos que la $E[\hat{\beta}] = \beta$ y por lo tanto, $E[\hat{\beta}) - \beta] = 0$. De esta forma, el estimador MCO con perturbaciones no esféricas sigue siendo insesgado y consistente.

Pero no será eficiente, dado $E(\epsilon_i \cdot \epsilon_j) = \sigma^2\Omega$, entonces la varianza de $\hat{\beta}$ es:

$$\begin{aligned} V(\hat{\beta}) &= E\left[(\hat{\beta}-\beta)(\hat{\beta}-\beta)'\right] \\ &= E[(X'X)^{-1}X'\epsilon\epsilon'X(X'X)^{-1}] \\ &= \sigma^2(X'X)^{-1}(X'\Omega X)(X'X)^{-1} \end{aligned}$$

De esta forma, solo si $E(\epsilon_i, \epsilon_j) = \sigma^2 I_n$ la matriz de covarianzas de $\hat{\beta}$ será igual a $\sigma^2(X'X)^{-1}$, por lo tanto el estimador MCO en presencia de perturbaciones no esféricas no tendrá varianza mínima, es decir, no será eficiente. Entonces cualquier inferencia basada en $\hat{\sigma}^2(X'X)^{-1}$ llevará a conclusiones erróneas.

Estimación Eficiente: Mínimos Cuadrados Generalizados

La estimación eficiente de β en el modelo generalizado, donde los errores pueden no ser esféricos, requiere el conocimiento de Ω. Para comenzar supondremos que Ω es una matriz conocida, simétrica y definida positiva.

Bajo estas condiciones el Método de Mínimos Cuadrados Generalizados nos permite estimar de manera eficiente los parámetros.

Dado que Ω es una matriz simétrica definida positiva, puede ser descompuesta de la siguiente manera:

$$\Omega = C\Lambda C'$$

donde las columnas de C son los vectores propios de Ω y los valores propios (λ_j) de Ω se encuentran en la diagonal de Λ. Entonces sea $\Lambda^{1/2}$, la matriz diagonal con el j-ésimo elemento igual a $\sqrt{\lambda_j}$ y sea $T = C\Lambda^{1/2}$. De esta forma, $\Omega = TT'$. Además sea $P' = C\Lambda^{-1/2}$ y por lo tanto, $\Omega^{-1} = P'P$.

Si pre multiplicamos $Y = X\beta + \epsilon$ por P obtenemos:

$$\begin{aligned} PY &= PX\beta + P\epsilon \\ Y_* &= X_*\beta + \epsilon_* \end{aligned}$$

Notemos que la ecuación anterior es un modelo transformado de forma tal que:

$$\begin{aligned} V(\epsilon_*) &= E(\epsilon_*\epsilon_*') \\ &= \sigma^2 P\Omega P' \\ &= \sigma^2 I_n \end{aligned}$$

Por lo tanto, el modelo transformado cumple con los supuestos del modelo clásico de regresión, y se puede utilizar MCO para estimar el parámetro β:

$$\begin{aligned} \hat{\beta}_{MCG} &= (X_*'X_*)^{-1}X_*'Y_* \\ &= (X'P'PX)^{-1}X'P'PY \\ \hat{\beta}_{MCG} &= (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y \end{aligned}$$

Como el estimador MCG de β es idéntico al estimador MCO aplicado al modelo transformado y que cumple con los supuestos, $\hat{\beta}_{MCG}$ es MELI.

Nuevamente como el estimador MCG es igual al estimador MCO sólo que se aplica al modelo transformado, todos los procesos para testear hipótesis y construir intervalos de confianza se mantienen.

Por ejemplo si queremos testear q hipótesis lineales $H_0 : Q'\beta = c$, se tiene el siguiente estadístico F:

$$\frac{(Q'\hat{\beta}_{MCG} - c)'[Q'\hat{\sigma}^2_{MCG}(X'_*X_*)^{-1}Q]^{-1}(Q'\hat{\beta}_{MCG} - c)}{q} \sim F_{q,n-k}$$

$$\frac{1}{q}\frac{(Q'\hat{\beta}_{MCG} - c)'[Q'(X'_*X_*)^{-1}Q]^{-1}(Q'\hat{\beta}_{MCG} - c)}{\hat{\sigma}^2_{MCG}} \sim F_{q,n-k}$$

donde $\hat{\sigma}^2_{MCG}$ es el estimador insesgado de σ^2 en presencia de perturbaciones no esféricas:

$$\hat{\sigma}^2_{MCG} = \frac{\hat{\epsilon}'_*\hat{\epsilon}_*}{n-k} = \frac{(Y - X\hat{\beta}_{MCG})'\Omega^{-1}(Y - X\hat{\beta}_{MCG})}{n-k}$$

Anteriormente asumimos que Ω era conocida, en este caso una simple transformación del modelo de regresión lineal lleva a una matriz de covarianza esférica. En la práctica, Ω es desconocida y es necesario estimar los parámetros al interior de esta matriz.

Entonces lo que debemos hacer es sustituir Ω por un estimador de ella $\hat{\Omega}$. Esto se denomina estimador Mínimos Cuadrados Factibles (MCF), donde el estimador de β se define de la siguiente forma:

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

El problema es que tenemos más incógnitas $(n(n+1)/2)$ en Ω que observaciones, para $n > 1$. En la práctica para lograr la estimación de Ω debemos asumir que es función de un número fijo y reducido de parámetros θ. El problema se reduce a encontrar $\hat{\theta}$ y usarlo para computar $\hat{\Omega} = \Omega(\hat{\theta})$.

Heterocedasticidad

La Heterocedasticidad surge cuando a pesar de que $Cov(\epsilon_i, \epsilon_j) = 0$ para $i \neq j$, las varianzas de cada observación son diferentes, es decir, $V(\epsilon_j) = \sigma_j^2$ para $j = 1, ..., n$.

La matriz de covarianzas en este caso es:

$$E[\epsilon\epsilon'] = \sigma^2\Omega = \begin{pmatrix} \sigma_1^2 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_n^2 \end{pmatrix} = \sigma^2 \begin{pmatrix} \omega_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \omega_n \end{pmatrix}$$

Algunas razones por las que ϵ_i puede variar son las siguientes:

- En los modelos de aprendizaje sobre errores, a medida que la gente aprende, sus errores de comportamiento son menores.
- A medida que aumentan los ingresos, la gente tiene más posibilidades de disponer de parte de ese ingreso de la forma que desee.
- Al omitir variables relevantes, a parte del sesgo que se produce en las estimaciones por esto, se produce Heterocedasticidad ya que este variable estará en el término de error.
- Asimetría en la distribución de una o más variables explicativas incluidas en el modelo, por ejemplo: ingreso, riqueza y educación.

Como mencionamos anteriormente en presencia de Heterocedasticidad el estimador MCO seguirá siendo insesgado, pero no tendrá varianza mínima. El estimador que si cumple con la propiedad de MELI es el de MCG.

Este último estimador requiere conocimiento de la matriz Ω. Sin embargo, White (1980) ha propuesto una aproximación a la matriz de covarianzas del estimador MCO:

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

que no requiere una representación específica de la forma funcional que adopta la heterocedasticidad, por lo que no tendremos riesgo de asumir una forma funcional incorrecta.

La sugerencia de White es que la varianza del estimador $\hat{\beta}_{MCO}$ se exprese de la siguiente forma:

$$V(\hat{\beta}|X) = n(X'X)^{-1}\left(\frac{1}{n}X'\sigma^2\Omega X\right)(X'X)^{-1}$$

se define:

$$\begin{aligned}\Sigma &= n^{-1}\sigma^2 X'\Omega X \\ &= n^{-1}\sum_{i=1}^{n}\sigma_i^2 x_i x_i'\end{aligned}$$

la que se estima de la siguiente forma:

$$\hat{\Sigma} = n^{-1}\sum_{i=1}^{n}\hat{\epsilon}_i^2 x_i x_i'$$

White demuestra bajo condiciones generales que:

$$\hat{\Sigma} = n^{-1} \sum_{i=1}^{n} \hat{\epsilon}_i^2 x_i x_i' \rightarrow_p \Sigma$$

De esta forma, una estimación consistente de la matriz de covarianzas es:

$$V(\hat{\beta}|X) = n(X'X)^{-1}\hat{\Sigma}(X'X)^{-1} \quad (\dagger)$$

su comparación con $\sigma^2(X'X)^{-1}$ puede dar noción del grado de heterocedasticidad.

1. El Contraste de White:
La hipótesis nula es de Homocedasticidad (al igual que en todos los contrastes que estudiaremos). Esto es, $H_0 : \sigma_i^2 = \sigma^2 \quad \forall i$. Bajo la hipótesis nula el estimador de la matriz de covarianzas de $\hat{\beta}$ es $\hat{V}(\hat{\beta}|X) = \hat{\sigma}^2(X'X)^{-1}$, pero bajo la hipótesis alternativa es (†).

Basado en la observación de esto, White propone un test que puede obtenerse al calcular nR^2 de una regresión de $\hat{\epsilon}_i^2$ contra todos los productos posibles entre las variables explicativas.

Demuestra que $nR^2 \sim \chi^2_{J-1}$, donde J es el número de regresores de esta ecuación.

Consideremos el siguiente modelo:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + \epsilon_i$$

Los pasos para realizar el test de White son:

a. Obtener $\hat{\beta}$ y los residuos de la estimación del modelo anterior por MCO $\{\hat{\epsilon}\}_{i=1}^{n}$.
b. Correr una regresión de $\hat{\epsilon}_i$ sobre una constante, x_i, z_i, x_i^2, z_i^2, $x_i z_i$.
c. Computar nR^2 de la regresión anterior.
d. Para el nivel de significancia escogido, comparar nR^2 con el valor crítico de una distribución chi cuadrado con 5 grados de libertad. Si nR^2 excede el valor crítico se rechaza la hipótesis nula de Homocedasticidad.

2. El contraste de Goldfeld y Quandt:
Este contraste parte del supuesto de que la magnitud de σ_i^2 depende de cierta variable z_i, la que generalmente es una variable explicativa pero no es necesario.

Supongamos que dicha relación es positiva, es decir, para valores más altos de z_i mayor es σ_i^2.

Las observaciones se dividen en dos grupos, bajo la hipótesis nula ambos grupos tienen la misma varianza, pero bajo la alternativa las varianzas difieren significativamente.

Entonces el contraste consiste en:

a. Ordenar las observaciones por los valores de la variable z_i, de menor a mayor.
b. Omitir p observaciones en la mitad de la muestra, se sugiere no eliminar más de la tercera parte de las observaciones.
c. Estimar dos veces el modelo original, una con las $\frac{n-p}{2}$ primeras observaciones muestrales y otra con las $\frac{n-p}{2}$ últimas observaciones en la muestra. Notar que p debe ser lo suficientemente pequeño de manera que $\frac{T-p}{2}$ sea mayor al número de parámetros.
d. Se calcula es estadístico:

$$\frac{\hat{\epsilon}_2'\hat{\epsilon}_2}{\hat{\epsilon}_1'\hat{\epsilon}_1} \sim F_{m,m} \quad \text{con} \quad m = \frac{n-p}{2} - k$$

Si se sospecha que la varianza del error depende inversamente de z_i, entonces las observaciones se deben ordenar de mayor a menor.

Si se llega a la conclusión de que el término de error del modelo no presenta heterocedasticidad, podría deberse a que hemos comenzado con una mala especificación del parámetro σ_i^2, que quizás depende de un variable diferente a la que hemos supuesto.

Por esta razón el contraste debería realizarse varias veces con distintas variables de las que tengamos sospechas pueda depender la varianza del término de error.

3. El contraste de Breusch y Pagan:

Supongamos que la varianza del término de error de cada observación depende de un vector de variables z_i de dimensión p, es decir:

$$\sigma_i^2 = h(z_i'\alpha) = h(\alpha_0 + \alpha_1 z_{1i} + \alpha_2 z_{2i} + ... + \alpha_p z_{pi})$$

Notemos que si todos los coeficientes α's excepto el correspondiente a α_0 fuesen cero, tendríamos una situación de Homocedasticidad.

Por lo tanto, si puedieramos estimar los coeficientes $\alpha_0, \alpha_1, ..., \alpha_p$ un contraste para la hipótesis nula de Homocedasticidad es:

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

Los pasos para realizar este contraste son:

a. Se estima por MCO el modelo original y se obtienen los residuos correspondientes.
b. Se obtiene la serie de residuos normalizados al cuadrado:

$$\hat{\varepsilon}_i^2 = \frac{\hat{\epsilon}_i^2}{\hat{\sigma}_\epsilon^2}, \quad i = 1, \dots, n \quad \text{donde} \quad \hat{\sigma}_\epsilon^2 = \frac{\sum \hat{\epsilon}_i^2}{n}$$

c. Se estima una regresión de $\hat{\varepsilon}_i^2$ sobre una constante y las variables $z_{1i}, z_{2i}, \dots, z_{pi}$ y se obtiene la suma explicada (SE) de dicha regresión.
d. Bajo la hipótesis nula de Homocedasticidad y dado el supuesto de normalidad del término de error, la razón $\frac{SE}{2}$ se distribuye χ_p^2.