Universidad Carlos III de Madrid

César Alonso

ECONOMETRIA

Tema 8: HETEROCEDASTICIDAD*

Índice

*Este material está basado en material previo elaborado por María Arrazola y José de Hevia.

1. Introducción

- El análisis de muchos fenómenos económicos exige relajar el supuesto de HOMOCEDASTICIDAD (varianza condicional del error constante) que se hace en el modelo de regresión clásico.

- Tanto en el contexto de datos de sección cruzada como de datos de series temporales es habitual encontrar situaciones con HETEROCEDASTICIDAD.

- El término **heterocedasticidad** se refiere a aquella situación en que la varianza de Y condicional a las variables del modelo no es constante para los distintos valores de las X's.

1.1. Ejemplos

- Supongamos que 100 estudiantes se apuntan a una clase de mecanografía. Supongamos que algunos han escrito a máquina antes y otros no lo han hecho nunca. Al acabar la primera clase, algunos estudiantes mecanografían fatal y otros lo hacen bastante bien.
  Si construyéramos un gráfico mostrando los errores tipográficos cometidos por los estudiantes con respecto a las horas acumuladas de mecanografía, veríamos que la dispersión de los errores tipográficos disminuye a medida que la variable explicativa (el número de horas acumuladas de mecanografía) aumenta.
  La diferencia en los errores cometidos entre el mejor y el peor de la clase será probablemente más pequeña después de la última clase que después de la primera clase.

- Supongamos que tenemos datos de renta y de gasto en alimentación para un número grande de familias. Si representamos en un gráfico el gasto en alimentación frente a la renta es de esperar que encontremos heterocedasticidad.

  Probablemente, la dispersión en el gasto en alimentación para diferentes niveles de renta aumente con la renta.

  Las familias pobres tienen menos flexibilidad en su nivel de gasto en alimentación, de manera que veremos poca dispersión en el gasto de alimentación para dichas familias. Por el contrario, cabe esperar que algunas familias ricas gasten mucho en alimentación (por ejemplo, comiendo caviar o cenando a menudo en restaurantes caros) y que otras familias ricas con preferencias diferentes gasten mucho menos en alimentación, destinando su renta a otros usos.

2. El modelo de regresión lineal con heterocedasticidad

- Sin pérdida de generalidad, consideraremos el caso del modelo de regresión simple:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

  donde:

$$\begin{aligned} & E(\varepsilon|X) = 0 \\ & \qquad \Rightarrow E(Y|X) = \beta_0 + \beta_1 X = PLO(Y|X) \\ & \qquad \Rightarrow \ \beta_0 = E(Y) - \beta_1 E(X) \quad \text{y} \quad \beta_1 = \frac{C(Y,X)}{V(X)} \end{aligned}$$

y:

$$V(\varepsilon|X) = V(Y|X) = \sigma^2(X) \qquad \text{HETEROCEDASTICIDAD}$$

- Veremos que:
  - MCO es insesgado y consistente, incluso en ausencia de homoscedasticidad.
  - Los errores estándar de las estimaciones basados en la expresión convencional son sesgados e inconsistentes en presencia de heteroscedasticidad.
    (y por tanto no podemos utilizarlos para hacer inferencia).

2.1. Consecuencias sobre el estimador MCO

- ¿Qué propiedades tendrá el estimador de MCO en este contexto?.
- Centrémonos en el análisis de la estimación de β_1.
- Suponiendo que tenemos una muestra aleatoria:
  - **Insesgadez**
    Sabemos que $\widehat{\beta}_1 = \sum_i c_i Y_i$, con $c_i = \dfrac{x_i}{\sum_i x_i^2}$.
    Dado que $\sum_i c_i = 0$ y $\sum_i c_i X_i = 1$, tenemos que:

$$
\begin{aligned}
E(\widehat{\beta}_1|X_i) &= E(\textstyle\sum_i c_i Y_i|X_i) = \textstyle\sum_i c_i E(Y_i|X_i) \\
&= \textstyle\sum_i c_i(\beta_0 + \beta_1 X_i) = \beta_1
\end{aligned}
$$

    Por tanto,

$$
E(\widehat{\beta}_1) = \beta_1 \Rightarrow \quad \text{Insesgadez}
$$

  - **Consistencia**
    Sabemos que

$$
\widehat{\beta}_1 = \frac{\sum_i x_i y_i}{\sum_i x_i^2} = \frac{\frac{1}{n}\sum_i x_i y_i}{\frac{1}{n}\sum_i x_i^2},
$$

luego

$$p\,\text{lím}\,\widehat{\beta}_1 = \frac{p\,\text{lím}\,\frac{1}{n}\sum_i x_i y_i}{p\,\text{lím}\,\frac{1}{n}\sum_i x_i^2} = \frac{C(X,Y)}{V(X)} = \beta_1 \Rightarrow \quad \text{Consistencia}$$

- **Expresión de la varianza**

$$\begin{aligned} V(\widehat{\beta}_1|X_i) = \sigma^2_{\widehat{\beta}_1} &= V(\textstyle\sum_i c_i Y_i|X_i) = \sum_i c_i^2 V(Y_i|X_i) \\ &= \textstyle\sum_i c_i^2 \sigma_i^2 = \dfrac{\sum_i x_i^2 \sigma_i^2}{\left(\sum_i x_i^2\right)^2} \\ &\Rightarrow V(\widehat{\beta}_1) = E\left[\dfrac{\sum_i x_i^2 \sigma_i^2}{\left(\sum_i x_i^2\right)^2}\right] \end{aligned}$$

  Nótese que esta expresión no es la habitual.

  Por tanto, la construcción de intervalos de confianza y la contrastación de hipótesis que se realiza a partir de la expresión habitual

$$\sigma^2_{\widehat{\beta}_1} = \frac{\sigma^2}{E\left(\sum_i x_i^2\right)}$$

  no son válidas. La importancia del "sesgo" dependerá de la magnitud de la Heterocedasticidad.

- **Eficiencia**

  MCO **no** es, en este contexto, el Estimador Lineal Insesgado de Mínima Varianza.

2.2. Consecuencias sobre los estimadores de variables instrumentales

- Al igual que en el caso del estimador MCO, la propiedad de consistencia no se ve afectada por la presencia de heterocedasticidad, dado que los supuestos sobre la varianza no se utilizan para derivar la propiedad de consistencia.

- No obstante, la expresión convencional de la varianza de este estimador es incorrecta, lo que invalida también toda inferencia basada en dicho estimador de la varianza.

3. Inferencia Robusta a la heterocedasticidad: Propuesta de White o de Eiker-White

- Nótese que en general desconocemos la forma que tiene σ_i^2.

- En cualquier caso, y a partir de resultados asintóticos, puede llevarse a cabo la construcción "aproximada" de intervalos de confianza y la contrastación de hipótesis, empleando la propuesta de estimación para $\sigma^2_{\widehat{\beta}_1} = E\left[\frac{\sum_i x_i^2 \sigma_i^2}{\left(\sum_i x_i^2\right)^2}\right]$ de White o Eiker-White.

- Esta propuesta es aplicable tanto para el estimador MCO como para estimadores de variables instrumentales.

  No obstante, vamos a centrarnos formalmente en el caso del estimador MCO.

- La principal ventaja es que no precisa conocer el patrón que caracteriza la heterocedasticidad.

- La propuesta consiste en calcular
$$s^2_{\widehat{\beta}_1} = \frac{\sum_i x_i^2 \widehat{\varepsilon}_i^2}{\left(\sum_i x_i^2\right)^2},$$
  en donde $\widehat{\varepsilon}_i$ son los residuos de MCO del modelo, es decir:
$$\widehat{\varepsilon}_i = Y_i - \widehat{Y}_i = Y_i - (\widehat{\beta}_0 + \widehat{\beta}_1 X_i)$$

- Los errores estándar de los estimadores calculados de esta forma se conocen como **errores estándar robustos**, o errores estándar consistentes a heterocedasticidad, o errores estándar de Eicker-White.

- Se puede demostrar que:

$$\frac{\widehat{\beta}_1 - \beta_1}{\sqrt{\frac{\sum_i x_i^2 \widehat{\varepsilon}_i^2}{\left(\sum_i x_i^2\right)^2}}} \stackrel{a}{\sim} N(0,1),$$

  es decir

$$\frac{\widehat{\beta}_1 - \beta_1}{s_{\widehat{\beta}_1}} \stackrel{a}{\sim} N(0,1),$$

  Se puede emplear este resultado para construir intervalos de confianza y realizar contrastes de hipótesis como se hace habitualmente.

- En la práctica, la mayor parte de los programas econométricos incorporan con la estimación MCO o de VI la opción de errores estándar robustos a heterocedasticidad.

- **Importante**: dado que en presencia de heterocedasticidad la varianza condicional de Y no es constante, no tiene sentido considerar el error estándar de la regresión o el R^2 como medidas de bondad del ajuste, puesto que sólo tienen sentido en un contexto homocedástico.