

Tema 5: Regresión con datos de sección cruzada

Universidad Complutense de
Madrid

2013

Introducción (I)

En este tema se trata el incumplimiento de dos hipótesis del MLG, aparentemente no relacionadas, que son: la **varianza no constante del error** (heteroscedasticidad) y la **no normalidad del error**. Un tercer problema relacionado con los dos anteriores, es la **presencia de datos “atípicos” en la muestra**.

Estos problemas están relacionados por diversos motivos:

- (1) Existen distribuciones no normales que producen heteroscedasticidad (por ejemplo, las variables log-normales que no son normales en nivel, pero sí en log)
- (2) Cuando contrastamos normalidad con datos heteroscedásticos, frecuentemente se rechaza la nula₂

Introducción (II)

- (3) Muchas veces, la presencia de datos atípicos (outliers) provoca la presencia de heteroscedasticidad y/o no normalidad. La solución aquí es la detección y el tratamiento adecuado de los atípicos.

El *esquema* a seguir en estos temas es el siguiente:

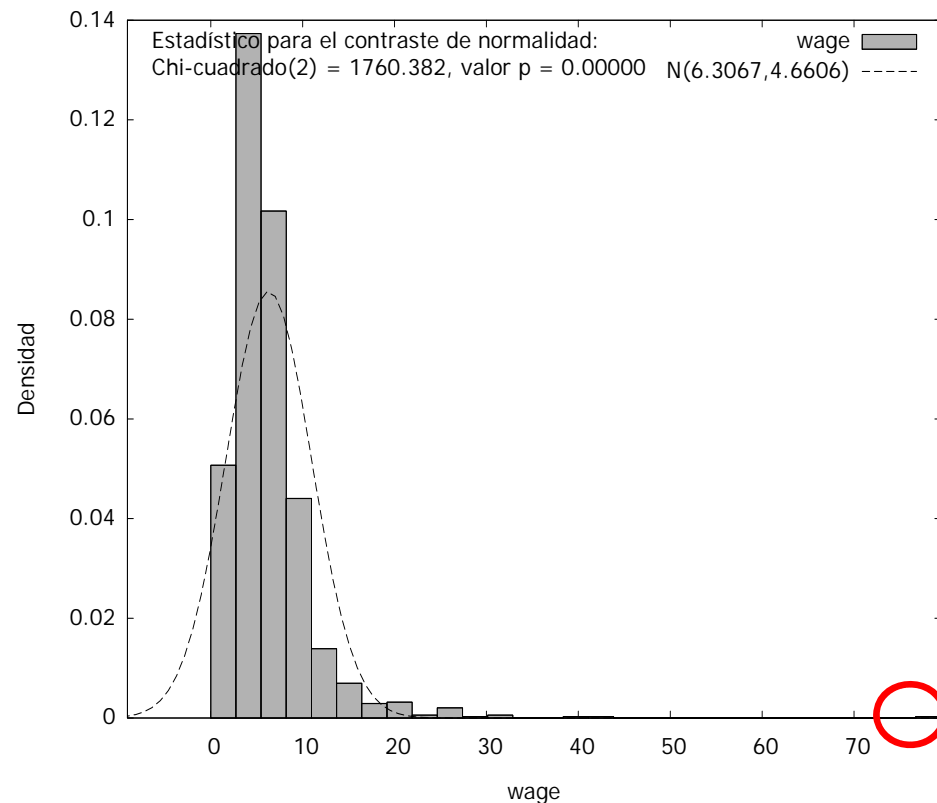
- (a) ¿Cuándo surgen típicamente estos problemas?
- (b) ¿Cuáles son sus consecuencias sobre el estimador MCO y sus propiedades estadísticas?
- (c) ¿Qué procedimientos gráficos y/o estadísticos existen para la detección?
- (d) ¿Cómo se resuelven?

No normalidad (I)

Cuando usamos datos económicos, **la no normalidad surge al modelizar variables que toman valores positivos** (que toman valores muy cerca y muy lejos del cero) Por ejemplo, variables como salarios, precios, tamaños de empresas, etc. Si los datos de la variable a explicar no son normales, es muy difícil que el error se distribuya como una normal. Las **propiedades de linealidad, insesgadez y eficiencia** (en el sentido del Teorema de Gauss-Markov) **se mantienen, pero ...**

Los contrastes de hipótesis habituales, los intervalos de confianza o la predicción por intervalo de la variable endógena dejan de ser válidos. Esta es la consecuencia más importante de que los datos no sean normales.

No normalidad (II)



Muchas variables de “tamaño”
presentan este tipo de histograma

Ejemplo: En la figura de la izquierda se muestra el histograma de la sección cruzada de salarios (wage) en dólares por hora de 1260 trabajadores americanos.

La línea de puntos muestra la función de distribución de una variable estrictamente Normal con la misma media y varianza que la variable de salarios.

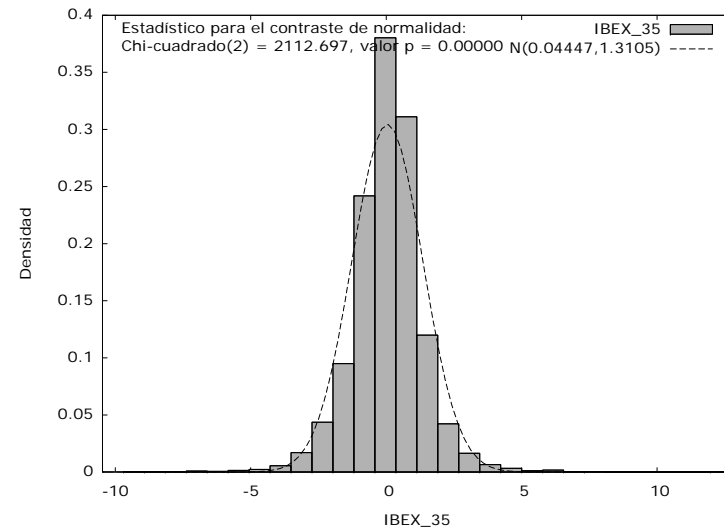
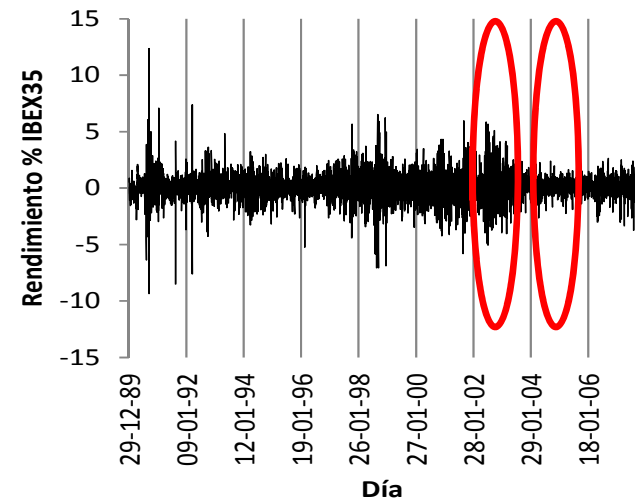
Se observa que la distribución de salarios presenta asimetría a la derecha, ya que han de ser siempre positivos y tiene más apuntamiento que la normal (es decir, tiene exceso de curtosis). Los residuos de una regresión que explique esta variable, frecuentemente no seguirán una normal.

No normalidad (III)

A veces, la heteroscedasticidad es una característica estructural de los datos. Por ejemplo, el rendimiento de los activos financieros cuando los observamos con alta frecuencia (datos diarios, intradiarios)

En la **primera figura** de la derecha se muestra la evolución diaria del rendimiento porcentual del IBEX-35 y en la **figura de abajo** se dibuja su distribución de frecuencias frente a la normal.

Lo relevante es observar el alto exceso de curtosis de los datos financieros, debido fundamentalmente a la alternancia entre períodos de alta volatilidad y baja volatilidad (ver los dos círculos del gráfico de arriba). Esto hace que no podamos hablar de homoscedasticidad.

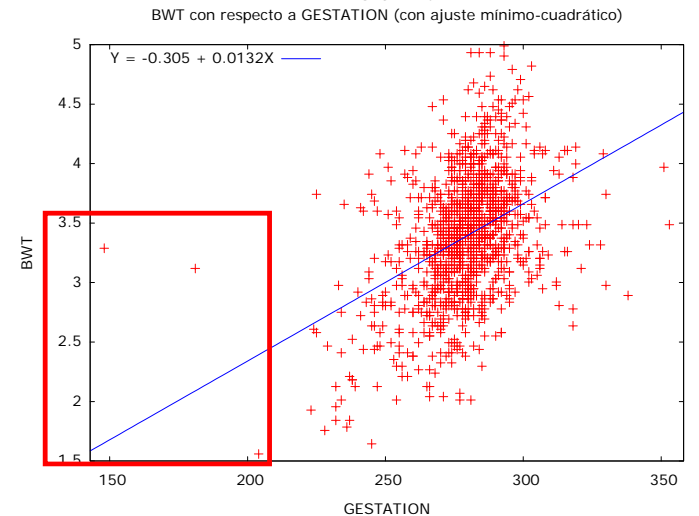
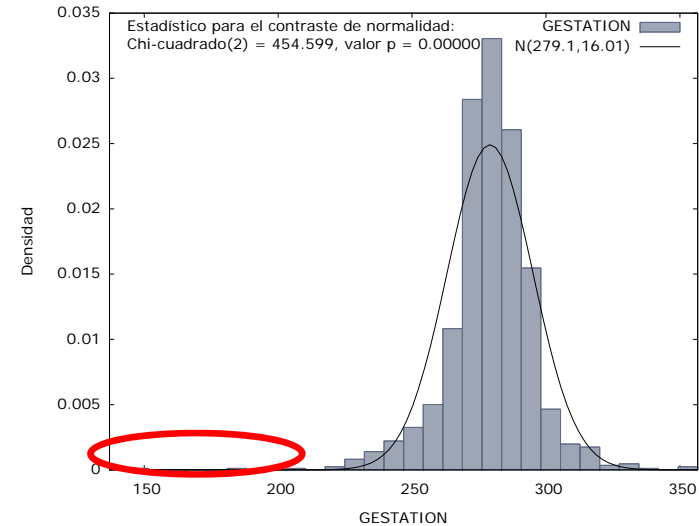


No normalidad (IV)

La presencia de outliers en la muestra crea la apariencia de heteroscedasticidad y/o no normalidad.

Por otro lado, los outliers son un problema en sí mismos, ya que unos pocos datos atípicos pueden cambiar significativamente las estimaciones y en este sentido, hablamos de observaciones influyentes.

Los dos gráficos de la derecha muestran la distribución de los datos de Días de Gestación (Gretl – Wooldrige) y su relación con el Peso del niño recién nacido. Se observan unos valores de Días de Gestación sospechosamente muy pequeños (lo que hace que la distribución tenga asimetría a la izquierda). A su vez, esos pocos datos atípicos en Gestación pueden tener un peso importante en la estimación por MCO. Si los eliminamos cambia la pendiente de la recta.



No normalidad (V)

El contraste más conocido y utilizado de normalidad es el que desarrollaron **Jarque-Bera (JB)**, en donde, bajo la hipótesis nula tenemos normalidad [en concreto, que el coeficiente de asimetría (CA) es cero y la curtosis (CK) es 3]. Estos momentos de 3º y 4º orden son:

$$\hat{CA} = \frac{1}{n} \sum_{i=1}^n \left(\frac{X_i - \hat{\mu}_x}{\hat{\sigma}_x} \right)^3 \quad \hat{CK} = \frac{1}{n} \sum_{i=1}^n \left(\frac{X_i - \hat{\mu}_x}{\hat{\sigma}_x} \right)^4$$

que son combinados en el estadístico de contraste siguiente:

$$JB = n \left(\frac{\hat{CA}^2}{6} + \frac{(\hat{CK} - 3)^2}{24} \right) \underset{H_0}{\sim} \chi_2^2$$

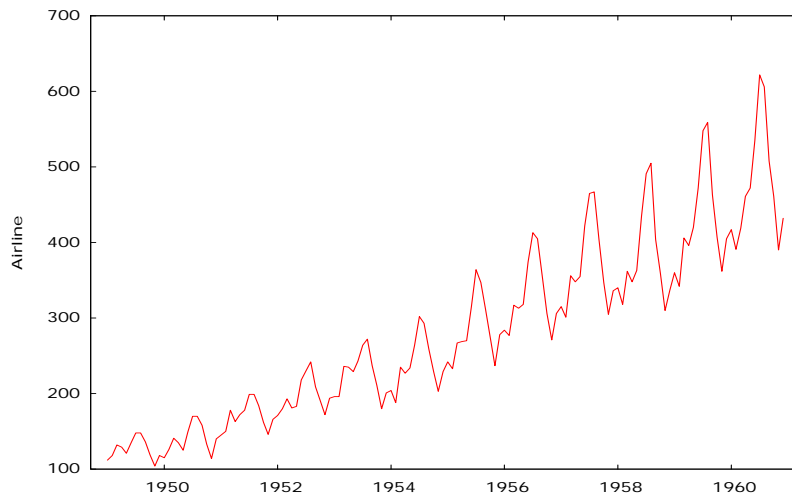
Si el valor del **JB** para la muestra supera el valor crítico de la distribución chi-cuadrado con dos grados de libertad (al nivel de significación escogido) **se rechaza la nula de normalidad**.

No normalidad (VI)

- La mejor forma de tratar la no normalidad de los datos, depende de la causa del mismo. Así,
- (1) Si la variable sólo toma valores positivos, la transformación logarítmica puede inducir normalidad
- (2) Si el problema es la heteroscedasticidad, el tratamiento de la misma es más complicado y se verá en las siguientes secciones del tema actual.
- (3) Si el problema es la presencia de outliers, la solución es detectarlos y, en ocasiones, eliminarlos de la muestra.

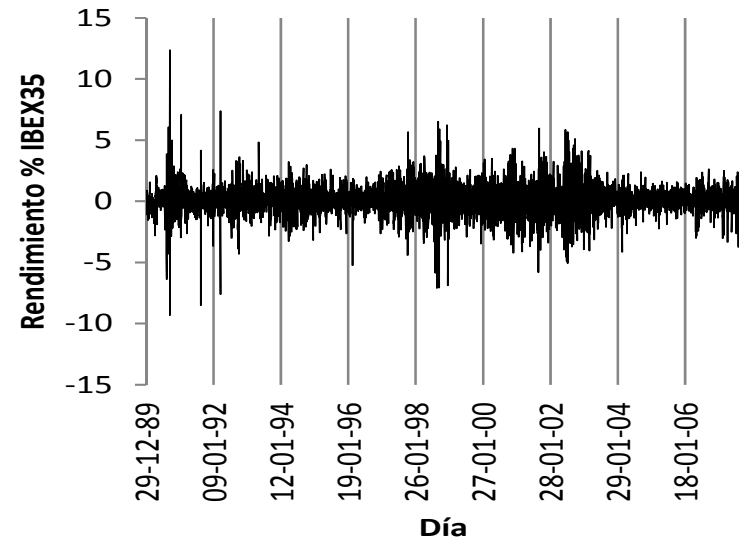
Heteroscedasticidad (I)

- Si los **datos son temporales**, las formas más comunes de heteroscedasticidad son:



Ejemplo: N^o de pasajeros que vuela mes a mes en una compañía aérea.

Se observa que la **varianza crece a medida que crece la media de la serie.**



Ejemplo: Rendimiento (en %) diario del IBEX-35

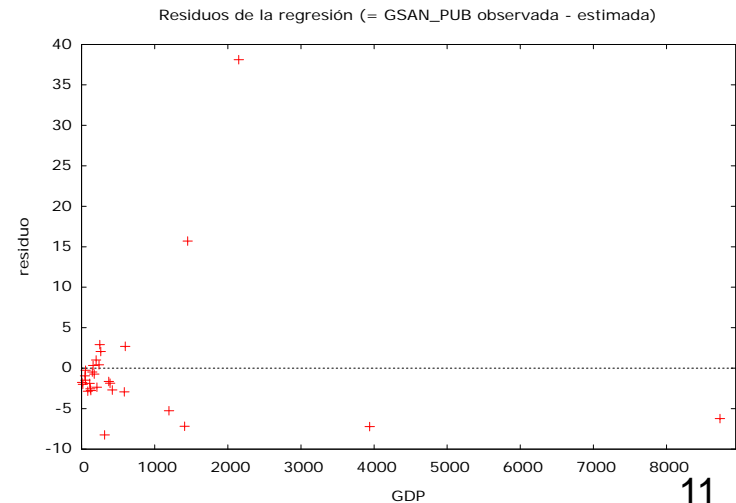
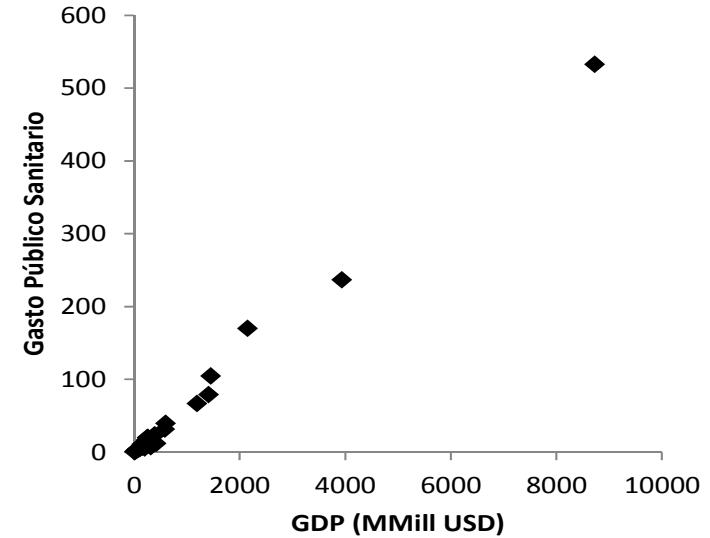
Se observan días de alta varianza seguidos de otros de poca volatilidad

Heteroscedasticidad (II)

Con **datos de sección cruzada**, la heteroscedasticidad surge muchas veces al tratar con “unidades” de diferente tamaño. Estas unidades pueden ser familias, empresas, países, etc.

Los dos gráficos de la derecha muestran algunos resultados de la estimación de la relación del Gasto Público sanitario y el PIB (GDP) de un conjunto de países de la OCDE. **Es importante notar la existencia de países grandes y pequeños en esta muestra.**

El gráfico (X-Y) de los residuos MCO resultantes de la regresión versus el PIB muestra cómo la dispersión de los mismos crece con el PIB (GDP).



Heteroscedasticidad (III)

Si el error del MLG no tiene varianza constante, el estimador MCO de los parámetros β es lineal e insesgado, pero no eficiente. Bajo las hipótesis deseables, escribimos:

$$Y = X\beta + \varepsilon \quad E[\varepsilon_i^2] = \sigma^2, i = 1, 2, \dots, n$$

y bajo heterocedasticidad $E[\varepsilon_i^2] = \sigma_i^2, i = 1, 2, \dots, n$

O en términos matriciales

$$E[\varepsilon\varepsilon^T] = \Omega = \begin{pmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_n^2 \end{pmatrix}$$

que sigue siendo una matriz diagonal si no hay autocorrelación en los errores, sólo heteroscedasticidad.

Heteroscedasticidad (IV)

El estimador MCO del MLG tiene la expresión:

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

O bien, $\hat{\beta} = \beta + (X^T X)^{-1} X^T \varepsilon$ y por tanto, es insesgado bajo heteroscedasticidad. Es decir, la esperanza del estimador no depende de la matriz de varianzas del error, sólo de la esperanza del error que es cero por hipótesis. **No obstante, la varianza del estimador ahora es:**

$$\text{var}[\hat{\beta}] = (X^T X)^{-1} X^T \Omega X (X^T X)^{-1}$$

que no coincide con $\text{var}[\hat{\beta}] = \sigma^2 (X^T X)^{-1}$

Heteroscedasticidad (V)

Implicaciones en la estimación: Si existe heteroscedasticidad,

$$E[\varepsilon_i^2] = \sigma_i^2, i = 1, 2, \dots, n \quad \text{y por tanto,} \quad \text{var}[y_i] = \sigma_i^2$$

En MCO minimizamos $\sum_{i=1}^n [y_i - x_i^T \hat{\beta}]^2$

pero si las varianzas de los datos difieren entre sí, sería mejor asignar un peso más pequeño a las observaciones con mayor varianza y un peso más grande a las que tienen menor varianza. Se puede usar un criterio de estimación de **Mínimos Cuadrados Ponderados** de la forma que el objetivo es minimizar la suma de cuadrados de residuos ponderados:

$$\sum_{i=1}^n w_i^2 [y_i - x_i^T \hat{\beta}]^2$$

donde las ponderaciones w_i decrecen a medida que crece la varianza del error. Habrá que ver la forma óptima de escoger esos pesos.

Heteroscedasticidad (VI)

Posibles modelos de heteroscedasticidad: Una posibilidad es que la varianza del error esté directamente relacionada (lineal o no linealmente) con una variable explicativa del modelo. Por ejemplo:

$$\text{var}[\varepsilon_i] = E[\varepsilon_i^2] = \sigma^2 x_{i1}^2, i = 1, 2, \dots, n$$

donde x_{i1} suele ser una variable de tamaño. En este caso, es muy fácil saber cómo ponderar el modelo y aplicar MCO al modelo ponderado será eficiente. Si el modelo original es:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i$$

El modelo ponderado se construye dividiéndolo por x_{i1} término a término.

Heteroscedasticidad (VII)

Es decir:
$$\frac{y_i}{x_{i1}} = \frac{\beta_0}{x_{i1}} + \beta_1 + \beta_2 \frac{x_{i2}}{x_{i1}} + \frac{\varepsilon_i}{x_{i1}}$$

donde el nuevo error $v_i = \frac{\varepsilon_i}{x_{i1}}$ tiene esperanza nula

y varianza constante $\text{var}[v_i] = \frac{\text{var}[\varepsilon_i]}{x_{i1}^2} = \frac{\sigma^2 x_{i1}^2}{x_{i1}^2} = \sigma^2$

Aplicar MCO en el modelo ponderado (transformado) es eficiente. Sólo hay que transformar adecuadamente los datos de la variable endógena y los de las exógenas y volver a aplicar MCO al nuevo modelo para recuperar la eficiencia.

Obsérvese que el término constante del modelo ponderado no coincide con el del modelo original.

Heteroscedasticidad (VIII)

Desviaciones típicas de White Si la forma funcional de la heterocedasticidad no es conocida, o tan simple como en el ejemplo anterior, sabemos que los contrastes habría que llevarlos a cabo usando una estimación de la matriz de varianzas-covarianzas siguiente:

$$\text{var}[\hat{\beta}] = (X^T X)^{-1} X^T \Omega X (X^T X)^{-1}$$

Si suponemos que no hay autocorrelación en los errores, aunque sí hay heterocedasticidad, la matriz Ω es diagonal con $\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2$ elementos en la diagonal principal. Entonces, la matriz de varianzas-covarianzas anterior se puede escribir:

$$\text{var}[\hat{\beta}] = (X^T X)^{-1} \left[\sum_{i=1}^n \sigma_i^2 x_i x_i^T \right] (X^T X)^{-1}$$

donde X_i es el vector ($k \times 1$) de valores de las explicativas en la observación i -ésima. 17

Heteroscedasticidad (IX)

Un estimador del parámetro σ_i^2 es el cuadrado de los residuos MCO obtenidos como $\hat{\varepsilon}_i = y_i - \mathbf{x}_i^T \hat{\beta}$. Por tanto, una estimación de esa matriz de varianzas de los estimadores MCO de β es:

$$\text{var}[\hat{\beta}] = (X^T X)^{-1} \left[\sum_{i=1}^n \hat{\varepsilon}_i^2 \mathbf{x}_i \mathbf{x}_i^T \right] (X^T X)^{-1}$$

denominada estimación de la matriz de varianzas-covarianzas de White. La raíz cuadrada de los elementos de su diagonal principal se denominan desviaciones típicas (o errores estándar) de White. La matriz de varianzas -covarianzas de White es un estimador consistente de la siguiente matriz:

$$\text{var}[\hat{\beta}] = (X^T X)^{-1} \left[\sum_{i=1}^n \sigma_i^2 \mathbf{x}_i \mathbf{x}_i^T \right] (X^T X)^{-1}$$

Heterocedasticidad (X)

Detección: Disponemos de procedimientos **gráficos y estadísticos** de detección de heteroscedasticidad.

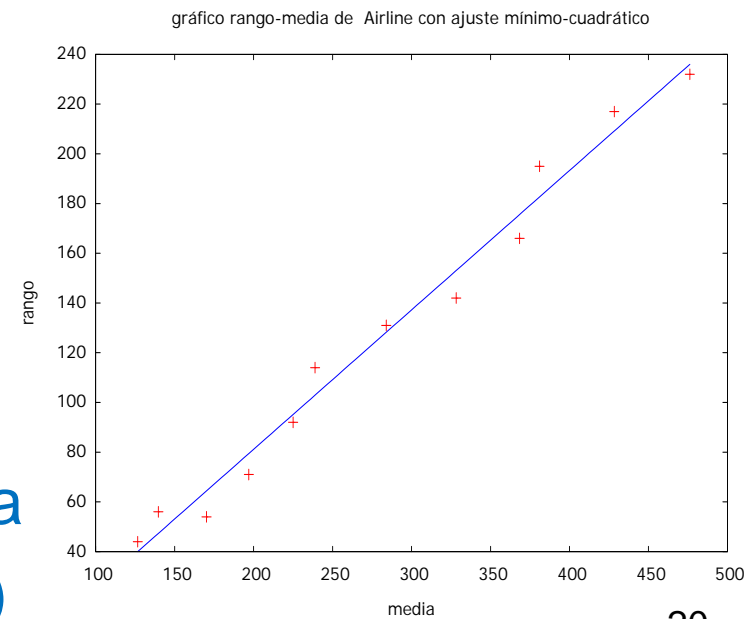
(A) Gráficos: Si los datos son de sección cruzada, se pueden **dibujar los residuos MCO** resultantes del modelo original versus alguna(s) variable(s) **explicativa(s)** (las candidatas a causar el problema) ordenadas de menor valor a mayor valor. Por ejemplo, con respecto al tamaño de los países, tamaño de las empresas, población o nivel de renta de las familias.

En estos gráficos, **lo habitual es ver mayor dispersión en los residuos a medida que crece el tamaño de la entidad o la renta familiar, etc.**

Heteroscedasticidad (XI)

(B) Gráficos: Si los datos son temporales, un gráfico muy usado es el denominado Rango-Media. En la figura de abajo se muestra el gráfico Rango-Media para la serie mensual de pasajeros de una compañía aérea.

Se divide la muestra en submuestras de idéntico tamaño, por ejemplo, de 12 meses. Se calcula para cada una de las submuestras la media y la desviación típica local. Se observa una asociación positiva (a mayor media, mayor dispersión)



Heteroscedasticidad (XII)

(B) Estadísticos: existen muchos, veremos sólo dos

(B.1) Contraste de Breusch-Pagan Está basado en que la forma de la heteroscedasticidad es del tipo

$$\sigma_i^2 = h \left[z_i^T \gamma \right]$$

donde las variables $z_i^T \equiv (1, z_{i2}, \dots, z_{ip})$ son las que explican las diferencias que hay en las varianzas de los errores y h es cualquier función (en principio, lineal). La hipótesis nula de ausencia de heteroscedasticidad se corresponde con las $(p-1)$ restricciones siguientes:

$$\gamma_2 = \gamma_3 = \dots = \gamma_p = 0$$

Heteroscedasticidad (XIII)

Pasos del contraste:

Paso 1: Se aplica MCO al modelo original $Y = X\beta + \varepsilon$ ignorando que existen problemas con el fin de recuperar los residuos $\hat{\varepsilon} = Y - X\hat{\beta}$

Paso 2: Se lleva a cabo una regresión auxiliar en donde se relacionan los residuos MCO al cuadrado en función de las variables z . Es decir, se usa como estimador de la varianza de los errores los residuos MCO al cuadrado:

$$\hat{\varepsilon}_i^2 = \gamma_1 + \gamma_2 z_{i2} + \dots + \gamma_p z_{ip} + \eta_i$$

Paso 3: Se calcula nR^2 de la regresión del Paso 2. Este estadístico sigue una χ_{p-1}^2 bajo la nula (es asintótico).

Heteroscedasticidad (XIV)

(B.2) Contraste de White. Un problema del test de Breusch-Pagan, es que deben conocerse las variables z que hacen que cambie la varianza del error. En el caso de que estas variables sean desconocidas, White propone sustituirlas por: (1) todas las variables explicativas del modelo, (2) sus cuadrados y (3) sus productos cruzados de segundo orden.

Ejemplo: Sea el modelo $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i$

Paso 1: Se estima el modelo anterior por MCO y se recuperan los residuos, $\hat{\varepsilon}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2}$

Paso 2: Se lleva a cabo la regresión auxiliar siguiente:

$$\hat{\varepsilon}_i^2 = \gamma_0 + \gamma_1 x_{i1} + \gamma_2 x_{i2} + \gamma_3 x_{i1}^2 + \gamma_4 x_{i2}^2 + \gamma_5 x_{i1} x_{i2} + \eta_i$$

Heteroscedasticidad (XV)

Paso 3: Se calcula nR^2 de la regresión auxiliar anterior, donde, en este caso:

$$nR^2 \sim \chi_5^2$$

Este test también es asintótico, es decir, su distribución es cierta si el tamaño de la muestra es suficientemente grande. Los grados de libertad son 5 en este caso, ya que en la regresión auxiliar hay 5 pendientes.

Intuición del contraste: Si el R-cuadrado de esa regresión auxiliar tiende a cero, las variables x 's y sus cuadrados, no explican apenas el cuadrado de los residuos, indicando que no existe heteroscedasticidad. Si el R-cuadrado tiende a 1, como n es muy grande, el valor del estadístico es grande y se rechaza la nula. Es decir, las x 's y sus cuadrados (y/o productos cruzados), explican bien el cuadrado de los residuos.

Datos atípicos e influyentes (I)

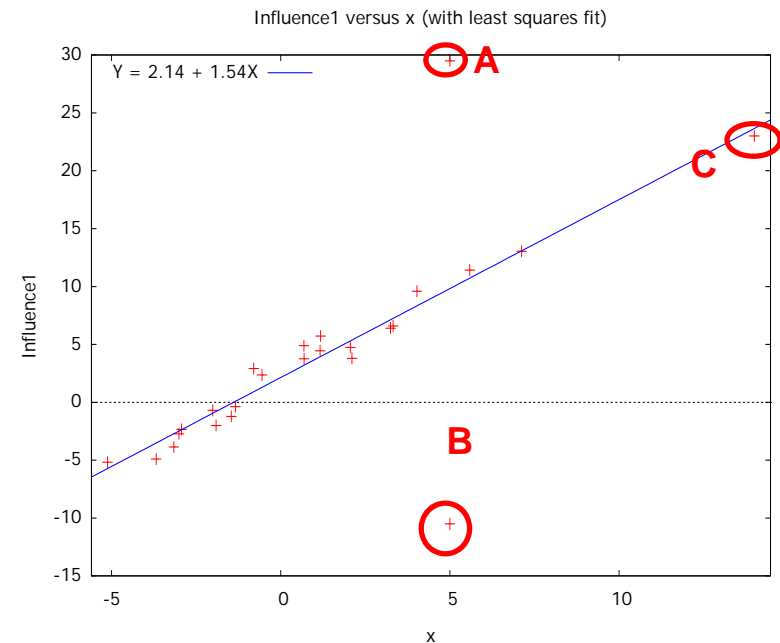
Una observación “ j ” se dice que es atípica si el valor de la variable dependiente y_j difiere mucho del que esperaríamos tener viendo el patrón general que siguen las demás.

Hay tres tipos de observaciones atípicas:

- (1) Relaciones atípicas en el eje de ordenadas (puntos A y B de la Figura de la derecha)
- (2) Relaciones atípicas en el eje de abcisas
- (3) Relaciones atípicas en el eje de ordenadas y de abcisas (punto C de la Figura)

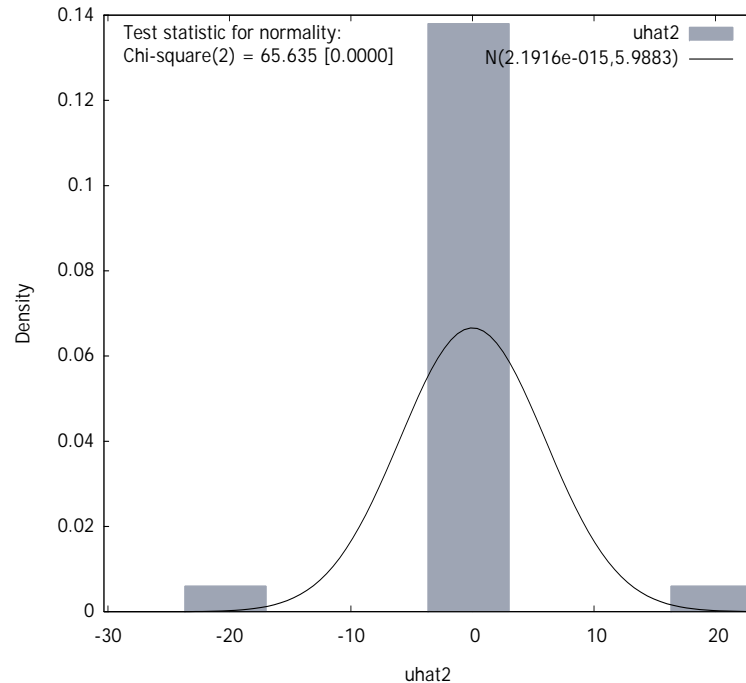
Pueden surgir varios problemas al existir datos atípicos:

- (a) La posible apariencia de no normalidad y/o heteroscedasticidad.
- (b) Se distorsionan los resultados MCO, al ser el peso de estos datos excesivo en la estimación.



Usando datos simulados, se dibuja en el plano X-Y la nube de puntos junto con los resultados de la estimación MCO de un modelo lineal simple.

Datos atípicos e influyentes (II)



Influencel:

OLS, using observations 1-25

Dependent variable: Influencel

	coefficient	std. error	t-ratio	p-value
const	2.14152	1.24486	1.720	0.0988 *
x	1.53666	0.290844	5.283	2.32e-05 ***

Mean dependent var	3.935364	S.D. dependent var	8.722149
Sum squared resid	824.7877	S.E. of regression	5.988350
R-squared	0.548265	Adjusted R-squared	0.528624
F(1, 23)	27.91479	P-value(F)	0.000023
Log-likelihood	-79.17659	Akaike criterion	162.3532
Schwarz criterion	164.7909	Hannan-Quinn	163.0293

Test for normality of residual -

Null hypothesis: error is normally distributed

Test statistic: Chi-square(2) = 65.6351

with p-value = 5.59143e-015

Como se muestra en la información de arriba, la estimación por MCO del modelo que relaciona Influence1 sobre x, genera dos residuos muy altos (uno de signo positivo y otro de signo negativo). Esto genera un exceso de curtosis en la distribución de los residuos.

Obsérvese cómo el contraste de normalidad a través del estadístico de Jarque-Bera rechaza contundentemente la nula de normalidad. 26

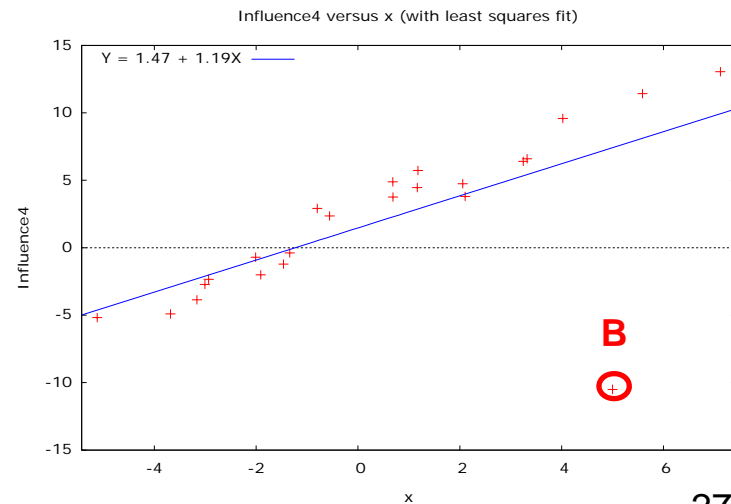
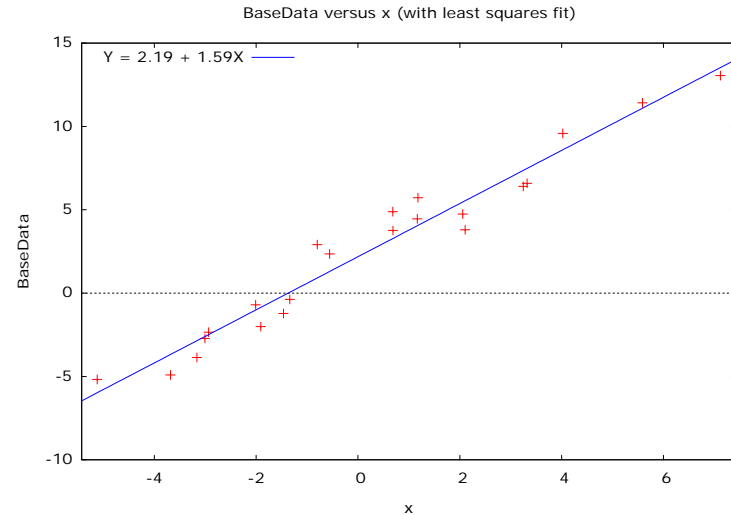
Datos atípicos e influyentes (III)

En el primer gráfico se muestra la recta de regresión estimada al usar la variable Influence1, eliminando los puntos A, B y C de la muestra. Cabe señalar que:

- (1) No se rechaza la hipótesis nula de normalidad de los residuos ($JB = 1.35$, p-valor = 0.51)
- (2) Los coeficientes estimados son muy similares a los obtenidos cuando se incluyen los puntos A, B y C. Por tanto, son datos atípicos pero no influyentes.

En el segundo gráfico se muestra la recta de regresión estimada usando la variable Influence4, eliminando sólo los puntos A y C. Cabe señalar que:

- (1) Se rechaza claramente la normalidad de los residuos ($JB = 65.64$, p-valor = 0.0)
- (2) Los coeficientes estimados cambian mucho, luego B es un dato influyente de la muestra, cuando se eliminan los puntos A y C.



Detección (I)

Existen distintos métodos:

(A) Dibujar el histograma y fijarse en los datos que se alejan mucho de la media. No obstante, en una distribución normal, un 5% de la muestra puede ser atípica (estar en las colas).

(B) También se puede caracterizar un “dato influyente” en la regresión usando la denominada matriz H , definida como:

$$H = X(X^T X)^{-1} X^T$$

La parte explicada de la Y : $\hat{Y} = X\hat{\beta} = X(X^T X)^{-1} X^T Y = HY$
y el elemento j -ésimo de la diagonal principal de H es:

$$h_j = x_j^T (X^T X)^{-1} x_j$$

donde x_j^T es un vector ($1 \times k$) que se corresponde con la fila j -ésima de la matriz X de tamaño ($n \times k$).

Detección (II)

Al valor h_j se le denomina “**influencia**” o “**apalancamiento**” **de la observación j** y cumple algunas propiedades como:

$$(1) \quad 0 \leq h_j \leq 1$$

$$(2) \quad \sum_{j=1}^n h_j = k$$

por lo que la influencia media es k/n . Un valor de h_j por encima de k/n indica que ese dato es influyente.

Un valor de h_j “grande” significa que los valores de las variables explicativas en el instante j son “inusuales” comparados con el valor medio de este valor h en toda la muestra.

Deteccción (III)

(C) Para contrastar si una observación “ j ” es atípica puede considerarse el siguiente modelo de regresión, donde se incluye una variable dummy para la propia observación:

$$y_i = \mathbf{x}_i^T \boldsymbol{\beta} + \gamma D_{ji} + \varepsilon_i, i = 1, 2, \dots, n$$

siendo $D_{jj} = 1$ y $D_{ji} = 0, \forall i \neq j$. La hipótesis nula de que el dato j -ésimo se ajusta al patrón que generó el resto de los datos es

$$H_0 : \gamma = 0$$

Detección (IV)

La regresión anterior puede escribirse de forma matricial como:

$$Y = X\beta + D_j\gamma + \varepsilon$$

donde D_j es un vector ($n \times 1$) con elementos $D_{ji}, i=1,2,\dots,n$

Por ejemplo,

$$D_1 = \begin{bmatrix} D_{11} \\ D_{12} \\ \cdot \\ D_{1n} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ \cdot \\ 0 \end{bmatrix}$$

El estimador MCO de γ en esa regresión tiene la expresión: $\hat{\gamma} = \frac{\hat{\varepsilon}_j}{1 - h_j}$ donde $\hat{\varepsilon}_j = D_j^T \hat{\varepsilon}$ y $\hat{\varepsilon}$ son los residuos MCO.

Detección (V)

El valor del estadístico t de significación individual del parámetro γ es:

$$t = \frac{\hat{\varepsilon}_j}{\hat{\sigma} \sqrt{1 - h_j}} \sim t_{n-k-1}$$

donde $\hat{\sigma}^2 = \frac{SR}{n-k-1}$, es decir, es la estimación MCO de la varianza residual del modelo. La observación j -ésima es un *outlier* si el parámetro γ es significativo, es decir, si el residuo $\hat{\varepsilon}_j$ o el valor h_j es suficientemente grande. Ese ratio t es llamado residuo “estudentizado”, $t = \hat{\varepsilon}_j^*$

Detección (VI)

Nótese que los *outliers* no siempre se pueden detectar a partir del gráfico de los residuos MCO $\hat{\varepsilon}$

Es decir, $\hat{\varepsilon}_j$ puede ser pequeño incluso si el residuo estudentizado es grande. En la regresión

$$y_i = \mathbf{x}_i^T \beta + \gamma D_{ji} + \varepsilon_i, i = 1, 2, \dots, n$$

las estimaciones MCO de los parámetros , la varianza residual y la suma de cuadrados de residuos son las mismas que obtendríamos si eliminamos la observación j de la muestra. El R -cuadrado no coincide ya que la ST con toda la muestra no es la misma que la ST sin ese dato. 33

Deteccción (VII)

Influencia sobre los parámetros estimados: La influencia de las observaciones individuales sobre las estimaciones de los β se puede calcular de la siguiente manera. En el modelo:

$$Y = X\beta + D_j\gamma + \varepsilon$$

sea $\hat{\beta}$ el estimador MCO de β bajo la restricción de que $\gamma = 0$ y el vector de residuos resultantes es $\hat{\varepsilon}$. Sean $\hat{\beta}(j)$ y $\hat{\gamma}$ los estimadores MCO de β y γ con la dummy incluida y los residuos son denotados por $\hat{\varepsilon}(j)$. El estadístico de contraste es el siguiente:

$$D_j = \frac{[\hat{\beta} - \hat{\beta}(j)]^T \mathbf{X}^T \mathbf{X} [\hat{\beta} - \hat{\beta}(j)]}{k \hat{\sigma}_\varepsilon^2} \sim F_{k, n-k}$$

y es conocido como el **estadístico de Cook**.

Detección (VIII)

Influencia sobre los valores ajustados “*dffits*”:

La influencia de la observación j -ésima sobre los valores ajustados de la Y viene dada por la distancia: $\hat{Y} - \hat{Y}(j)$ donde $\hat{Y} = X\hat{\beta}$ y $\hat{Y}(j) = X\hat{\beta}(j)$

En el instante j -ésimo es fácil probar que la distancia:

$$\hat{y}_j - \hat{y}_j(j) = \frac{h_j}{1 - h_j} \hat{\varepsilon}_j$$

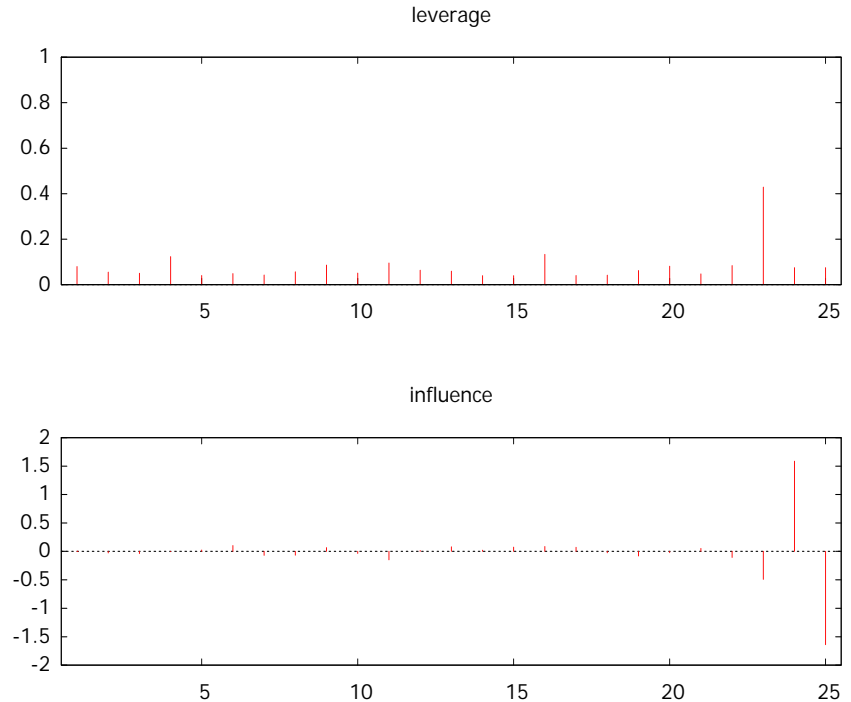
y como la varianza de \hat{y}_j es igual a $\hat{\sigma}^2 h_j$ una medida sin escala de la diferencia en los valores ajustados se puede calcular como:

$$dffits_j = \frac{\hat{y}_j - \hat{y}_j(j)}{\hat{\sigma} \sqrt{h_j}} = \hat{\varepsilon}_j^* \sqrt{\frac{h_j}{1 - h_j}}$$

Detección (IX)

	residual u	leverage 0<=h<=1	influence u*h/(1-h)	DFFITs
1	0.021741	0.080	0.0018818	0.001
2	-0.46028	0.055	-0.026704	-0.019
3	-0.72557	0.050	-0.038345	-0.028
4	-0.03139	0.124	-0.0044239	-0.002
5	0.56151	0.041	0.023727	0.019
6	2.0001	0.049	0.10334	0.076
7	-1.576	0.042	-0.06922	-0.055
8	-1.1105	0.056	-0.066274	-0.046
9	0.69156	0.086	0.065141	0.036
10	-0.65314	0.051	-0.035055	-0.025
11	-1.3878	0.095	-0.14651	-0.078
12	0.2478	0.064	0.01689	0.011
13	1.2642	0.059	0.079598	0.053
14	0.5238	0.040	0.021825	0.018
15	1.7726	0.040	0.073857	0.060
16	0.55074	0.133	0.08471	0.038
17	1.6979	0.041	0.071764	0.058
18	-0.55677	0.042	-0.024329	-0.019
19	-1.2112	0.062	-0.080532	-0.053
20	-0.24236	0.081	-0.021396	-0.012
21	1.0692	0.047	0.052748	0.040
22	-1.1419	0.084	-0.10501	-0.059
23	-0.65472	0.428*	-0.4908	-0.123
24	19.675	0.075	1.5872	1.352
25	-20.325	0.075	-1.6396	-1.447

(** indicates a leverage point)



En la regresión de X sobre Y , Gretl muestra algunos resultados de “apalancamiento” e influencia. La observación 23 es detectada como atípica (leverage alto, $h = 0.428$). Las observaciones 24 y 25 son influyentes, donde influence es $\hat{y}_j - \hat{y}_j(j) = \frac{h_j}{1-h_j} \hat{\varepsilon}_j$ aunque no son atípicas (leverage bajo, $h = 0.075$).

Resumen (I)

En MCO, las desviaciones de la relación que se postula entre la variable dependiente y las variables explicativas, se penalizan de forma cuadrática. Esto implica que **las observaciones que se desvían mucho del “patrón general” pueden tener una “influencia” excesiva en la estimación de los parámetros.** Para investigar la existencia de datos influyentes y reducir su influencia, **se pueden hacer varias cosas:**

- (1) **Inspeccionar el histograma de los residuos MCO y contrastar su normalidad con el estadístico de Jarque-Bera.**
- (2) **Detectar estos datos calculando la “influencia”, como se ha visto antes, en el método de detección (C)**

Resumen (II)

(3) Si hay datos que se alejan del patrón general, se debe intentar entender las causas. A veces, estos datos se eliminan, pero puede que sugieran que hay variables explicativas relevantes que se han omitido o que los errores siguen otra distribución diferente a la normal.

(4) A veces, la existencia de datos atípicos y/o influyentes supone un aspecto relevante de los datos y la solución no es eliminarlos, sino limitar su peso o su influencia en la estimación del modelo usando **métodos robustos**.