

# **Estudio y Aplicación de la Distribución Log-SNP a Variables Económicas y Financieras**

Lina M. Cortés

Estudiante Doctorado en Economía, Departamento de Economía e Historia Económica e IME, Universidad de Salamanca, Salamanca, España.

Profesora, Departamento de Finanzas, Escuela de Economía y Finanzas, Universidad EAFIT, Medellín, Colombia. E-mail: [lcortesd@eafit.edu.co](mailto:lcortesd@eafit.edu.co)

Javier Perote

Director de Tesis

Profesor, Departamento de Economía e Historia Económica e IME, Universidad de Salamanca, Salamanca, España. E-mail: [perote@usal.es](mailto:perote@usal.es)

Andrés Mora-Valencia

Codirector de Tesis

Profesor, Facultad de Administración, Universidad de los Andes, Bogotá, Colombia. E-mail: [a.mora262@uniandes.edu.co](mailto:a.mora262@uniandes.edu.co)

## **Resumen**

En esta propuesta consideramos un enfoque semi-noparamétrico, en donde la función de distribución de probabilidad desconocida se modela a partir de una expansión de series de polinomios ortogonales. Analizamos las expansiones de Gram-Charlier y en particular, proponemos el uso de una distribución log semi-noparamétrica (log-SNP) que anida a la lognormal. Se busca mostrar que la distribución log-SNP permite mejoras de ajuste significativas al modelar variables económicas y financieras. Esto se debe a que la distribución es muy flexible para representar diferentes formas, a partir de la incorporación de parámetros adicionales a los de una distribución paramétrica tradicional como la lognormal. Adicionalmente, se expone una aplicación al modelar la productividad de investigadores top en diferentes campos de conocimiento. Los resultados muestran claramente la mayor precisión de la distribución log-SNP en los ajustes en las colas y los errores inducidos en la estimación de la productividad de los investigadores top mediante el uso de distribuciones paramétricas tradicionales como la lognormal.

## **I. Introducción**

La caracterización de una variable aleatoria mediante su función de densidad (pdf) y su ajuste a la distribución empírica de una serie puede realizarse mediante distintos enfoques que van desde una perspectiva paramétrica basada en una distribución de frecuencias con forma funcional conocida a un

enfoque no paramétrico puro. Una posibilidad intermedia es el uso de aproximaciones semi-noparamétricas (SNP) en el que la forma funcional se parametriza sólo en parte y siendo el resto una función desconocida (Chen 2007).

Investigaciones que han incorporado este tipo de aproximaciones son las de Mauleón y Perote (2000) y Verhoeven y McCaller (2004) quienes analizan el comportamiento de series financieras de alta frecuencia desde su asimetría y leptocurtosis. Por su parte, Corrado y Su (1996) y Rubinstein (1998) utilizan estas densidades en estudios de opciones financieras. Asimismo, Perote y Del Brío (2003) y Del Brío, Mora-Valencia y Perote (2014a, b) se basan en estas densidades para mejorar medias del Valor en Riesgo (VaR). Desde un enfoque más económico, Ñíguez, Paya, Peel y Perote (2012) muestran que bajo distribuciones SNP los modelos de equilibrio general son estables.

En este trabajo consideramos un enfoque SNP en donde la función desconocida se modela a partir de una expansión de series de polinomios ortogonales. Analizaremos expansiones de Gram-Charlier que se han mostrado como aproximaciones asintóticas válidas a cualquier distribución empírica bajo condiciones de regularidad relativamente débiles (Sargan 1975; Phillips 1977). En particular proponemos transformaciones logarítmicas de una distribución SNP (a las que nos referiremos como log-SNP) que son extensiones de una distribución lognormal que permiten aproximar cualquier distribución empírica mediante la introducción de parámetros adicionales.

## **II. Revisión de Literatura**

Algunas de las variables que se estudian en economía y finanzas se pueden ajustar a funciones de densidad con una forma de campana (p.e. la t-Student o la Gaussiana), sin embargo las colas que describen los datos pueden ser muy pesadas (Cont 2001). Estudios realizados por Jondeau y Rockinger (2006, 2009) proporcionan evidencia sobre la importancia que tiene representar correctamente las variables financieras, especialmente por la forma de las colas.

Por ello, la aproximación a distribuciones con colas pesadas ha sido un campo muy estudiado desde su aparición (Embrechts, Klüppelberg y Mikosch, 1997), y hoy en día se sigue intentando encontrar un método para aproximarlas de la manera más ajustada. Por ello, investigadores han buscado modelar variables económicas y financieras expandiendo funciones de distribución a partir de series de funciones ortogonales. La mayoría de los estudios han considerado distribuciones a partir de la expansión Gram-Charlier.

A continuación se define la distribución SNP basada en series de Gram-Charlier y su transformación logarítmica, analizando sus propiedades básicas.

**a. La distribución SNP**

Sea  $\{P_s(x)\}$ ,  $x \in \mathbb{R}$  y  $s \in \mathbb{N}$  una familia de polinomios ortogonales con respecto a una función de densidad  $w(x)$  que satisface la siguiente relación<sup>1</sup>

$$\int_{-\infty}^{\infty} P_s(x)P_j(x)w(x)dx = 0, \quad \forall s \neq j, \quad s, j = 0, 1, 2, \dots \quad (1)$$

Dentro de esta familia, los polinomios de Hermite (HP) son aquellos que utilizan como peso una función de densidad normal estándar  $\phi(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}$ . En particular el polinomio de Hermite de orden  $s$ ,  $H_s(x)$ , se puede obtener en términos de derivada de orden  $s$  de la función de densidad la distribución normal estándar tal como se expresa en la ecuación (2):

$$H_s(x) = \frac{(-1)^s}{\phi(x)} \frac{d^s \phi(x)}{dx^s} \quad (2)$$

A continuación se muestran los ocho primeros HP:

$$H_0(x) = 1 \quad (3)$$

$$H_1(x) = x \quad (4)$$

$$H_2(x) = x^2 - 1 \quad (5)$$

$$H_3(x) = x^3 - 3x \quad (6)$$

$$H_4(x) = x^4 - 6x^2 + 3 \quad (7)$$

$$H_5(x) = x^5 - 10x^3 + 15x \quad (8)$$

$$H_6(x) = x^6 - 15x^4 + 45x^2 - 15 \quad (9)$$

$$H_7(x) = x^7 - 21x^5 + 105x^3 - 105x \quad (10)$$

$$H_8(x) = x^8 - 28x^6 + 210x^4 - 420x^2 + 105 \quad (11)$$

Es sencillo comprobar que estos polinomios satisfacen la mencionada propiedad de ortogonalidad dado que  $\forall s, j = 0, 1, 2, \dots$

$$\int_{-\infty}^{\infty} H_s(x)H_j(x)\phi(x)dx = \begin{cases} 0, & s \neq j \\ s!, & s = j \end{cases} \quad (12)$$

Los HP constituyen además la base de las series de Edgeworth y Gram-Charlier (Type A) que permiten, bajo ciertas condiciones de regularidad (Cramér 1925), expresar cualquier pdf,  $f(x)$ , en términos de una una serie infinita (Wallace, 1958) de la siguiente forma<sup>2</sup>

$$f(x) = \sum_{s=0}^{\infty} \delta_s H_s(x)\phi(x), \quad \text{donde } \delta_s = \frac{1}{s!} \int_{-\infty}^{\infty} H_s(x)\phi(x)dx \quad (13)$$

<sup>1</sup> Diferentes funciones de pesos pueden ser usadas  $w(x)$ , para detalles ver Abramowitz y Stegun (1972, pp. 774-775). Por convenio consideraremos  $P_0(x) = 1$ .

<sup>2</sup> Para más detalles sobre las series de Edgeworth y Gram-Charlier véase Kendall y Stuart 1977 pp. 167-172.

Incluso, gracias a la ortogonalidad de los HP, la truncación de las series en un determinado orden  $n$  de la expansión permite definir una familia de distribuciones SNP,  $g(x; \mathbf{d})$ , donde  $\mathbf{d} = (d_1, \dots, d_n)' \in \mathbb{R}^n$  denota el vector de parámetros.<sup>3</sup>

$$g(x; \mathbf{d}) = [1 + \sum_{s=1}^n d_s H_s(x)] \phi(x) \xrightarrow{n \rightarrow \infty} f(x) \quad (14)$$

Sin embargo la distribución SNP definida en la ecuación (14) sólo es una función de densidad para el subconjunto de valores de  $\mathbf{d}$  que garanticen  $g(x; \mathbf{d}) \geq 0$ . Para solucionar este problema se han propuesto distintos tipos de restricciones o transformaciones de positividad (Gallant y Nychka 1987), si bien éstas suponen introducir una complejidad innecesaria para aplicaciones empíricas que implementan algoritmos de máxima verosimilitud (dado que en el óptimo éstos conducen a estimaciones que garantizan la positividad).

La gran ventaja de estas distribución SNP frente a otras especificaciones paramétricas radica precisamente en las mejoras en los ajustes a los que conducen dada su gran flexibilidad paramétrica. Además este mayor número de parámetros no conlleva una mayor dificultad ni en términos teóricos ni empíricos. También se puede obtener una fórmula cerrada para la función de distribución acumulada (cdf) de la distribución SNP en función de la cdf de la distribución normal, tal como se muestra en la ecuación (15). Esto permite de forma sencilla el cálculo de las probabilidades y los cuantiles de la distribución SNP.

$$\begin{aligned} G_x(a) &= \int_{-\infty}^a g(x; \mathbf{d}) dx \\ &= \int_{-\infty}^a \phi(x) dx - \phi(a) \sum_{s=1}^n d_s H_{s-1}(a) \end{aligned} \quad (15)$$

### ***b. La distribución log-SNP***

Ñíguez et al. (2012) definen una variable  $z > 0$  como log-SNP (estándar) si la variable  $x = \log(z)$  tienen como pdf la distribución SNP presentada en la ecuación (14). La distribución resultante hereda todas las buenas propiedades de la distribución SNP y, especialmente, su flexibilidad para recoger los valores extremos de la distribución. Nosotros daremos un paso más y definiremos una distribución log-SNP de forma similar, pero sobre una transformación lineal  $y = \sigma x + \mu$ .

**Definición:** Diremos que la variable  $z > 0$  se distribuye log-SNP con parámetros de centralidad  $\mu \in \mathbb{R}$ , escala  $\sigma^2 \in \mathbb{R}$  y forma  $\mathbf{d} = (d_1, \dots, d_n)' \in \mathbb{R}^n$  si su pdf puede expresarse como

---

<sup>3</sup>Hay que notar que dado un orden de truncación la distribución resultante es puramente paramétrica, pero con carácter general el orden de truncación puede variar para admitir una aproximación más precisa a una distribución dada. Sin pérdida de generalidad asumiremos  $d_0 = 1$ .

$$h(z; \mu, \sigma^2, \mathbf{d}) = \left[ 1 + \sum_{s=1}^n d_s H_s \left( \frac{\log(z) - \mu}{\sigma} \right) \right] \left( \frac{1}{z\sigma\sqrt{2\pi}} e^{-\frac{(\log(z) - \mu)^2}{2\sigma^2}} \right). \quad (16)$$

Definida de esta manera, la distribución lognormal es un caso particular de la log-SNP (para  $d_s = 0, \forall s$ ) lo que permite contrastar la mejora en los ajustes de ésta última frente a la lognormal mediante un contraste de restricciones lineales como el de razón de verosimilitudes.

### III. Discusión de la propuesta

En diferentes campos del conocimiento, los seres humanos se han interesado por analizar el comportamiento de diferentes fenómenos con el fin de comprenderlos y anticiparse al futuro. En ese sentido, una idea importante es que la ciencia es un medio por el cual se logra el aprendizaje, con una interacción entre la teoría y la práctica.

Sin embargo, en ocasiones para alcanzar esta interacción se hacen suposiciones tentativas acerca del mundo real, que se saben son falsas, pero que pueden ser de utilidad. Por ejemplo, el estadístico sabe que en la naturaleza nunca hubo una distribución normal, nunca hubo una línea recta, pero con supuestos normales y lineales a menudo se puede derivar resultados que coinciden, a una aproximación útil, con las que se encuentran en el mundo real (Box 1976). Dos campos de gran interés para el análisis de fenómenos son los de la economía y las finanzas. Por ejemplo, gran parte del trabajo en matemática financiera supone que los precios siguen un paseo aleatorio lognormal geométrico (o su análogo de tiempo continuo, el movimiento browniano geométrico). Así que una pregunta natural es si este supuesto es generalmente cierto. La respuesta rápida es no.

La presente propuesta busca mostrar que efectivamente la flexibilidad paramétrica de la log-SNP permite mejoras de ajuste significativas para ajustar variables económicas y financieras al ser muy flexible para representar diferentes formas (incluyendo saltos en la masa probabilística y colas pesadas) a partir de la incorporación de parámetros adicionales a los de una distribución paramétrica tradicional como la lognormal.

### IV. Metodología

Se analizarán algunas variables de los campos económico y financiero con el fin de extraer conclusiones de carácter general. Se recopilarán datos cuantitativos y se empleará el análisis estadístico como característica resaltante.

Adicionalmente, la mayoría de estudios que han aplicado el enfoque SNP a partir de la expansión Gram-Charlier, han tomado como función de peso la distribución normal estándar. En la presente investigación, todas las pruebas se realizarán tomando como función de peso la distribución lognormal.

Un factor decisivo a la hora de realizar las estimaciones sobre lognormales es el importante uso que tienen en aplicaciones de interés práctico en la economía y las finanzas.

## **V. Hallazgos**

A continuación se presenta una aplicación del enfoque SNP. Proponemos el uso de la log-SNP que anida a la lognormal y permite mejorar los ajustes en la cola mediante la introducción de nuevos parámetros.

### ***a. Aplicación. Productividad investigadora: Un enfoque semi-noparamétrico***

En los últimos años la valoración de la productividad de los investigadores académicos en los diferentes campos del conocimiento, ha estado relacionada con la medición del impacto de los resultados de la producción científica (Abramo et al. 2008; Sabharwal 2013; Campanario 2015). La motivación de estudiar la productividad radica en el deseo de promover la excelencia académica, y hacer de la investigación dentro de cada país más competitiva a nivel mundial (Frandsen 2005; Kocher et al. 2006; Abramo y D'Angelo 2014).

Una manera de valorar la actividad científica es, en términos de productividad, a partir de análisis bibliométrico. Este método consiste, principalmente, en cuantificar el número de documentos publicados por un país, institución, grupo de investigación o individuo, así como las citas recibidas por dichos documentos (Broadus 1987; Borokhovich et al. 1995; Abramo et al. 2008; Heberger et al. 2010; Finardi 2013; Bertocchi et al. 2015). Las medidas bibliométricas más comunes son las basadas en las publicaciones y en las citas y esta información proviene de diferentes bases de datos como Web of Science (WoS), Scopus, Google Scholar, entre otras.

Los modelos de distribución de probabilidad que han sido más aplicados en la literatura sobre productividad investigadora son los que atienden a las siguientes leyes: la ley de Lotka (Lotka 1926; Nicholls 1986; Chung y Cox 1990; Kretschmer y Kretschmer 2007), ley potencial (Price 1976; Egghe 2005; Albarrán et al. 2011; Aguinis et al., 2015) y la ley de Bradford (Garfield 1980; Rousseau 1994; Nicolaisen y Hjørland 2007; Campanario 2015). Estas leyes están basadas principalmente en funciones de distribución como la exponencial o la de Pareto. Sin embargo, investigaciones como las de Kumar et al. 1998, Radicchi et al. 2008, Perc 2010, Eom y Fortunato 2011 y Birkmaier y Wohlrabe 2014 han propuesto la aplicación de la distribución lognormal para estudiar la actividad investigadora.

En esta aplicación planteamos por primera vez el uso del enfoque log-SNP para medir la productividad investigadora y determinar con mayor precisión los cuantiles que delimitan a los investigadores más productivos en cada área de conocimiento como medición de la dificultad que conlleva

ser un investigador estrella en cada ámbito. Dado que los indicadores bibliométricos, usualmente, exhiben colas relativamente largas y con multimodalidad (Guerrero-Bote et al. 2007; Lancho-Barrantes 2010; Sabharwal 2013), nosotros mostramos que, en contraste a la distribución lognormal, la distribución log-SNP provee un mejor ajuste a la hora de caracterizar el desempeño investigador.

### ***b. Datos de aplicación***

Para probar si una distribución lognormal o una log-SNP ajusta mejor a la distribución de desempeño de 140,971 investigadores que han producido 253,634 publicaciones en 18 campos de conocimiento, hemos utilizado los datos de O’Boyle y Aguinis (2012).<sup>4</sup> Estos autores clasificaron los campos de conocimiento a partir del Journal Citation Reports (JCR), que proporciona los factores de impacto (IF) en distintas áreas de conocimiento clasificadas dentro de las categorías de “ciencias” y “ciencias sociales”. El periodo de tiempo tomado para la muestra fue de enero de 2000 a junio de 2009.<sup>5</sup> De esta manera, como medida de productividad investigadora, se tomó el número de artículos publicados por un autor en cada uno de los campos de conocimiento durante el período de observación de 9.5 años.

### ***c. Metodología de aplicación***

Sea  $z_i$  el número de artículos publicados por un autor en uno de los campos de conocimiento seleccionados, la función de log-verosimilitud para una observación distribuida log-SNP( $\mu, \sigma^2, \mathbf{d}$ ) truncada en el octavo momento viene dada por:

$$\log L(\mu, \sigma^2, \mathbf{d} | z_i) = -\frac{1}{2} \log(2\pi\sigma^2 z_i^2) - \frac{1}{2} \left( \frac{\log(z_i) - \mu}{\sigma} \right)^2 + \log \left[ 1 + \sum_{s=1}^8 d_s H_s \left( \frac{\log(z_i) - \mu}{\sigma} \right) \right] \quad (17)$$

La estimación secuencial comienza con la más simple densidad anidada, la lognormal, y recursivamente se agregan los parámetros  $d_s$  cuyos valores iniciales se seleccionan consistentemente con los momentos muestrales. La inclusión de nuevos parámetros en la distribución de productividad se realiza en función de criterios de precisión de los ajustes de la log-verosimilitud (logL) y el Criterio de Información de Akaike (AIC) y los contrastes de restricciones lineales proporcionados por la razón de verosimilitud (LR). Atendiendo a estos criterios se seleccionó  $n=8$  como el orden óptimo de truncación óptimo y se seleccionaron únicamente los parámetros pares  $d_2, d_4, d_6$  y  $d_8$ .

### ***d. Discusión y conclusiones de la aplicación***

La Tabla 1 presenta la estimación ML para cada uno de los campos seleccionados. El Panel A muestra los parámetros estimados para la distribución lognormal y el Panel B muestra los parámetros

<sup>4</sup> Los autores agradecen a Herman Aguinis y Ernest O’Boyle por permitirnos el uso de su base de datos sobre productividad académica recopilada en O’Boyle y Aguinis (2012).

<sup>5</sup> Para detalles sobre la forma de depuración de los datos ver O’Boyle y Aguinis (2012), p.86.

estimados para la distribución log-SNP. En el Panel C encontramos el estadístico LR para el contraste de distribución log-SNP frente a lognormal.

[Insertar Tabla 1]

Los resultados de la estimación muestran que todos los modelos capturan de manera adecuada la media y la desviación estándar de cada uno de los campos, capturada por los parámetros  $\mu$  y  $\sigma$  respectivamente. Como se observa, los p-valores indican que estos parámetros son altamente significativos para ambas distribuciones. Sin embargo, como se muestra en el Panel B, para la distribución log-SNP, los parámetros  $d_s$  también resultan altamente significativos para la mayoría de los campos de conocimiento. Al analizar el AIC (que penaliza la inclusión de parámetros adicionales) para las dos distribuciones, se encuentra que este criterio resulta consistentemente inferior para la distribución log-SNP lo que nos sugiere la modelación a partir de esta distribución resulta claramente superior. Además, los estadísticos LR incluidos en el Panel C, concluyen que, para todos los campos seleccionados, la incorporación de los parámetros  $d_s$  mejoran la razón de verosimilitudes del modelo.

Un ejemplo de la calidad de los ajustes obtenidos para el campo Finance, se recoge en la Figura 1. En esta figura se muestra un detalle de las colas derechas de la distribución que incluye los investigadores de alta productividad. Como se observa, la especificación log-SNP permite caracterizar mejor la actividad investigadora.

[Insertar Figura 1]

Alternativamente, la Figura 2 muestra la comparación de los ajustes de la distribución para Finance en términos de la cdf empírica versus la teórica, la log-SNP y la lognormal. Esta última parece infravalorar la probabilidad acumulada en comparación con la log-SNP.

[Insertar Figura 2]

Las figuras anteriores muestran como la distribución lognormal subestima la productividad investigadora especialmente para los valores más extremos (bajo la distribución lognormal un investigador debe publicar menos artículos para ser considerado top). La Tabla 2 ilustra<sup>6</sup> estos efectos para las distintas áreas de conocimiento calculando los cuantiles empíricos y estimados bajo lognormal y log-SNP para niveles de confianza del 5%, 1%, 0.1% y 0.05%.

[Insertar Tabla 2]

---

<sup>6</sup> Para obtener los cuantiles de la distribución log-SNP nos basamos en la cdf presentada en la ecuación (15) y en el Método de la Transformada Inversa (ITM).



La interpretación de los valores de dicha tabla muestra claramente la mayor precisión de la distribución log-SNP en los ajustes en las colas y los errores inducidos en la estimación de la productividad de los investigadores top mediante el uso de distribuciones paramétricas tradicionales como la lognormal. Estos resultados son consistentes con las investigaciones de Kumar et al. (1998), Perc (2010) y Eom y Fortunato (2011) quienes al aplicar la distribución lognormal a indicadores bibliométricos encontraron que se queda corta a la hora de modelar series con colas muy largas.

## **VI. Conclusiones**

La literatura indica que en algunos casos, las distribuciones habitualmente utilizadas para la caracterización de variables económicas y financieras se han mostrado incapaces de representarlas. Esto se debe a que algunas variables parecen regirse por una distribución que presenta colas muy pesadas. Así, se pone de manifiesto la necesidad de proponer otras distribuciones y metodologías adecuadas para tales propósitos.

En este trabajo proponemos una novedosa metodología basada en la distribución log-SNP. Dicha distribución anida la lognormal incluyendo nuevos parámetros capaces de recoger mejor el comportamiento de las colas permitiendo contrastar las deficiencias de la distribución lognormal en esta dirección. Como siguientes aplicaciones se trabajará en la modelación de Riesgo Operacional, en donde su medición ocupa una de las preocupaciones clave de las instituciones financieras. También se realizará una aplicación desde los mercados financieros, en los cuales las series de tiempo también exhiben colas pesadas que requieren de especial atención en su modelación.

## **Referencias**

- Abramo, G., & D'Angelo, C. A. (2014). Assessing national strengths and weaknesses in research fields. *Journal of Informetrics*, 8(3), 766–775.
- Abramo, G., D'Angelo, A. C., & Pugini, F. (2008). The measurement of Italian universities' research productivity by a non parametric-bibliometric methodology. *Scientometrics*, 76(2), 225–244.
- Abramowitz, M., & Stegun, I. A. (1972). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. New York: Dover Publications.
- Aguinis, H., O'Boyle, E., Gonzalez-Mulé, E., & Joo, H. (2015). Cumulative advantage: Conductors and insulators of heavy-tailed productivity distributions and productivity tars. *Personnel Psychology*, <http://dx.doi.org/10.1111/peps.12095> (in press).
- Albarrán, P., Juan, A. C., Ortuño, I., & Ruiz-Castillo, J. (2011). The skewness of science in 219 sub-fields and a number of aggregates. *Scientometrics*, 88(2), 385-397.
- Bertocchi, G., Gambardella, A., Jappelli, T., Nappi, C. A., & Peracchi, F. (2015). Bibliometric evaluation vs. informed peer review: Evidence from Italy. *Research Policy*, 44(2), 451-466.

- Birkmaier, D., & Wohlrabe, K. (2014). The Matthew effect in economics reconsidered. *Journal of Informetrics*, 8(4), 880–889.
- Borokhovich, K. A., Bricker, R. J., Brunarski, K. R., & Simkins, B. J. (1995). Finance research productivity and influence. *The Journal of Finance*, 50(5), 1691-1717.
- Box, G. (1976). Science and statistics. *Journal of the American Statistical Association*, 71(356), 791-799.
- Broadus, R. N. (1987). Toward a definition of ‘bibliometrics’. *Scientometrics*, 12(5-6), 373-379.
- Campanario, J. M. (2015). Providing impact: The distribution of JCR journals according to references they contribute to the 2-year and 5-year journal impact factors. *Journal of Informetrics*, 9(2), 398–407.
- Chen, X. (2007). Large Sample Sieve Estimation of Semi-Nonparametric Models. En J. Heckman, & E. Leamer, *Handbook of Econometrics*, Vol. 6, Ch. 76, Part B (págs. 5549-5632). Elsevier.
- Chung, K. H., & Cox, R. A. (1990). Patterns of productivity in the finance literature: a study of the bibliometric distributions. *The Journal of Finance*, 45(1), 301-309.
- Cont, R. (2001). Empirical properties of asset returns: Stylized facts and statistical issues. *Quantitative Finance*, 1(2), 223-236.
- Corrado, C., & Su, T. (1996). Skewness and Kurtosis in S&P 500 index returns implied by option prices. *Journal of Financial Research*, 19(2), 175-192.
- Cramér, H. (1925). On some classes of series used in mathematical statistics. *Sixth Scandinavian Congress of Mathematicians*, (págs. 399-425). Copenhagen.
- Del Brio, E., Mora-Valencia, A., & Perote, J. (2014(a)). Semi-nonparametric VaR forecasts for hedge funds during the recent crisis. *Physica A: Statistical Mechanics and its Applications*, 401(1), 330-343.
- Del Brio, E., Mora-Valencia, A., & Perote, J. (2014(b)). VaR performance during the subprime and sovereign debt crises: An application to emerging markets. *Emerging Markets Review*, 20(C), 23-41.
- Egghe, L. (2005). *Power laws in the information production process: Lotkaian informetrics*. Kidlington, UK: Elsevier Academic Press.
- Embrechts, P., Klüppelberg, C., & Mikosch, T. (1997). *Modelling extremal events. For insurance and finance*. Berlin: Springer-Verlag Inc.
- Eom, Y. H., & Fortunato, S. (2011). Characterizing and modeling citation dynamics. *PLoS ONE*, 6(9), e24926.
- Finardi, U. (2013). Correlation between Journal Impact Factor and Citation Performance: An experimental study. *Journal of Informetrics*, 7(2), 357–370.
- Frandsen, T. F. (2005). Geographical concentration. The case of economics journals. *Scientometrics*, 63(1), 69-85.
- Gallant, A. R., & Nychka, D. W. (1987). Semiparametric maximum likelihood estimation. *Econometrica*, 55(2), 363–390.
- Garfield, E. (1980). Bradford’s Law and related statistical pattern. *Essays of an Information Scientist*, 4(19), 476-483.

- Guerrero-Bote, V. P., Zapico-Alonso, F., Espinosa-Calvo, M. E., Gomez-Crisostomo, R., & Moya-Anegon, F. (2007). Import–export of knowledge between scientific subject categories: The iceberg hypothesis. *Scientometrics*, *71*(3), 423–441.
- Heberger, A. E., Christie, C. A., & Alkin, M. C. (2010). A bibliometric analysis of the academic influences of and on evaluation theorists' published works. *American Journal of Evaluation*, *31*(1), 24-44.
- Jondeau, E., & Rockinger, M. (2006). Optimal allocation under higher moments. *European Financial Management*, *12*(1), 29-55.
- Jondeau, E., & Rockinger, M. (2009). The impact of shocks on higher moments. *Journal of Financial Econometrics*, *7*(2), 77-105.
- Kendall, M., & Stuart, A. (1977). *The Advanced Theory of Statistics, Vol. I, 4th ed.* London: C. Griffin.
- Kocher, M. G., Luptacik, M., & Sutter, M. (2006). Measuring productivity of research in economics: A cross-country study using DEA. *Socio-Economic Planning Sciences*, *40*(4), 314-332.
- Kretschmer, H., & Kretschmer, T. (2007). Lotka's distribution and distribution of co-author pairs' frequencies. *Journal of Informetrics*, *1*(4), 308–337.
- Kumar, S., Sharma, P., & Garg, K. C. (1998). Lotka's law and institutional productivity. *Information Processing & Management*, *34*(6), 775–783.
- Lancho-Barrantes, B. S., Guerrero-Bote, V. P., & Moya-Anegón, F. (2010). The iceberg hypothesis revisited. *Scientometrics*, *85*(2), 443–461.
- Lotka, A. J. (1926). The frequency distribution of scientific productivity. *Journal of the Washington Academy of Science*, *16*(12), 317-323.
- Mauleón, I., & Perote, J. (2000). Testing densities with financial data: an empirical comparison of the Edgeworth-Sargan density to the Student's t. *European Journal of Finance*, *6*(2), 225-239.
- Nicholls, T. P. (1986). Empirical validation of Lotka's law. *Information Processing & Management*, *22*(5), 417–419.
- Nicolaisen, J., & Hjørland, B. (2007). Practical potentials of Bradford's law: a critical examination of the received view. *Journal of Documentation*, *63*(3), 359 - 377.
- Ñíguez, T.-M., Paya, I., Peel, D., & Perote, J. (2012). On the stability of the constant relative risk aversion (CRRA) utility under high degrees of uncertainty. *Economics Letters*, *115*(2), 244-248.
- O'Boyle, E., & Aguinis, H. (2012). The best and the rest: Revisiting the norm of normality of individual performance. *Personnel Psychology*, *65*(1), 79–119.
- Perc, M. (2010). Zipf's law and log-normal distributions in measures of scientific output across fields and institutions: 40 years of Slovenia's research as an example. *Journal of Informetrics*, *4*(2), 358–364.
- Perote, J., & Del Brio, E. (2003). Measuring Value-at-Risk under the conditional Edgeworth-Sargan distribution. *Finance Letters*, *1*(3), 23-40.
- Phillips, P. B. (1977). A general theorem in the theory of asymptotic expansions as approximations to the finite sample distributions of econometric estimators. *Econometrica*, *45*(6), 1517–1534.

- Price, D. S. (1976). A general theory of bibliometric and other cumulative advantage processes. *Journal of the American Society for Information Science*, 27(5), 292–306.
- Radicchi, F., Fortunado, S., & Castellano, C. (2008). Universality of citation distribution: Towards an objective measure of scientific impact. *Proceedings of the National Academy of Sciences of the United States of America*, 105(45), 17268–17272.
- Rousseau, R. (1994). Bradford curves. *Information Processing and Management*, 30(2), 267–277.
- Rubinstein, M. (1998). Edgeworth binomial trees. *Journal of Derivatives*, 5(3), 20-27.
- Sabharwal, M. (2013). Comparing research productivity across disciplines and career stages. *Journal of Comparative Policy Analysis: Research and Practice*, 15(2), 141-163.
- Sargan, D. (1975). Gram-Charlier approximation applied t ratios or k-class estimators. *Econometrica*, 43(2), 327-346.
- Verhoeven, P., & McAleer, M. (2004). Fat tails in financial volatility models. *Mathematics and Computers Simulation*, 64(3-4), 351-362.
- Wallace, D. L. (1958). Asymptotic approximations to distributions. *Annals of Mathematical Statistics*, 29(3), 635-654.

**TABLA 1. Resultados de la estimación**

Field of knowledge	Panel A Lognormal				Panel B Log-SNP								Panel C LR
	$\mu$	$\sigma$	logL	AIC	$\mu$	$\Sigma$	$d_2$	$d_4$	$d_6$	$d_8$	logL	AIC	
Agronomy	0.2143 (<.0001)	0.4368 (<.0001)	-3359.52	6723.04	0.1182 (0.000)	0.4771 (<.0001)	-0.0786 (0.000)	0.1448 (<.0001)	0.0252 (<.0001)	0.0042 (<.0001)	-1890.49	3792.98	2938.07 (<.0001)
Anthropology	0.3753 (<.0001)	0.6024 (<.0001)	-3089.70	6183.40	0.1693 (<.0001)	0.5438 (<.0001)	0.1912 (<.0001)	0.2733 (<.0001)	0.0408 (<.0001)	0.0050 (<.0001)	-2259.29	4530.58	1660.83 (<.0001)
Clinical psychology	0.3791 (<.0001)	0.5994 (<.0001)	-5501.26	11006.52	0.1689 (<.0001)	0.5556 (<.0001)	0.1535 (<.0001)	0.2611 (<.0001)	0.0444 (<.0001)	0.0055 (<.0001)	-4236.31	8484.63	2529.90 (<.0001)
Dentistry	0.4934 (<.0001)	0.6763 (<.0001)	-6598.224	13200.45	0.2959 (<.0001)	0.6913 (<.0001)	0.0194 (0.1765)	0.1481 (<.0001)	0.0157 (<.0001)	0.0027 (<.0001)	-5740.93	11493.86	1714.58 (<.0001)
Dermatology	0.4553 (<.0001)	0.6914 (<.0001)	-18154.32	36312.64	0.8375 (<.0001)	0.4294 (<.0001)	1.1923 (<.0001)	0.3812 (<.0001)	0.1092 (<.0001)	0.0179 (<.0001)	-7262.16	14536.32	21784.32 (<.0001)
Ecology	0.3335 (<.0001)	0.5445 (<.0001)	-2736.83	5477.66	0.1653 (<.0001)	0.5435 (<.0001)	0.0499 (0.0023)	0.1708 (<.0001)	0.0174 (<.0001)	0.0037 (<.0001)	-2027.75	4067.50	1418.16 (<.0001)
Economics	0.2887 (<.0001)	0.5198 (<.0001)	-1450.68	2905.37	0.1418 (<.0001)	0.5133 (<.0001)	0.0538 (0.0819)	0.2073 (<.0001)	0.0277 (<.0001)	0.0041 (<.0001)	-935.65	1883.29	1030.08 (<.0001)
Educational psychology	0.3404 (<.0001)	0.5320 (<.0001)	-1356.60	2717.21	0.1764 (<.0001)	0.5367 (<.0001)	0.0381 (0.0900)	0.1614 (<.0001)	0.0194 (<.0001)	0.0034 (<.0001)	-1108.26	2228.51	496.70 (<.0001)
Ethics	0.2952 (<.0001)	0.5262 (<.0001)	-516.72	1037.45	0.1556 (0.0028)	0.5301 (<.0001)	0.0282 (0.4423)	0.2231 (<.0001)	0.0351 (0.0017)	0.0048 (<.0001)	-338.55	689.11	356.34 (<.0001)
Ethnic studies	0.2287 (<.0001)	0.4647 (<.0001)	-849.22	1702.44	0.1290 (0.0038)	0.5045 (<.0001)	-0.0854 (0.0011)	0.1877 (<.0001)	0.0347 (<.0001)	0.0050 (<.0001)	-511.39	1034.78	675.66 (<.0001)
Finance	0.4560 (<.0001)	0.6688 (<.0001)	-1692.96	3389.92	0.1693 (<.0001)	0.5763 (<.0001)	0.2975 (<.0001)	0.2992 (<.0001)	0.0484 (<.0001)	0.0060 (<.0001)	-1390.41	2792.82	605.10 (<.0001)

continúa

**TABLA 1 (continuación)**

Field of knowledge	Panel A Lognormal				Panel B Log-SNP								Panel C LR
	$\mu$	$\sigma$	logL	AIC	$\mu$	$\sigma$	$d_2$	$d_4$	$d_6$	$d_8$	logL	AIC	
Forestry	0.3785 (<.0001)	0.5755 (<.0001)	-5958.31	11920.63	0.1797 (<.0001)	0.5490 (<.0001)	0.1149 (<.0001)	0.1942 (<.0001)	0.0232 (<.0001)	0.0037 (<.0001)	-4879.51	9771.02	2157.61 (<.0001)
Genetics	0.3338 (<.0001)	0.5350 (<.0001)	-7617.37	15238.74	0.1720 (<.0001)	0.5379 (<.0001)	0.0399 (<.0001)	0.1748 (<.0001)	0.0224 (<.0001)	0.0037 (<.0001)	-6015.27	12042.54	3204.20 (<.0001)
History	0.3080 (<.0001)	0.4570 (<.0001)	-2198.69	4401.39	0.1984 (<.0001)	0.5112 (<.0001)	-0.0776 (<.0001)	0.0627 (<.0001)	-0.0004 (0.8251)	0.0013 (<.0001)	-2095.15	4202.29	207.10 (<.0001)
Law	0.2788 (<.0001)	0.4908 (<.0001)	-578.29	1160.59	0.1507 (<.0001)	0.4953 (<.0001)	0.0244 (0.6560)	0.1747 (<.0001)	0.0163 (0.0272)	0.0027 (<.0001)	-389.59	791.18	377.40 (<.0001)
Linguistics	0.3307 (<.0001)	0.5556 (<.0001)	-1801.66	3607.31	0.1558 (<.0001)	0.5395 (<.0001)	0.0844 (<.0001)	0.2007 (<.0001)	0.0246 (<.0001)	0.0042 (<.0001)	-1270.77	2553.54	1061.77 (<.0001)
Mathematics	0.2458 (<.0001)	0.4342 (<.0001)	-1346.20	2696.39	0.1652 (<.0001)	0.4945 (<.0001)	-0.1013 (<.0001)	0.1159 (<.0001)	0.0071 (0.0210)	0.0019 (<.0001)	-971.81	1955.62	748.77 (<.0001)
Statistics	0.4510 (<.0001)	0.6390 (<.0001)	-5553.69	11111.38	0.2429 (<.0001)	0.6251 (<.0001)	0.0779 (<.0001)	0.1858 (<.0001)	0.0253 (<.0001)	0.0036 (<.0001)	-4758.50	1590.38	1590.38 (<.0001)

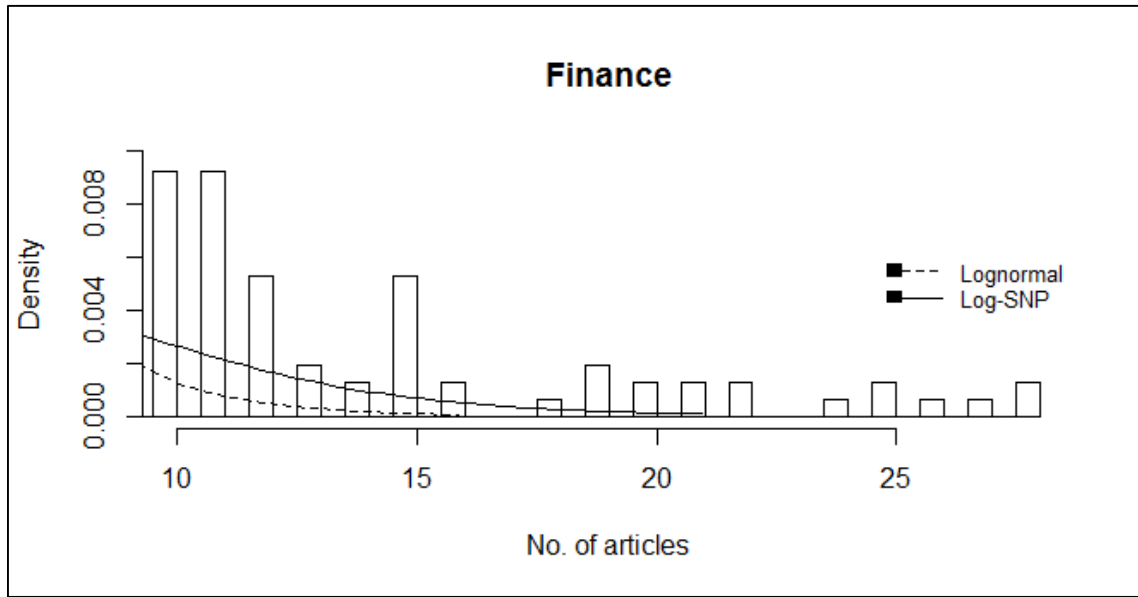
Esta tabla reporta la estimación ML para cada uno de los campos seleccionados. El Panel A muestra los parámetros estimados para la distribución lognormal. El Panel B muestra los parámetros estimados para la distribución log-SNP. El Panel C muestra la razón de verosimilitudes aplicada entre ambas distribuciones.  $\mu$  y  $\sigma$  son los parámetros de localización y escala y  $d_2, d_4, d_6$  y  $d_8$  los parámetros de peso de los polinomios de Hermite. logL=log-verosimilitud, AIC=Criterio de Información Akaike, LR=razón de verosimilitudes para el contraste de log-SNP frente a lognormal. P-valores en paréntesis. El estudio corresponde a 18 campos de conocimiento pertenecientes a las categorías JCR de las ciencias y las ciencias sociales entre los años 2000 y 2009.

**TABLA 2. Número de artículos observados empíricamente versus esperados teóricamente bajo lognormal y log-SNP**

Field of knowledge	N	Observed No. of articles				Expected number of articles							
		Top				Lognormal				Log-SNP			
		5%	1%	0.1%	0.05%	5%	1%	0.1%	0.05%	5%	1%	0.1%	0.05%
Agronomy	8,923	3	7	13	15	3	4	5	6	3	4	10	12
Anthropology	5,755	5	10	19	22	4	6	10	11	4	9	16	17
Clinical psychology	10,418	5	11	27	35	4	6	10	11	4	9	17	19
Dentistry	12,345	7	15	32	36	5	8	14	16	5	11	29	34
Dermatology	30,531	7	16	40	50	5	8	14	16	7	14	20	22
Ecology	5,730	4	8	17	20	4	5	8	9	4	7	14	16
Economics	3,048	4	8	25	26	4	5	7	8	3	7	13	14
Educational psychology	3,032	4	8	18	18	4	5	8	9	4	7	14	16
Ethics	1,073	4	9	24	25	4	5	7	8	3	8	14	16
Ethnic studies	2,003	3	8	16	16	3	4	6	6	3	5	12	14
Finance	3,019	6	13	26	28	5	8	13	15	5	11	19	21
Forestry	12,211	5	9	18	22	4	6	9	10	4	8	15	17
Genetics	16,574	4	8	18	23	4	5	8	9	4	7	14	16
History	6,708	3	5	8	12	3	4	6	7	3	5	8	11
Law	1,350	4	7	13	13	3	5	7	7	3	6	11	12
Linguistics	3,600	5	9	22	23	4	6	8	9	4	7	14	16
Mathematics	3,972	3	6	13	14	3	4	5	6	3	5	10	11
Statistics	10,679	6	13	26	35	5	7	12	13	5	10	22	26

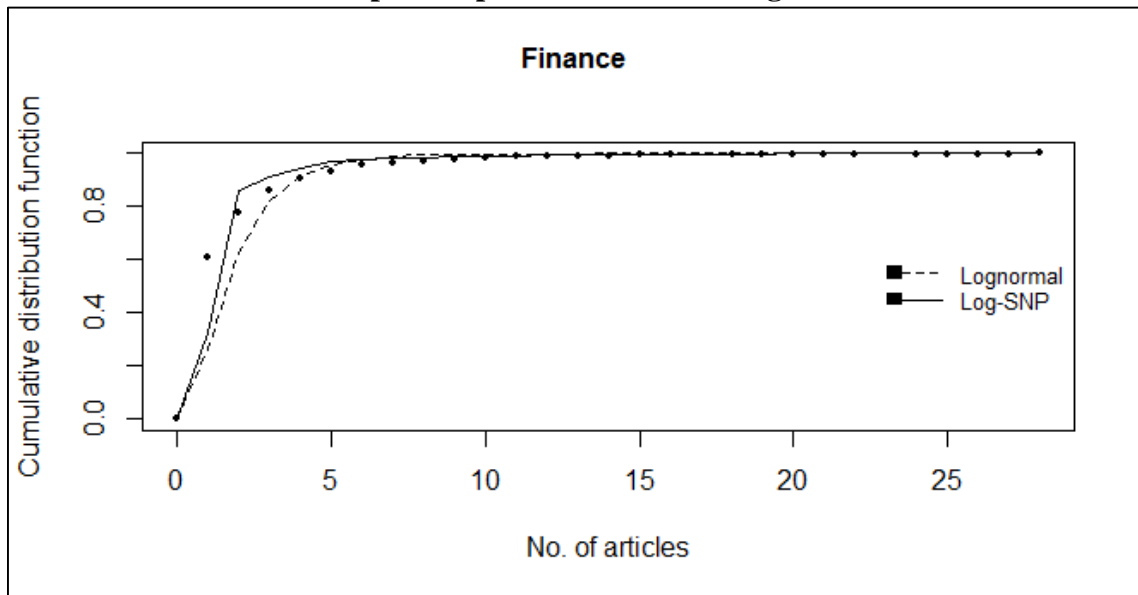
Esta tabla compara el número de artículos observados empíricamente en cada uno de los campos versus los esperados teóricamente bajo las distribuciones lognormal y log-SNP. N=número de investigadores por campo de conocimiento. Los valores 5%, 1%, 0.1% y 0.05% son percentiles de las distribuciones. El estudio corresponde a 18 campos de conocimiento pertenecientes a las categorías JCR de las ciencias y las ciencias sociales entre los años 2000 y 2009.

**FIGURA 1. PDF para la productividad investigadora en Finance**



La figura muestra la cola derecha para la distribución de frecuencias empírica (histograma) para la productividad de los investigadores que publicaron en las 5 mejores revistas (en términos JCR-2007) de Finance durante el periodo 2000-2009. Sobrepuestas se encuentran las pdf estimadas bajo una especificación lognormal y log-SNP.

**FIGURA 2. CDF para la productividad investigadora en Finance**



La figura muestra la función de distribución acumulada (cdf) empírica para la productividad de los investigadores que publicaron en las 5 mejores revistas (en términos JCR-2007) de Finance durante el periodo 2000-2009. Sobrepuestas se encuentran las pdf estimadas bajo una especificación lognormal y log-SNP.