



Universidad
Zaragoza

Trabajo Fin de Grado

Estudio del comportamiento de dos contrastes de normalidad a través de experimentos de Monte Carlo.

Study of the behaviour of two Normality tests through Monte Carlo experiments.

Autor/es

Natalia Gascón Vicente

Director/es

Inmaculada Villanúa

Facultad de Economía y Empresa
Año 2017

ESTUDIO DEL COMPORTAMIENTO DE DOS CONTRASTES DE NORMALIDAD A TRAVÉS DE EXPERIMENTOS DE MONTE CARLO.

Autor: Natalia Gascón Vicente

Director: Inmaculada Villanúa

Facultad de Economía y Empresa Zaragoza

RESUMEN

El Trabajo de Fin de Grado lleva a cabo un estudio sobre el comportamiento de dos contrastes de Normalidad de Jarque-Bera y Lilliefors, a través de experimentos de Monte Carlo. Tras definir el contexto de nuestro trabajo, esto es, un modelo econométrico lineal (Modelo Lineal General), y establecer las hipótesis básicas cuyo cumplimiento suele suponerse, en este documento se presentan los problemas derivados de la falta de normalidad del término de perturbación en un modelo econométrico, así como la forma de detectarla. Nuestro objetivo es estudiar los dos contrastes propuestos de normalidad de forma comparada, y analizar su capacidad para detectar este problema. En las simulaciones realizadas consideramos modelos en los que la perturbación aleatoria (u_i) sigue una distribución diferente a la normal, para así poder observar si los resultados que muestran un comportamiento adecuado de los contrastes, esto es, si detectan esta ausencia de normalidad. La potencia y el tamaño de los contrastes son utilizados para analizar la adecuación de los contrastes, e intentar establecer una comparación entre ellos.

Palabras clave: econometría, experimentos de Monte Carlo, Normalidad, Jarque-Bera, Lilliefors.

STUDY OF THE BEHAVIOUR OF TWO NORMALITY TESTS THROUGH MONTE CARLO EXPERIMENTS.

ABSTRACT

The final project studies the behaviour of two Normality tests by Jarque-Bera and Lilliefors, through Monte Carlo experiments. Once the context of our work is defined, that is, a linear econometric model (General Linear Model), and the basic hypothesis, whose compliance is usually supposed, are established, this document presents the problems deriving from the lack of normality of the disturbance term in an econometric model, as well as the way to detect it. Our aim is to comparatively study both of the suggested normality tests and to analyze their ability to detect this issue. In our simulations we considered models in which the aleatory disturbance (u_i) follows a different distribution as the usual one, in order to observe the results that show a suitable behaviour of the tests, that is, if they detect this absence of normality. The power and the size of the tests are used to analyze their adequacy and to try to establish a comparison between them.

Key words: econometrics, Monte Carlo experiment, Normality, Jarque-Bera, Lilliefors.

ÍNDICE:

1. INTRODUCCIÓN	5
2. LA ECONOMETRÍA: CONCEPTO, ETAPAS Y ESPECIFICACIÓN DEL MODELO LINEAL GENERAL.....	6
2.1 Concepto de la Econometría	6
2.2 Etapas de la econometría	6
2.3 Hipótesis Básicas del Modelo Lineal General	7
2.4 Propiedades de los estimadores MCO	9
3. NORMALIDAD DE LOS MODELOS	11
3.1 Causas, consecuencias y soluciones	11
3.2 Contrastes de Normalidad	13
3.2.1 Contraste de Jarque-Bera	13
3.2.1.1 Contraste de asimetría	14
3.2.1.2 Contraste de apuntamiento	14
3.2.2 Contraste de Kolmogorov – Smirnov – Lilliefors	15
3.2.3 Contraste de Shapiro – Wilk	17
3.2.4 Contraste de Doornik – Hansen	17
4. EXPERIMENTO DE MONTE CARLO	18
4.1 Explicación del Experimento	18
4.2 Nuestro estudio de Monte Carlo.	19
5. RESULTADOS DEL EXPERIMENTO	22
6. CONCLUSIONES	35
7. BIBLIOGRAFÍA	37

1. INTRODUCCIÓN

Este trabajo de fin de grado está enmarcado en la línea de simulación y experimentos en econometría.

El objetivo general será mostrar la forma de trabajar con experimentos de Monte Carlo para analizar determinado instrumento econométrico.

Nuestro objetivo específico es comparar el comportamiento de los contrastes de normalidad utilizando para ello experimentos de Monte Carlo. En concreto comparamos el contraste de Jarque-Bera, que es el contraste más utilizado en econometría, con otro de los clásicos de los libros de estadística, Kolmogorov – Smirnov – Lilliefors. Se justificará la comparación realizada por el uso tradicional de ambos, y porque aparecen, junto a otros dos, en el programa Gretl, que es muy utilizado en aplicaciones econométricas. Con este objetivo, el trabajo se estructura de la siguiente forma.

En el segundo apartado se recuerda el concepto y etapas de la Econometría, así como las hipótesis básicas que consideramos en un Modelo Lineal General. Se recuerdan los procedimientos de estimación que suelen utilizarse, y las propiedades que cumplen los estimadores correspondientes. En el tercer apartado, nos centramos en el supuesto de normalidad, analizando las consecuencias de su incumplimiento y los mecanismos para su detección.

En el cuarto apartado, diseñamos un experimento de Monte Carlo para estudiar el comportamiento de los dos contrastes de normalidad que hemos seleccionado. Posteriormente se obtienen resultados en el apartado quinto, y se interpretan con el fin de obtener las conclusiones, que son presentadas en el último apartado.

2. LA ECONOMETRÍA: CONCEPTO, ETAPAS Y ESPECIFICACIÓN DEL MODELO LINEAL GENERAL

2.1. CONCEPTO DE ECONOMETRÍA

El concepto de Econometría proviene de dos palabras griegas: oikonomía (economía) y metron (medida) (Trivez, 2004). Por lo tanto, etimológicamente econometría significa medición de la economía.

Hay que tener en cuenta que el concepto de esta disciplina ha ido desarrollándose y evolucionando a lo largo del tiempo, pero hay ciertos aspectos que la definen.

Una definición que englobe todos esos aspectos es la de *Otero (1978)*, quien estableció que “*La econometría se ocupa de la cuantificación de los fenómenos económicos y de la verificación de las teorías económicas, haciendo uso para ello de las estadísticas económicas y aplicando a tal fin métodos especiales de inferencia estadística*”.

2.2. ETAPAS DE LA ECONOMETRÍA

La metodología de la econometría se puede ver reflejada a través de cuatro fases fundamentales:

1. Especificación. Consiste en la construcción de un modelo en términos probabilísticos, el cual creemos que subyace a los datos observados. Esta fase acapara tres aspectos generales. En primer lugar, explica cuáles son las variables elegidas para ser explicadas (que en ciertas ocasiones sólo es una variable) o la delimitación de la realidad que se desea estudiar. En segundo lugar, determina las variables que participan en el modelo. Diferenciamos entre la parte sistemática, donde aparecen las variables fundamentales, y la parte aleatoria o perturbación aleatoria. Finalmente, establece la forma funcional, que relaciona las variables, esto es, las que explican a la endógena.
2. Estimación. Consiste en la cuantificación de los parámetros que aparecen en el modelo. Para ello, es necesario tener información numérica sobre todas las variables que aparecen en el modelo, lo que nos permitirá obtener el modelo estimado. Los métodos de estimación más habituales son: el de los Mínimos Cuadrados Ordinarios (MCO) y el de la Máxima Verosimilitud (MV). Para un modelo lineal, ambos coinciden.

3. Validación. La validación de un modelo nos permite conocer si aceptamos o rechazamos la teoría inicialmente considerada, aunque hay que tener en cuenta que esta conclusión no es definitiva, ya que cualquier aceptación puede acabar en rechazo si cambian las condiciones. Es una etapa muy amplia. En cualquier aplicación es crucial analizar la significatividad, tanto individual como conjunta, de las variables, así como, analizar las medidas de bondad de ajuste ($R^2 \dots$). También se incluiría en esta etapa toda una batería de contrastes que tratan de asegurar el cumplimiento de las hipótesis básicas, incluidas aquellas relativas a la perturbación aleatoria.
4. Explotación. Cuando en la fase anterior, validación, se establece que el modelo es adecuado y coherente con los datos, se puede pasar a la última fase de la econometría, la explotación. Esta etapa puede tener tres finalidades, pudiendo abarcar una o varias de ellas:
- El análisis estructural se encarga de analizar el signo y magnitud de los parámetros, y por tanto estudia la relación entre las variables. Nos permite analizar la sensibilidad de la variable endógena antes cambios en las variables explicativas (X_i), y puede servir para la toma de decisiones de tipo económico.
 - La predicción consiste en, dados los valores de las variables explicativas, determinar valores extramuestrales de las variables objetivo. Si más de un modelo ha superado la etapa de validación, utilizando la predicción podemos seleccionar el modelo que nos parece mejor, esto es, el que de predicciones más exactas.
 - La simulación intenta originar la trayectoria de las variables objetivo a través de cambios en las variables explicativas.

2.3. HIPÓTESIS BÁSICAS DEL MODELO LINEAL GENERAL

Para poder explicar las condiciones ideales que completan un modelo econométrico, hay que partir del modelo clásico de regresión lineal. Dado un modelo general, su expresión matricial es:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$$

Donde y es el vector $T \times 1$ de observaciones de la variable endógena, X es una matriz $T \times k$ que recoge las T observaciones de las k variables explicativas, β el vector de coeficientes de orden $k \times 1$, y u el vector $T \times 1$ de perturbaciones aleatorias:

$$y = \begin{pmatrix} Y_1 \\ Y_2 \\ \dots \\ Y_T \end{pmatrix} \quad X = \begin{pmatrix} 1 & X_{21} & \dots & X_{k1} \\ 1 & X_{22} & \dots & X_{k2} \\ \dots & \dots & \dots & \dots \\ 1 & X_{2T} & \dots & X_{kT} \end{pmatrix} \quad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \dots \\ \beta_k \end{pmatrix} \quad u = \begin{pmatrix} u_1 \\ u_2 \\ \dots \\ u_T \end{pmatrix}$$

La relación que siguen las variables y parámetros de posición puede escribirse también como:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + u_i, (i=1,2,\dots, T)$$

donde la variable endógena (Y) y las variables exógenas (X_2, X_3, \dots, X_k) son magnitudes numéricas. Y sus conjuntos de valores (Y_1, Y_2, \dots, Y_T), ($X_{21}, X_{22}, \dots, X_{2T}$), ..., (X_{k1}, \dots, X_{kT}) son el resultado de la observación de una muestra aleatoria de tamaño T .

En la anterior relación cabe destacar que la parte sistemática es $\beta_1 + \beta_2 X_{2i} + \dots + \beta_k X_{ki}$, y la parte aleatoria es u_i .

Una especificación correcta de un modelo econométrico no está completa sin la especificación de la distribución de probabilidad de la perturbación así como, una indicación sobre la forma en que se han obtenido los valores de las variables explicativas. Este conjunto de información constituye las llamadas hipótesis básicas o clásicas.

A continuación, sintetizamos los supuestos de partida que habitualmente se realizan para aplicar las etapas de la econometría, que se denominan las hipótesis básicas del Modelo Lineal General:

- 1) La relación entre la variable endógena y las variables exógenas es lineal. La perturbación aleatoria se distribuye idéntica e independientemente como una normal:

$$u_i \sim \text{iiN}(0, \sigma^2)$$

Se cumple por tanto:

- $E(u_i) = 0$
- Homoscedasticidad: $\text{Var}(u_i) = \sigma^2, \forall i$

- No autocorrelación: $\text{Cov}(u_i, u_j) = E(u_i u_j) = 0, \forall i \neq j$
- 2) No hay ninguna restricción sobre los parámetros del modelo $(\beta_1, \beta_2, \dots, \beta_k, \sigma^2)$.
 $\text{Cov}(X, u) = 0$, que implica independencia entre la parte sistemática y aleatoria.
 Pero si hacemos más restrictivo este supuesto y consideramos las X no estocásticas, ello ya garantiza la comentada independencia.
 - 3) Grados de libertad positivos: El número de observaciones debe de ser superior al número de parámetros: $T > k$. Esta hipótesis verifica la existencia de un número suficiente de grados de libertad en la estimación.
 - 4) La ausencia de multicolinealidad establece que todas las variables explicativas son linealmente independientes, y la hipótesis de convergencia garantiza la estabilidad a las X (oscilan alrededor de una constante, y su variabilidad también es constante).

2.4. PROPIEDADES DE LOS ESTIMADORES MCO

A la hora de estimar los parámetros de posición que aparecen en el Modelo Lineal General se pueden utilizar dos métodos: Mínimos Cuadrados Ordinarios (MCO) y el de la Máxima Verosimilitud (MV). En nuestro caso, dado que vamos a analizar la normalidad, conocemos que cuando esta hipótesis se cumple, ambos métodos de estimación serían iguales.

La estimación MCO de los parámetros de posición vendrá dada por:

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

El parámetro de dispersión vendrá determinado por:

$$\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{T-2}.$$

Bajo el cumplimiento de las hipótesis básicas, la matriz de varianzas y covarianzas de $\hat{\beta}$ será: $\text{Var}(\hat{\beta}) = \hat{\sigma}^2 (X'X)^{-1}$.

Las principales propiedades de los estimadores tanto para muestras pequeñas como para muestras grandes ó asintóticas se sintetizan a continuación.

Para **muestras pequeñas o finitas** se cumple que los estimadores MCO de los parámetros de posición son:

- 1) Lineales.
- 2) Inssegados. $E(\hat{\beta}) = \beta$
- 3) ELIO (Estimadores Lineales Inssegados Óptimos). Se considera que un estimador $(\hat{\beta})$ es ELIO de β , si es el estimador con menor varianza entre todos los estimadores de β que sean lineales e inssegados.
- 4) Eficiente. Se considera que un estimador $(\hat{\beta})$ es eficiente de β cuando tenga la menor varianza de entre todos los estimadores inssegados de β . Para obtener la eficiencia de un estimador, debemos basarnos en un teorema conocido como el Teorema de Cramer-Rao. Dicho teorema establece que el límite inferior de la varianza de los estimadores inssegados viene dado en la diagonal principal de la inversa de la matriz de información.

$$\left(-E \left(\frac{\partial^2 \ell}{\partial \theta \partial \theta'} \right) \right)^{-1}$$

siendo ℓ , la función logarítmica de verosimilitud de una muestra determinada, y teniendo en cuenta que θ incluye β y σ^2 .

En cuanto al parámetro de dispersión, éste cumple que es inssegado: $E(\hat{\sigma}^2) = \sigma^2$, pero no es ELIO ya que este parámetro no es lineal.

Las **propiedades de los estimadores para muestras grandes ó propiedades asintóticas** son las que se limitan únicamente a las distribuciones muestrales que proceden de muestras con tamaño que crecen sin límite.

Se nombraran las dos propiedades más utilizadas para los estimadores MCO de los parámetros de posición:

- 1) Inssegadez asintótica. Se establece que un estimador es inssegado asintóticamente si se transforma en inssegado cuando el tamaño de la muestra tiende a infinito. Es decir:

$$\lim_{T \rightarrow \infty} E(\hat{\beta}) = \beta$$

Además, hay que tener en cuenta, que si un estimador es inssegado, también será asintóticamente inssegado, pero no a la inversa.

- 2) Consistencia. La consistencia es el mínimo requisito que se exige a cualquier estimador para poder considerarlo adecuado. *Un estimador $\hat{\beta}$ será consistente si, a medida que T tiende a infinito, la distribución de este estimador se concentra en el punto correspondiente al verdadero valor del parámetro.* (Trivez, 2004). Hablar de que $\hat{\beta}$ es consistente es equivalente a decir que converge en probabilidad al verdadero valor del parámetro.

En cuanto al parámetro de dispersión, hay que decir que cumple que es asintóticamente insesgado debido a su insesgadez, y cumple la propiedad de consistencia.

3. NORMALIDAD DE LOS MODELOS

Las propiedades de los estimadores MCO se cumplen bajo los supuestos de las hipótesis básicas. Analizaremos ahora la normalidad. ¿Qué ocurre si no se cumple? ¿Cómo podemos saber si se cumple? Si no se cumple, ¿podemos resolver los problemas de algún modo?.

En este trabajo se va a analizar la no normalidad de la perturbación aleatoria de los modelos. Para ello es conveniente explicar en un primer momento cuales son las causas que provocan que no haya normalidad en la perturbación en los modelos, las consecuencias que esto conlleva, y las posibles soluciones a los problemas encontrados. Y en segundo lugar se explicarán los instrumentos para detectar su incumplimiento. Concretamente, los contrastes más importantes de normalidad y los que se utilizarán en el Experimento de Monte Carlo.

3.1. CAUSAS, CONSECUENCIAS Y SOLUCIONES

Existen varias causas que provocan la no normalidad de la perturbación en un Modelo Lineal General. Por un lado se encuentra la forma funcional incorrecta del modelo específico. Por otro lado, puede ser causado por la existencia de datos atípicos, lo que quiere decir que algunas observaciones de las que se disponen son muy dispares, es decir, o muy grandes o muy pequeñas, con respecto al resto de observaciones de la muestra.

En cuanto a las consecuencias que presenta la no normalidad, podemos distinguir:

- 1) La inferencia estadística no es válida. Si la variable aleatoria u , sigue una distribución normal, la variable endógena (Y) también seguirá esa distribución según las hipótesis básicas nombradas anteriormente. Los parámetros de posición estimados son:

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

por tanto seguirán una distribución normal. La normalidad de $\hat{\beta}_j$, permite utilizar la distribución

$$\frac{\hat{\beta}_j - \beta_j}{\hat{\sigma}_{\hat{\beta}_j}} \sim t_{T-k} \quad (2)$$

para hacer contrastes de hipótesis. Uno de los más habituales es el de significatividad individual de los parámetros. Si u no sigue una normal, la distribución (2) no se cumple, por lo que no podemos fiarnos de los t-ratios para contrastar. Lo mismo ocurre con cualquier otro contraste que utilice el estadístico t , y también con los contrastes de restricciones lineales, u otros, que usan la F de Snedecor, que también se basa en la normalidad de $\hat{\beta}$.

- 2) Los estimadores MCO dejan de coincidir con los estimadores por máxima verosimilitud. Lo que ocurre es que, si maximizamos ℓ , siendo ℓ el logaritmo de la función de densidad conjunta de la variable endógena y , cuando la función de densidad es normal, es decir, $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$, ambos estimadores son iguales. Pero si la función de densidad no es normal, la función a maximizar será diferente, y los estimadores MV obtenidos no coincidirán con los MCO.
- 3) Los estimadores mínimo-cuadráticos no son eficientes. La eficiencia nos la da la cota de Crammer Rao, que depende de ℓ , y ahora ℓ cambia, por lo que no podemos asegurar que sean eficientes.
- 4) La estimación por intervalo no sería válida. Sigue la misma explicación que la inferencia estadística, ya que la construcción de un intervalo de confianza consiste es: $IC = \hat{\beta} \pm t_{\varepsilon/2} \hat{\sigma}_{\hat{\beta}}$

Como soluciones al problema existente de la no normalidad, es decir, cuando la perturbación no sigue una distribución normal, si conociéramos la distribución correcta,

estimaríamos por MV, pero si no la conocemos podemos intentar paliar el problema de alguna de las formas siguientes:

- 1) Realizar una mejor especificación del modelo, que es la solución que se toma de manera general, y lo que se hace en esta situación es introducir variables ficticias, tomar logaritmos,.. entre otros.
- 2) Como soluciones más completas que no están a nuestro alcance, se puede mencionar, tal y como establece Peña en su libro Estadística Modelos y Métodos, que *si la distribución es bimodal, para solucionar el problema podemos segmentar la población en subpoblaciones homogéneas.*

3.2. CONTRASTES DE NORMALIDAD

Una vez analizadas las causas y las soluciones a la ausencia de normalidad, nos queda saber cómo podemos detectar su incumplimiento. Utilizaremos contrastes de hipótesis, y en este apartado explicamos los contrastes más importantes y los habituales en la literatura estadística y econométrica.

3.2.1. Contraste de Jarque-Bera

El contraste de Jarque- Bera (1987) es el más utilizado en la metodología econométrica. *Es un contraste de grandes muestras (asintótico) y fundamentado en los residuos MCO. Esta prueba evalúa la asimetría y la kurtosis (o apuntamiento) de los residuos MCO.* (Gujarati, 2003).

Las distribuciones normales se caracterizan a través de los momentos hasta el cuarto orden. El primer y segundo momento son la media y la varianza, el momento de tercer orden es asimetría, y el momento de cuarto orden es kurtosis, los cuales son 0 y 3 respectivamente.

La hipótesis nula y la alternativa del contraste son:

$H_0 : u \sim \text{Normalidad (los residuos están normalmente distribuidos). Con } g_1=0 \text{ y } g_2 = 3.$

$H_A : u \sim \text{No normalidad (los residuos no están normalmente distribuidos)}$

El estadístico de contraste es de la forma: $JB = T \left[\frac{g_1^2}{6} + \frac{(g_2-3)^2}{24} \right]$

siendo: T el tamaño de la muestra, g_1 el coeficiente de asimetría y g_2 el coeficiente de kurtosis.

Dicho estadístico asintóticamente sigue siempre una distribución chi-cuadrado con dos grados de libertad: X_2^2

Tras haber explicado el Contraste de Jarque-Bera, sería conveniente explicar dos contrastes, cuya combinación entre ambos, da lugar a Jarque-Bera: los contrastes de asimetría y de apuntamiento (ó kurtosis).

3.2.1.1. El contraste de asimetría

Tal y como establece Cao Abad y otros (2008), *en el caso de ser normal la distribución, el coeficiente de asimetría sigue una distribución asintótica $N(0, \sqrt{6/T})$* , y para un tamaño de muestra grande (50 o más datos), $g_1 \sqrt{T/6}$ sigue una distribución asintótica $N(0,1)$, que es la que se utiliza para contrastar $H_0 : g_1 = 0$

El contraste de asimetría establece que si se cumple la hipótesis de normalidad, dicho contraste estima un parámetro poblacional de asimetría que es cero. Por ello, el estadístico de contraste de una normal es el coeficiente de asimetría muestral:

$$g_1 = \frac{\sum \hat{u}_t^3}{\tilde{\sigma}^3}$$

donde g_1 es el momento muestral de tercer orden, u_t es la perturbación aleatoria (o los residuos del modelo), y $\tilde{\sigma}$ es la varianza por estimación de Máxima Verosimilitud.

3.2.1.2. El contraste de apuntamiento ó kurtosis

En el caso de ser normal la distribución, el coeficiente de apuntamiento toma el valor 3, y sigue una distribución asintótica $N(0, \sqrt{24/T})$. (Cao Abad y otros, 2008). Pero para tamaños muestrales grandes, $(g_2-3) \sqrt{T/24}$ sigue una distribución asintótica $N(0,1)$, que es la que se utiliza para contrastar la hipótesis nula como $g_2 - 3 = 0$.

El contraste de apuntamiento considera, tal y como ocurría en el contraste de asimetría, que el coeficiente de apuntamiento de la distribución normal es cero. Por ello, el estadístico de contraste para la bondad de ajuste de una normal es el coeficiente de apuntamiento muestral.

$$g_2 = \frac{\sum \hat{u}_t^4}{\tilde{\sigma}^4} - 3$$

donde g_2 es el momento muestral de cuarto orden, u_t es la perturbación aleatoria (o los residuos del modelo), y $\tilde{\sigma}$ es la varianza por estimación de máxima verosimilitud.

3.2.2. El contraste de Kolmogorov – Smirnov - Lilliefors

El contraste de Kolmogorov- Smirnov-Lilliefors es una modificación del contraste de Kolmogorov-Smirnov. Y por ello es conveniente explicar primero el contraste Kolmogorov-Smirnov, y más tarde su modificación.

El contraste KS está diseñado para el contraste de ajuste a distribuciones continuas, el cual está basado en la distribución del estadístico D_n , que compara la distribución muestral y la poblacional supuesta:

$$D_n = \sup_{x \in R} |F_n(x) - F(x)|$$

donde: $F_n(x)$ es una función de distribución empírica muestral, y $F(x)$ es la función teórica normal de la población que queremos contrastar en nuestro caso.

El contraste de Kolmogorov – Smirnov – Lilliefors establece que se debe estimar la media poblacional mediante la media muestral y la varianza poblacional a través de la varianza muestral. De este modo, construir el siguiente estadístico:

$$D_n = \sup_{x \in R} |F_n(x) - \Phi\left(\frac{x - \bar{X}}{S}\right)|$$

donde: $F_n(x)$ es una función de distribución empírica muestral, S es la desviación típica media, \bar{X} es la media.

La hipótesis nula y la alternativa serán las siguientes:

- H_0 : $u \sim$ Normalidad (los residuos están normalmente distribuidos)
- H_A : $u \sim$ No Normalidad (los residuos no están normalmente distribuidos)

La distribución de este estadístico fue tabulada por Lilliefors, y será respecto a dicha tabulación como debemos juzgar la significación del valor obtenido para este estadístico. (Cao Abad y otros, 2008)

Podemos entender mejor este contraste mediante el ejemplo que aparece en Peña (1991).

Contrastar la hipótesis de que los datos siguientes provienen de una distribución normal: (20, 22, 24, 30, 31, 32, 38).

En primer lugar calculamos la media y varianza muestrales para obtener las probabilidades teóricas. La media de los siete datos es: $\bar{x} = 28,14$, y la desviación típica corregida por grados de libertad: $\hat{S} = \frac{\sum (x_i - \bar{x})^2}{T-1} = 6,39$.

Para efectuar el contraste construimos la siguiente tabla:

Nº de Obs.	X	$F_n(x)$	F(x)	D ₁	D ₂	D _n
1	20	0,1429	0,1020	0,1020	0,0409	0,1020
2	22	0,2857	0,1685	0,0256	0,1172	0,1172
3	24	0,4286	0,2578	0,0279	0,1718	0,1718
4	30	0,5714	0,6141	0,1855	0,0427	0,1855
5	31	0,7143	0,6736	0,1018	0,0407	0,1018
6	32	0,8571	0,7258	0,1115	0,0313	0,1115
7	38	1,0000	0,9382	0,0811	0,0618	0,0811

Los valores de $F_n(x)$ se han obtenido simplemente mediante: $F_n(x) = i/T$

donde i representa el índice ordinal de la observación, y los valores F(x) se han calculado tipificando los siete datos y mirando en tablas de la normal estándar.

Para ello se calcula: $z_i = \frac{x_i - \bar{x}}{\hat{S}}$ y se obtiene en tablas el valor de F(z_i). En la tabla,

D₁ representa el valor: $D_1 = |F_n(x_{i-1}) - F(x_i)|$. Mientras que D₂ = |F_n(x_i) – F(x_i)|.

La tabla de Lilliefors, indica que el valor de D crítico para un nivel de significación del 5% y un tamaño muestral de 7 es: D_C = (0,05;7) = 0,300

Y como la máxima distancia obtenida en nuestros datos es D_n = 0,1855, concluimos que no hay evidencia suficiente en los datos para rechazar la hipótesis de normalidad.

3.2.3. El contraste de Shapiro – Wilk

El contraste se construye a partir de un estadístico que tiene en cuenta la distancia entre la mayor y menor observación muestral, entre la segunda y la penúltima, y así sucesivamente. Es un estadístico mucho menos intuitivo que el anterior. El estadístico es el siguiente:

$$w = \frac{A^2}{T S^2} \quad \text{siendo } A = \sum_{j=1}^{\left[\frac{T}{2}\right]} a_{j,T} (x_{(T-j+1)} - x_{(j)})$$

donde: $T S^2 = \sum (X_i - \bar{X})^2$, los coeficientes de $a_{j,T}$ están tabulados y $x_{(j)}$ es el valor ordenado en la muestra con lugar j.

Shapiro y Wilk han tabulado los valores de estas constantes, así como la distribución del estadístico w. (Cao Abad y otros, 2008)

Cabría destacar que el estadístico bajo la hipótesis de normalidad se rechazará para valores pequeños del estadístico.

3.2.4 El contraste de Doornik - Hansen

Este es un contraste general, en el sentido de que considera tanto el coeficiente de asimetría como el coeficiente de kurtosis, igual que hace el contraste de normalidad de Jarque-Bera. Puede considerarse una modificación de Jarque-Bera.

Doornik y Hansen (2008) establecen que si los coeficientes de asimetría y kurtosis se estiman a partir de sus correspondientes contrapartidas muestrales, tales coeficientes no están independientemente distribuidos en muestras finitas; el coeficiente de kurtosis especialmente se aproxima a la normalidad muy lentamente y el uso de la distribución asintótica conduce a un sobre rechazo de la hipótesis nula de normalidad.

Su planteamiento se basa en calcular tales coeficientes a partir de una transformación de los datos originales, lo que da lugar al coeficiente de asimetría transformado (z_1) y al coeficiente de kurtosis transformado (z_2) que están más próximos a la distribución Normal.

El estadístico que proponen viene dado por:

$$DH = z_1^2 + z_2^2$$

el cual bajo la hipótesis nula de normalidad se distribuye asintóticamente como una X^2 con 2 grados de libertad.

4. EXPERIMENTO DE MONTE CARLO

4.1 EXPLICACIÓN DEL EXPERIMENTO

Los experimentos de Monte Carlo se pueden utilizar con diversos objetivos. En ocasiones se utilizan para entender problemas con datos reales y medir la influencia de todos los factores que se hayan considerado como importantes.

Habitualmente, un experimento de Monte Carlo sigue la siguiente estructura:

- 1) Especificar un PGD (Proceso Generador de los datos) o modelo verdadero.

$$Y_i = \beta_1 + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + u_i \quad i = 1, \dots, T$$

siendo la variable dependiente (Y_i), los coeficientes ($\beta_1, \beta_2, \dots, \beta_k$), la perturbación aleatoria (u_i) y las variables explicativas (X_{2i}, \dots, X_{ki}).

En el caso de que el modelo verdadero sea el Modelo Lineal General (MLG) que cumple las hipótesis básicas, generamos la perturbación como una normal $N(0, \sigma^2)$, y establecemos las variables explicativas, los coeficientes y el tamaño muestral.

- 2) Generar un conjunto de datos de Y_i usando el PGD anterior. Esto es, una muestra de Y_i a partir de lo anterior.
- 3) Estimar el modelo para obtener el estimador o estadístico de contraste objeto de estudio, con la finalidad de analizarlo con esta muestra generada artificialmente.
- 4) Repetir las etapas 2 y 3 un elevado número de veces. Cada una de ellas se denomina réplica, iteración o simulación. Por tanto tendremos R muestras de Y a partir de R muestras generadas de u , manteniendo fijas las muestras de las variables explicativas y también los parámetros.
- 5) Evaluar los resultados obtenidos.

Como nuestro objetivo es analizar el funcionamiento de un contraste, necesitamos de algunos conceptos e instrumentos sobre inferencia estadística. En el caso de que nuestro objetivo fuera analizar el posible sesgo de los estimadores, no sería necesario.

Un contraste de hipótesis se puede definir como la regla de decisión a través de la cual optamos creer en una o en otras hipótesis. En la toma de decisiones para cada uno de los contrastes realizados utilizaremos dos tipos de errores:

- Error de tipo I: es el error que se comete al rechazar la hipótesis nula cuando realmente es cierta.
- Error de tipo II: es el error que se produce al aceptar la hipótesis nula cuando realmente es falsa.

Puesto que en un contraste de hipótesis se pueden cometer estos dos tipos de errores, se tienen que cuantificar de alguna forma. Para ello, se pueden calcular las probabilidades de que ocurran (son probabilidades condicionadas puesto que no se conoce la autenticidad de las hipótesis). Estas probabilidades son las que se denominan “tamaño del error” tipo I y tipo II respectivamente.

Tamaño error tipo I = $P \{ \text{Rechazar } H_0 \mid H_0 \text{ es cierta} \} = \varepsilon$

Tamaño error tipo II = $P \{ \text{Aceptar } H_0 \mid H_0 \text{ es falsa} \} = \delta$

Se denominará ε al nivel de significación o tamaño del contraste, $1 - \varepsilon$ al nivel de confianza, $1 - \delta$ a la potencia del contraste, esto es, la probabilidad de rechazar H_0 cuando H_0 es falsa.

Para realizar un contraste de hipótesis, el investigador fija de antemano ε , por lo tanto, la probabilidad del error del tipo I es conocida.

A partir de aquí, podemos llevar a cabo experimentos que nos ayuden a determinar el comportamiento de un estadístico de contraste utilizando para ello la proporción de rechazos de la hipótesis nula (tanto si es cierta como si es falsa).

Como fijamos ε , si el contraste funciona bien es de esperar que cuando la hipótesis nula sea cierta, el porcentaje de veces, de las R iteraciones que se rechace una H_0 que es cierta se aproxima a $(\varepsilon * 100)\%$.

Otro concepto muy importante es el de potencia, definido como la probabilidad de rechazar la hipótesis nula cuando realmente es falsa. Si hemos generado el PGD de forma que H_0 es falsa, el contraste será adecuado si capta dicha falsedad un porcentaje de veces muy elevado, que tienden al 100% (potencia).

4.2 NUESTRO ESTUDIO DE MONTE CARLO

Supongamos que nuestro modelo verdadero o PGD (proceso generador de los datos o modelo verdadero) es:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$$

donde la Y_i es la variable dependiente, X_i las variables explicativas y la u_i es la perturbación aleatoria del modelo.

Para comenzar:

- 1) Generamos las variables X_{2i} y X_{3i} para mantenerlas fijas en las diferentes réplicas, siendo X_{2i} una chi-cuadrado con 5 grados de libertad y X_{3i} una $u(0,1)$. Estos valores suponen una variabilidad similar para las dos variables, aunque los intervalos de valores que pueden tomar sean diferentes.
- 2) Damos valor a los parámetros: β_1 , β_2 y β_3 . Concretamente cada uno de ellos tiene valor igual a la unidad.

A partir de aquí, generaremos 1000 muestras de la perturbación u_i , es decir, hacemos 1000 réplicas, y en cada una de ellas, la muestra de las variables X_2 y X_3 es la misma, esto es, son variables fijas para muestras repetidas. Cada muestra de u da lugar a una Y . Estimaremos el modelo 1000 veces.

En nuestro caso, el objetivo es generar el modelo con diferentes distribuciones para la perturbación (t de student, Normal, Gamma,...), y aplicar contrastes de normalidad. El interés se centra en dos contrastes de normalidad: el habitual de la literatura econométrica, Jarque-Bera, y el ampliamente utilizado en la literatura estadística, contraste de Lilliefors. Llevaremos a cabo varios experimentos en los que la distribución de la perturbación sea diferente.

Se presenta en la tabla 1 las características de todas las distribuciones que vamos a considerar, para así poder ver la similitud o diferencia con la Normal.

En la tabla 1 se han expuesto las principales características de las distribuciones que se van a utilizar como perturbación aleatoria en nuestro experimento de Monte Carlo. Vamos a explicar los resultados del experimento en grupos de distribuciones. En el primer grupo se encuentran las distribuciones t de student y logística, dado que son las que tienen el mismo rango que la Normal, además de su simetría. En el segundo grupo, se encuentran las distribuciones F de Snedecor y chi-cuadrado, dado que son aquellas que se construyen a partir de la Normal, pero tienen rango sólo positivo y no son simétricas. Otro grupo será el formado por la distribución gamma, y su caso particular, exponencial, dado que es una distribución muy diferente de la Normal. Y aparecerán dos distribuciones solas debido a las características que tienen: log-normal y bimodal.

Tabla 1. Características de distribuciones.

	Rango de la variable	Esperanza y varianza	Otras características relevantes
Normal ($N_{0,2}$)	$(-\infty, +\infty)$	$E(X) = 0$ $Var(X) = 2$	Simetría
t de Student (t_n)	$(-\infty, +\infty)$	$E(X) = 0$ $Var(X) = \frac{n}{(n-2)}$	Simetría $n > 30$ se parece a la Normal.
Logística	$(-\infty, +\infty)$	$E(X) = 0$ $Var(X) = \frac{\pi^2}{3}$	Simetría
F de Snedecor ($F_{n,m}$)	$(0, +\infty)$	$E(X) = \frac{m}{m-2}$ $Var(X) = \frac{2m^2(n+m-2)}{n(m-4)(m-2)^2}$	Se construye a partir de la Normal. No es simétrica.
Chi-cuadrado (X_n^2)	$(0, +\infty)$	$E(X) = n$ $Var(X) = 2n$	Se construye a partir de la Normal. No es simétrica
Log-normal [$\log(x) \sim N(\mu, \sigma^2)$]	$(0, +\infty)$	$E(X) = e^{\mu + \frac{\sigma^2}{2}}$ $Var(X) = e^{2\mu}(e^{2\sigma^2} - e^{\sigma^2})$	El logaritmo de la variable es Normal.
Gamma ($G_{a,p}$)	$(0, +\infty)$	$E(X) = \frac{a}{p}$ $Var(X) = \frac{a}{p^2}$	Si $a=1$, y p cualquier valor, tenemos la distribución exponencial (E_λ), ó Gamma($1, \lambda$).
Bimodal	$(-\infty, +\infty)$	$E(X) = \pm\mu$ $Var(X) = \sigma^2$	Combina 2 normales (50%). Consideramos esperanzas de magnitud igual y signo contrario, y mismas varianzas.

Las preguntas que nos planteamos a la hora de explicar los resultados son, entre otras: ¿Captan los dos contrastes la no normalidad sea cual sea la distribución de u , o por el contrario, depende del tipo de distribución que sigue la perturbación aleatoria? ¿Puedo fiarme del contraste en muestras pequeñas? ¿Puedo decir que uno de ellos es mejor?.

En el siguiente apartado trataremos de dar respuesta a estas preguntas a partir de los resultados de los experimentos.

5. RESULTADOS DEL EXPERIMENTO DE MONTE CARLO

En este apartado, se van a mostrar los resultados obtenidos en el experimento de Montecarlo. En cada una de las tablas presentadas tendremos en la parte superior el tamaño muestral utilizado ($T=50, 100$ y 500), así como el nivel de significación que hemos fijado previamente ($0.01, 0.05$ y 0.1). En el margen izquierdo de la tabla se explica la distribución de la que tenemos los resultados, y los contrastes que hemos utilizado. El contenido de cada celda nos da la proporción (tanto por uno) de rechazos de H_0 sobre el total de réplicas ($\frac{n^{\circ} \text{ rechazos}}{1000}$).

La tabla 2 presenta los resultados obtenidos cuando la $u_i \sim N(0,2)$.

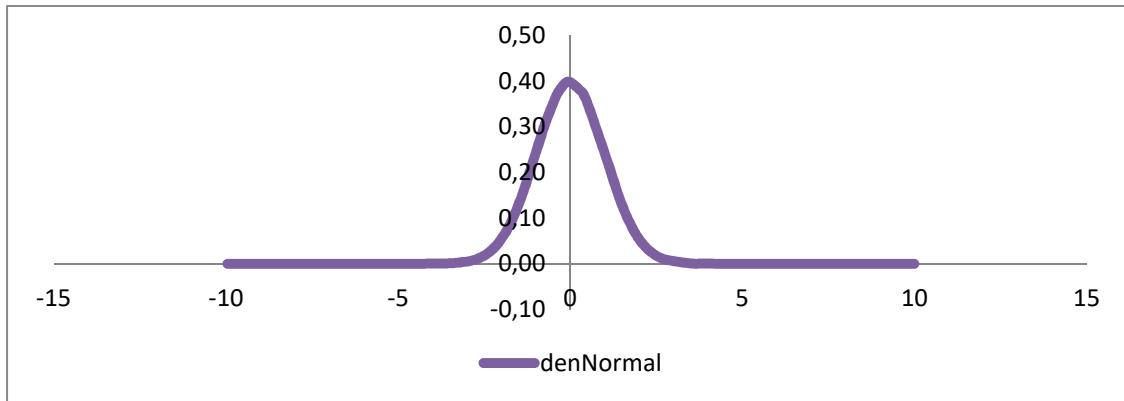
Tabla 2. Proporción de rechazos de H_0 si $u_i \sim N(0,2)$

Estadístico contraste	T=50			T=100			T=500		
	1%	5%	10%	1%	5%	10%	1%	5%	10%
JB	0,018	0,045	0,066	0,018	0,038	0,068	0,011	0,034	0,084
LI	0,006	0,046	0,098	0,007	0,052	0,1	0,008	0,049	0,102

Teniendo en cuenta que siempre la hipótesis nula es normalidad, esta tabla, a diferencia de las demás, no nos muestra la potencia, sino el tamaño del error de tipo I. Nosotros queremos que dicho error se aproxime al fijado previamente ($0.01, 0.05$ ó 0.1) para así poder decir que capta la normalidad. Es decir, si nos fijamos en la tabla, lo ideal que buscamos es que los valores de las celdas se aproximen a $0.01, 0.05$ ó 0.1 (según ε). Esto ocurre para un tamaño muestral de 100 y 500 , cuando ε es 5% y 10% para el contraste de Lilliefors. Vemos que para los tres tamaños muestrales analizados, ambos

contrastes detectan la normalidad (no rechazan H_0), pero es el contraste de Lilliefors el que mejor la capta en comparación con Jarque-Bera.

Gráfico 1. Función de densidad Normal cuando $u \sim N(0,2)$



El gráfico 1 muestra cómo es la función de densidad de una Normal. Así, vemos gráficamente cuál es la estructura que toma, y posteriormente, podemos compararla con otras distribuciones.

En la tabla 3 se muestran la proporción de rechazos de la hipótesis nula siendo la perturbación aleatoria una t de Student o una logística, conociendo que estas distribuciones son similares a la normal.

Tabla 3. Proporción de rechazos de H_0 si $u_i \sim t$ de Student (t_n) y logística.

Distribución y estadístico		T=50			T=100			T=500		
		1%	5%	10%	1%	5%	10%	1%	5%	10%
t_1	JB	0,983	0,988	0,991	0,999	1	1	1	1	1
	LI	0,961	0,984	0,991	0,999	1	1	1	1	1
t_{15}	JB	0,07	0,112	0,147	0,106	0,185	0,222	0,311	0,436	0,527
	LI	0,008	0,057	0,116	0,005	0,068	0,139	0,022	0,123	0,228
t_{40}	JB	0,031	0,058	0,088	0,042	0,082	0,116	0,056	0,127	0,197
	LI	0,008	0,048	0,092	0,008	0,05	0,114	0,011	0,052	0,118
Logística	JB	0,145	0,204	0,248	0,286	0,384	0,445	0,804	0,889	0,911
	LI	0,015	0,086	0,168	0,031	0,153	0,251	0,204	0,483	0,652

Sabemos que aquellas distribuciones t de Student con grados de libertad superiores a 30, se aproximarán a una $N(0,1)$. En la tabla 3 podemos ver cómo conforme aumentan estos grados de libertad, la potencia es cada vez menor, lo que quiere decir que no rechaza la normalidad dada la similitud entre ambas distribuciones. Es decir, la similitud entre ambas distribuciones hace que el contraste no discrimine entre ellas.

Si comparamos gráficamente la distribución Normal con la t de Student, podemos observar cómo, a medida que los grados de la t aumentan, la diferencia entre la distribución de probabilidad de ambas se hace menor. Esto lo podemos observar en los gráficos 2, 3 y 4 presentados a continuación. Destacamos que la diferencia cuando toma grado de libertad 15 y 40, es muy pequeña, tanto en tablas como gráficamente.

Gráfico 2. Función de densidad $u \sim t_1$

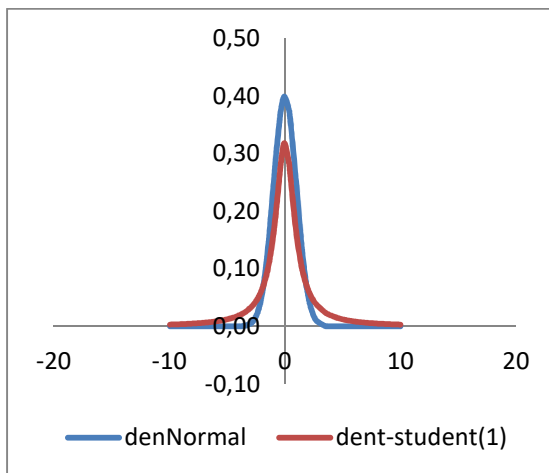


Gráfico 3. Función de densidad $u \sim t_{15}$

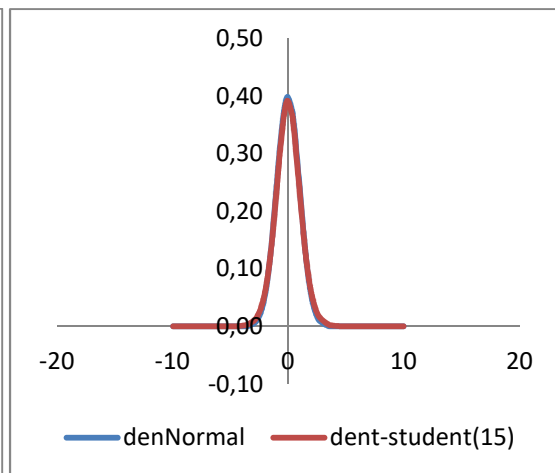
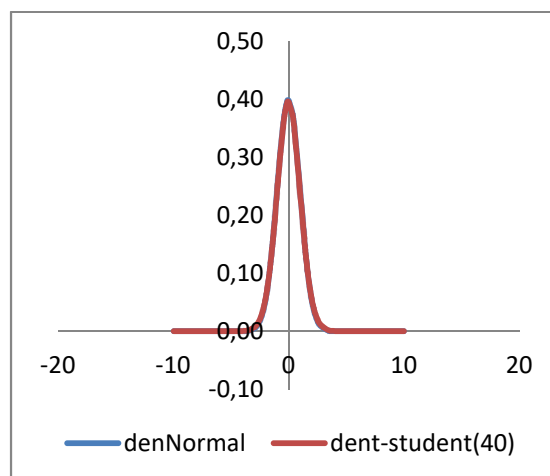


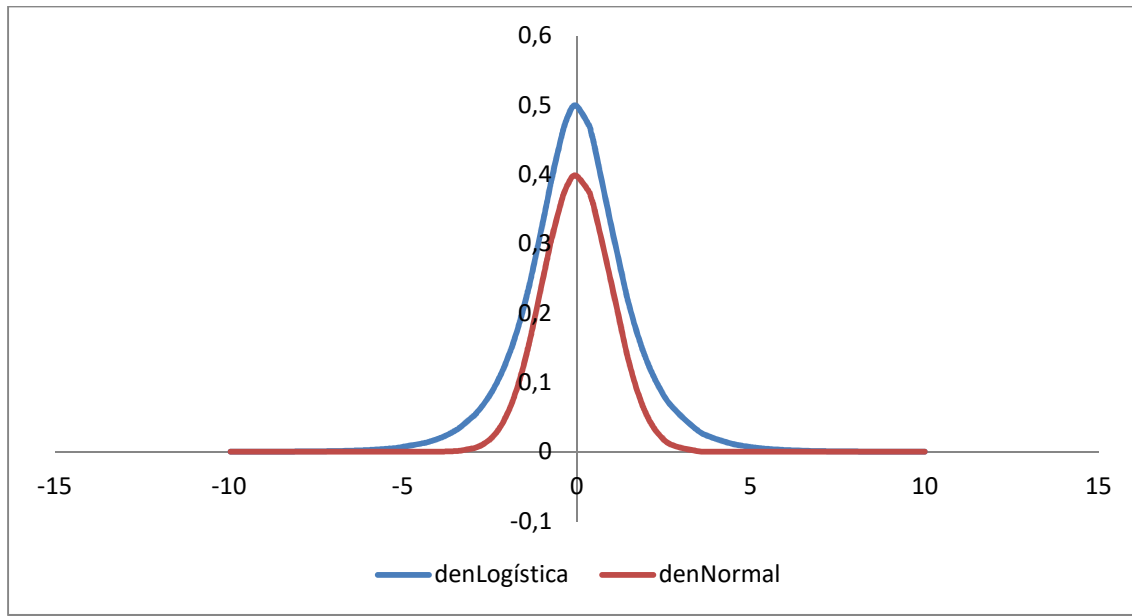
Gráfico 4. Función de densidad $u \sim t_{40}$



En la tabla 3 vemos que cuando u sigue una distribución logística, las potencias no son tan bajas como en las distribuciones t_{15} y t_{40} .

El gráfico 5 muestra la función de densidad de una distribución logística comparada con una distribución Normal.

Gráfico 5. Función de densidad de la distribución logística.



A pesar de que la distribución logística es similar a la normal, el gráfico 5 muestra diferencias que justifican las proporciones de rechazos que hemos obtenido. En muestras grandes el contraste capta la no normalidad cuando la perturbación es logística. No obstante, Lilliefors sigue mostrando una potencia más baja. Por tanto, en este caso el estadístico Jarque-Bera se comporta mucho mejor que Lilliefors.

Observamos en la tabla 3 que para ambas distribuciones el contraste más potente es el de Jarque-Bera. Cabría destacar que en la función t de student, a pesar de que Jarque-Bera es más potente, para niveles de significación de 0,1, vemos que las diferencias entre éste contraste y Lilliefors son muy pequeñas.

La tabla 4 presenta los resultados obtenidos cuando la perturbación aleatoria sigue una distribución F de Snedecor o Chi-cuadrado.

Por lo explicado anteriormente sobre la toma de decisiones de cómo hemos dividido las variables en tablas, conocemos que estas dos distribuciones, F y X^2 , se forman a partir de la normal. Sin embargo, el rango de valores que pueden tomar las variables es muy diferente al de la Normal, ya que sólo adopta valores positivos.

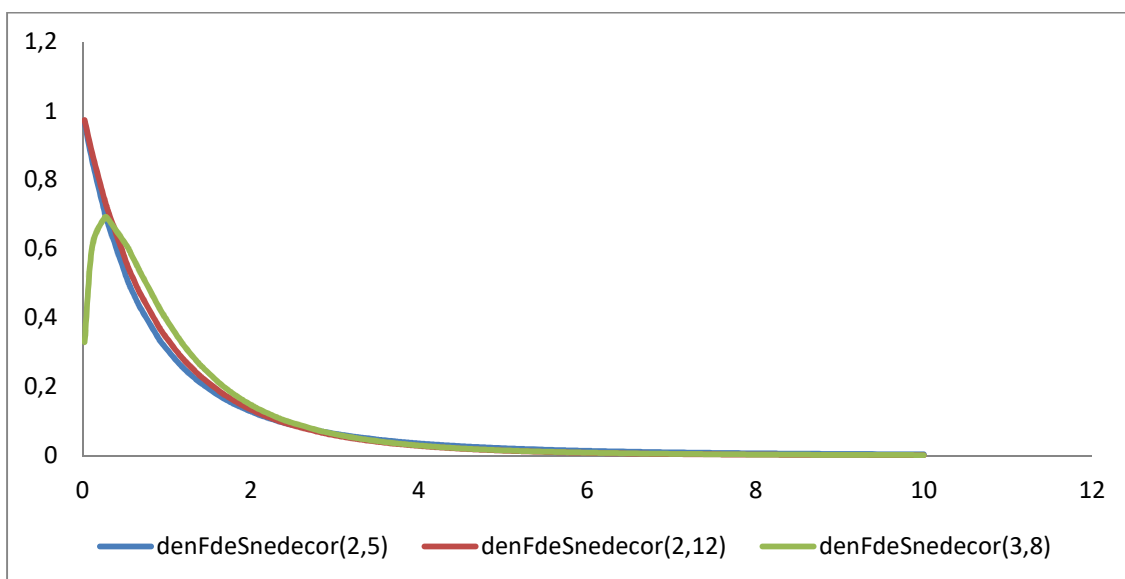
Tabla 4. Proporción de rechazos de H_0 si $u_i \sim F$ de Snedecor ($F_{n,m}$) y Chi-cuadrado (X^2_n)

Distribución y estadístico		T=50			T=100			T=500		
		1%	5%	10%	1%	5%	10%	1%	5%	10%
$F_{2,5}$	JB	0,987	0,997	0,999	1	1	1	1	1	1
	LI	0,957	0,995	0,997	1	1	1	1	1	1
$F_{2,12}$	JB	0,943	0,98	0,99	1	1	1	1	1	1
	LI	0,842	0,96	0,984	0,998	0,999	1	1	1	1
$F_{3,8}$	JB	0,938	0,978	0,987	1	1	1	1	1	1
	LI	0,824	0,952	0,978	1	1	1	1	1	1
X^2_1	JB	0,97	0,991	0,999	1	1	1	1	1	1
	LI	0,943	0,992	0,998	1	1	1	1	1	1
X^2_5	JB	0,505	0,632	0,718	0,872	0,949	0,982	1	1	1
	LI	0,215	0,499	0,636	0,57	0,829	0,899	1	1	1
X^2_{15}	JB	0,178	0,284	0,367	0,407	0,577	0,672	1	1	1
	LI	0,037	0,209	0,325	0,141	0,392	0,571	0,914	0,987	0,996

La distribución F de Snedecor destaca por sus elevadas potencias aun con tamaños muestrales y niveles de significación pequeños, lo que quiere decir que captan la no normalidad de las variables. Pero en el caso de la distribución chi-cuadrado, para grados de libertad pequeños tiene potencias muy elevadas (captando así la no normalidad), pero conforme aumenta los grados de libertad, la potencia va disminuyendo.

El gráfico 6 muestra la función de densidad de una variable que sigue una F de Snedecor, comparando diferentes grados de libertad. De esta forma comprobamos que son funciones muy diferentes a la Normal. No obstante se observa que para la $F_{3,8}$, aun tomando sólo valores positivos, es la que más podríamos decir que se parece a la Normal, aún siendo bastante diferente.

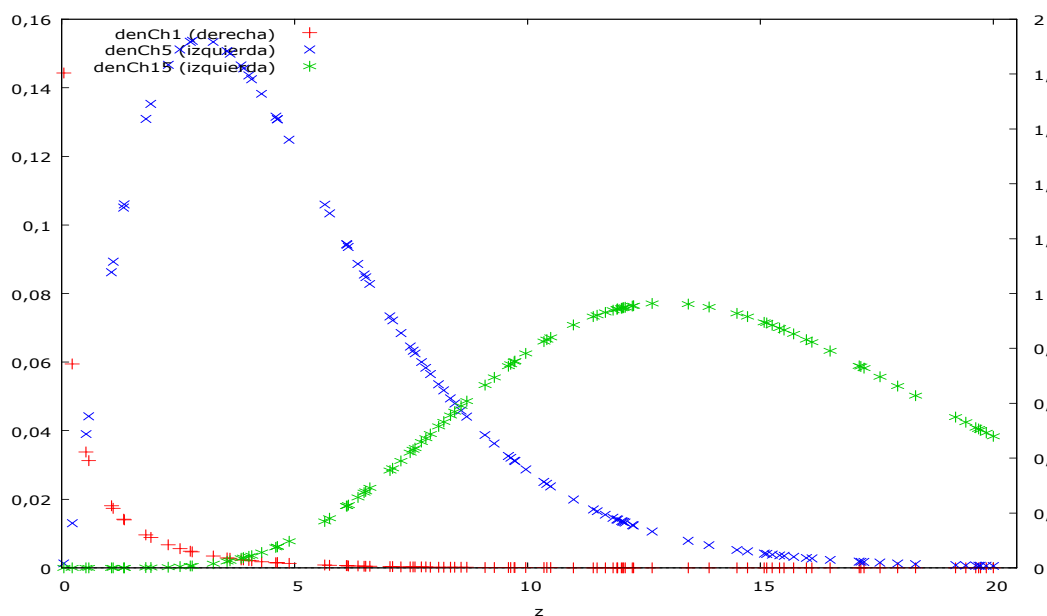
Gráfico 6. Función de densidad de la distribución F de Snedecor



Estas grandes diferencias justifican los resultados de la tabla 4.

En el gráfico 7 se muestran la evolución de las distribuciones chi-cuadrado conforme aumentan los grados de libertad.

Gráfico 7. Función de densidad de la distribución Chi-cuadrado.¹



¹ "denCh1", "denCh5" y "denCh15" son las densidades de distribuciones X_1^2 , X_5^2 , y X_{15}^2 respectivamente.

Sabemos que la distribución chi-cuadrado se forma a partir de la Normal, pero aun con eso, es muy diferente a la Normal, característica que los gráficos muestran. Con ello podríamos decir que, aunque para tamaños muestrales pequeños (50), y cuando la

distribución tiene 15 grados de libertad, tanto Jarque-Bera como Lilliefors no captan la falta de normalidad, los gráficos muestran la gran diferencia con respecto a la Normal. El gráfico podría justificar en parte este resultado, ya que muestra (línea verde) que dicha distribución sigue una forma parecida a la Normal sólo que no está centrada en el cero.

En la tabla 4 se observa que Jarque-Bera se comporta mejor, porque tiene una potencia más alta que Lilliefors. Es importante tenerlo en cuenta en muestras pequeñas, porque cuando aumenta el tamaño muestral, ambos son muy potentes, no habría diferencia.

Por último, respecto a este grupo de distribuciones, cabría destacar que para $T=500$, la potencia es muy alta en todos los casos.

La tabla 5 nos muestra la proporción de rechazos de la hipótesis nula cuando la u_i sigue una distribución log-normal. Una variable aleatoria X sigue una distribución log-normal si su logaritmo sigue una normal, y en este caso, sigue una normal con media 3 y varianza 0.5.

Tabla 5. Proporción de rechazos de H_0 si $u_i \sim \text{Log-Normal} [\log(x) \sim N(\mu, \sigma^2)]$

Distribución y estadístico		T=50			T=100			T=500		
		1%	5%	10%	1%	5%	10%	1%	5%	10%
log(x) ~ N _{3, 0.5}	JB	0,626	0,744	0,81	0,949	0,987	0,993	1	1	1
	LI	0,358	0,646	0,755	0,735	0,93	0,97	1	1	1

Observamos que es más potente Jarque-Bera que Lilliefors, aunque cabe destacar que conforme aumentamos el tamaño muestral y el nivel de significación las diferencias entre ambos contrastes desaparecen.

Es importante mencionar, que a pesar de ser una distribución que proviene de la Normal, las potencias son bastante elevadas, lo que quiere decir que ambos contrastes detectan la no normalidad de la distribución.

En la tabla 6 presentamos los resultados cuando $u_i \sim \text{Gamma}(G_{a,p})$ y su caso particular, exponencial (E_λ).

Tabla 6. Proporción de rechazos de H_0 si $u_i \sim \text{Gamma}(G_{a,p})$ y Exponencial (E_λ)

Distribución y estadístico		T=50			T=100			T=500		
		1%	5%	10%	1%	5%	10%	1%	5%	10%
G_{0,1,1}	JB	1	1	1	1	1	1	1	1	1
	LI	1	1	1	1	1	1	1	1	1
G_{1,1} ó E₁	JB	0,827	0,926	0,965	0,997	1	1	1	1	1
	LI	0,664	0,889	0,939	0,986	0,998	1	1	1	1
G_{3,1}	JB	0,401	0,558	0,651	0,813	0,92	0,961	1	1	1
	LI	0,18	0,466	0,605	0,47	0,782	0,873	1	1	1
G_{4,1}	JB	0,345	0,474	0,568	0,671	0,841	0,912	1	1	1
	LI	0,128	0,365	0,477	0,327	0,647	0,765	0,999	1	1

Se trata de dos distribuciones muy diferentes a la Normal, de ahí que las potencias sean muy altas aun cuando miramos tamaños muestrales pequeños, y cuando aumentamos los parámetros que las componen. Se puede resaltar que en la distribución Gamma cuyo parámetro a es 0,1 y cuyo parámetro p es 1, ambos estadísticos detectan la no normalidad para cualquier tamaño muestral. En general, en tamaños muestrales grandes se detecta la no normalidad sea cual sean los parámetros, pero en tamaños muestrales pequeños la potencia es más baja cuando mayor es el parámetro a .

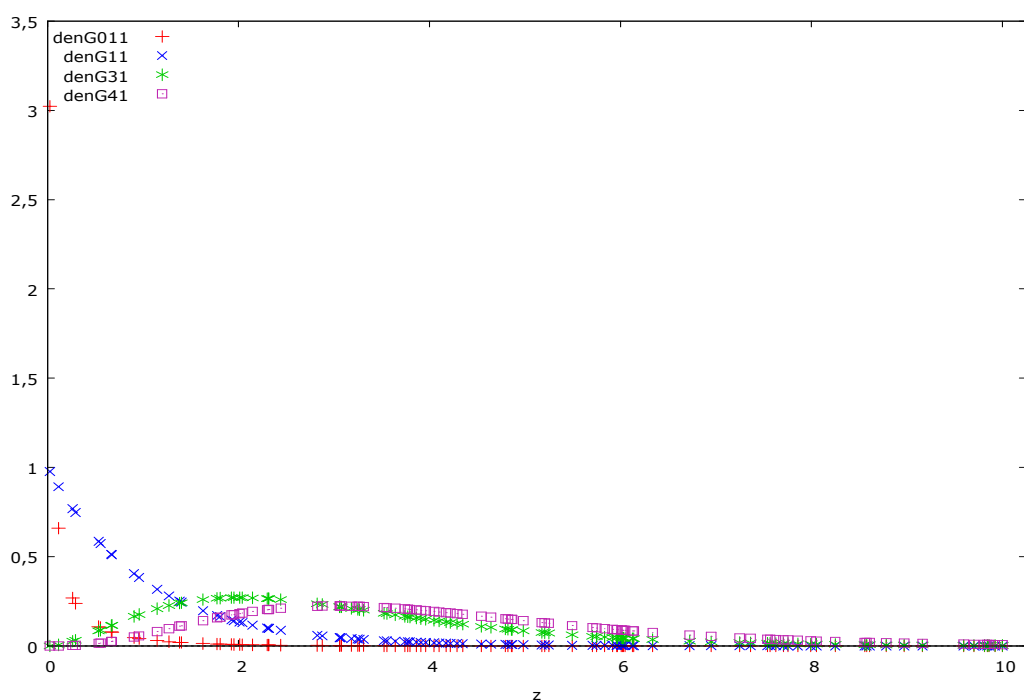
Conforme el parámetro a aumenta, ambos contrastes son menos potentes. Esto ocurre en $T = 50$ y $T = 100$, no es $T = 500$, donde la potencia es siempre la máxima.

El contraste de Jarque-Bera es más potente que el de Lilliefors, aspecto que se nota sobre todo para tamaños muestrales pequeños ya que conforme éstos aumentan, las diferencias disminuyen.

El gráfico 8 muestra las distribuciones Gamma que hemos analizado en la tabla 6.

Vemos que tanto la Gamma (3,1) como la Gamma(4,1) tienen una forma algo parecida a una “campana” como la Normal, tal y como mostraba la tabla 6. Asimismo, las otras dos distribuciones Gamma analizadas, vemos gráficamente que toman una forma diferente a la Normal.

Gráfico 8. Función de densidad de la distribución Gamma. ²



² "denG011", "denG11", "denG31" y "denG41" son las densidades de distribuciones $G_{0,1,1}$, $G_{1,1}$, $G_{3,1}$ y $G_{4,1}$ respectivamente.

La tabla 7 muestra el caso en el que la distribución de u es bimodal. Hemos generado las distribuciones bimodales 50-50 [$N(\mu, \sigma^2)$, $N(\mu, \sigma^2)$].

Tabla 7. Proporción de rechazos de H_0 si $u_i \sim \text{Bimodal}$

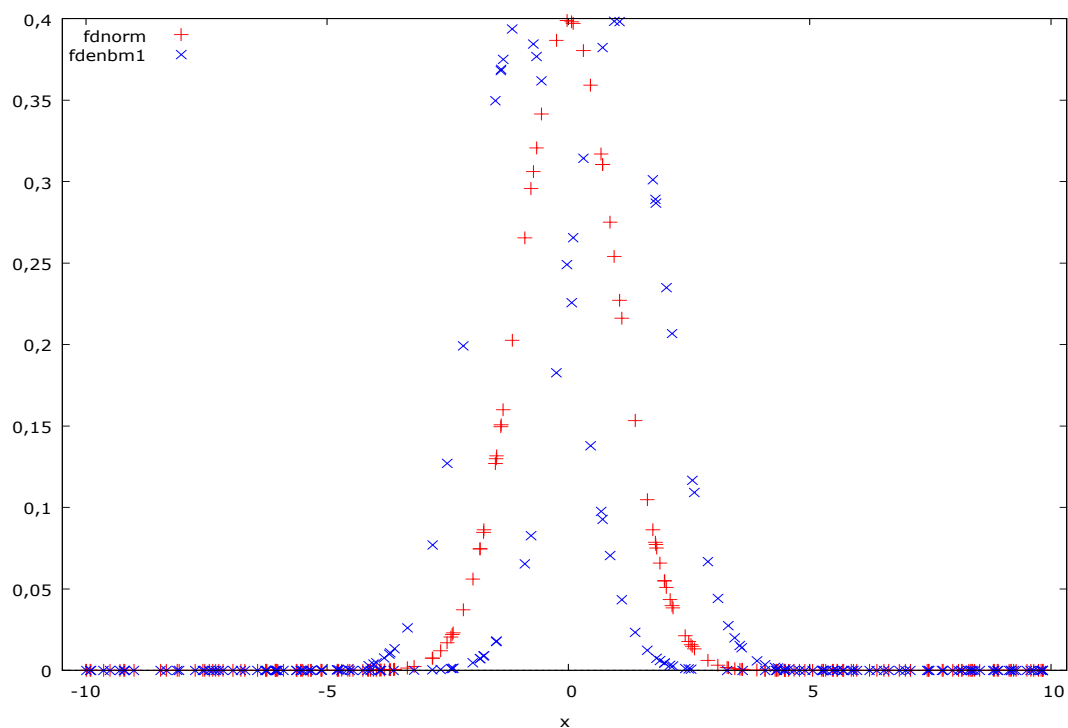
Distribución y estadístico		T=50			T=100			T=500		
		1%	5%	10%	1%	5%	10%	1%	5%	10%
N(-1,1), N(1,1)	JB	0	0,001	0,006	0	0,003	0,019	0,128	0,479	0,687
	LI	0,006	0,049	0,109	0,01	0,084	0,154	0,118	0,425	0,604
N(-3,1), N(3,1)	JB	0	0,017	0,78	0,822	1	1	1	1	1
	LI	0,937	0,999	1	1	1	1	1	1	1
N(-1,5), N(1,5)	JB	0,014	0,038	0,053	0,022	0,038	0,061	0,008	0,038	0,082
	LI	0,005	0,051	0,101	0,01	0,05	0,089	0,006	0,047	0,104
N(-3,5), N(3,5)	JB	0,009	0,02	0,038	0,009	0,021	0,035	0,002	0,024	0,068
	LI	0,005	0,048	0,093	0,007	0,056	0,104	0,012	0,062	0,113

Notar que hemos mantenido siempre $\pm\mu$, y la misma varianza. Esto podría haber sido de otro modo, pero la lógica nos lleva a hacerlo así, para que al menos exista cierta similitud con la normal, en cuanto a “simetría” y a varianza constante.

En la tabla 7, se observa que cuando la varianza es pequeña (en nuestro caso, 1), conforme se aumenta la media, la potencia aumenta, y por tanto captan la no normalidad. Sin embargo, cuando aumentamos la varianza (en nuestro caso, 5), conforme aumenta la media, vemos que la potencia es más pequeña. A nivel general, con varianza 5, ambos contrastes no captan la no normalidad.

Los gráficos 9 y 10 muestran las distribuciones bimodales con varianza igual a la unidad y con distinta media (en el caso del gráfico 9, es con media igual a ± 1 , y en el caso del gráfico 10, es con media igual a ± 3).

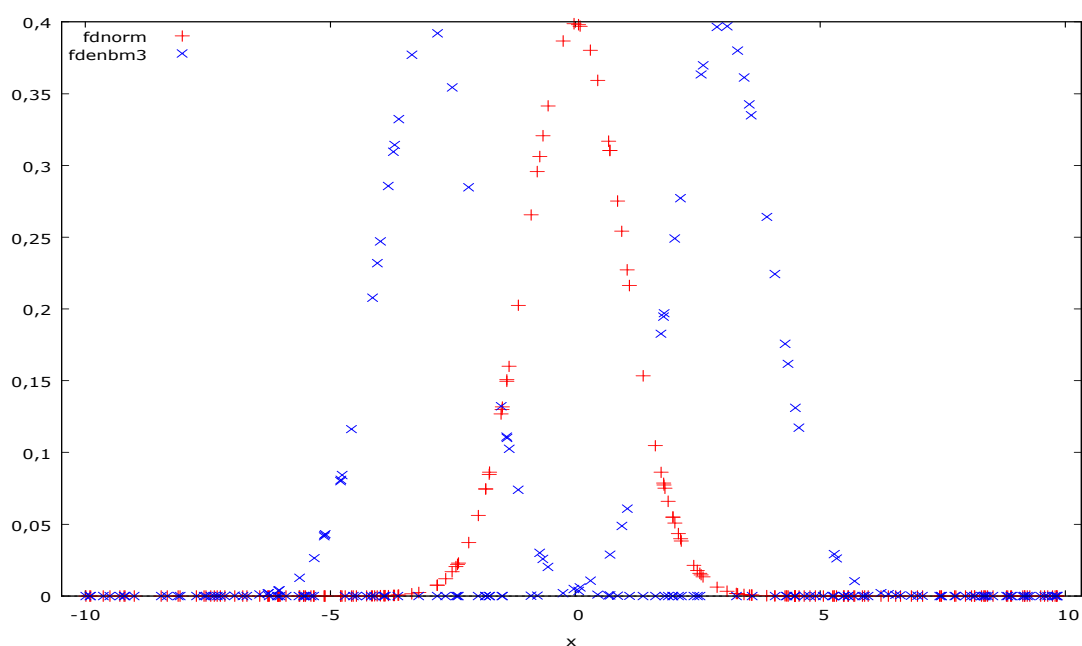
Gráfico 9. Función de densidad Bimodal con $\mu=\pm 1$ y $\sigma^2 = 1$.³



³“fdnorm” y “fdenbm1” son las densidades de distribuciones $N_{0,1}$ y Bimodal (50%) $N(-1,1), N(1,1)$.

Vemos que cuando la media es 3, gráficamente es mucho más diferente (tal y como vemos en la tabla 7 con las potencias tan elevadas), por tanto, detectan la falta de normalidad, en cambio, cuando toma media 1, lo que ocurre es que no discrimina por la cercanía de las funciones de densidad.

Gráfico 10. Función de densidad Bimodal con $\mu=\pm 3$ y $\sigma^2 = 1$.⁴

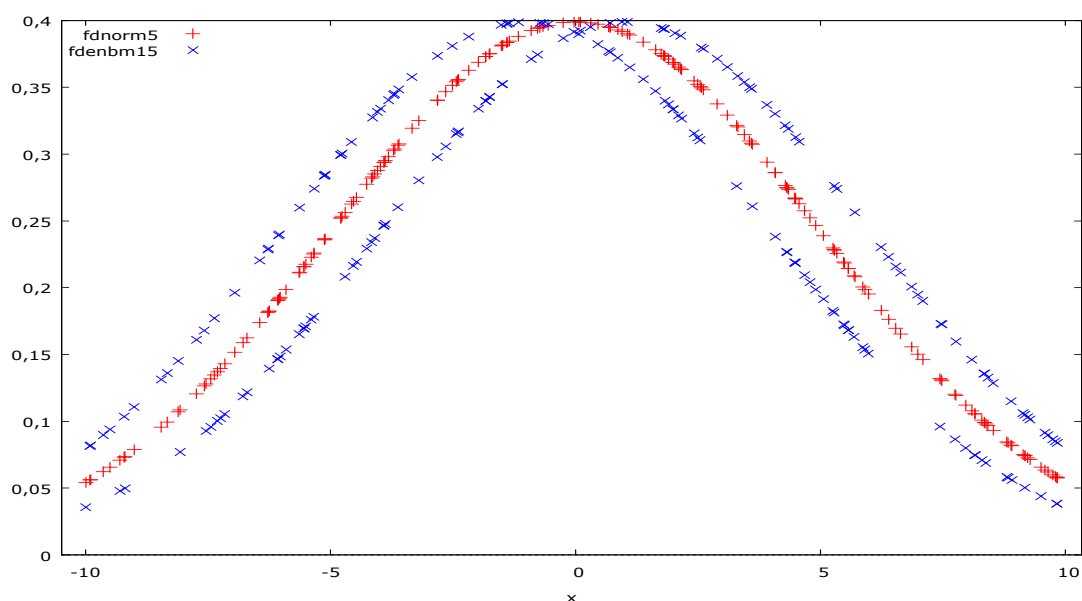


⁴“fdnorm” y “fdenbm3” son las densidades de distribuciones $N_{0,1}$ y Bimodal (50%) $N(-3,1), N(3,1)$.

Por otro lado, cuando aumentamos la varianza (en nuestro caso a 5), lo que ocurre es que ambos contrastes detectan la normalidad siendo más potente Lilliefors.

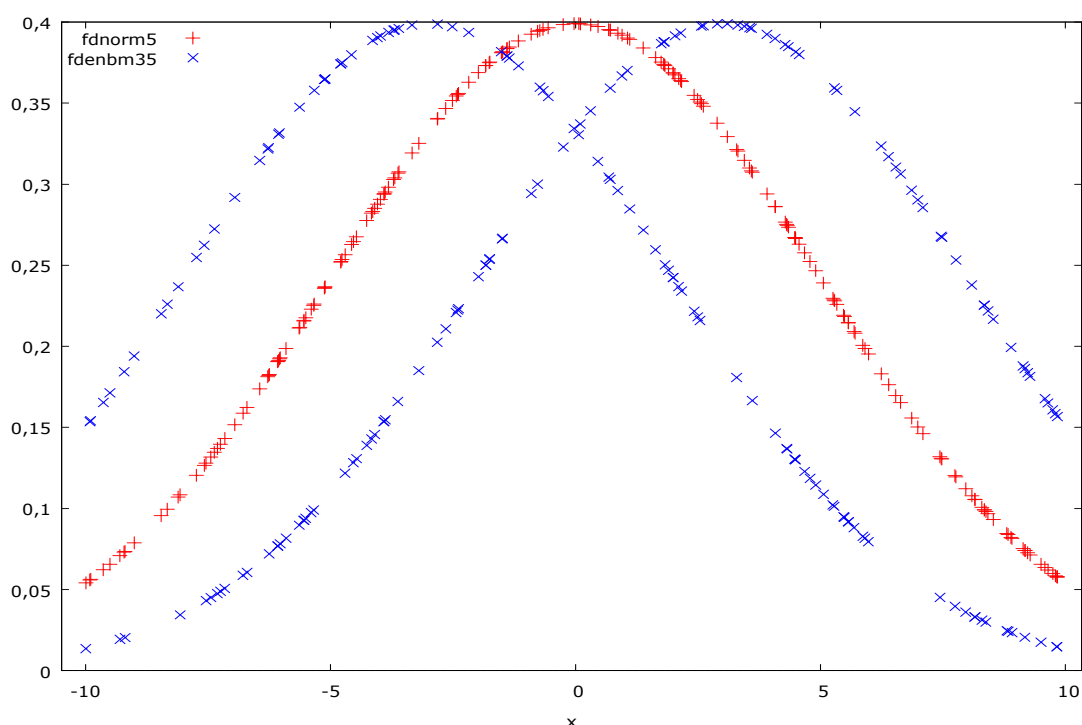
Los gráficos 11 y 12 muestran las distribuciones bimodales con varianza igual a la cinco y con distinta media (en el caso del gráfico 11, la media es igual a ± 1 , y para el gráfico 12, la media es igual a ± 3).

Gráfico 11. Función de densidad Bimodal con $\mu=\pm 1$ y $\sigma^2 = 5$.⁵



⁵“fdnorm5” y “fdenbm15” son las densidades de distribuciones $N_{0,5}$ y Bimodal (50%) $N(-1,5), N(1,5)$.

Gráfico 12. Función de densidad Bimodal con $\mu=\pm 3$ y $\sigma^2 = 5$.⁶



⁶“fdnorm5” y “fdenbm35” son las densidades de distribuciones $N_{0,5}$ y Bimodal (50%) $N(-3,5), N(3,5)$.

Vemos que ambos gráficos muestran lo que los resultados de la tabla 7 exponían, y es la similitud que comparten las distribuciones bimodales con varianza grande (5), con la función de distribución Normal. Normalidad en el aspecto de que a pesar de ser dos subpoblaciones, las probabilidades que toman en función de la perturbación aleatoria son muy parecidas.

Como resumen general, hemos realizado una tabla (tabla 8) en la que, de mayor a menor potencia de los contrastes, se ordenan las distribuciones señalando en cada caso qué contraste (Jarque-Bera ó Lilliefors) es más potente. Hemos establecido las potencias para cada tamaño muestral al 5% de nivel de significación, así como que las hemos ordenado, resaltando que las cuatro últimas distribuciones, cambian de orden cuando el tamaño muestral pasa de 50 a 100, pero se mantienen en el mismo orden cuando se pasa de tamaño muestral 100 a 500.

Tabla 8. Cuadro resumen de los resultados de Experimento de Monte Carlo.⁷

T=50			T=100			T=500		
Distrib	Pot	CMP	Distrib	Pot	CMP	Distrib	Pot	CMP
$G_{0.1, 1}$	1	JB y LI	$G_{0.1, 1}$	1	JB y LI	$G_{0.1, 1}$	1	JB y LI
$N_{(-3,1)(3,1)}$	0,999	LI	$N_{(-3,1)(3,1)}$	1	LI	$N_{(-3,1)(3,1)}$	1	LI
$F_{2,5}$	0,997	JB	$F_{2,5}$	1	JB	$F_{2,5}$	1	JB
X^2_1	0,991	JB	X^2_1	1	JB	X^2_1	1	JB
$F_{2,12}$	0,98	JB	$F_{2,12}$	1	JB	$F_{2,12}$	1	JB
$F_{3,8}$	0,978	JB	$F_{3,8}$	1	JB	$F_{3,8}$	1	JB
t_1	0,988	JB	t_1	1	JB	t_1	1	JB
$G_{1,1}$	0,926	JB	$G_{1,1}$	1	JB	$G_{1,1}$	1	JB
log-normal	0,744	JB	log-normal	0,987	JB	log-normal	1	JB
X^2_5	0,632	JB	X^2_5	0,949	JB	X^2_5	1	JB
$G_{3,1}$	0,558	JB	$G_{3,1}$	0,92	JB	$G_{3,1}$	1	JB
$G_{4,1}$	0,474	JB	$G_{4,1}$	0,841	JB	$G_{4,1}$	1	JB
X^2_{15}	0,284	JB	X^2_{15}	0,577	JB	X^2_{15}	1	JB
Logística	0,204	JB	Logística	0,384	JB	Logística	0,889	JB
t_{15}	0,112	JB	t_{15}	0,185	JB	t_{15}	0,436	JB
t_{40}	0,058	JB	$N_{(-1,1)(1,1)}$	0,084	LI	$N_{(-1,1)(1,1)}$	0,425	LI
$N_{(-1,5)(1,5)}$	0,051	LI	t_{40}	0,082	JB	t_{40}	0,127	JB
$N_{(-1,1)(1,1)}$	0,049	LI	$N_{(-3,5)(3,5)}$	0,056	LI	$N_{(-3,5)(3,5)}$	0,062	LI
$N_{(-3,5)(3,5)}$	0,048	LI	$N_{(-1,5)(1,5)}$	0,05	LI	$N_{(-1,5)(1,5)}$	0,047	LI

⁷“Distrib”, “Pot” y “CMP” son la distribución de la u_i en el PGD, la potencia para cada distribución y el contraste de mayor potencia.

A través de la tabla 8, vemos claramente, que exceptuando en las distribuciones bimodales, en el resto de distribuciones es más potente Jarque-Bera que Lilliefors. Se puede añadir que, las distribuciones que mejor captan la falta de normalidad son muy distintas entre sí, es decir, no son todas las distribuciones iguales pero con distintos grados de libertad. También resalta que, las distribuciones que no discriminan la normalidad, son la distribución logística, la distribución t de student (con 15 y 40 grados de libertad), y finalmente las distribuciones bimodales analizadas excepto la que tiene media 3 y varianza 1. Para entender esta excepción podemos observar los cuatro gráficos de las bimodales. Claramente, el correspondiente a la bimodal $N(-3,1), N(3,1)$, al tener medias muy distanciadas y dispersión pequeña, concentra las probabilidades altas en valores más extremos que en la normal. En cambio, si las medias están más próximas las probabilidades altas se darán para valores similares de la variable, y más todavía si la dispersión es mayor (varianza 5).

6. CONCLUSIONES

El objetivo de este trabajo ha consistido en mostrar la forma en la que se trabaja con experimentos de Monte Carlo a la hora de analizar si dos contrastes captan o no la falta de normalidad dependiendo de las distribuciones que se estudian. Teniendo en cuenta que en el software econométrico que usamos habitualmente Gretl aparecen varios estadísticos de contraste, de ellos en este trabajo hemos comparado Jarque-Bera con Lilliefors.

A la vista de los resultados se pueden obtener algunas conclusiones. En primer lugar, cuando la distribución que establecemos como perturbación aleatoria es Normal, el contraste que mejor se comporta es Lilliefors, que tiende a aceptar la hipótesis de normalidad una proporción de veces muy próxima al nivel de significación. Por otro lado, cuando analizamos distribuciones muy similares a la normal (simétricas y que toman valores entre $-\infty$ y $+\infty$), que en nuestro caso han sido la logística y la distribución t de Student, el contraste más adecuado es Jarque-Bera. En tercer lugar, hemos observado que, para todas aquellas distribuciones que son muy diferentes a la Normal, es mucho más potente Jarque-Bera que el contraste de Lilliefors, salvo en el caso de la Bimodal, ya que en este caso es más potente Lilliefors que Jarque-Bera.

Aunque en nuestro trabajo hemos generado la perturbación en cada caso con distribuciones diferentes a la Normal, en realidad en la práctica el hecho de que u_i no siga una Normal puede deberse a un error en la especificación de la parte sistemática. La diferencia entre la parte sistemática especificada y la verdadera, se supone que irá a la perturbación que puede dejar de ser Normal. La solución que más aparece en los libros para corregir la falta de normalidad consiste en aplicar logaritmos, y hemos comprobado en los casos de nuestros experimentos que no siempre resuelve el problema. Únicamente si el error es que estimamos un modelo lineal pero el PGD es logarítmico o semilogarítmico, la no normalidad que captan los contrastes se resolvería aplicando logaritmos en la variable dependiente y/o en las explicativas.

Es decir, la solución de aplicar logaritmos para corregir la no normalidad sería efectiva en situaciones en las que hemos especificado un modelo lineal:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i}$$

y sin embargo, el modelo verdadero es uno semilogarítmico, o doblemente logarítmico, es decir:

$$\ln Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} \quad \text{ó}$$

$$\ln Y_i = \beta_1 + \beta_2 \ln X_{2i} + \beta_3 \ln X_{3i}$$

ó incluso una mezcla de ambos.

Hemos comprobado mediante simulaciones que sólo en este caso, aplicar logaritmos a la Y, y/o a las X, puede resolver el problema, ya que vemos como los contrastes pasan de potencias altas a bajas. Pero en una situación más general, como cualquiera de los experimentos recogidos en el apartado anterior de este trabajo, aplicar logaritmos no resuelve el problema. No obstante, el error que se comete no debería ser importante cuando las distribuciones son similares.

Para terminar las conclusiones, se intentarán contestar a las preguntas formuladas anteriormente (véase página 22). En primer lugar, es destacable que Jarque-Bera es mucho más potente en casi todas las distribuciones excepto en las Bimodales, en las que resalta Lilliefors. En segundo lugar, se ha corroborado que son más fiables los resultados contemplados en tamaños muestrales más grandes. Esto es, siempre se cumple el hecho de que conforme aumenta el tamaño muestral, la potencia es más

grande. Además, hemos podido comprobar que cuanto mayor es el nivel de significación mayor es la potencia.

7. BIBLIOGRAFÍA

Trivez Bielsa, F.J. (2004): *Introducción a la Econometría*. Ed. Pirámide, Madrid.

Otero, J.M (1978): *Lógica y limitaciones de la Econometría*. Ed. ICE, Madrid.

Peña Sánchez de Rivera, D (1991): *Estadística Modelos y Métodos, 1, Fundamentos*. Ed. Alianza, Madrid.

Gujarati, D.N. (2003): *Econometría*.

Cao Abad, R. y otros (2008): *Introducción a la estadística y sus aplicaciones*. Ed. Pirámide, Madrid.

C.M Jarque y A. K. Bera (1987): “A test for Normality of Observations and Regression Residuals”. *International Statistical Review*, Vol. 55, páginas 163-172.

Doornik, J.A. y Hansen, H. (2008): “An Omnibus Test for Univariate and Multivariate Normality”, *Oxford Bulletin of Economics and Statistics*, Vol. 70, páginas 927–939