
ANÁLISIS DE LA DISTRIBUCIÓN DE LAS TASAS DE RETORNO ACCIONARIAS HACIENDO USO DE LA DISTRIBUCIÓN G Y H DE TUKEY

Andrés Mauricio Mendoza Piñeros¹
José Alfredo Jiménez Moscoso²

A nivel teórico y práctico, el modelo de Movimiento Browniano Geométrico (*MBG*) y el Modelo de Black and Scholes (*MBS*) son muy utilizados en el mercado de acciones para conocer el comportamiento de un activo y medir las expectativas de la volatilidad del mercado. Particularmente, en el *MBG* los logaritmos de los cocientes de precios sucesivos son independientes y siguen una distribución normal. Por su parte, los retornos futuros dependen de los valores presentes (propiedad de Markov).

¹Estadístico y Magíster en Economía, investigador del Centro de Investigaciones para el Desarrollo de la Universidad Nacional de Colombia (Bogotá, Colombia). E-mail: ammendozap@unal.edu.co. Dirección de correspondencia: Calle 44 No 45-67, Unidad Camilo Torres, Bloque B4 - CID (Bogotá, Colombia).

²Magíster en Banca y Finanzas y en Ciencias Estadística. Profesor asociado del Departamento de Matemáticas de la Universidad Nacional de Colombia (Bogotá, Colombia). E-mail: josajimenezm@unal.edu.co. Dirección de correspondencia: Cra 30 No 45-03, Facultad de Ciencias, Universidad Nacional de Colombia (Bogotá, Colombia).

Los autores agradecen los valiosos comentarios de los árbitros anónimos de la Revista Cuadernos de Economía.

Este artículo fue recibido el 28 de noviembre de 2008 y su publicación aprobada el 12 de diciembre de 2009.

La forma de la distribución de los cambios de la tasa de retorno es un factor importante en la determinación del menor riesgo de inversión en acciones, además tiene gran importancia desde un punto de vista académico, puesto que ésta provee información descriptiva concerniente a la naturaleza del proceso generado por cambios de la tasa de retorno³. Asimismo, es de gran utilidad para el trabajo empírico en esta área.

Por las razones antes expuestas, el objetivo de este artículo es conocer la distribución del retorno de las acciones haciendo uso de la distribución g y h de Tukey, siendo ésta una herramienta eficiente en el modelamiento de fenómenos financieros, al recoger la asimetría y la curtosis.

El artículo está organizado de la siguiente manera: en la primera sección se presenta la distribución g y h de Tukey y se hace una breve revisión de sus propiedades. En el apartado siguiente, se muestra la metodología de estimación de parámetros. En la tercera parte se expone un caso aplicado a un activo particular, en el cual la función de densidad tiene parámetros no constantes y se evidencia un ajuste significativo sobre la distribución empírica de los datos. Finalmente, se presentan las conclusiones.

PRELIMINARES

La distribución g y h ya ha sido utilizada en finanzas por autores como: Badrinath y Chatterjee (1988), Mills (1995), y Badrinath y Chatterjee (1991), con el fin de modelar el retorno en un índice accionario y el retorno en acciones en varios mercados. En Dutta y Babbel (2002) y Dutta y Babbel (2005) se emplea la distribución para modelar tasas de interés y opciones en tasas de interés, mientras que en Dutta y Perry (2007) la g y la h son usadas para estimar riesgo operativo. Sin embargo, en estos artículos no se presenta de manera explícita la función de densidad.

A continuación se presenta su definición básica y sus propiedades como lo hizo Tukey (1977).

³En este artículo, la tasa de retorno durante el período t se consigue suponiendo un proceso de capitalización continua, es decir:

$$R_t = \ln \left(\frac{P_t}{P_{t-1}} \right) \quad t = 1, 2, \dots, T$$

Siendo P_t el precio en el tiempo t .

Familia de distribuciones g y h de Tukey

Sea Z una variable aleatoria con distribución normal estándar y sean g y h dos constantes (parámetros). La variable aleatoria Y dada por:

$$Y = T_{g,h}(Z) = \frac{1}{g} (\exp(gZ) - 1) \exp(hZ^2/2) \quad \text{con } g \neq 0, h \in \mathbb{R} \quad (1)$$

tiene distribución g y h de Tukey.

Los parámetros g y h miden respectivamente el sesgo y alargamiento de las colas de la distribución g y h de Tukey.

Cuando $h = 0$ la distribución g y h de Tukey se reduce a:

$$T_{g,0}(Z) = \frac{1}{g} (\exp(gZ) - 1) \quad (2)$$

Ésta tiene una **distribución g de Tukey**, también conocida como familia de distribuciones lognormales, porque presentan un alargamiento de las colas distinto a la distribución normal estándar y son sesgadas con respecto a esta última.

De forma similar, cuando $g \rightarrow 0$ la distribución g y h de Tukey esta dada por:

$$T_{0,h}(Z) = Z(\exp(hZ^2/2)) \quad (3)$$

la cual tiene una **distribución h de Tukey**. Este tipo de distribuciones, tienen la característica de ser simétricas, pero con colas más pesadas que la distribución normal estándar a medida que h crece.

En la práctica, los parámetros g y h son no constantes, por ello, en este artículo se plantea la transformación dada en (1) como sigue:

$$Y_p = \frac{1}{g(z_p)} [\exp\{g(z_p)Z_p\} - 1] \exp\{h(z_p)Z_p^2/2\}, \quad p > 0,5 \quad (4)$$

Siendo p el p -ésimo percentil de la distribución.

Por otra parte

$$g(z_p) = \sum_{i=0}^{\infty} c_i |Z_p|^i \quad \text{y} \quad h(z_p) = \sum_{i=0}^{\infty} d_i |Z_p|^{2i} \quad (5)$$

es decir $g(z_p)$ y $h(z_p)$ son polinomios de $|Z_p|$ y ambos determinan funciones pares.

Con el fin de determinar la distribución de una variable aleatoria X , se introducen parámetros de localización (A) y escala (B), y se estiman cuatro parámetros atendiendo la siguiente relación:

$$X_p = A + BY_p, \quad p > 0,5 \quad (6)$$

Función de densidad

La función de densidad se evalúo considerando que los parámetros que controlan el sesgo y el alargamiento de las colas de la distribución son constantes (g y h), Jiménez M. (2004).

$$t_{g,h}(y_p) = \frac{\varphi(z_p)}{\frac{dY_p}{dZ_p}}, \quad (7)$$

Siendo $\varphi(x)$ la función de densidad normal estándar. A partir de 7 en Jiménez M. (2004) se establece que la función de densidad $t_{g,h}(y_p)$ de una variable aleatoria Y con distribución g y h de Tukey, está dada por:

$$t_{g,h}(y_p) = \frac{\exp(-\frac{1}{2}(h+1)z_p^2)}{\sqrt{2\pi} \left[\exp(gz_p) + \frac{h}{g}z_p(\exp(gz_p) - 1) \right]} \quad (8)$$

y_p y z_p denotan el p -ésimo cuantil de la transformación $Y = T_{g,h}(Z)$ y de la distribución normal estándar, respectivamente.

Siguiendo la misma metodología se obtiene que la función de densidad de la expresión (4), viene dada por

$$t(Y_p) = \frac{\varphi(Z_p)}{T'(Z_p)} \quad (9)$$

donde

$$T'(Z_p) = \left(g'(Z_p) - \frac{g'(Z_p)}{Z_p g(Z_p)} + h(Z_p) + \frac{1}{2}Z_p h'(Z_p) \right) Z_p T(Z_p) + \left(e^{g(Z_p)Z_p} + Z_p \frac{g'(Z_p)}{g(Z_p)} \right) e^{\frac{1}{2}h(Z_p)Z_p^2}$$

Adicionalmente, en Jiménez M. (2004) a partir de la ecuación (8), se deriva la función de densidad para las distribuciones g de Tukey y h de Tukey.

La función de densidad para la variable aleatoria X , dada en la expresión (6) viene dada en Jiménez M. (2004) como una función de la distribución g y h de Tukey:

$$f_X(x_p) = f_X(A + By_p) = \frac{1}{B} t_{g,h}(y_p) \tag{10}$$

ESTIMACIÓN DE LOS PARÁMETROS DE LA FAMILIA DE DISTRIBUCIONES g Y h DE TUKEY

Método de cuantiles introducido por Hoaglin (1985)

En Hoaglin (1985), se presenta la estimación de los parámetros de la familia de distribuciones g y h de Tukey.

Usando las siguientes ecuaciones, con $Z_p = -Z_{1-p}$ y $p < 0,5$:

$$X_p - A = \frac{B}{g} (\exp(gZ_p) - 1) (\exp(hZ_p^2/2)) \tag{11}$$

$$X_{1-p} - A = \frac{B}{g} (\exp(gZ_{1-p}) - 1) (\exp(hZ_{1-p}^2/2)) \tag{12}$$

Para la estimación de los parámetros, se tienen en cuenta las siguientes propiedades, las cuales fueron recopiladas en Dutta y Babbal (2002):

1. X es una transformación estrictamente creciente de Z . Es decir, la transformación de una normal estándar a una g y h es uno a uno.
2. El parámetro de localización de la distribución g y h de Tukey, es estimado por la mediana de los datos; es decir, $A = X_{0.5}$.
3. La estimación del parámetro que controla el sesgo de la distribución (g), se estima generalmente, por la mediana de los g_p dados por la siguiente expresión:

$$g_p = - \left(\frac{1}{Z_p} \right) \ln \left(\frac{X_{1-p} - X_{0.5}}{X_{0.5} - X_p} \right)$$

4. Cuando, $g \neq 0$, el parámetro que controla el alargamiento de las colas (h), se estima mediante

$$\ln \left(\frac{g(X_p - X_{1-p})}{\exp(gZ_p) - \exp(-gZ_p)} \right) = \ln(B) + h \left(\frac{Z_p^2}{2} \right)$$

El valor de h se obtiene haciendo uso de una regresión lineal; en la cual, el coeficiente de la regresión estima el valor de h . El valor de B se obtiene tomando la función exponencial del intercepto de la regresión. Dado que la distribución no es simétrica $g \neq 0$, se pueden obtener dos estimaciones más para h y por ende, dos estimaciones diferentes del alargamiento para las dos colas de la distribución:

$$\ln \left(\frac{g(X_{1-p} - X_{0.5})}{\exp(-gZ_p) - 1} \right) = \ln(B) + h \frac{Z_p^2}{2}$$

$$\ln \left(\frac{g(X_{0.5} - X_p)}{1 - \exp(gZ_p)} \right) = \ln(B) + h \frac{Z_p^2}{2}$$

La estimación del parámetro h debe considerar las tres estimaciones.

5. Cuando, $g \rightarrow 0$, h se estima teniendo en cuenta la siguiente regresión lineal:

$$\ln \left(\frac{X_p - X_{1-p}}{2Z_p} \right) = \ln(B) + h \frac{Z_p^2}{2} \quad (13)$$

6. En la práctica, el comportamiento de los parámetros g y h puede tener patrones más complejos que los estipulados anteriormente; por ello, una forma de estimarlos es tratarlos como polinomios de Z_p^2 ,

$$g(Z) = k_0 + k_1 Z^2 + k_2 Z^4 + k_3 Z^6 + \dots \quad (14)$$

$$h(Z) = a_0 + a_1 Z^2 + a_2 Z^4 + a_3 Z^6 + \dots \quad (15)$$

La característica de estos patrones y el grado del polinomio, pueden ser determinados graficando g_p y h_p contra Z_p^2 , respectivamente.

Modificación del método de cuantiles de Hoaglin (1985)

En esta sección se presenta una nueva propuesta para estimar los parámetros de la distribución g y h de Tukey, partiendo del método propuesto por Hoaglin (1985) y Jiménez M. (2004).

La estimación de los parámetros de la distribución g y h de Tukey, se lleva a cabo teniendo en cuenta que el sesgo puede inducir a patrones de alargamiento de las colas en los datos. Por tanto es importante *investigar la naturaleza del alargamiento de las colas luego de que el sesgo haya sido considerado*. La nueva estimación es análoga a la propuesta en Hoaglin (1985), pero asumiendo que $p > 0, 5$ y considerando:

1. Cuando $g \neq 0$, el parámetro que controla el alargamiento de las colas (h), se estima mediante

$$\ln(X_{0.5} - \theta_p) = \ln\left(\frac{B}{g}\right) + h\frac{Z_p^2}{2}, \tag{16}$$

Con $\theta_p < X_{0.5}$, para todo $p > 0, 5$ y

$$\theta_p = \frac{X_p X_{1-p} - X_{0.5}^2}{(X_p - X_{0.5}) - (X_{0.5} - X_{1-p})} \quad \text{para } p \in (0, 1), p \neq 0, 5, \tag{17}$$

nótese que $\theta_p = \theta_{1-p}$. El valor de h se obtiene haciendo uso de una regresión lineal; donde el coeficiente de la regresión estima el valor de h y el valor de B se cacula multiplicando el valor de g por la función exponencial del intercepto de la regresión. En Jiménez M. y Martínez C. (2006) se propone que si $\theta_p = 0$ entonces h tiende a cero (*distribuciones g de Tukey*).

2. Cuando, $g \rightarrow 0$ (*distribuciones h de Tukey*), h se estima teniendo en cuenta la siguiente regresión lineal:

$$\ln\left(\frac{X_p - X_{0.5}}{Z_p}\right) = \ln(B) + \frac{hZ_p^2}{2} \tag{18}$$

3. Un caso particular de las distribuciones g y h de Tukey, son las *distribuciones g de Tukey*, donde $g \neq 0$ y $h = 0$. Para esta familia de distribuciones, el parámetro g se calcula como se indicó anteriormente y la estimación del parámetro de escala se obtiene al promediar los percentiles de la variable X dados por las ecuaciones 11 y 12:

$$\frac{1}{2}(X_p + X_{1-p}) = A + \frac{B}{2g}(\exp(gZ_p) + \exp(-gZ_p) - 2) \tag{19}$$

Ahora, expresando $\exp(gZ_p)$ y $\exp(-gZ_p)$, en series de potencias:

$$\exp(gZ_p) + \exp(-gZ_p) - 2 = 2 \sum_{k=1}^{\infty} \frac{(gZ_p)^{2k}}{(2k)!}$$

y se expresa $2k!$ en términos de la función Gamma obteniéndose que $(2k)! = \Gamma(2k + 1) = (2k)\Gamma(2k)$, por tanto la expresión (19) queda determinada por:

$$\frac{1}{2}(X_p + X_{1-p}) = A + \frac{B}{2g} \sum_{k=1}^{\infty} \frac{(gZ_p)^{2k}}{(k\Gamma(2k))} \tag{20}$$

Para valores de $|g| \leq 0,5$, la ecuación (20) toma la siguiente forma

$$\frac{1}{2}(X_p + X_{1-p}) = A + \frac{1}{2}BgZ_p^2 \quad (21)$$

Para la implementación práctica se hace uso de las siguientes relaciones:

$$\exp(gZ_p) = \frac{X_p - X_{0.5}}{X_{0.5} - X_{1-p}} \quad \text{y} \quad \exp(-gZ_p) = \frac{X_{0.5} - X_{1-p}}{X_p - X_{0.5}}$$

B se estima utilizando una regresión lineal sin intercepto, con base en la siguiente ecuación:

$$\frac{1}{2}(X_p + X_{1-p}) = A + \frac{B}{2g} \left[\frac{X_p - X_{0.5}}{X_{0.5} - X_{1-p}} + \frac{X_{0.5} - X_{1-p}}{X_p - X_{0.5}} - 2 \right] \quad (22)$$

4. Adicionalmente, el comportamiento de los parámetros g y h son estimados mediante las expresiones dadas en (5).

¿Por qué estimar los parámetros de la distribución g y h de Tukey utilizando el método de cuantiles de Hoaglin?

En la estimación de los parámetros de la distribución g y h (con parámetros constantes), métodos alternativos pueden ser usados: máxima verosimilitud o de momentos (Majumder y Ali, 2008). Estos métodos limitan el uso de la riqueza estructural de la distribución g y h de Tukey y aunque el método de máxima verosimilitud genera estimaciones con menor nivel de error, el costo computacional del cálculo de los parámetros de la g y h es muy alto.

El método basado en cuantiles usa la estructura de la distribución g y h de Tukey, utilizando el comportamiento de las colas de la distribución empírica de los datos y no la información del centro de esta distribución como lo realiza el método de Momentos o Máxima Verosimilitud (Hoaglin, 1985).

La metodología empleada en este artículo, para estimar los parámetros no constantes (polinomios de Z) de la distribución g y h que se ajusten a los datos, es la propuesta por Hoaglin (1985) –método basado en cuantiles. Varios autores argumentan que un método de estimación de parámetros mediante máxima verosimilitud asigna pesos iguales a todos los datos usados en el ajuste, mientras que el método basado en cuantiles puede poner más peso en las colas de la distribución. No obstante, las diferencias entre los

métodos de estimación dependen del tamaño de muestra, en el caso particular de las series financieras, es suficientemente grande y por lo tanto, la pérdida de eficiencia es mínima (He y Raghunathan, 2006).

APLICACIÓN

Muestra

En este apartado, se utiliza la distribución g y h de Tukey como herramienta para describir la distribución de las tasas de retorno de las acciones de la compañía **Boeing**⁴. Se tomaron los precios diarios de las acciones ajustados por los dividendos, entre enero 3 de 2000 y noviembre 7 de 2005, para un total de 1.471 precios históricos. El precio del activo está dado en dólares para los días bursátiles del año.

Histograma de frecuencias y coeficientes de asimetría y curtosis

En el Cuadro 1 se presentan las estadísticas descriptivas de los datos:

CUADRO 1
ESTADÍSTICAS DESCRIPTIVAS

N	1.470
Media	0,00039789
Desv. Estándar	0,02160789
Sesgo	-0,6295872
Curtosis	5,99
Rango	0,27970
Rango Intercuartílico	0,02497

Fuente: elaboración propia.

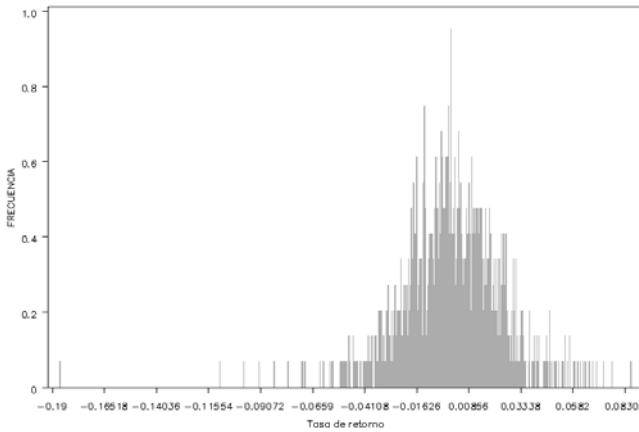
A partir de los gráficos de secuencia (comportamiento de la variable a través del tiempo), presentados para la tasa de rendimiento y la tasa de retorno, se evidencia que tanto el comportamiento como los valores que toma cada una de las variables son los mismos. Por tanto, se analizará la tasa de retorno ya que es la variable que utiliza Fama (1965) en sus estudios y con base en su comportamiento se estima el retorno promedio y la volatilidad de un activo. La característica de leptocurtosis se nota claramente en la Gráfica 1.

El histograma de frecuencias (Gráfica 1) y los coeficientes de asimetría y curtosis indican que la distribución de la variable de estudio tiene un leve

⁴Los datos históricos de este activo fueron tomados de <http://finance.yahoo.com>.

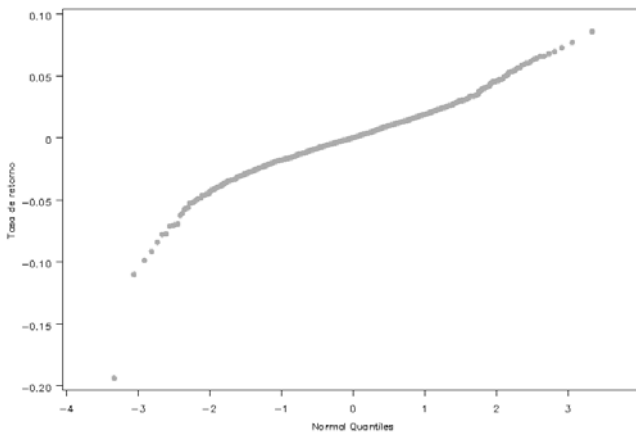
grado de sesgo a la izquierda y es leptocúrtica. En este gráfico se observa un comportamiento irregular en las colas de la distribución, lo cual Fama denominó colas pesadas o “Heavy tails”. Adicionalmente, las pruebas de normalidad indican que la variable de estudio no sigue una distribución normal, con un nivel de significancia de 1 %.

GRÁFICA 1
HISTOGRAMA DE FRECUENCIAS



Fuente: elaboración propia.

GRÁFICA 2
GRÁFICO DE CUANTILES, Q-Q PLOT

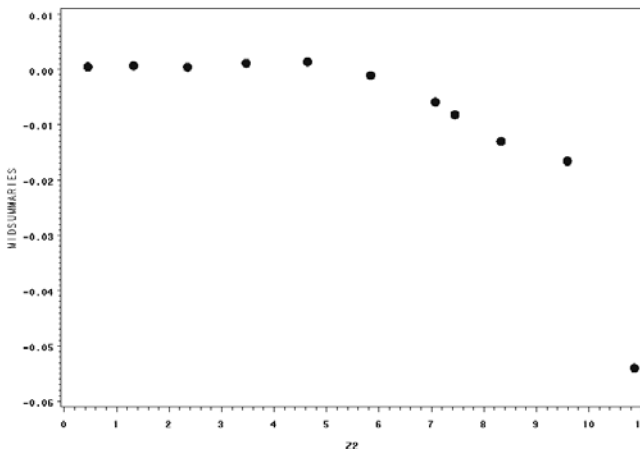


Fuente: elaboración propia.

Descripción del comportamiento del sesgo y alargamiento de las colas

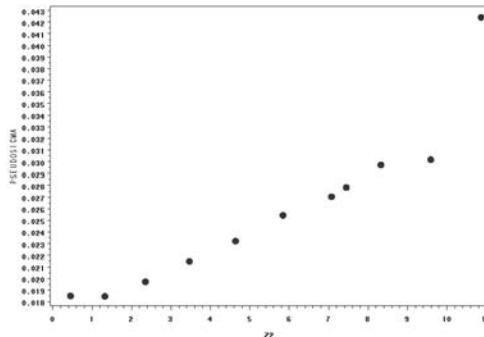
A continuación se realiza un análisis descriptivo del comportamiento de la forma de la distribución de la tasa de retorno, utilizando medidas de sesgo y alargamiento de las colas descritas en Hoaglin, Mosteller y Tukey (1985), las cuales permiten observar y evaluar el grado de sesgo y alargamiento de las colas de una manera más detallada y no tan global como los coeficientes de asimetría y curtosis (gráficas 3 y 4).

GRÁFICA 3
COMPORTAMIENTO DEL SESGO



Fuente: elaboración propia.

GRÁFICA 4
ALARGAMIENTO DE LAS COLAS



Fuente: elaboración propia.

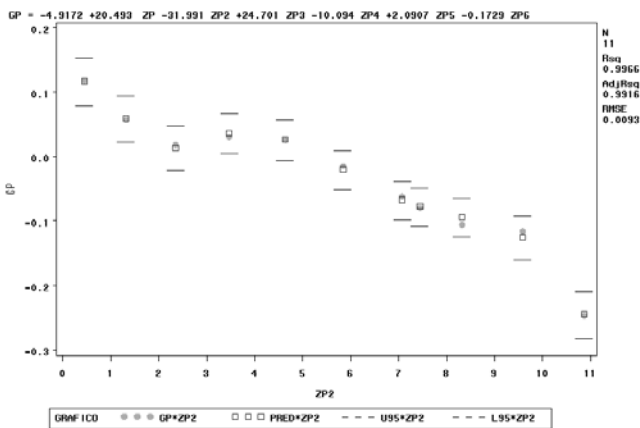
En la Gráfica 3 se observa que la distribución es sesgada a izquierda al igual que lo indicaba el coeficiente de asimetría. En las gráficas 3 y 4 se evidencia un comportamiento constante alrededor de cero en el centro de la distribución, lo que indica que esta parte de la distribución tiene un comportamiento simétrico alrededor de la mediana. La asimetría de la distribución se agudiza en las colas de la distribución, notando que para el periodo de estudio se encontraron tasas de rendimiento con pérdidas inferiores a -10% , en tanto que la mayor tasa de ganancia estaba alrededor de 8% .

Como en las gráficas 3 y 4 no se observa un patrón constante, se evidencia que *la forma de la distribución* varía a través de los percentiles. Por tanto, se utilizará un análisis de regresión para determinar el comportamiento del sesgo y el alargamiento de las colas, utilizando como variables explicativas un polinomio construido con los valores absolutos de los percentiles de la distribución normal estándar ($|Z_p|$).

Comportamiento del sesgo de la distribución

A partir del análisis descriptivo del sesgo y al observar que las estimaciones de $g(z_p)$ varían a través de los percentiles, se estima el grado de sesgo considerando que $g(z_p)$ es explicado por un polinomio de $|Z_p|$. Para ello se hace uso de un análisis de regresión (Gráfica 5).

GRÁFICA 5
ANÁLISIS DE REGRESIÓN PARA SESGO DE LA DISTRIBUCIÓN



Fuente: elaboración propia.

El ajuste del modelo es bastante bueno. El coeficiente de determinación es cercano a 1, la estadística F indica una significancia global del modelo

y los valores predichos para el grado de sesgo son bastante cercanos a los niveles de sesgo observados para los datos. En este análisis de regresión todos los parámetros son significativos al 1 %.

Un gran avance en este artículo es la descripción del comportamiento del sesgo en términos de un polinomio de los percentiles de la distribución normal estándar Z_p ; debido a que en la literatura sólo se evalúan polinomios de Z_p^2 . En el caso aquí analizado, se evidenciaron mejores resultados al considerar la relación entre el nivel de sesgo con Z_p que con Z_p^2 .

Comportamiento del alargamiento de las colas de la distribución

Con base en el comportamiento no constante del alargamiento de las colas en el análisis descriptivo y bajo la estimación de $g(z_p)$ en el análisis de regresión para el sesgo, se implementó un análisis de regresión tomando como variables respuesta las descritas en los métodos de estimación, y las variables explicativas correspondientes a un polinomio de Z_p^2 . Este análisis se realizó de igual manera para las dos colas de la distribución, teniendo en cuenta otra variable respuesta para cada cola.

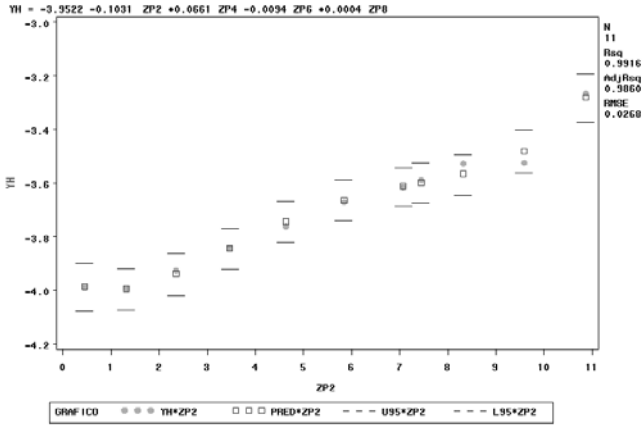
En el análisis de regresión para el alargamiento de las colas sin desagregar el análisis para cada cola, se evidenció un ajuste acertado (Gráfica 6). La tabla de análisis de varianza presenta una significancia global aceptable. El coeficiente de determinación es cercano a la unidad, indicando que la explicación del modelo es próxima a 100 %. La mayoría de los parámetros son significativos a un nivel de significancia de 1 %, excepto el coeficiente asociado a Z_p^2 el cual es significativo al 11 %. La exclusión de esta última variable implicaba predicciones menos acertadas que si era incluida.

Al observar el comportamiento del alargamiento de las colas individualmente (gráficas 7a y 7b), se aprecia que los modelos de regresión son similares en significancia, tanto global como marginal, y los coeficientes estimados son muy cercanos al análisis de regresión en el cual se contempla toda la distribución.

Se evidencia que el comportamiento del alargamiento de las colas no difiere significativamente de una cola a otra. Además, analizar este comportamiento para cada cola de la distribución presenta problemas de sobreestimación, tal como lo indican algunos autores entre ellos Dutta y Babbel (2002), Mills (1995), y Badrinath y Chatterjee (1988). El parámetro de localización es estimado por la mediana de los registros históricos, esta medida tuvo el valor de 0. El parámetro de escala es estimado por la exponencial del in-

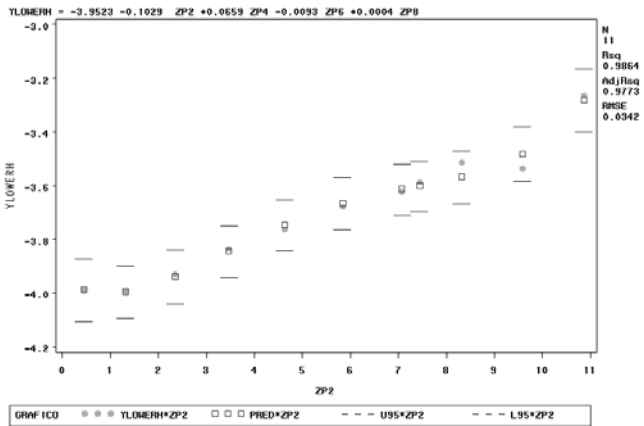
tercepto de la regresión aplicada, para determinar el comportamiento del alargamiento de las colas de la distribución, este valor fue 0,019212.

GRÁFICA 6
ANÁLISIS DE REGRESIÓN PARA ALARGAMIENTO DE LAS COLAS DE LA DISTRIBUCIÓN



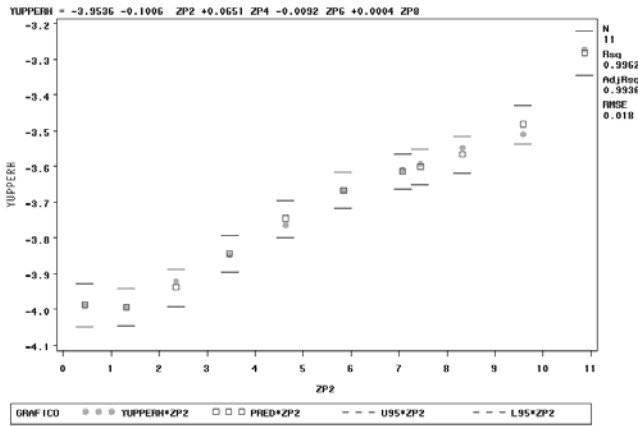
Fuente: elaboración propia.

GRÁFICA 7a
ALARGAMIENTO DE LA COLA INFERIOR



Fuente: elaboración propia.

GRÁFICA 7b
ALARGAMIENTO DE LA COLA SUPERIOR



Fuente: elaboración propia.

Percentiles estimados

El análisis de la distribución descrito en este artículo utiliza percentiles de la forma $p = 2^{-k}$ con $k = 1, 2, 3, \dots$, teniendo en cuenta una progresión geométrica que va desde el percentil 0,5 hasta el 0,999511719 para la cola superior, y de manera simétrica para cola inferior. En el Cuadro 2 se muestra la estimación de estos percentiles.

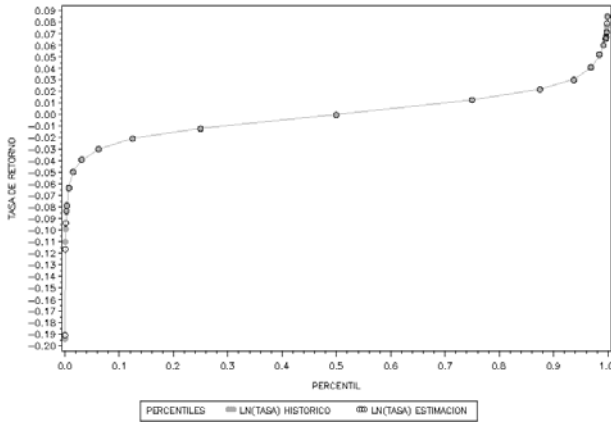
CUADRO 2
ESTIMACIÓN DE CUANTILES

p	X_p	X_{1-p}
0,50	0,0000000	0,0000000
0,75	0,0129810	-0,0119921
0,90	0,0246123	-0,0236596
0,95	0,0335290	-0,0332374
0,99	0,0567053	-0,0566497
1,00	0,0858794	-0,1938202

Fuente: elaboración propia.

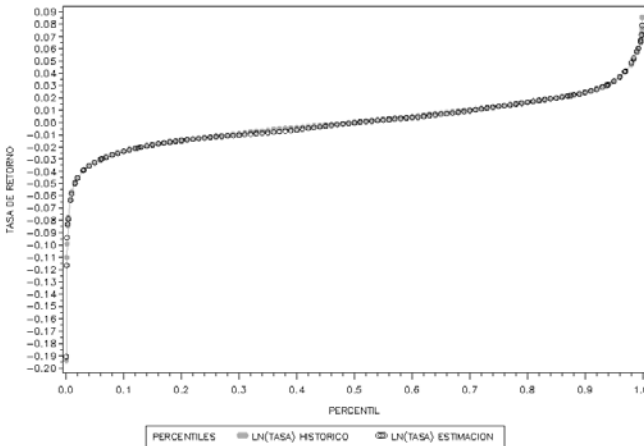
Para estos percentiles se analizó la bondad de ajuste de la distribución g y h de Tukey para la tasa de retorno. Los gráficos 8 y 9 muestran que la estimación de estos percentiles comparada con los datos históricos del activo para el periodo de estudio, es bastante acertada.

GRÁFICA 8
 PERCENTILES ESTIMADOS DE LA FORMA 2^{-k}



Fuente: elaboración propia.

GRÁFICA 9
 PERCENTILES ESTIMADOS EN PROGRESIÓN ARITMÉTICA Y GEOMÉTRICA

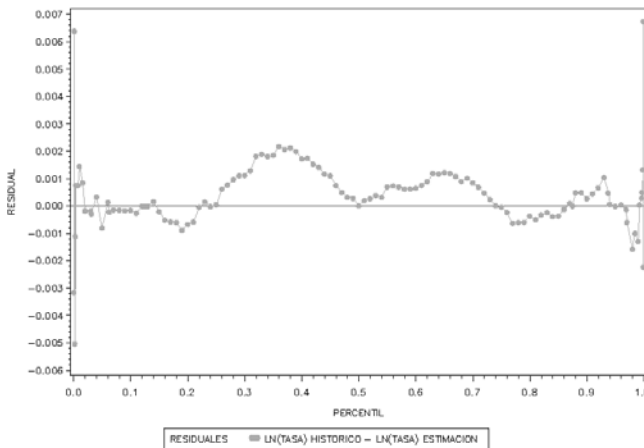


Fuente: elaboración propia.

En la bondad de ajuste del análisis de la distribución también se estudiaron los residuales estimados. La Gráfica 10 muestra el comportamiento de los residuales del modelo estimado (asumiendo g y h como polinomios de $|Z_p|$), evidenciando que los residuales de los percentiles del centro de la distribución están entre $-0,001$ y $0,002$, en tanto que los de las colas de la distribución oscilan entre $-0,005$ y $0,0065$. Es claro que las diferencias

entre los valores históricos y estimados son pequeñas, a lo más se tiene una diferencia cercana a 0,5 unidades porcentuales con respecto al valor histórico registrado para la tasa de retorno.

GRÁFICA 10
RESIDUALES ESTIMADOS



Fuente: elaboración propia.

Función de densidad de la tasa de retorno

Una vez determinado el comportamiento del sesgo y alargamiento de las colas de la distribución, y los parámetros de localización (mediana) y escala, se puede describir la distribución de la variable de estudio. Lo anterior fue utilizado en la bondad de ajuste de la distribución g y h de Tukey estimada y estableciéndose que ésta describe adecuadamente los registros históricos del activo. La Gráfica 11 corresponde a la distribución que representa la función de densidad de la distribución g y h de Tukey para la variable *tasa de retorno*.

Se evidenció que los parámetros que determinan la forma de la distribución de la tasa de retorno, no son constantes y varían a través de los percentiles. Por tanto, se debe desarrollar la función de densidad para la distribución g y h de Tukey en presencia de esta situación.

La estimación de los parámetros y la variable Y_p para la tasa de retorno de la compañía Boeing, está dada por:

$$g(z_p) = -4,9172 + 20,493|Z_p| - 31,991Z_p^2 + 24,701|Z_p^3| - 10,094Z_p^4 + 2,0907|Z_p^5| - 0,1729Z_p^6 \quad (23)$$

$$h(z_p) = -0,2062 + 0,1322Z_p^2 - 0,0188Z_p^4 + 0,0008Z_p^6 \quad (24)$$

Aquí se usa el hecho que $|Z_p^{2n}| = Z_p^{2n}$, teniendo en cuenta que $g(z_p)$ y $h(z_p)$, notados en las ecuaciones (23) y (24), representan el parámetro del sesgo y alargamiento de las colas como función de los percentiles de la distribución normal estándar.

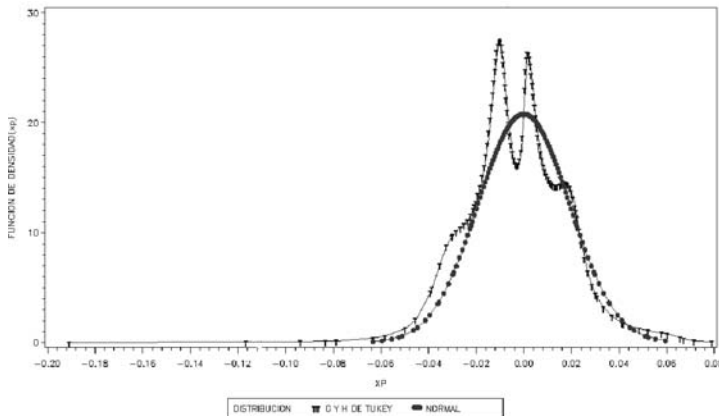
Luego la expresión (6) queda de la siguiente manera:

$$X_p = 0 + 0.019212 Y_p \quad (25)$$

Siendo Y_p dada por la expresión (4). Como se nota en la Gráfica 11, la función de densidad tiene un comportamiento similar al histograma de los datos históricos de la variable de estudio presentado en la Gráfica 1.

GRÁFICA 11

FUNCIÓN DE DENSIDAD CON PARÁMETROS NO CONSTANTES



Fuente: elaboración propia.

CONCLUSIONES

En este artículo se mostró que la distribución g y h de Tukey es una herramienta útil en la descripción del comportamiento de las tasas de retorno de un activo.

En la aplicación se aprecia que la tasa de retorno *no sigue la distribución normal* y por tanto, su distribución no dependen únicamente de la volatilidad y del valor promedio de la tasa de retorno en un periodo, sino también de la forma de la distribución: sesgo y alargamiento de las colas.

Adicionalmente, se encontró que el comportamiento del sesgo y alargamiento de las colas de la distribución de la tasa de retorno de la compañía **Boeing**, varía a través de los percentiles, lo cual indica que la volatilidad de esta tasa *no es constante*.

Los desarrollos teóricos realizados en este artículo, permiten hacer uso de la familia de distribuciones g y h de Tukey en un amplio campo de acción, tanto en el cálculo de probabilidades a partir de la función de distribución como en la descripción gráfica y analítica de la función de densidad, aún cuando los parámetros de sesgo y alargamiento de las colas *no son constantes*.

La estimación de parámetros constantes de la distribución g y h de Tukey puede realizarse por distintos métodos, pero para utilizarlos es necesario considerar los supuestos y restricciones de los modelos econométricos. Para el caso en que los parámetros sean no constantes la modelación puede realizarse mediante la metodología de Hoaglin o la propuesta en este artículo dependiendo del comportamiento de la tasa de retorno.

REFERENCIAS BIBLIOGRÁFICAS

- [1] Badrinath, S. y Chatterjee, S. (1988). On measuring skewness and elongation in common stock return distributions: The case of the market index. *The Journal of Business*, 61(4), 451-472.
- [2] Badrinath, S. y Chatterjee, S. (1991). A data analytic look at skewness and elongation in common stock return distributions. *Journal of Business & Economic Statistics*, 9(2), 223-233.
- [3] Dutta, K. K. y Babbel, D. F. (2002). On measuring skewness and kurtosis in short rate distributions: The case of the us dollar london inter bank offer rates. *Center for Financial Institutions Working Papers 02-25*. Philadelphia: University of Pennsylvania.
- [4] Dutta, K. K. y Babbel, D. F. (2005). Extracting probabilistic information from the prices of interest rate options: Tests of distributional assumptions. *The Journal of Business*, 78(3), 841-870.
- [5] Dutta, K. K. y Perry, J. (2007). A tale of tails: An empirical analysis of loss distribution models for estimating operational risk capital. *Working Paper No. 06-13*. Boston: Federal Reserve Bank of Boston.
- [6] Fama, E. F. (1965). The behavior of stock market prices. *The behavior of stock market prices*, 38, 34-105.

- [7] He, Y. y Raghunathan, T. E. (2006). Tukeys gh distribution for multiple imputation. *The American Statistician*, 60(3), 251-256.
- [8] Hoaglin, D.C. (1985). Summarizing shape numerically: the g-and-h distributions. En: Hoaglin, D.C., Mosteller, F., Tukey, J.W. (Eds.), *Exploring Data Tables, Trends, and Shapes* (461-513). New York: John Wiley & Sons.
- [9] Hoaglin, D. C., Mosteller, F. y Tukey, J. W. (1985). *Exploring Data Tables, Trends, and Shapes*. New York: John Wiley & Sons.
- [10] Jiménez M., J. A. (2004). *Aproximaciones de las funciones de riesgo del tiempo de sobrevivencia mediante la distribución g – h de Tukey*. Trabajo de grado no publicado (Especialización en Actuaría). Universidad Nacional de Colombia, Facultad de Ciencias, Departamento de Matemáticas.
- [11] Jiménez M., J. A. y Martínez C., J. (2006). Una estimación del parámetro de la distribución g de Tukey. *Revista Colombiana de Estadística*, 29(1), 1-16.
- [12] Majumder, M. M. A. y Ali, M. M. (2008). A comparison of methods of estimation of parameters of tukeys gh family of distributions. *Pakistan Journal of Statistics*, 24(2), 135-144.
- [13] Mills, T.C. (1995). Modelling skewness and kurtosis in the London stock exchange ft-se index return distributions. *The Statistician*, 44(3), 323-332.
- [14] Tukey, J.W. (1977). Modern techniques in data analysis, NSF-sponsored regional research conference at Southeastern Massachusetts University. North Dartmouth, MA.

ANEXO 1.

RESULTADOS DE LAS SIMULACIONES

Tablas de análisis de varianza y efectos marginales

CUADRO A1

ANAVA COMPORTAMIENTO DEL SESGO

Análisis de Varianza					
Causa	Gl	Suma de Cuadr.	Cuadrado Medio	Valor F	Pr > F
Modelo	6	0,10193	0,01699	197,28	<,0001
Residual	4	0,00034444	0,00008611		
Total	10	0,10227			

Fuente: elaboración propia.

CUADRO A2
EFECTOS MARGINALES

Parámetros Estimados					
Variable	Gl	Parámetro Estimado	Error Estd	Valor t	Pr > t
Intercepto	1	-4,91723	1,12035	-4,39	0,0118
ZP	1	20,49335	4,35227	4,71	0,0092
ZP2	1	-31,99104	6,50704	-4,92	0,0080
ZP3	1	24,70092	4,86378	5,08	0,0071
ZP4	1	-10,09423	1,93815	-5,21	0,0065
ZP5	1	2,09068	0,39370	5,31	0,0060
ZP6	1	-0,17295	0,03206	-5,39	0,0057

Fuente: elaboración propia.

Estimación del alargamiento de las colas de la distribución

CUADRO A3
ANAVA COMPORTAMIENTO DEL ALARGAMIENTO DE LAS COLAS

Análisis de Varianza					
Causa	Gl	Suma de Cuadr.	Cuadrado Medio	Valor F	Pr > F
Modelo	4	0,50875	0,12719	177,49	<,0001
Residual	6	0,00430	0,00071660		
Total	10	0,51305			

Fuente: elaboración propia.

CUADRO A4
EFECTOS MARGINALES

Parámetros Estimados					
Variable	Gl	Parámetro Estimado	Error Estd	Valor t	Pr > t
Intercepto	1	-3,95219	0,04245	-93,09	<,0001
ZP2	1	-0,10309	0,05464	-1,89	0,1081
ZP4	1	0,06606	0,01991	3,32	0,0160
ZP6	1	-0,00937	0,00266	-3,52	0,0126
ZP8	1	0,00043103	0,00011715	3,68	0,0103

Fuente: elaboración propia.

Estimación del alargamiento de la cola inferior de la distribución

CUADRO A5

ANAVA COMPORTAMIENTO DEL ALARGAMIENTO DE LA COLA INFERIOR

Análisis de Varianza					
Causa	Gl	Suma de Cuadr.	Cuadrado Medio	Valor F	Pr > F
Modelo	4	0,50818	0,12705	108,50	<,0001
Residual	6	0,00703	0,00117		
Total	10	0,51521			

Fuente: elaboración propia.

CUADRO A6

EFFECTOS MARGINALES

Parámetros Estimados					
Variable	Gl	Parámetro Estimado	Error Estd	Valor t	Pr > t
Intercepto	1	-3,95228	0,05427	-72,83	<,0001
ZP2	1	-0,10286	0,06985	-1,47	0,1913
ZP4	1	0,06595	0,02545	2,59	0,0411
ZP6	1	-0,00935	0,00341	-2,75	0,0335
ZP8	1	0,00043022	0,00014975	2,87	0,0283

Fuente: elaboración propia.

Estimación del alargamiento de la cola superior de la distribución

CUADRO A7

ANAVA COMPORTAMIENTO DEL ALARGAMIENTO DE LA COLA SUPERIOR

Análisis de Varianza					
Causa	Gl	Suma de Cuadr.	Cuadrado Medio	Valor F	Pr > F
Modelo	4	0,50794	0,12698	391,72	<,0001
Residual	6	0,00195	0,00032417		
Total	10	0,50988			

Fuente: elaboración propia.

CUADRO A8
EFECTOS MARGINALES

Parametros Estimados					
Variable	Gl	Parámetro Estimado	Error Estd	Valor t	Pr > t
Intercepto	1	-3,95358	0,02855	-138,46	<,0001
ZP2	1	-0,10061	0,03675	-2,74	0,0338
ZP4	1	0,06505	0,01339	4,86	0,0028
ZP6	1	-0,00923	0,00179	-5,15	0,0021
ZP8	1	0,00042472	0,00007880	5,39	0,0017

Fuente: elaboración propia.