

## Artículo

# Estudio de Monte Carlo para comparar 8 pruebas de normalidad sobre residuos de mínimos cuadrados ordinarios en presencia de procesos autorregresivos de primer orden



Julio César Alonso<sup>a</sup> y Sebastián Montenegro<sup>b,\*</sup>

<sup>a</sup> Profesor tiempo completo, Director del CIENFI – Universidad Icesi, Cali, Colombia

<sup>b</sup> Asistente de investigación, CIENFI – Universidad Icesi, Cali, Colombia

## INFORMACIÓN DEL ARTÍCULO

*Historia del artículo:*

Recibido el 17 de junio de 2014

Aceptado el 19 de diciembre de 2014

On-line el 7 de marzo de 2015

*Códigos JEL:*

C12

C15

C90

*Palabras clave:*

Pruebas de normalidad

Autocorrelación

Simulación de Monte Carlo

Tamaño estadístico

Poder estadístico

## RESUMEN

Este estudio tiene como objetivo evaluar el poder y tamaño de 8 pruebas de normalidad en la presencia de errores autorregresivos de orden uno y diferentes tamaños de muestra. Para lograr este objetivo se emplean simulaciones de Monte Carlo evaluando las siguientes pruebas: Cramér-von Mises, Anderson-Darling, Lilliefors, Pearson, Shapiro-Wilk, Shapiro-Francia, Jarque-Bera y D'Agostino-Pearson. Los resultados muestran 4 aspectos importantes: primero, el efecto de la autocorrelación sobre el tamaño y el poder de las pruebas es asimétrico. Segundo, el tamaño de todas las pruebas se distorsionan en presencia de autocorrelación fuerte. Tercero, ninguna de las pruebas tiene un mejor poder que las demás. Cuarto, cuando la muestra es pequeña, las pruebas de normalidad estudiadas tienen un poder muy bajo.

© 2015 Universidad ICESI. Publicado por Elsevier España, S.L.U. Este es un artículo Open Access bajo la licencia CC BY (<http://creativecommons.org/licenses/by/4.0/>).

## A Monte Carlo Study to compare 8 normality tests for least-squares residuals following a first order autoregressive process

## ABSTRACT

The objective of this study is to assess the statistical power and size of 8 normality tests in presence of first-order autoregressive errors and different simple sizes. Using a Monte Carlo experiment, the following tests were compared: Cramér-von Mises, Anderson-Darling, Lilliefors, Pearson, Shapiro-Wilk, Shapiro-Francia, Jarque-Bera and D'Agostino-Pearson. Our results show 4 relevant findings: First, an asymmetrical effect of autocorrelation on the power and size of the tests. Second, the statistical size of all tests is affected by the autocorrelation. Third, none of the tests has greater power than the others. Fourth, the power of the normality test decreases as sample size decreases.

© 2015 Universidad ICESI. Published by Elsevier España, S.L.U. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

*JEL classification:*

C12

C15

C90

*Keywords:*

Normality tests

Autocorrelation

Monte Carlo experiment

Statistical size

Statistical power

\* Autor para correspondencia: Calle 18 # 122-135, Universidad Icesi, Cali, Colombia.

Correo electrónico: [smontenegro@icesi.edu.co](mailto:smontenegro@icesi.edu.co) (S. Montenegro).

## Estudo de Monte Carlo para comparar 8 provas de normalidade sobre resíduos de mínimos quadrados ordinários em presença de processos autoregressivos de primeira ordem

R E S U M O

Classificações JEL:

C12  
C15  
C90

Palavras-chave:

Provas de normalidade  
Autocorrelação  
Simulação de Monte Carlo  
Tamanho estatístico  
Poder estatístico

Este estudo tem como objectivo avaliar o poder e tamanho de oito provas de normalidade na presença de erros autorregressivos de ordem um e diferentes tamanhos de amostra. Para alcançar este objectivo empregam-se simulações de Monte Carlo avaliando as seguintes provas: Cramér-von Mises, Anderson-Darling, Lilliefors, Pearson, Shapiro-Wilk, Shapiro-Francia, Jarque-Bera e D'Agostino-Pearson. Os resultados mostram quatro aspectos importantes: primeiro, o efeito da autocorrelação sobre o tamanho e o poder das provas é assimétrico. Segundo, o tamanho de todas as provas distorce-se na presença de autocorrelação forte. Terceiro, nenhuma das provas tem um melhor poder que as demais. Quarto, quando a amostra é pequena, as provas de normalidade estudadas têm um poder muito baixo.

© 2015 Universidad ICESI. Publicado por Elsevier España, S.L.U. Este é um artigo Open Access sob a licença de CC BY (<http://creativecommons.org/licenses/by/4.0/>).

### 1. Introducción

En trabajos econométricos empíricos es común que el investigador se vea en la necesidad de probar si los errores de una regresión provienen de una distribución normal. En este sentido, las pruebas de normalidad son necesarias para saber si es apropiado hacer inferencia sobre los coeficientes estimados de una regresión con muestras pequeñas, para el análisis de la regresión o para crear intervalos de confianza para la proyección de la variable dependiente. En la literatura se encuentran diferentes aproximaciones para probar la normalidad o no de una muestra o de los residuos de una regresión. Por ejemplo, Kilian y Demiroglu (2000) presentan una lista no exhaustiva de casos en los cuales un econométrista estaría interesado en probar la normalidad de los residuos de un modelo estimado. Ahora bien, en los trabajos empíricos se presentan comúnmente 3 procedimientos para determinar la normalidad de una muestra: los gráficos como los histogramas y los gráficos Q-Q, las estadísticas descriptivas como la curtosis y el coeficiente de asimetría y las pruebas formales de normalidad (Razali y Wah, 2011).

Centrando la atención en las pruebas formales de normalidad, se encuentra que estas presentan diferentes formas de acercarse al problema, tales como verificar la curtosis y simetría y compararla con las de una distribución normal o ejecutar pruebas de bondad de ajuste para verificar el ajuste de la distribución. Estas diferentes aproximaciones al problema implican que las pruebas presentan comportamientos diferentes en su tamaño y poder con el cambio del tamaño de la muestra o cuando la matriz de varianzas y covarianzas de la muestra se aleja de un comportamiento esférico (varianza constante y no autocorrelación).

El desempeño de las pruebas de normalidad tiene una documentación relativamente amplia. No obstante, la mayor parte de los estudios del poder de las diferentes pruebas se realizan bajo el supuesto de que las observaciones son independientes entre sí y mantienen la misma varianza. Por ejemplo, Shapiro, Wilk y Chen (1968) mostraron que la prueba de Shapiro y Wilk (1964) tiene un mejor desempeño que las pruebas de Kolmogorov-Smirnov (Kolmogorov, 1933), por su parte, Cramér-von Mises (Cramér, 1928) y Anderson y Darling (1952) encuentran que la prueba tiene un mejor desempeño cuando la muestra cumple el supuesto de esfericidad. Asimismo, Thadewald y Büning (2007) demostraron que la prueba de Jarque y Bera (1987) tiene un alto poder frente a distribuciones simétricas o ligeramente sesgadas que tengan colas largas y que cumplan el supuesto de esfericidad. Sin embargo, estos autores muestran que cuando la distribución tiene colas cortas es recomendable emplear modificaciones a la pruebas de Cramér-von Mises o Shapiro-Wilk.

Por otra parte, Razali y Wah (2011), empleando simulaciones de Monte Carlo, compararon el comportamiento de 4 pruebas

de normalidad: Shapiro y Wilk (1964), Kolmogorov-Smirnov (Kolmogorov, 1933), Lilliefors (1967) y Anderson y Darling (1952). Para lo anterior, los autores diseñaron un experimento en el que se generaban muestras con observaciones independientes y homocedásticas provenientes de 7 diferentes distribuciones y 16 diferentes tamaños de muestras. De esta forma, Razali y Wah (2011) encontraron que la prueba de Shapiro y Wilk (1964) es la que presenta el poder más grande para todas las distribuciones y tamaños de muestras evaluadas.

Referente al estudio de Mbah y Paothong (2014), también realizaron un experimento de Monte Carlo para comparar el desempeño de la prueba de Shapiro-Francia con Kolmogorov-Smirnov (Kolmogorov, 1933), Anderson y Darling (1952), Cramér-von Mises (Cramér, 1928), Lilliefors (1967), Shapiro y Wilk (1964), Chi-cuadrado de Pearson (Pearson, 1900), Jarque y Bera (1987) y D'Agostino (D'Agostino y Pearson, 1973). El experimento de Monte Carlo de estos autores implicó generar muestras de 5 distribuciones diferentes y 12 tamaños, cuyas observaciones son independientes y homocedásticas. Dado lo anterior, Mbah y Paothong (2014) encontraron que la prueba de Shapiro-Francia se comporta mejor que las otras 8 empleadas.

A partir de lo anterior, hasta donde llega el conocimiento de los autores del presente artículo, solamente existe un estudio reciente que compara pruebas de normalidad en presencia de errores autorregresivos. Este se refiere al trabajo de Lobato y Velasco (2004) en el que comparan el comportamiento de una prueba sugerida por ellos y la prueba Epps (1987), con un test para curtosis y simetría, para muestras producidas por un proceso autorregresivo de orden 1 (AR(1)). Los autores estudian el tamaño de las 2 pruebas empleando simulaciones de Monte Carlo en las que las muestras provienen de un proceso AR(1) con errores normales y 8 diferentes valores de la autocorrelación entre las observaciones. Lobato y Velasco (2004) encuentran que su prueba presenta distorsiones grandes en el tamaño cuando la muestra es pequeña; no obstante, en ese estudio no se presenta un análisis comparativo con las diferentes pruebas de normalidad, incluyendo la prueba de Jarque y Bera (1980) que considera datos que podrían estar autocorrelacionados y ser heterocedásticos.

En este sentido, el presente estudio tiene la intención de continuar la línea de investigación de Lobato y Velasco (2004), al evaluar el comportamiento de diferentes pruebas de normalidad en la presencia de errores autocorrelacionados. En este se pretende extender los resultados de Lobato y Velasco (2004) en 2 sentidos: primero, se estudia tanto el tamaño de las pruebas como el poder; y segundo, se comparan más pruebas siguiendo a Mbah y Paothong (2014).

Como lo documentan Mbah y Paothong (2014) y Razali y Wah (2011), existen más de 40 pruebas de normalidad. Tal vez la primera de estas fue la propuesta por Pearson (1894) y desde entonces se encuentran en la literatura diferentes tests que intentan

aproximarse al problema de identificar si una muestra proviene de una distribución normal o no mediante diferentes vías. De esta forma, las pruebas se pueden clasificar de diferentes maneras; por ejemplo, según Shapiro et al. (1968), se pueden clasificar en tests de distancia (*distance tests*) y de bondad de ajuste (*goodness of fit*).

Seguendo a Mbah y Paothong (2014), en el presente estudio comparativo se emplearán 8 de las pruebas de normalidad más populares en la literatura, recurriendo a las siguientes pruebas de distancia: Cramér-von Mises (Cramér, 1928), Anderson y Darling (1952) y Lilliefors (1967). La prueba Anderson y Darlin emplea el estadístico Cramér-von Mises ponderado por su función de probabilidad acumulada y el test de Lilliefors corresponde a una modificación de la prueba Kolmogorov-Smirnov para el caso en el que los parámetros poblacionales son estimados. Por eso, en el estudio no se incluyó la prueba de Kolmogorov-Smirnov (Kolmogorov, 1933).

Adicionalmente, se consideran las siguientes pruebas de bondad de ajuste: Pearson (1900), Shapiro y Wilk (1964), Francia y Shapiro (1972), Jarque y Bera (1987) y D'Agostino (D'Agostino y Pearson, 1973; D'Agostino, Belanger y D'Agostino, 1990).

Así, el objetivo de este documento es determinar el poder y el tamaño de las 8 pruebas de normalidad antes relacionadas, cuando estas se aplican a residuos estimados a partir de una regresión múltiple por mínimos cuadrados ordinarios (MCO) en presencia de autocorrelación en los errores.

Este documento consta de 3 secciones más. La segunda sección describe el diseño del ejercicio de Monte Carlo para determinar el tamaño y poder de las diferentes pruebas. La tercera sección presenta los resultados del ejercicio. La última sección concluye con unos comentarios finales y las implicaciones prácticas de los resultados.

## 2. Diseño del experimento

Antes de describir el experimento, vale la pena recapitular las 8 pruebas que se evaluarán (nombres abreviados entre paréntesis): i) Cramér-von Mises (CM), ii) Anderson y Darling (1952) (WCM), iii) Lilliefors (1967) (LKS), iv) Pearson (1900) (Chi), v) Shapiro y Wilk (1964) (SW), vi) Francia y Shapiro (1972) (SF), vii) Jarque y Bera (1987) (JB) y viii) D'Agostino (D'Agostino y Pearson, 1973) (K).

Para verificar el comportamiento de las 8 pruebas del estudio se diseñó un experimento de Monte Carlo que se ejecuta empleando el lenguaje estadístico R versión 3.0.1 (R Core Team, 2013). El procedimiento para las simulaciones implica los siguientes pasos. Primero, se genera un vector aleatorizado y de dimensiones  $T \times 1$  a partir del siguiente proceso generador de datos:

$$\mathbf{y}_{T \times 1} = \begin{bmatrix} \mathbf{1}_{T \times 1} & \mathbf{X}_{T \times 1} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \mathbf{v}_{T \times 1} \quad (1)$$

Donde  $\mathbf{x}$  corresponde al vector de observaciones de una variable fija (no varía entre iteraciones) que es generado a partir de una distribución uniforme entre 0 y 1 y  $\mathbf{v}$  es el vector aleatorizado de error que es homocedástico y sigue un proceso AR(1), es decir,  $v_t = \rho v_{t-1} + \varepsilon_t$  donde  $\varepsilon_t$  es un ruido blanco. El segundo paso es estimar la regresión de  $y$  en función de  $[\mathbf{1} \quad \mathbf{x}]$  y encontrar el vector de residuos  $\hat{\mathbf{v}}_{T \times 1} = \mathbf{y}_{T \times 1} - \hat{\beta}[\mathbf{1} \quad \mathbf{x}]$ . El tercer paso es emplear las 8 pruebas de normalidad seleccionadas para determinar si se puede rechazar la hipótesis nula de que  $\hat{\mathbf{v}}_{T \times 1}$  proviene de una distribución normal. Para cada prueba se guarda la decisión de rechazar o no la hipótesis nula. Estos 3 pasos se repiten 10.000 veces.

El experimento de Monte Carlo se realiza para diferentes distribuciones de  $\mathbf{v}$ . Se considera la distribución estándar normal ( $N[0,1]$ ), la distribución t de Student con 3, 4, 5 y 6 grados de libertad, la distribución Chi-cuadrado con grados de libertad 1, 5, 10, 15

y 20 y una distribución uniforme entre cero y uno ( $U[0,1]$ ). En total se consideran 6 distribuciones simétricas y 5 asimétricas.

Por otro lado, para cada distribución se consideran diferentes grados de autocorrelación. Específicamente,  $\rho$  tomará valores entre  $-0,9$  y  $0,9$  a una razón de cambio de  $0,1$ ; es decir, el vector de coeficientes de autocorrelación será  $\rho = (-0,9; -0,8; \dots; 0; \dots; 0,8; 0,9)$ . De esta manera, se tendrá un conjunto de simulaciones con errores autocorrelacionados negativamente ( $\rho < 0$ ), otro conjunto donde no están autocorrelacionados ( $\rho = 0$ ) y un tercero con errores autocorrelacionados positivamente ( $\rho > 0$ ).

Finalmente, para cada combinación de distribución y de  $\rho$  se consideran 5 diferentes tamaños de muestra ( $T$ ); en especial, se tendrá que  $T = 25, 60, 100, 200, 500$ .

Para el caso en que las simulaciones se realizan sobre errores que siguen una distribución diferente a la normal, el poder de cada prueba se calcula como la proporción de veces que la prueba rechaza la hipótesis nula de normalidad (usando un nivel de significación del 5%)<sup>1</sup>. Para el caso en que las simulaciones se realicen con errores normales, realizando el mismo procedimiento, se obtiene el tamaño de la prueba (probabilidad de cometer el error tipo I).

## 3. Resultados

En esta sección se discutirán los resultados de las simulaciones de Monte Carlo, empleando una subsección para cada una de las 4 familias de distribuciones consideradas: distribución normal, distribución t de Student (y los diferentes grados de libertad), la distribución Chi-cuadrado (y los diferentes grados de libertad) y la distribución uniforme.

### 3.1. Distribución normal

En el caso en el que los errores provienen de una distribución estándar normal, todas las pruebas presentan un tamaño cercano al nominal ( $\alpha = 0,05$ ) cuando no hay autocorrelación ( $\rho = 0$ ) y la muestra es relativamente grande (superior a 100) (tabla 1). No obstante, la prueba JB presenta un tamaño levemente más bajo que el nominal cuando la muestra es de tamaño 200. Para el caso de no autocorrelación y muestras pequeñas, se encuentra que con la prueba JB la probabilidad de cometer el error tipo I es menor a  $0,05$ . En general, la prueba JB es la que presenta el menor tamaño para las otras muestras. La prueba K tiene la probabilidad de cometer el error tipo I más grande para muestras de 25 y 60 observaciones. Es decir, la probabilidad de rechazar erróneamente la hipótesis nula de normalidad es más grande que lo teóricamente esperado ( $\alpha = 0,05$ ). Referente a las demás pruebas, estas presentan un tamaño entre  $0,05$  y  $0,058$  que corresponde con el diseño del test (tabla 1).

Cuando se presenta una autocorrelación positiva, la probabilidad de rechazar la hipótesis nula de normalidad, siendo esta verdadera, es más grande que el tamaño nominal y que el tamaño que se presenta con la autocorrelación de igual valor absoluto pero con signo negativo; es decir, se presenta una trayectoria asimétrica en el tamaño de la prueba entre el caso de autocorrelación positiva y autocorrelación negativa. Además, el tamaño de la prueba es más grande a medida que aumenta la muestra (tabla 1 y fig. 1).

Cuando la autocorrelación negativa es más fuerte, se observa un resultado similar al caso en el que la autocorrelación es positiva. Sin embargo, el tamaño de las pruebas en presencia de autocorrelación positiva es mayor que cuando existe autocorrelación negativa (tabla 1 y fig. 1). La prueba que presenta una menor distorsión en su

<sup>1</sup> Es decir, si  $\gamma$  corresponde al número de veces que se rechazó la hipótesis nula de normalidad para cada prueba, entonces el poder de cada prueba se calcula como  $H = \gamma/10.000$ .

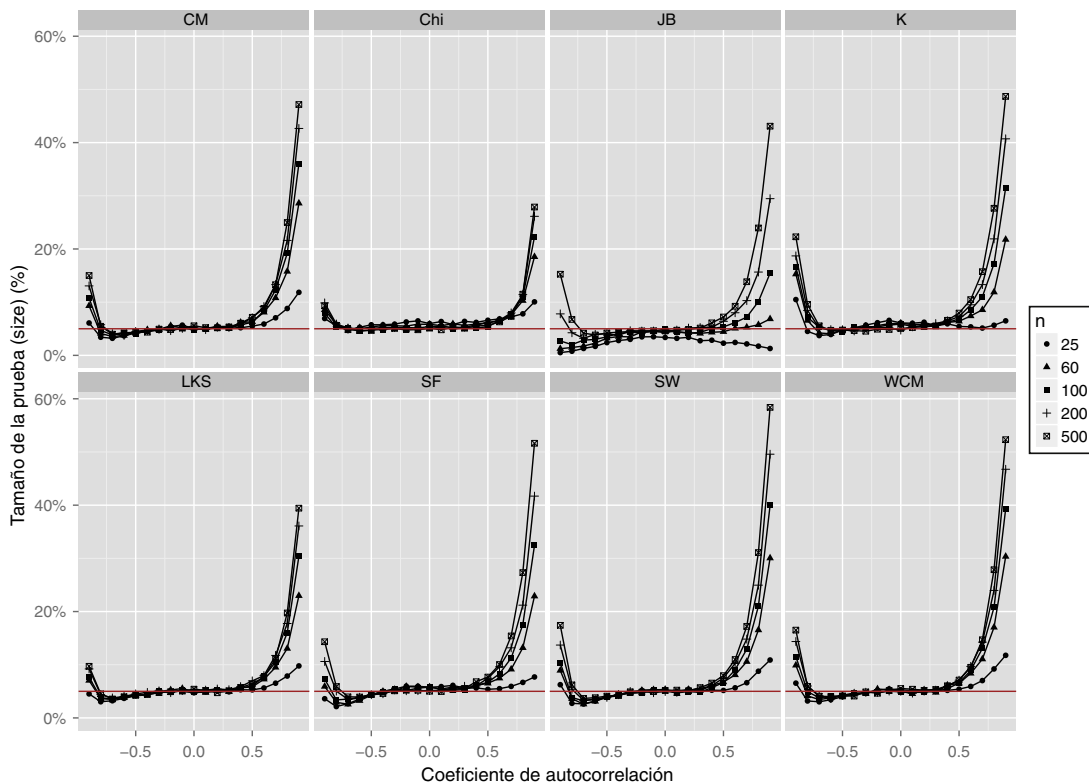
**Tabla 1**  
Tamaño de las pruebas de normalidad

$\rho$	Muestra (T)	Pruebas de normalidad							
		CM	K	JB	LKS	WCM	Chi	SF	SW
-0,9	25	0,061	0,105	<b>0,005</b>	0,045	0,065	0,069	0,036	0,063
	60	0,093	0,153	<b>0,013</b>	0,070	0,099	0,086	0,059	0,089
	100	0,108	0,167	<b>0,027</b>	0,076	0,115	0,079	0,073	0,104
	200	0,131	0,187	<b>0,078</b>	0,092	0,144	0,099	0,106	0,137
	500	0,150	0,223	0,153	0,097	0,165	<b>0,093</b>	0,143	0,174
-0,4	25	0,046	0,052	<b>0,028</b>	0,044	0,045	0,058	0,050	0,042
	60	0,048	0,051	<b>0,035</b>	0,044	0,047	0,056	0,050	0,044
	100	0,047	0,053	<b>0,041</b>	0,044	0,046	0,048	0,049	0,045
	200	0,046	0,046	<b>0,038</b>	0,048	0,044	0,052	0,045	0,044
	500	0,044	0,047	0,044	0,044	<b>0,041</b>	0,052	0,046	0,042
0	25	0,050	0,061	<b>0,034</b>	0,047	0,053	0,060	0,058	0,053
	60	0,051	0,061	<b>0,045</b>	0,050	0,050	0,058	0,058	0,053
	100	<b>0,048</b>	0,059	0,049	0,049	0,049	0,051	0,058	0,053
	200	0,052	0,048	<b>0,043</b>	0,051	0,050	0,054	0,051	0,050
	500	0,056	0,051	<b>0,046</b>	0,054	0,055	0,051	0,050	0,051
0,4	25	0,052	0,059	<b>0,029</b>	0,053	0,052	0,062	0,056	0,052
	60	0,062	0,062	<b>0,043</b>	0,057	0,061	0,059	0,061	0,061
	100	0,057	0,062	<b>0,050</b>	0,057	0,055	0,051	0,061	0,057
	200	0,059	0,065	<b>0,053</b>	0,059	0,059	0,057	0,060	0,061
	500	0,061	0,066	0,061	0,057	0,060	<b>0,055</b>	0,068	0,066
0,9	25	0,119	0,065	<b>0,013</b>	0,098	0,118	0,101	0,077	0,109
	60	0,286	0,218	<b>0,069</b>	0,230	0,304	0,185	0,229	0,301
	100	0,361	0,315	<b>0,155</b>	0,304	0,393	0,223	0,324	0,400
	200	0,427	0,407	0,295	0,361	0,468	<b>0,261</b>	0,417	0,496
	500	0,472	0,487	0,431	0,394	0,523	<b>0,279</b>	0,516	0,584

En la tabla se reporta en negrita y cursiva el tamaño de la prueba más bajo para cada muestra. Por ejemplo, se resalta en la primera línea de la tabla el número 0,005 que corresponde al tamaño de la prueba JB, que es el más pequeño entre todas las pruebas cuando se emplean 25 observaciones y un  $\rho = -0,9$ . Fuente: elaboración propia.

tamaño es la prueba JB. No obstante, su tamaño puede llegar a ser de 0,431 en presencia de errores con una autocorrelación positiva muy fuerte ( $\rho = 0,9$ ). En otras palabras, la probabilidad de cometer el error de rechazar la hipótesis nula cuando es verdadera alcanza el 43,1%, en presencia de autocorrelación positiva fuerte.

Cuando el coeficiente de autocorrelación ( $\rho$ ) —en valor absoluto— crece, y la muestra se hace más grande, las pruebas experimentan una gran distorsión en su tamaño. Sin embargo, los resultados muestran que las pruebas mantienen un tamaño relativamente cercano al nominal para el intervalo  $\rho \in [-0,5; 0,5]$ .



**Figura 1.** Tamaño de las pruebas de normalidad sobre errores con distribución normal. Fuente: elaboración propia.

3.2. Distribución t de Student

En la [tabla 2](#) y la [figura 2](#) se reportan los resultados del ejercicio de simulación cuando los errores provienen de una distribución t de Student con 3 grados de libertad. De manera similar, en la [tabla 3](#) y la [figura 3](#) se reportan los resultados cuando los grados de libertad son 6. En general, los resultados muestran que el poder disminuye drásticamente a medida que la autocorrelación se hace más fuerte (ya sea positiva o negativa, es decir,  $\rho$  en valor absoluto crece). Asimismo, todas las pruebas muestran una caída en su poder cuando el tamaño de la muestra es relativamente pequeño (25 observaciones).

Por otro lado, para la distribución Chi-cuadrado con 3 grados de libertad, todas las pruebas tienen un poder similar; pero la prueba SF parece tener un poder ligeramente mayor que las demás cuando no hay autocorrelación ([tabla 2](#)).

En general, semejante a los resultados de la distribución normal, las pruebas funcionan relativamente bien cuando  $\rho$  toma valores entre  $\pm 0,5$  ([fig. 2](#)) y las muestras poseen más de 60 observaciones. Sin embargo, SW parece ser afectada en menor medida por la dependencia lineal de los errores en comparación con las demás pruebas. Adicionalmente, al igual que en el caso del tamaño de las pruebas, cuando se presenta una autocorrelación positiva, la probabilidad de rechazar la hipótesis nula de normalidad cuando esta es falsa es más pequeña que cuando existe autocorrelación con el mismo valor pero con signo negativo. Es decir, se presenta un comportamiento asimétrico entre el poder de la prueba cuando la autocorrelación es positiva y cuando la autocorrelación es negativa. Siendo el poder más bajo cuando existe una autocorrelación negativa más fuerte.

Los resultados del poder de la prueba comienzan a cambiar su comportamiento cuando los grados de libertad aumentan. En la sección de anexos se presentan los resultados cuando se consideran 4 y 5 grados de libertad ([anexo, figs. A1 y A2](#)). Para el caso de 6 grados de libertad ([tabla 3](#) y [fig. 3](#)), se puede observar que el poder

de las pruebas es menor que en el caso de una distribución con 3 grados de libertad. Este resultado es apenas natural, pues a medida que crecen los grados de libertad, la distribución t se parece más a la distribución normal. De esta manera, se hace más difícil que la prueba rechace la hipótesis nula de una distribución normal. Es más, para el caso de no autocorrelación ( $\rho = 0$ ), todas las pruebas presentan un poder relativamente bajo para muestras pequeñas. El poder de las pruebas solo mejora cuando la muestra es grande (T=500).

Para los casos en los que existe autocorrelación, las pruebas JB, K y SF muestran el mejor poder relativo, dejando todo lo demás constante. Por el contrario, la prueba Chi es la que presenta el poder más bajo ([fig. 3](#)).

En este caso también se presenta el comportamiento asimétrico del poder de la prueba cuando se comparan los escenarios con autocorrelación positiva y negativa. Por otro lado, se observa un desempeño interesante para autocorrelaciones positivas grandes. El poder de las pruebas comienza a aumentar, después de haber caído, cuando  $\rho$  toma valores superiores a 0,7. Este comportamiento se presenta en los resultados de todas las pruebas empleadas.

3.3. Distribución Chi-cuadrado

El ejercicio de Monte Carlo incluye simulaciones para errores que provienen de una distribución Chi-cuadrado con grados de libertad 1, 5, 10, 15 y 20. En esta subsección se reportan los resultados para el caso de 10 grados de libertad ([tabla 4](#) y [fig. 4](#)), los otros resultados para esta misma distribución con grados de libertad 1, 5, 15 y 20 se encuentran en los anexos ([figs. A3–A6](#)).

En el caso de la distribución Chi-cuadrado con grados de libertad 1 y 5, los resultados son similares a los obtenidos con la distribución t de Student ([figs. A3 y A4](#)). Sin embargo, cuando los grados de libertad son 10, el poder de la prueba es mayor en todos los casos, a diferencia de lo encontrado para la distribución t. Para esta distribución también se observa una caída en el poder a medida que la

**Tabla 2**  
Poder de las pruebas de normalidad para errores con distribución  $t_3$

$\rho$	Muestra (T)	Pruebas de normalidad							
		CM	K	JB	LKS	WCM	Chi	SF	SW
-0,9	25	0,100	<b>0,119</b>	0,018	0,072	0,098	0,100	0,069	0,087
	60	0,169	<b>0,176</b>	0,079	0,124	<b>0,176</b>	0,138	0,145	0,154
	100	0,222	0,223	0,160	0,171	<b>0,254</b>	0,164	0,213	0,215
	200	0,323	0,328	0,318	0,253	<b>0,345</b>	0,220	0,351	0,343
	500	0,501	0,566	0,588	0,409	0,536	0,322	<b>0,596</b>	0,574
-0,4	25	0,241	<b>0,345</b>	0,294	0,196	0,260	0,149	0,341	0,290
	60	0,471	0,598	0,608	0,383	0,567	0,261	<b>0,626</b>	0,562
	100	0,666	0,783	0,808	0,561	0,705	0,367	<b>0,811</b>	0,764
	200	0,894	0,952	0,963	0,812	0,912	0,601	<b>0,964</b>	0,948
	500	0,999	<b>1</b>	<b>1</b>	0	0,999	0,926	<b>1</b>	<b>1</b>
0	25	0,347	<b>0,433</b>	0,374	0,285	0,364	0,206	<b>0,433</b>	0,382
	60	0,643	0,726	0,737	0,544	0,676	0,381	<b>0,762</b>	0,708
	100	0,834	0,883	0,902	0,738	0,855	0,532	<b>0,912</b>	0,884
	200	0,978	0,987	0,991	0,945	0,983	0,799	<b>0,992</b>	0,989
	500	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	0,993	<b>1</b>	<b>1</b>
0,4	25	0,261	<b>0,349</b>	0,290	0,213	0,276	0,161	0,346	0,302
	60	0,506	0,628	0,630	0,416	0,536	0,276	<b>0,647</b>	0,586
	100	0,676	0,781	0,804	0,572	0,707	0,373	<b>0,810</b>	0,761
	200	0,907	0,958	<b>0,968</b>	0,835	0,926	0,604	<b>0,968</b>	0,954
	500	0,999	<b>1</b>	<b>1</b>	0,994	0,999	0,933	<b>1</b>	<b>1</b>
0,9	25	0,235	0,143	0,057	0,199	<b>0,238</b>	0,176	0,192	0,233
	60	0,456	0,344	0,209	0,393	<b>0,481</b>	0,344	0,416	0,477
	100	0,544	0,456	0,358	0,490	<b>0,581</b>	0,407	0,533	<b>0,581</b>
	200	0,651	0,586	0,543	0,591	<b>0,695</b>	0,490	0,675	0,709
	500	0,768	0,752	0,754	0,709	0,815	0,587	0,832	<b>0,845</b>

En la tabla se reporta en negrita y cursiva el tamaño de la prueba más alto para cada tamaño de muestra. Por ejemplo, se resalta en la primera línea de la tabla el número 0,119 que corresponde al poder de la prueba K, que es el más grande entre todas las pruebas cuando se emplean 25 observaciones y un  $\rho = -0,9$ .

Fuente: elaboración propia.

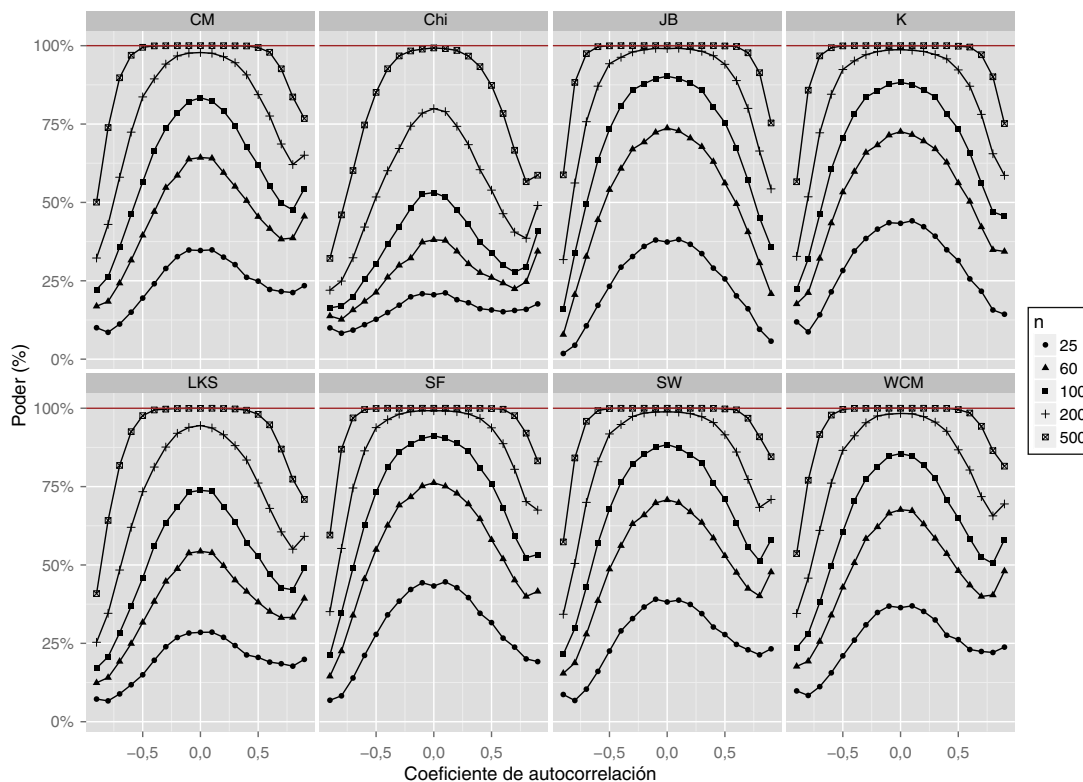


Figura 2. Poder de las pruebas de normalidad sobre errores con distribución  $t_3$ . Fuente: elaboración propia.

muestra disminuye y cada vez que la autocorrelación se hace más negativa. Para el caso de autocorrelación positiva fuerte ( $\rho=0,9$ ) el poder de todas las pruebas es igual o mayor que cuando no hay autocorrelación.

### 3.4. Distribución uniforme

Para esta distribución, el comportamiento del poder de las pruebas de normalidad es relativamente parecido a lo que

Tabla 3 Poder de las pruebas de normalidad para errores con distribución  $t_6$

$\rho$	Muestra (T)	Pruebas de normalidad							
		CM	K	JB	LKS	WCM	Chi	SF	SW
-0,9	25	0,074	<b>0,107</b>	0,010	0,051	0,074	0,073	0,043	0,068
	60	0,111	<b>0,153</b>	0,025	0,079	0,116	0,095	0,079	0,104
	100	0,133	<b>0,163</b>	0,051	0,094	0,140	0,097	0,100	0,119
	200	0,166	<b>0,194</b>	0,123	0,119	0,181	0,116	0,156	0,171
	500	0,203	<b>0,243</b>	0,226	0,143	0,225	0,121	0,233	0,237
-0,4	25	0,104	<b>0,167</b>	0,122	0,084	0,111	0,079	0,159	0,123
	60	0,154	0,274	0,270	0,122	0,176	0,089	<b>0,275</b>	0,227
	100	0,213	0,370	<b>0,393</b>	0,164	0,239	0,097	0,385	0,322
	200	0,341	0,565	<b>0,614</b>	0,242	0,387	0,121	0,596	0,525
	500	0,668	0,882	<b>0,912</b>	0,498	0,733	0,215	0,893	0,856
0	25	0,146	<b>0,219</b>	0,169	0,121	0,159	0,095	0,214	0,175
	60	0,236	0,360	0,359	0,183	0,263	0,116	<b>0,373</b>	0,312
	100	0,333	0,506	<b>0,533</b>	0,249	0,372	0,134	0,530	0,462
	200	0,531	0,712	<b>0,757</b>	0,402	0,583	0,196	0,743	0,685
	500	0,878	0,961	<b>0,973</b>	0,744	0,914	0,390	0,969	0,957
0,4	25	0,122	<b>0,174</b>	0,125	0,101	0,129	0,087	0,166	0,142
	60	0,186	0,292	0,283	0,144	0,202	0,098	<b>0,297</b>	0,251
	100	0,243	0,391	<b>0,407</b>	0,188	0,270	0,107	0,405	0,342
	200	0,374	0,586	<b>0,628</b>	0,277	0,424	0,129	0,612	0,552
	500	0,689	0,884	<b>0,911</b>	0,524	0,754	0,223	0,894	0,863
0,9	25	0,163	0,093	0,028	0,136	<b>0,167</b>	0,129	0,119	0,160
	60	0,334	0,253	0,106	0,277	<b>0,359</b>	0,235	0,288	0,356
	100	0,418	0,353	0,218	0,351	0,454	0,276	0,391	<b>0,458</b>
	200	0,484	0,440	0,358	0,416	<b>0,534</b>	0,320	0,495	0,552
	500	0,553	0,531	0,501	0,472	0,607	0,350	0,609	<b>0,651</b>

En la tabla se pone en negrita y cursiva el tamaño de la prueba más alto para cada tamaño de muestra. Por ejemplo, se resalta en la primera línea de la tabla el número 0,107 que corresponde al poder de la prueba K, que es el más grande entre todas las pruebas cuando se emplean 25 observaciones y un  $\rho=-0,9$ . Fuente: elaboración propia.

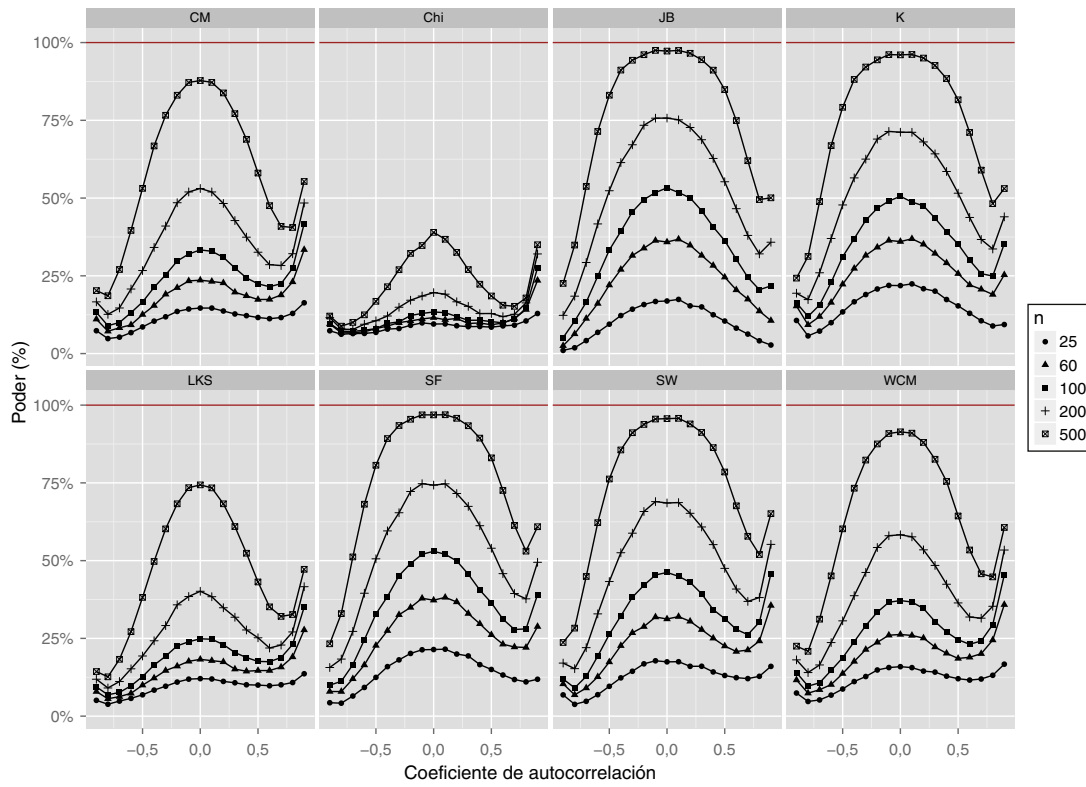


Figura 3. Poder de las pruebas de normalidad sobre errores con distribución  $t_6$ . Fuente: elaboración propia.

ocurre con las anteriores distribuciones (tabla 5 y fig. 5). De esta forma, el poder disminuye cuanto más fuerte es la autocorrelación negativa, mientras que el poder se deteriora en gran medida cuando la muestra es pequeña. Adicionalmente,

a medida que la autocorrelación es más positiva el poder de la prueba mejora. La figura 5 muestra en detalle lo indicado anteriormente para todos los coeficientes de autocorrelación.

Tabla 4 Poder de las pruebas de normalidad para errores con distribución  $\chi^2_{10}$

$\rho$	Muestra (T)	Pruebas de normalidad							
		CM	K	JB	LKS	WCM	Chi	SF	SW
-0,9	25	0,100	<b>0,146</b>	0,006	0,068	0,102	0,088	0,055	0,093
	60	0,125	<b>0,178</b>	0,016	0,085	0,125	0,101	0,077	0,109
	100	0,127	<b>0,175</b>	0,042	0,090	0,133	0,093	0,092	0,117
	200	0,158	<b>0,200</b>	0,103	0,118	0,169	0,108	0,138	0,161
	500	0,180	<b>0,223</b>	0,176	0,124	0,197	0,107	0,183	0,199
-0,4	25	0,086	<b>0,139</b>	0,086	0,081	0,093	0,071	0,122	0,118
	60	0,221	0,323	0,277	0,182	0,253	0,102	0,332	<b>0,339</b>
	100	0,400	0,522	0,491	0,325	0,451	0,145	0,553	<b>0,570</b>
	200	0,734	0,863	0,854	0,606	0,800	0,309	0,887	<b>0,898</b>
	500	0,993	<b>1</b>	<b>1</b>	0,963	0,998	0,749	<b>1</b>	<b>1</b>
0	25	0,120	<b>0,143</b>	0,099	0,099	0,120	0,088	0,137	0,103
	60	0,364	0,398	0,362	0,284	0,388	0,184	<b>0,422</b>	0,394
	100	0,640	0,672	0,650	0,516	0,692	0,306	<b>0,730</b>	0,727
	200	0,951	0,978	0,976	0,860	0,975	0,674	0,987	<b>0,990</b>
	500	<b>1</b>	<b>1</b>	<b>1</b>	0,999	<b>1</b>	0,993	<b>1</b>	<b>1</b>
0,4	25	0,116	<b>0,301</b>	0,222	0,081	0,138	0,097	0,291	0,178
	60	0,213	0,287	0,331	0,143	0,255	0,130	<b>0,438</b>	0,292
	100	0,377	0,402	0,441	0,272	0,440	0,187	<b>0,614</b>	0,476
	200	0,745	0,747	0,747	0,604	0,812	0,397	<b>0,902</b>	0,850
	500	0,995	0,998	0,998	0,973	0,998	0,874	<b>1</b>	<b>1</b>
0,9	25	<b>0,357</b>	0,063	0,015	0,293	0,346	0,201	0,229	0,310
	60	0,989	0,981	0,978	0,976	0,992	0,954	0,994	<b>0,995</b>
	100	0,999	<b>1</b>	<b>1</b>	0,999	<b>1</b>	0,994	<b>1</b>	<b>1</b>
	200	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	0,999	<b>1</b>	<b>1</b>
	500	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	0,998	<b>1</b>	<b>1</b>

En la tabla se reporta en negrita y cursiva el tamaño de la prueba más alto para cada tamaño de muestra. Por ejemplo, se resalta en la primera línea de la tabla el número 0,146 que corresponde al poder de la prueba K, que es el más grande entre todas las pruebas cuando se emplean 25 observaciones y un  $\rho=-0,9$ . Fuente: elaboración propia.

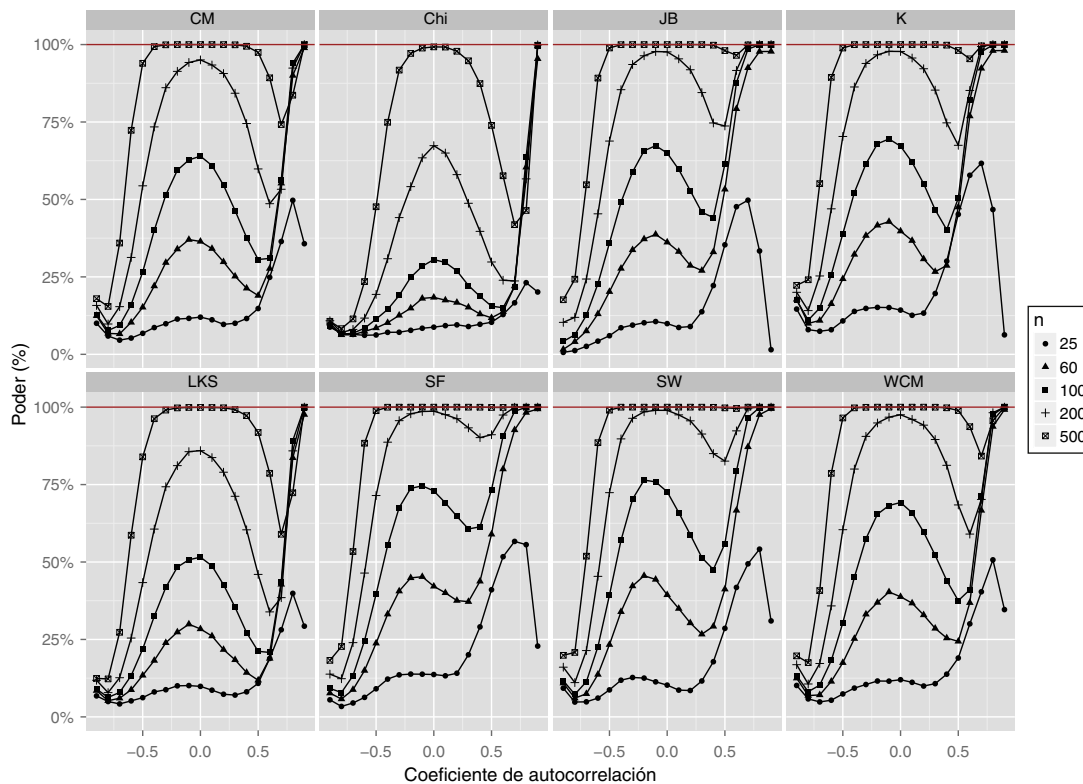


Figura 4. Poder de las pruebas de normalidad sobre errores con distribución  $\chi^2_{10}$ . Fuente: elaboración propia.

4. Comentarios finales

Los resultados de las simulaciones de Monte Carlo permiten encontrar 4 regularidades interesantes, con implicaciones para el

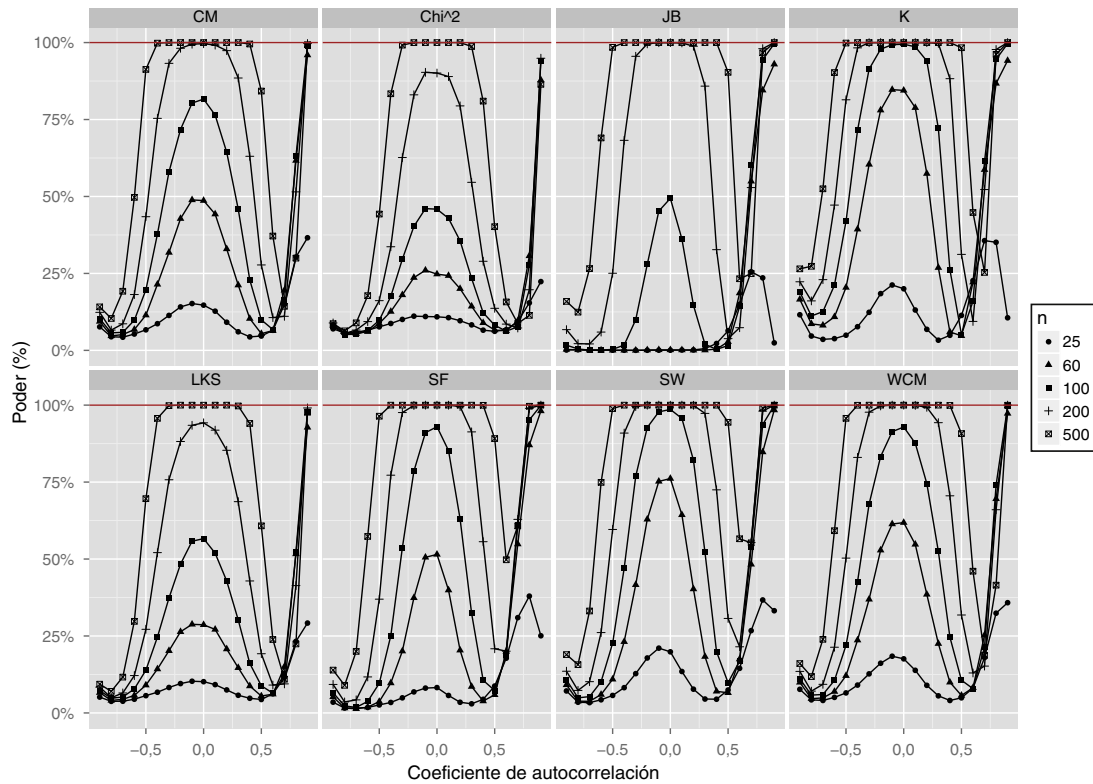
trabajo empírico con series de tiempo. Primero, el efecto de la presencia de autocorrelación sobre el tamaño y el poder de las pruebas es asimétrico; es decir, el efecto que tiene una autocorrelación, de igual tamaño pero de signo contrario, sobre el tamaño o el poder,

Tabla 5 Poder de las pruebas de normalidad para errores con distribución uniforme

$\rho$	Muestra (T)	Pruebas de normalidad							
		CM	K	JB	LKS	WCM	Chi	SF	SW
-0,9	25	0,077	<b>0,116</b>	0,001	0,052	0,077	0,070	0,035	0,072
	60	0,094	<b>0,166</b>	0,003	0,066	0,099	0,081	0,052	0,091
	100	0,103	<b>0,190</b>	0,017	0,073	0,111	0,079	0,066	0,107
	200	0,124	<b>0,223</b>	0,067	0,084	0,135	0,089	0,093	0,136
	500	0,141	<b>0,265</b>	0,159	0,094	0,161	0,087	0,139	0,190
-0,4	25	0,087	0,077	0,001	0,067	<b>0,091</b>	0,087	0,034	0,082
	60	0,215	<b>0,394</b>	0,	0,143	0,237	0,126	0,097	0,231
	100	0,378	<b>0,715</b>	0,017	0,247	0,427	0,175	0,251	0,470
	200	0,754	<b>0,983</b>	0,682	0,521	0,830	0,337	0,772	0,909
	500	0,998	<b>1</b>	<b>1</b>	0,957	<b>1</b>	0,834	<b>1</b>	<b>1</b>
0	25	0,147	<b>0,200</b>	0,001	0,102	0,176	0,109	0,082	0,199
	60	0,487	<b>0,845</b>	0,002	0,286	0,618	0,248	0,515	0,762
	100	0,816	<b>0,995</b>	0,494	0,566	0,929	0,459	0,930	0,986
	200	0,996	<b>1</b>	<b>1</b>	0,943	<b>1</b>	0,901	<b>1</b>	<b>1</b>
	500	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>
0,4	25	0,044	<b>0,049</b>	0,022	0,048	0,041	0,066	0,044	0,045
	60	<b>0,103</b>	0,060	0,004	0,088	0,100	0,090	0,039	0,070
	100	0,229	<b>0,262</b>	0,003	0,161	0,246	0,122	0,106	0,199
	200	0,631	<b>0,883</b>	0,327	0,429	0,705	0,290	0,556	0,725
	500	0,995	<b>1</b>	<b>1</b>	0,941	0,999	0,810	<b>1</b>	<b>1</b>
0,9	25	<b>0,366</b>	0,106	0,025	0,292	0,358	0,224	0,251	0,332
	60	0,960	0,941	0,930	0,928	0,973	0,878	0,981	<b>0,984</b>
	100	0,990	0,997	0,996	0,979	0,996	0,939	<b>0,999</b>	<b>0,999</b>
	200	0,997	<b>1</b>	<b>1</b>	0,991	<b>1</b>	0,949	<b>1</b>	<b>1</b>
	500	0,993	<b>1</b>	<b>1</b>	0,977	<b>1</b>	0,864	<b>1</b>	<b>1</b>

En la tabla se reporta en negrita y cursiva el tamaño de la prueba más alto para cada tamaño de muestra. Por ejemplo, se resalta en la primera línea de la tabla el número 0,116 que corresponde al poder de la prueba K, que es el más grande entre todas las pruebas cuando se emplean 25 observaciones y un  $\rho = -0,9$ . Fuente: elaboración propia.





**Figura 5.** Poder de las pruebas de normalidad sobre errores con distribución uniforme. Fuente: elaboración propia.

es diferente independientemente de la distribución y la prueba considerada. Por ejemplo, cuando se presenta una autocorrelación positiva, la probabilidad de rechazar erróneamente la hipótesis nula de normalidad (tamaño de la prueba) es más grande que el tamaño nominal y que el tamaño con la misma autocorrelación en valor absoluto pero con signo contrario. En el caso del poder, este es más grande cuando existe autocorrelación positiva en comparación con el escenario con autocorrelación negativa de la misma magnitud.

Segundo, la prueba que presenta una menor distorsión en su tamaño es la JB. No obstante, su tamaño puede llegar a ser de 0,431 en presencia de errores con una autocorrelación positiva muy fuerte ( $\rho=0,9$ ). En este sentido, todas las pruebas ven distorsionado su tamaño en presencia de autocorrelación fuerte (mayor a 0,5 en valor absoluto). Por otro lado, el tamaño no se ve fuertemente distorsionado por el tamaño de la muestra cuando la autocorrelación es moderada (menor a 0,4 en valor absoluto).

Tercero, si bien la prueba Chi tiene el poder más bajo en todos los escenarios estudiados, ninguna de las 8 pruebas estudiadas tiene consistentemente un mejor poder que las demás en todos los escenarios analizados.

Cuarto, cuando la muestra es pequeña ( $T=25$ ) y existe autocorrelación, así sea moderada, las pruebas de normalidad estudiadas tienen un poder muy bajo. El poder no supera el 37%.

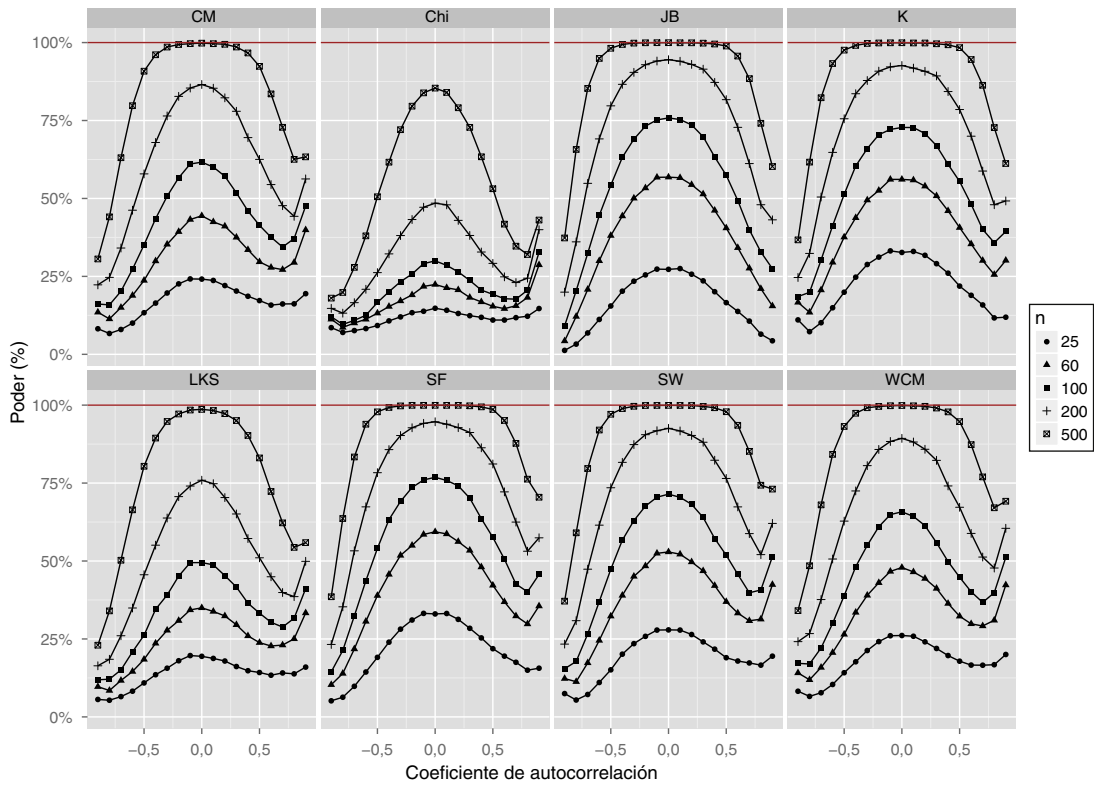
Finalmente, en presencia de errores que siguen un proceso AR(1), las 8 pruebas más comunes de normalidad no parecen ser una buena opción. Por un lado, si el investigador se enfrenta sin saberlo a errores que provienen de una distribución normal, las 8 pruebas presentan una probabilidad relativamente alta de rechazar la hipótesis nula erróneamente. Por otro lado, si el investigador se enfrenta sin saberlo a unos errores que no proceden de una distribución normal, las 8 pruebas presentan una relativa baja probabilidad de no rechazar la hipótesis nula de normalidad. En especial esta probabilidad es más baja para muestras pequeñas. En este orden de ideas, se debe ser muy cuidadoso al concluir sobre la distribución de los errores MCO si se emplean las 8 pruebas más comunes. Este resultado debe llamar la atención a los investigadores que emplean series de tiempo que estén autocorrelacionadas.

**Conflicto de intereses**

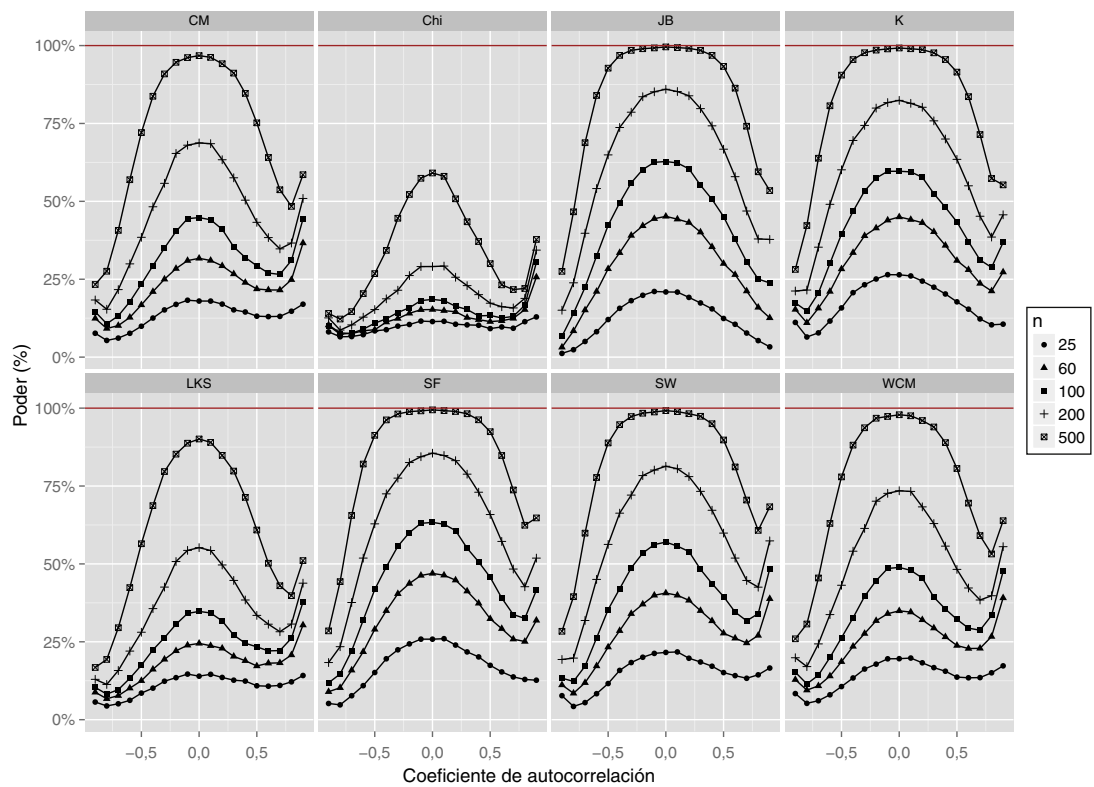
Los autores declaran no tener ningún conflicto de intereses.

**Anexo.**

figs. A1–A6



**Figura A1.** Poder de las pruebas de normalidad sobre errores con distribución  $t_4$ .  
Fuente: elaboración propia.



**Figura A2.** Poder de las pruebas de normalidad sobre errores con distribución  $t_5$ .  
Fuente: elaboración propia.

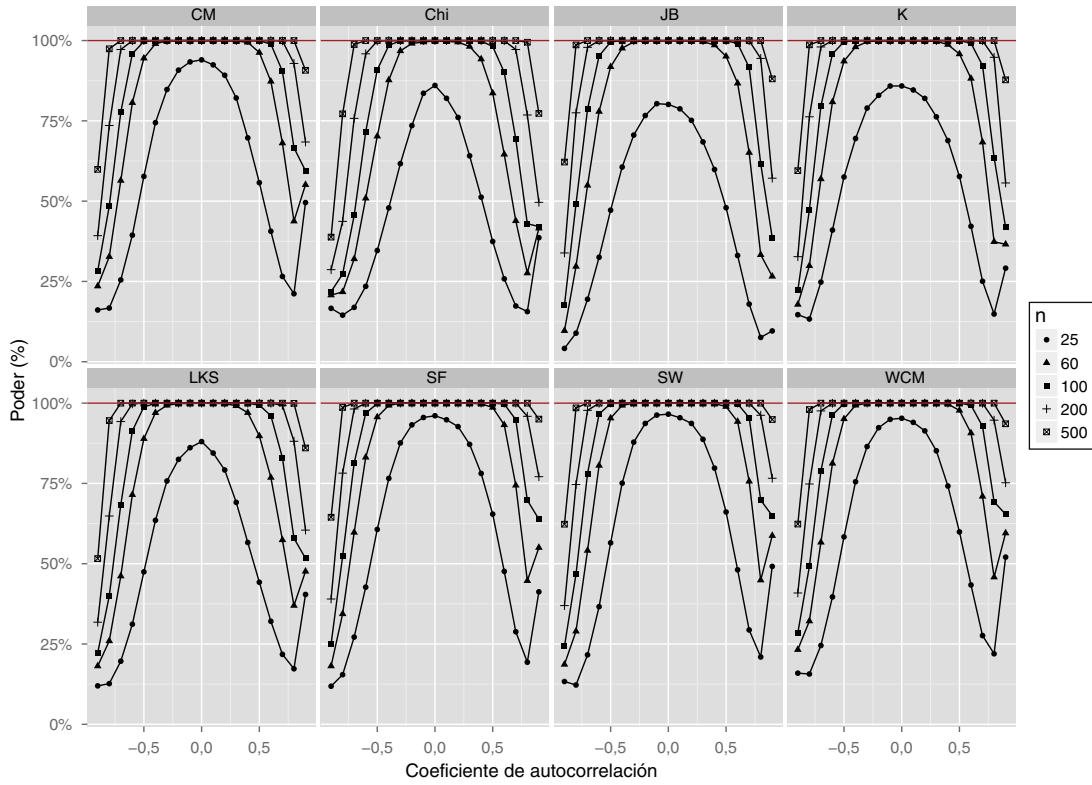


Figura A3. Poder de las pruebas de normalidad sobre errores con distribución  $\chi^2_1$ . Fuente: elaboración propia.

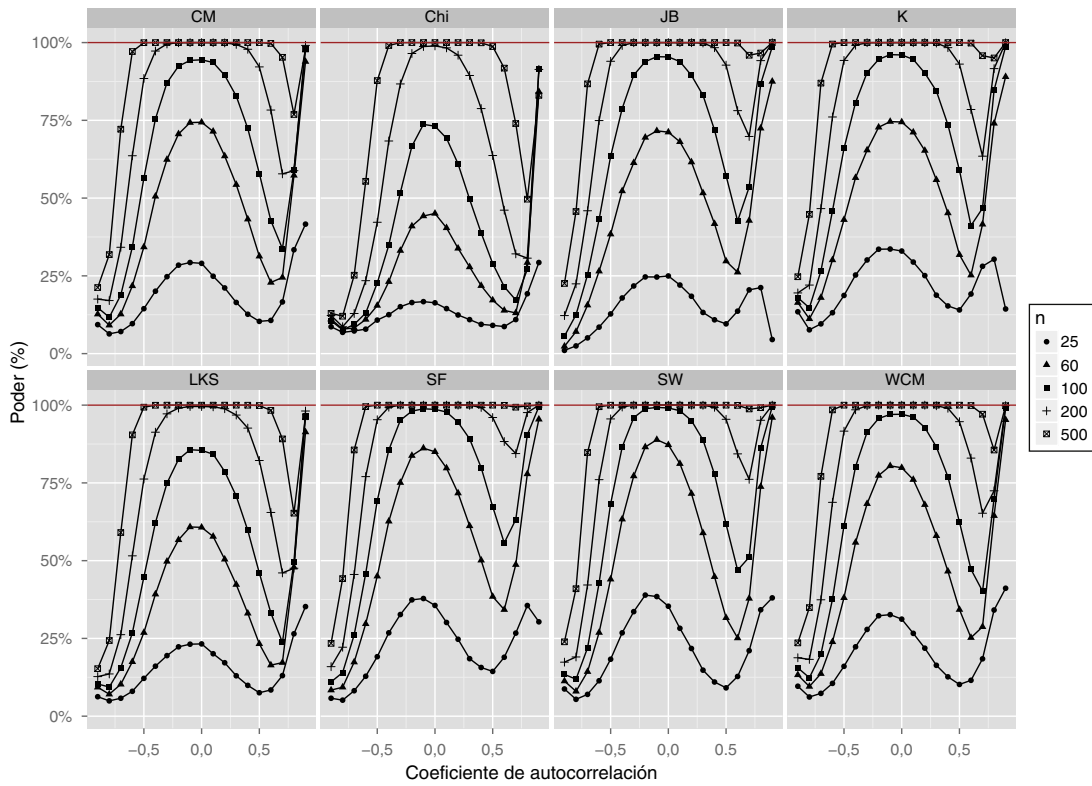
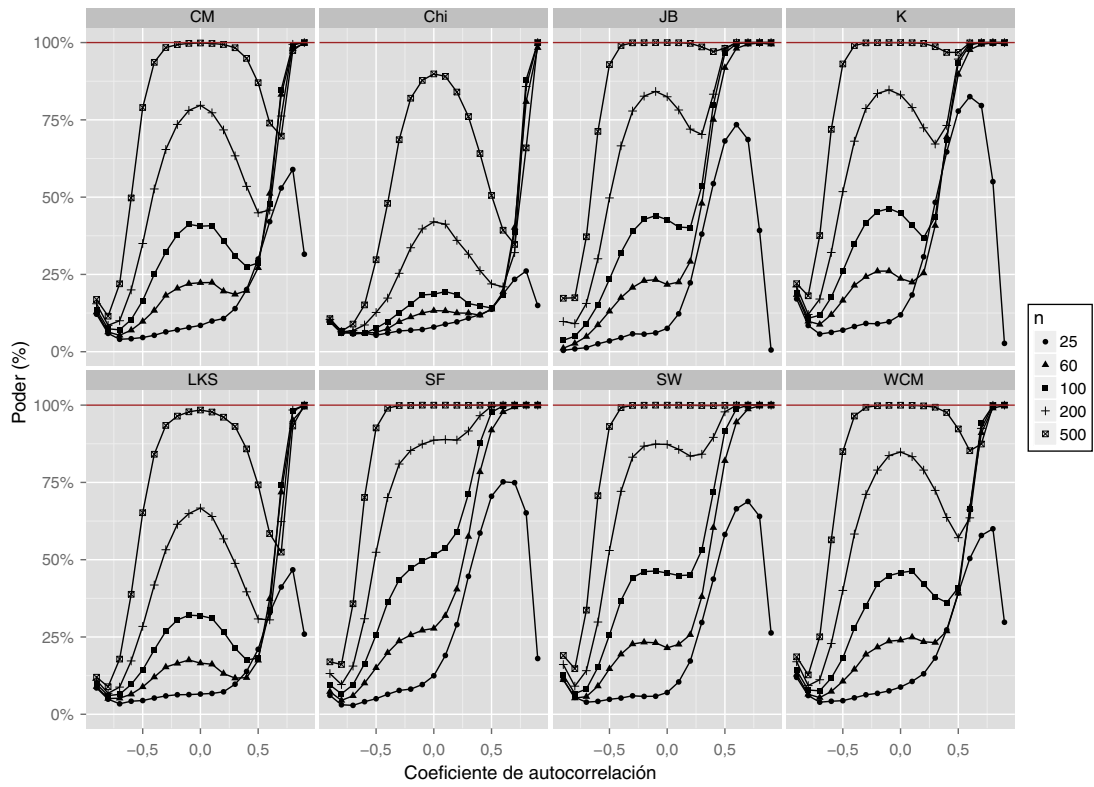
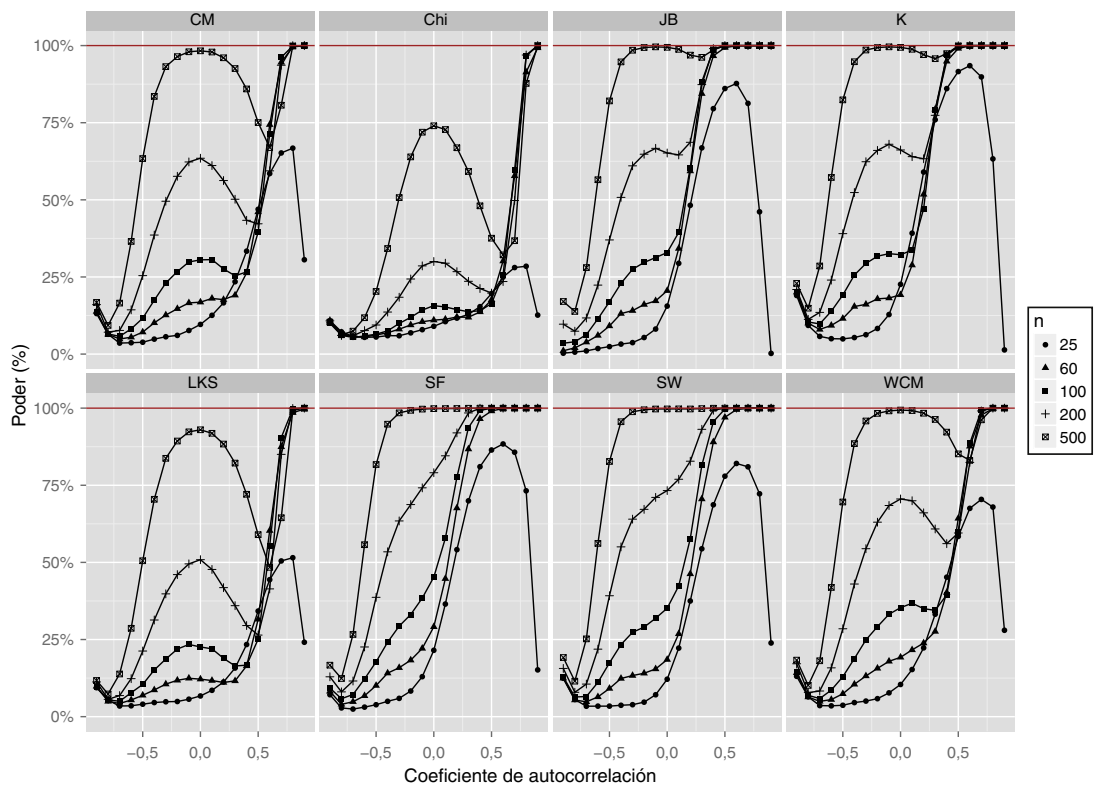


Figura A4. Poder de las pruebas de normalidad sobre errores con distribución  $\chi^2_5$ . Fuente: elaboración propia.



**Figura A5.** Poder de las pruebas de normalidad sobre errores con distribución  $\chi^2_{15}$ .  
Fuente: elaboración propia.



**Figura A6.** Poder de las pruebas de normalidad sobre errores con distribución  $\chi^2_{20}$ .  
Fuente: elaboración propia.

## Bibliografía

- Anderson, T. W. y Darling, D. A. (1952). Asymptotic theory of certain «goodness of fit» criteria based on stochastic processes. *The annals of mathematical statistics*, 193–212.
- Cramér, H. (1928). On the composition of elementary errors: First paper: Mathematical deductions. *Scandinavian Actuarial Journal*, 1, 13–74.
- D'Agostino, R. B., Belanger, A. y D'Agostino, R. B., Jr. (1990). A suggestion for using powerful and informative tests of normality. *The American Statistician*, 44(4), 316–321.
- D'Agostino, R. y Pearson, E. S. (1973). Tests for departure from results for the normality of  $b_2$  and  $b_1^2/2$ . *Biometrika*, 60(3), 613–689.
- Francia, R. S. y Shapiro, S. S. (1972). An approximate analysis of variance test for normality. *Journal of the American Statistical Association*, 67(337), 215–216.
- Jarque, C. y Bera, A. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economic Letters*, 55(2), 255–259.
- Jarque, C. y Bera, A. (1987). A test for normality of observations and regression residuals. *International Statistical Review*, 55(2), 163–172.
- Kilian, L. y Demiroglu, U. (2000). Residual-based tests for normality in autoregressions: Asymptotic theory and simulation evidence. *Journal of Business and Economic Statistics*, 18(1), 40–50.
- Kolmogorov, A.N. (1933). Sulla determinazione empirica di una legge di distribuzione. *NA*.
- Lilliefors, H. (1967). On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association*, 62(318), 399–402.
- Lobato, I. N. y Velasco, C. (2004). A simple test of normality for time series. *Econometric Theory*, 20(4), 671–689.
- Mbah, A. K. y Paothong, A. (2014). Shapiro-Francia test compared to other normality test using expected p-value. *Journal of Statistical Computation and Simulation*, 1–15.
- Pearson, K. (1894). Contributions to the mathematical theory of evolution. *Philosophical Transactions of the Royal Society of London*, 71–110.
- Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 50(302), 157–175.
- R Core Team (2013). R: A language and environment for statistical computing. Vienna, Austria.
- Razali, N. y Wah, Y. (2011). Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *Journal of Statistical Modeling and Analytics*, 2(1), 21–33.
- Shapiro, S. y Wilk, M. (1964). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3), 591–611.
- Shapiro, S., Wilk, M. y Chen, H. (1968). A comparative study of various tests for normality. *Journal of the American Statistical Association*, 63(324), 1343–1372.
- Thadewald, T. y Büning, H. (2007). Jarque-Bera test and its competitors for testing normality: A power comparison. *Journal of Applied Statistics*, 34(1), 87–105.