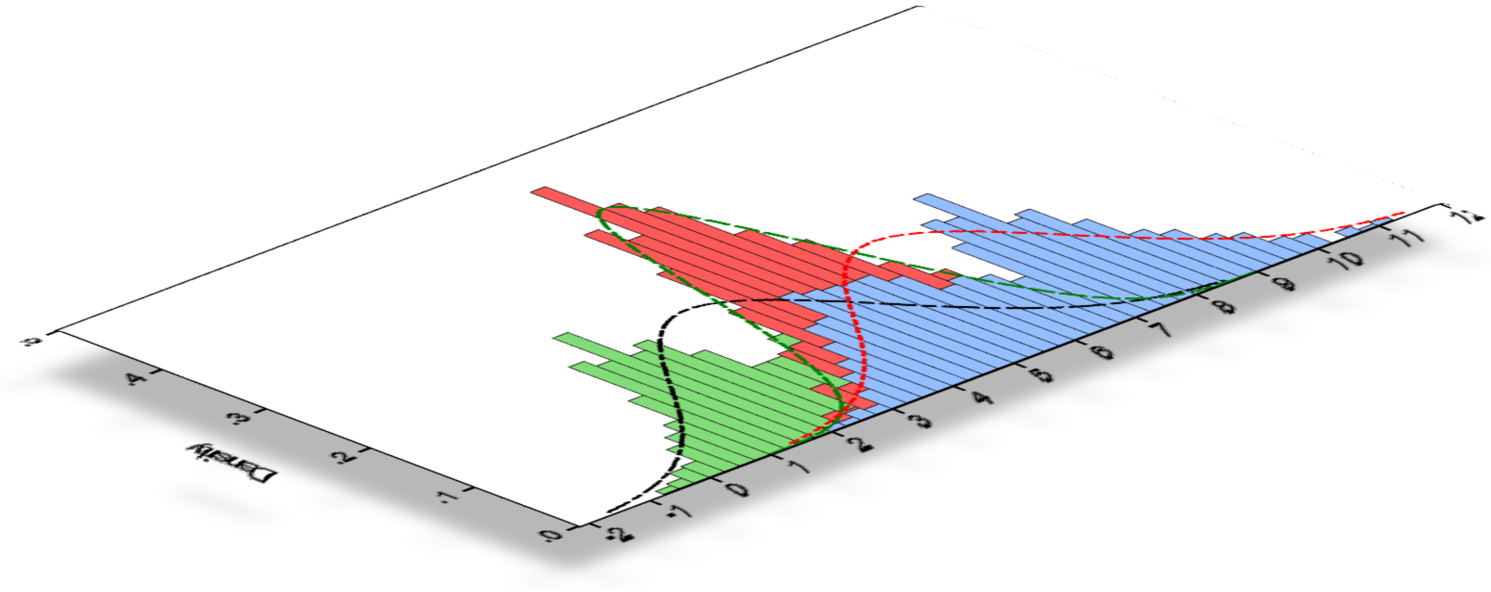
Pruebas de Validación

NORMALIDAD



Prueba de Normalidad

International Statistical Review (1987), **55**, 2, pp. 163–172. Printed in Great Britain

A Test for Normality of Observations and Regression Residuals

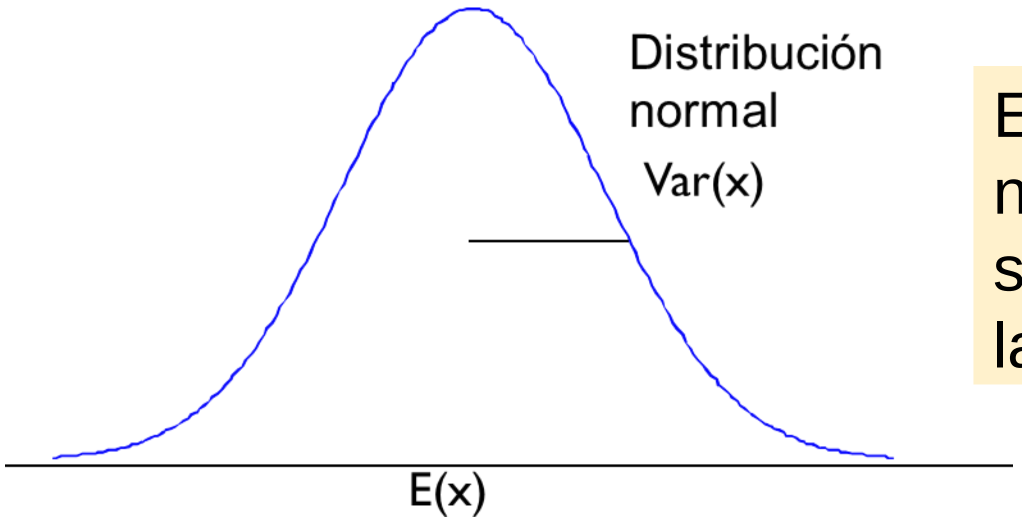
Carlos M. Jarque[1] and Anil K. Bera[2]

Se basa en el tercer y cuarto momento de la distribución de los errores. Es decir el sesgo y la curtosis

Test de Jarque –Bera (JB)

Tercero y cuarto momento de la distribución: Sesgo o Simetría y curtosis

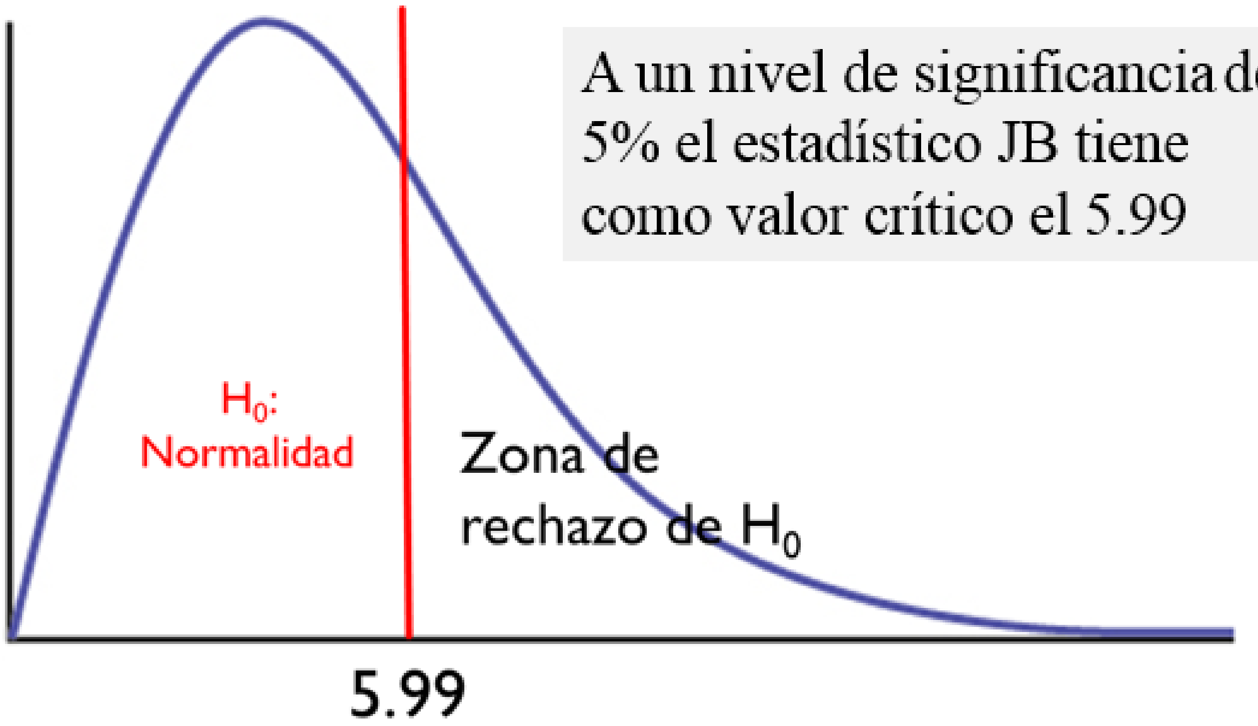
$$JB = n\left[\frac{A^2}{6} + \frac{(K-3)^2}{24}\right]$$



En una distribución normal el coeficiente de sesgo es igual a cero y la curtosis 3

Bajo la hipótesis nula el estadístico JB se distribuye como una ji-cuadrada con dos grados de libertad, ya que es la suma de dos variables aleatorias normalizadas

A un nivel de significancia del 5% el estadístico JB tiene como valor crítico el 5.99



Multicolinealidad

Multicolinealidad: ¿qué pasa si las regresoras stán correlacionadas?

Uno de los supuestos del *modelo clásico de regresión lineal* (MCRL) plantea que no existe **multicolinealidad** entre las regresoras incluidas en el modelo de regresión.

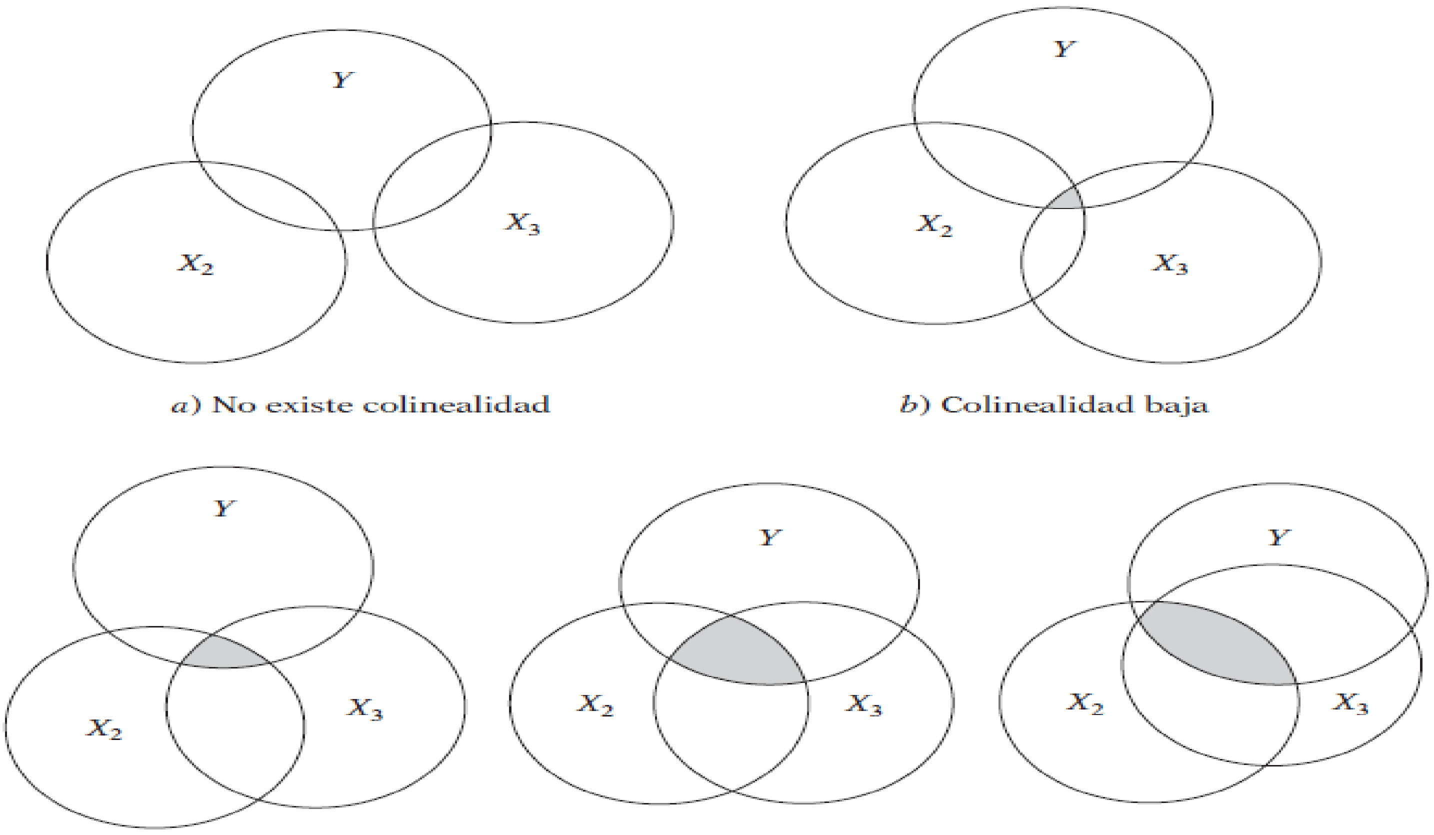
Consideramos en forma crítica el supuesto de no multicolinealidad en busca de respuestas a las siguientes preguntas

1. ¿Cuál es la naturaleza de la multicolinealidad?
2. ¿Es la multicolinealidad realmente un problema?
3. ¿Cuáles son sus consecuencias prácticas?
4. ¿Cómo se detecta?
5. ¿Qué medidas pueden tomarse para aliviar el problema de multicolinealidad?

Cuál es la naturaleza de la multicolinealidad

El término *multicolinealidad* se atribuye a Ragnar Frisch. Originalmente, designaba una relación lineal “perfecta” o exacta entre algunas o todas las variables explicativas de un modelo de regresión

Consecuencias de la multicolinealidad



a) No existe colinealidad

b) Colinealidad baja

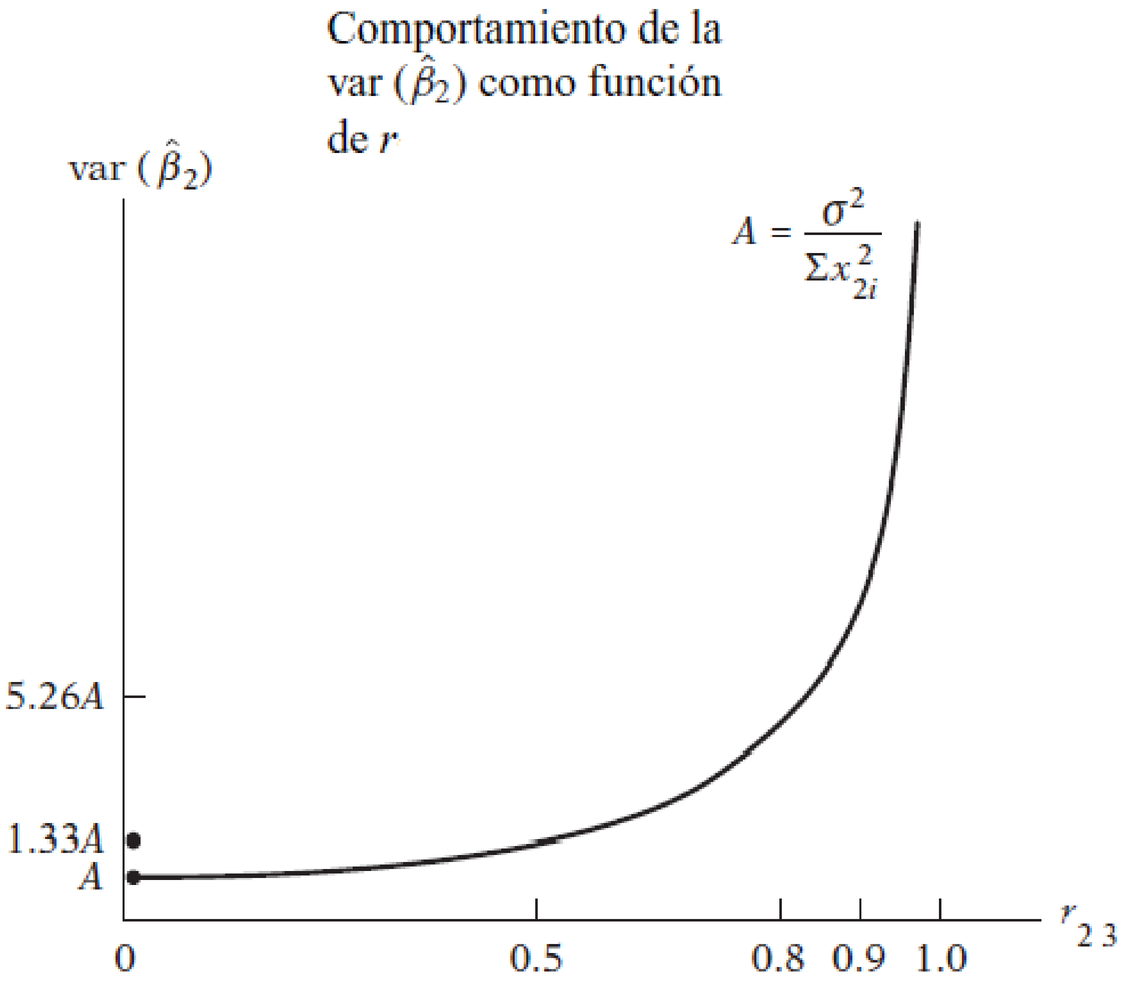
c) Colinealidad moderada

d) Colinealidad alta

e) Colinealidad muy alta

Si la multicolinealidad es perfecta los coeficientes de regresión de las variables *X* son indeterminados, y sus errores estándar, infinitos.

Si la multicolinealidad es menos que perfecta, los coeficientes de regresión, aunque sean indeterminados, poseen grandes errores estándar (en relación con los coeficientes mismos), lo cual significa que los coeficientes no pueden ser estimados con gran precisión o exactitud.



La velocidad con que se incrementan las varianzas y covarianzas se ve con el **factor inflacionario de la varianza** (**FIV**), que se define como

El FIV muestra la forma como la varianza de un estimador se ***infla*** por la presencia de la multicolinealidad

$$FIV = \frac{1}{(1 - r_{X1X2})}$$

FIV	Interpretación
1	No hay multicolinealidad
1 - 5	Baja a moderada multicolinealidad
> 5	Potencial multicolinealidad
> 10	Alta multicolinealidad, preocupante

Detección de la multicolinealidad

1. Una *R*2 elevada pero pocas razones *t* significativas
2. Altas correlaciones entre parejas de regresoras
3. Regresiones auxiliares
4. Diagrama de dispersión
5. Valores propios e índice de condición.

$$K = \frac{\text{Valor propio máximo}}{\text{Valor propio mínimo}}$$

K=número de condición

$$IC = \sqrt{K}$$

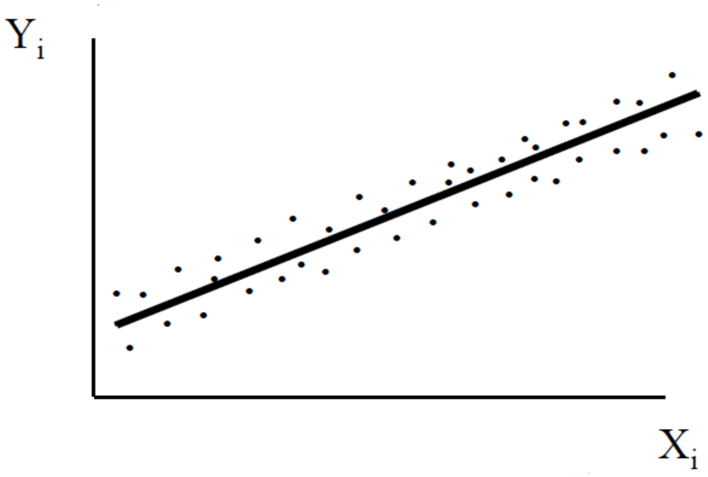
Heterocedasticidad

Un supuesto importante del modelo clásico de regresión lineal (supuesto 4) es que las perturbaciones E*i* que aparecen en la función de regresión poblacional son homocedásticas; es decir, que todas tienen la misma varianza. Se examina la validez de este supuesto y también lo que sucede si no se cumple. T20.xls

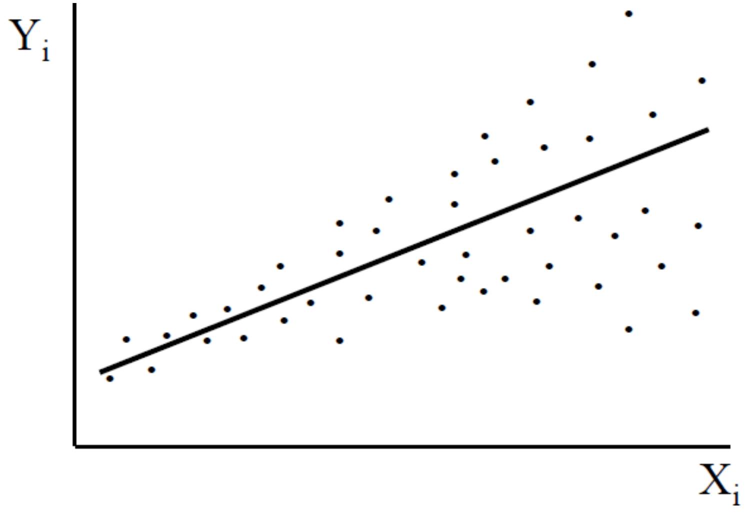
$$E(\varepsilon_i^2) = \sigma^2 \qquad i=1,2,....n$$

$$E(\varepsilon_i^2) = \sigma_i^2$$

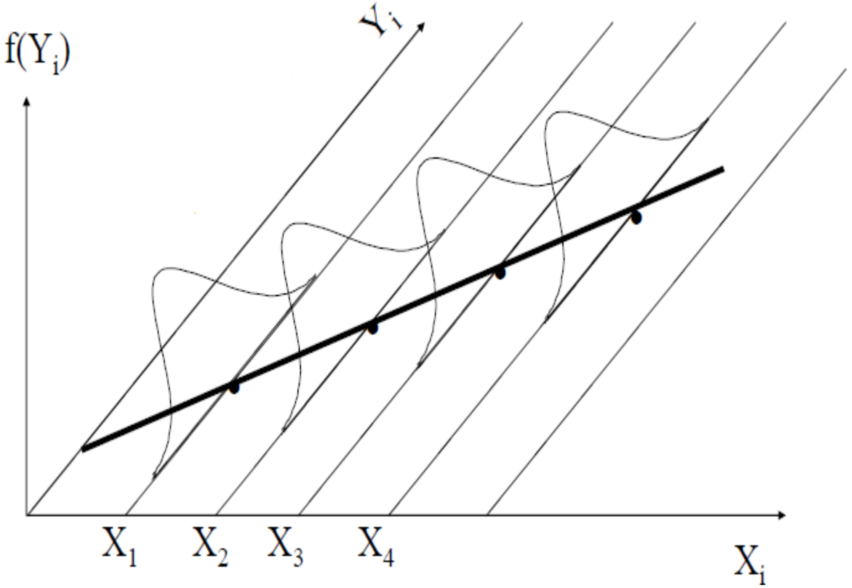
Varianza constante



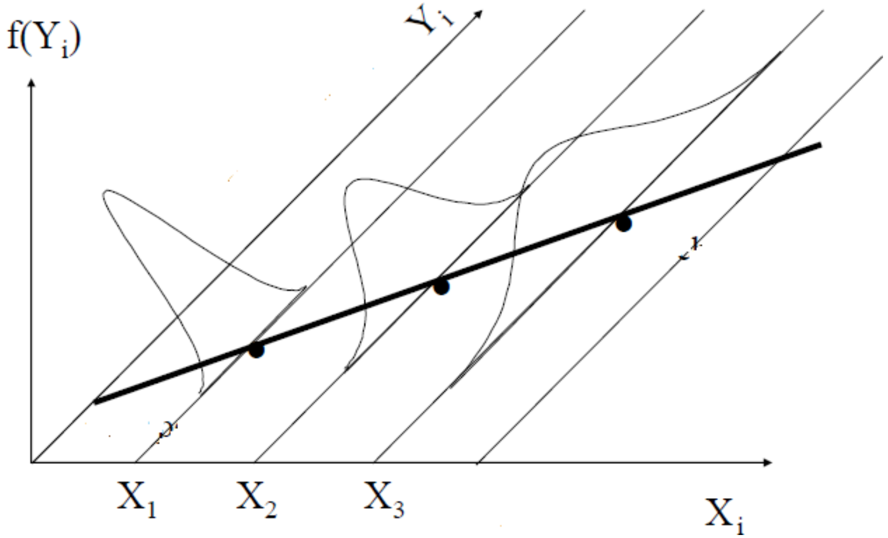
Heteroscedasticidad



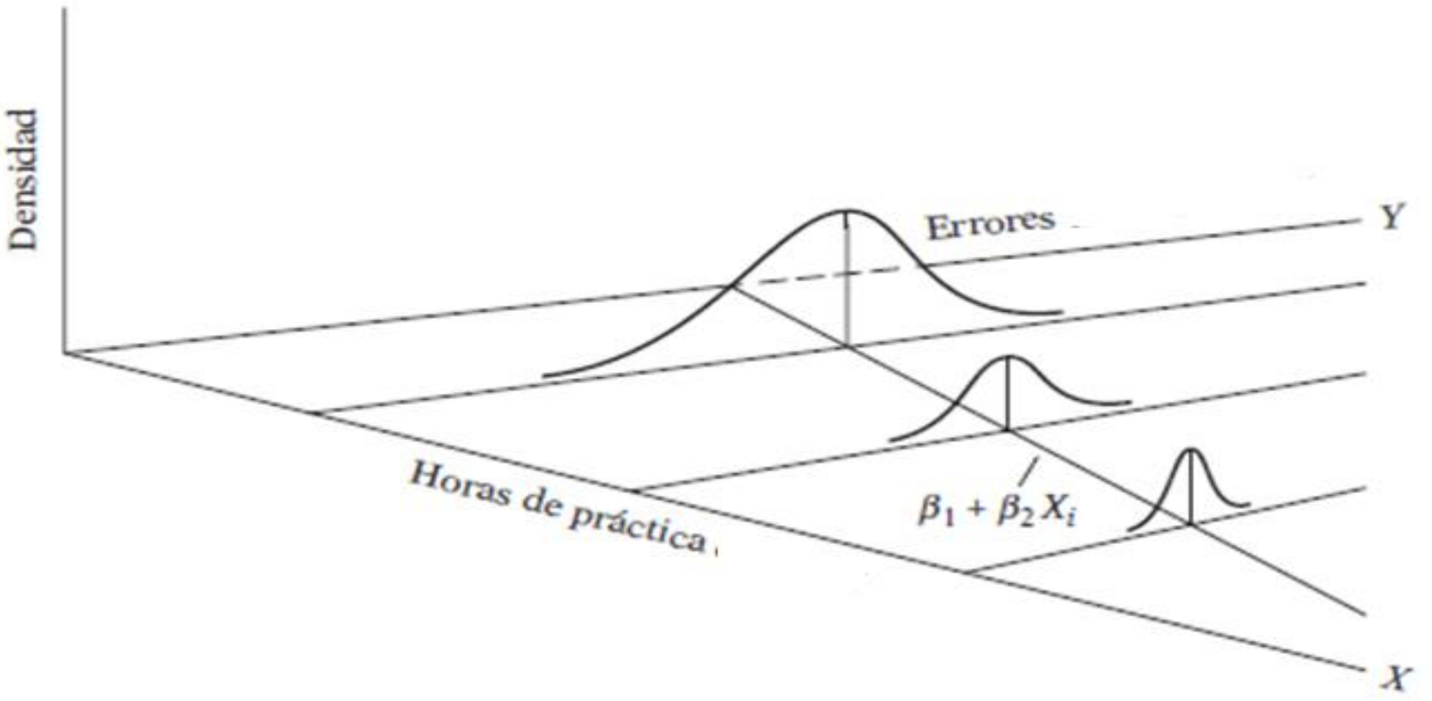
Varianza constante



Heteroscedasticidad



Aprendizaje de los errores



Consecuencias de la heteroscedasticidad

- La principal consecuencia de la heteroscedasticidad es que los estimadores de MCO pierden eficiencia
- La construcción de los intervalos de confianza de los estimadores utiliza el error estándar de los errores, en consecuencia no son apropiados
- La prueba t-Student pierde potencia por que también utiliza el error estándar del estimador

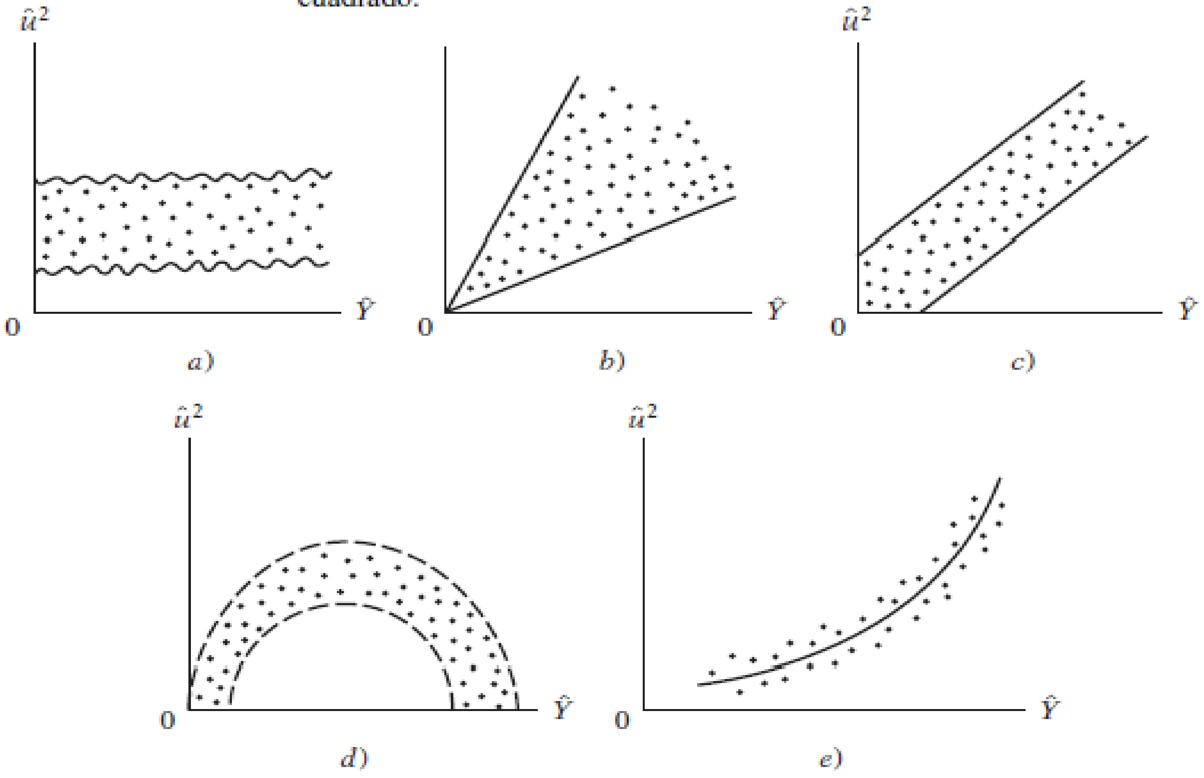
Detección de la heterocedasticidad

Métodos informales

Naturaleza del problema

Método gráfico

Patrones hipotéticos de
los residuos estimados al
cuadrado.



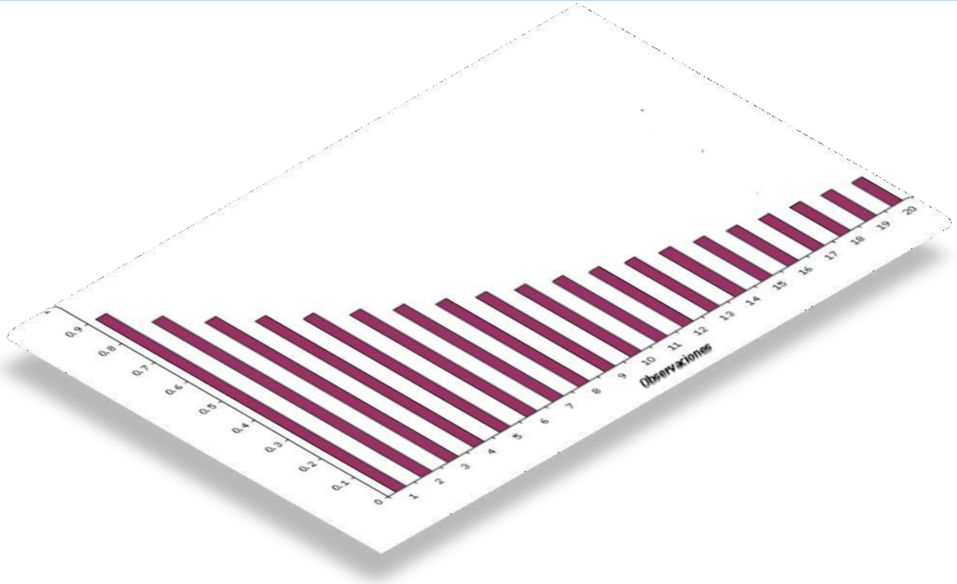
Detección de la heterocedasticidad

Métodos formales

Prueba Breusch-Pagan-Godfrey (Consultar este test)

Prueba general de heteroscedasticidad de White

AUTOCORRELACIÓN



SUPUESTO: La covarianza de los términos de error es igual a cero

$$Cov(\varepsilon_i \varepsilon_j) = 0$$

No existe autocorrelación. Los términos de error son estadísticamente independientes, no existe relación entre los errores

Cuando el supuesto no se cumple el modelo presenta problemas de Autocorrelación

$$Cov(\varepsilon_i \varepsilon_j) \neq 0$$

Se asume que la relación entre los errores describe un proceso autorregresivo de orden uno $AR(1)$

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

Donde $|\rho| < 1$

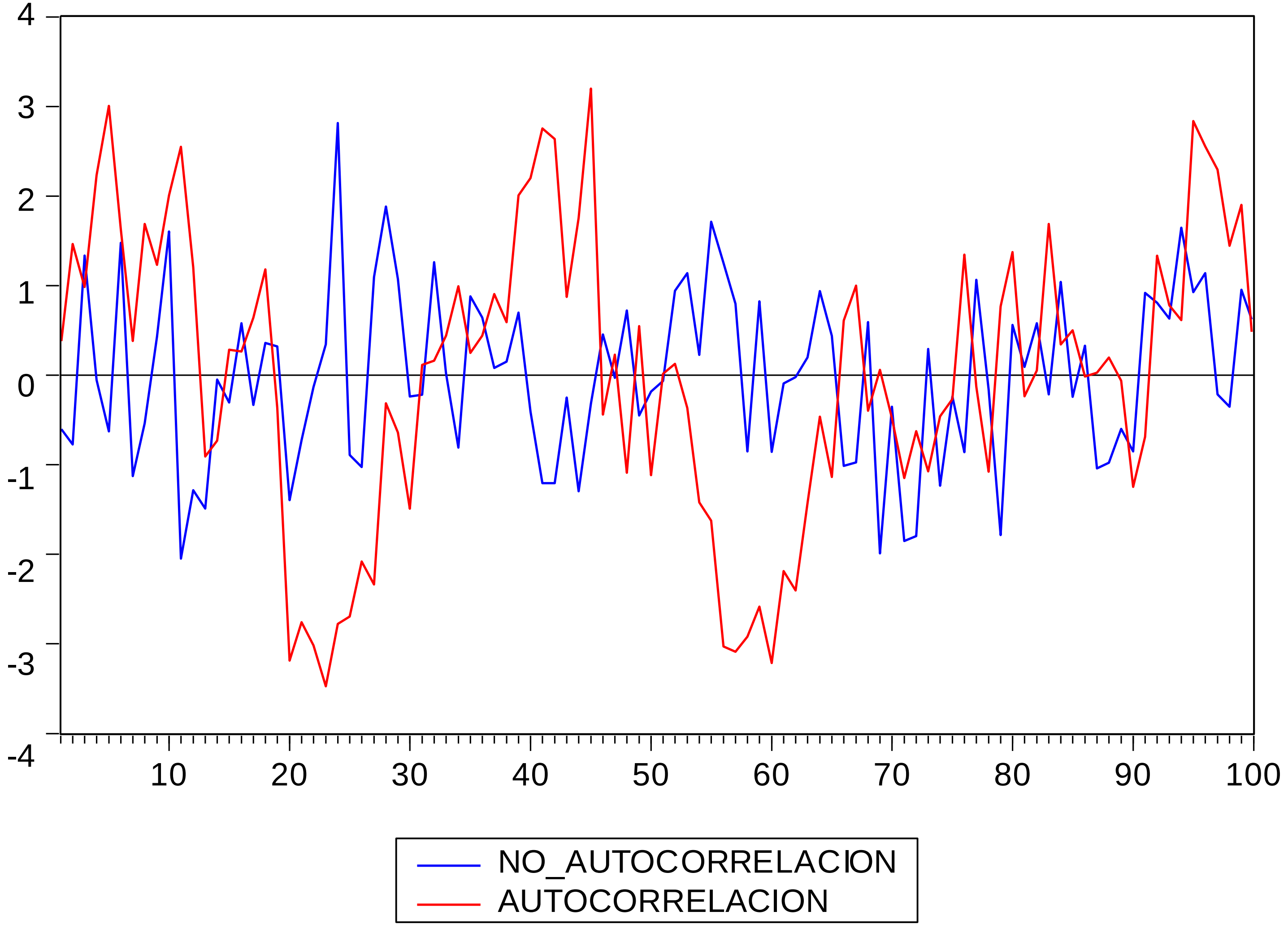
Los supuestos $\mathrm{E}(\nu_t) = 0$ $\mathrm{Var}(\nu_t) = \sigma^2$

No están correlacionados los términos

$$E(\varepsilon_{t-1}\upsilon_t) = 0$$

Cúales son las características del término de error bajo el problema de autocorrelación:

Comparativo errores con y sin autocorrelación



Estadístico Durbin-Watson

Testing for Serial Correlation in Least Squares Regression: I
Author(s): J. Durbin and G. S. Watson
Source: *Biometrika*, Vol. 37, No. 3/4 (Dec., 1950), pp. 409-428

Testing for Serial Correlation in Least Squares Regression. II
Author(s): J. Durbin and G. S. Watson
Source: *Biometrika*, Vol. 38, No. 1/2 (Jun., 1951), pp. 159-177

Se asume que los errores dependen del periodo anterior

$$\hat{\varepsilon}_t = \rho\hat{\varepsilon}_{t-1} + \upsilon_t$$

La hipótesis nula (H_0) es que no existe autocorrelación.

$$H_0 : \rho = 0$$

La hipótesis alternativa es que el parámetro rho sea diferente de cero

$$H_a : \rho \neq 0$$

Si rho es diferente de cero se puede plantear el caso de que $\rho < 0$, también $\rho > 0$. En este contexto Durbin y Watson plantean un estadístico cuya distribución tenga una hipótesis alternativa con dos alternativas

Durbin y Watson proponen un estadístico cuya distribución permita manejar dos límites: uno superior y otro inferior

El estadístico propuesto es el denominado *DW* estadístico Durbin Watson, que se define como la razón de la suma del cuadrado de la primera diferencia de los residuales con respecto a la suma del cuadrado de los residuales

$$DW = \frac{\sum_{t=2}^{T}(\hat{\varepsilon}_t - \varepsilon_{t-1})^2}{\sum_{t=1}^{T}\hat{\varepsilon}_t^2}$$

Las zonas de H_0 y de la hipótesis alternativa de se definen a partir de los límites inferior y superior

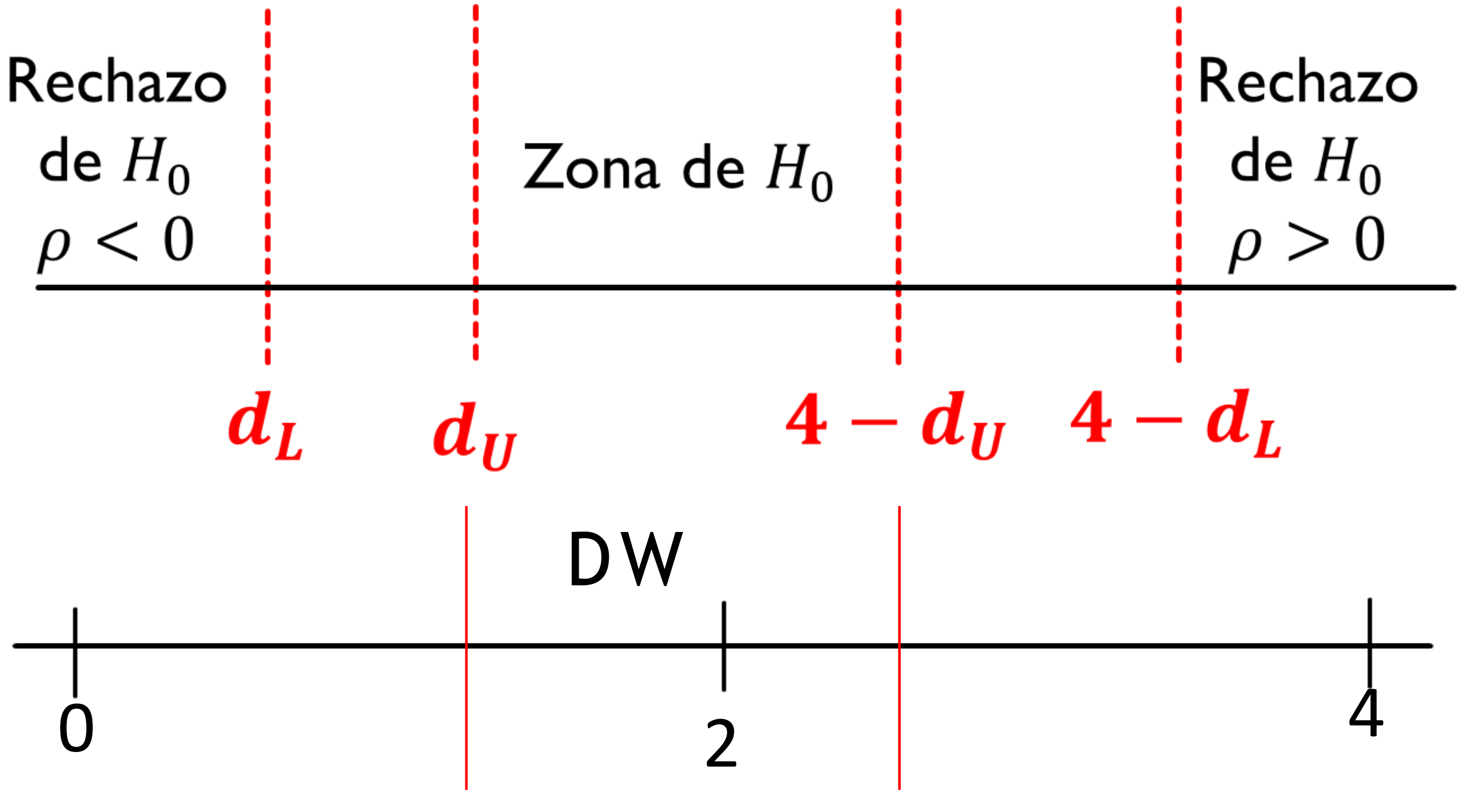


Tabla Estadístico Durbin-Watson

Table 4. *Significance points of d_L and d_U: 5 %*

n	$k'=1$		$k'=2$		$k'=3$		$k'=4$		$k'=5$	
	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
15	1·08	1·36	0·95	1·54	0·82	1·75	0·69	1·97	0·56	2·21
16	1·10	1·37	0·98	1·54	0·86	1·73	0·74	1·93	0·62	2·15
17	1·13	1·38	1·02	1·54	0·90	1·71	0·78	1·90	0·67	2·10
18	1·16	1·39	1·05	1·53	0·93	1·69	0·82	1·87	0·71	2·06
19	1·18	1·40	1·08	1·53	0·97	1·68	0·86	1·85	0·75	2·02
20	1·20	1·41	1·10	1·54	1·00	1·68	0·90	1·83	0·79	1·99
21	1·22	1·42	1·13	1·54	1·03	1·67	0·93	1·81	0·83	1·96
22	1·24	1·43	1·15	1·54	1·05	1·66	0·96	1·80	0·86	1·94
23	1·26	1·44	1·17	1·54	1·08	1·66	0·99	1·79	0·90	1·92
24	1·27	1·45	1·19	1·55	1·10	1·66	1·01	1·78	0·93	1·90
25	1·29	1·45	1·21	1·55	1·12	1·66	1·04	1·77	0·95	1·89
26	1·30	1·46	1·22	1·55	1·14	1·65	1·06	1·76	0·98	1·88
27	1·32	1·47	1·24	1·56	1·16	1·65	1·08	1·76	1·01	1·86
28	1·33	1·48	1·26	1·56	1·18	1·65	1·10	1·75	1·03	1·85
29	1·34	1·48	1·27	1·56	1·20	1·65	1·12	1·74	1·05	1·84
30	1·35	1·49	1·28	1·57	1·21	1·65	1·14	1·74	1·07	1·83
31	1·36	1·50	1·30	1·57	1·23	1·65	1·16	1·74	1·09	1·83
32	1·37	1·50	1·31	1·57	1·24	1·65	1·18	1·73	1·11	1·82
33	1·38	1·51	1·32	1·58	1·26	1·65	1·19	1·73	1·13	1·81
34	1·39	1·51	1·33	1·58	1·27	1·65	1·21	1·73	1·15	1·81
35	1·40	1·52	1·34	1·58	1·28	1·65	1·22	1·73	1·16	1·80
36	1·41	1·52	1·35	1·59	1·29	1·65	1·24	1·73	1·18	1·80