Scale-free tails in Colombian financial indexes: a primer

Por: Carlos León

Borradores de
ECONOMÍA

Núm. 812
2014

Scale-free tails in Colombian financial indexes:
a primer[1]

Carlos León[2]
Banco de la República

Abstract

A maximum likelihood method for estimating the power-law exponent verifies that the positive and negative tails of the Colombian stock market index (IGBC) and the Colombian peso exchange rate (TRM) approximate a scale-free distribution, whereas none of the heavy tails of a local sovereign securities index (IDXTES) are a plausible case for such distribution. Results also (i) support critiques regarding the flaws of ordinary least squares estimation methods for scale-free distributions; (ii) question the validity of Zipf's law; (iii) suggest that IGBC and TRM display the scale-free nature documented as a stylized fact of financial returns, and that they may be following a *gradually truncated Lévy flight*; and (iv) suggest that local financial markets are self-organized systems.

JEL: C46; C58; G32

Keywords: scale-free, power-law, Zipf's law, financial returns.

[2] Research and Development Section Manager, Financial Infrastructure Oversight Department, Payment Systems and Banking Operation Division, Banco de la República, cleonrin@banrep.gov.co or carlosleonr@hotmail.com.

1 Traditional evidence of scaling laws

The magnitude of earthquakes, the population of cities, the intensity of wars, the level of rivers, the size of avalanches, the number of connections in most social and biological networks, and the usage of words in written share a common feature. All these phenomena exhibit extremely skewed distributions, where the most immediate consequence is the absence of a typical or average observation that may properly describe the whole sample.

Two examples are provided in Figure 1. The left panel shows that the distribution of inhabitants across the Colombian territory is particularly inhomogeneous, with municipalities' population extremely varying across observations. Due to the extreme skewness, the average population of municipalities in Colombia (i.e. 0.04 million) and its standard deviation (i.e. 0.23 million) are uninformative about the distribution of population across the territory. Moreover, the standard deviation falls short to account for the observed inhomogeneity, where Bogotá's and Medellín's population are 27.62 and 9.36 standard deviations away from the mean. The right panel of Figure 1, which contains the distribution of Colombian financial institutions' assets size, exhibits similar patterns; Bancolombia, the biggest local bank by assets' value, is more than 7 standard deviations away from the mean.

Figure 1
Distribution of Colombian municipalities and financial institutions[3]
a. Municipalities' population   b. Financial institutions' assets size
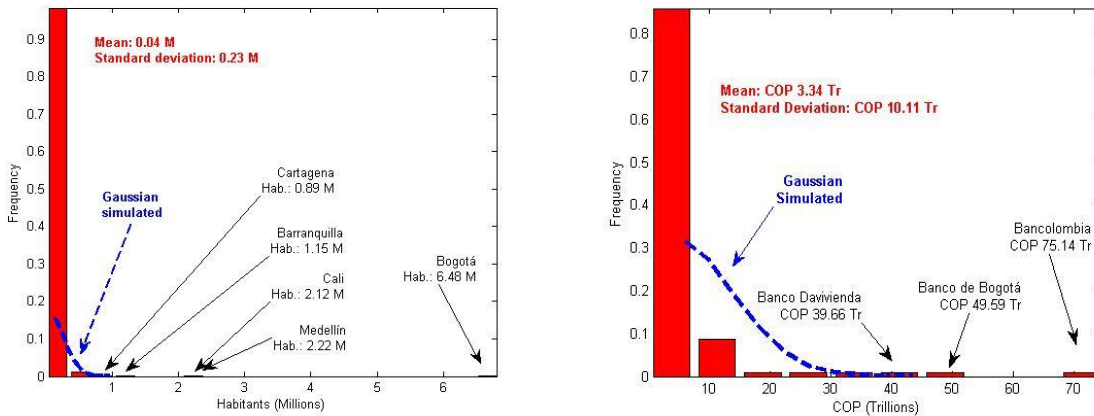
Source: author's calculations with data from DANE (panel a.) and Financial Superintendence of Colombia (b.).

The inadequacy of the first two moments of a Gaussian distribution to fit the population of Colombian municipalities or the assets of local financial institutions is even more prominent when using a standard (i.e. Gaussian) Monte Carlo method for simulating municipalities or financial institutions. After simulating 20,000 municipalities based on the estimated mean and standard deviation the largest municipality would have 1.07 million inhabitants, about the

---

[3] Panel (a.) data correspond to the last population census (June 2005) for 1,119 municipalities, based on public reports by DANE. Panel (b.) data correspond to financial reports for 114 financial institutions as of December 2012, based on public information from the Financial Superintendence of Colombia.

size of the fourth largest (i.e. Barranquilla), which is about one fifth the size of the largest (i.e. Bogotá); likewise, the largest financial institution by Gaussian simulation would have COP 45.57 trillion in assets, less than two thirds the size of the biggest.[4] In this sense, under such a skewed distribution the mathematical expectation is not informative of the shape of the distribution, and the dispersion (e.g. variance or standard deviation) is unable to account for the major heterogeneity of the observed data.

The particularities of the distribution of population have been documented long ago, and have resulted in what is called the *rank-size rule* or Zipf's law, which states that the population of a city is inversely proportional to its rank (Krugman, 1996). The most basic form to test for the *rank-size rule* is to plot size (x) and rank ($\Pr(x)$) using a double logarithmic scale, and to estimate the slope of the line relating both size and frequency across the whole distribution by standard ordinary least squares (OLS) regression. Figure 2 exhibits such basic test for the Colombian municipalities and financial institutions, which yields an estimated slope of -1.07 and -2.41, respectively.

Figure 2
Distribution of Colombian municipalities and financial institutions
(OLS on double logarithmic scale)

a. Municipalities' population

b. Financial institutions' assets size
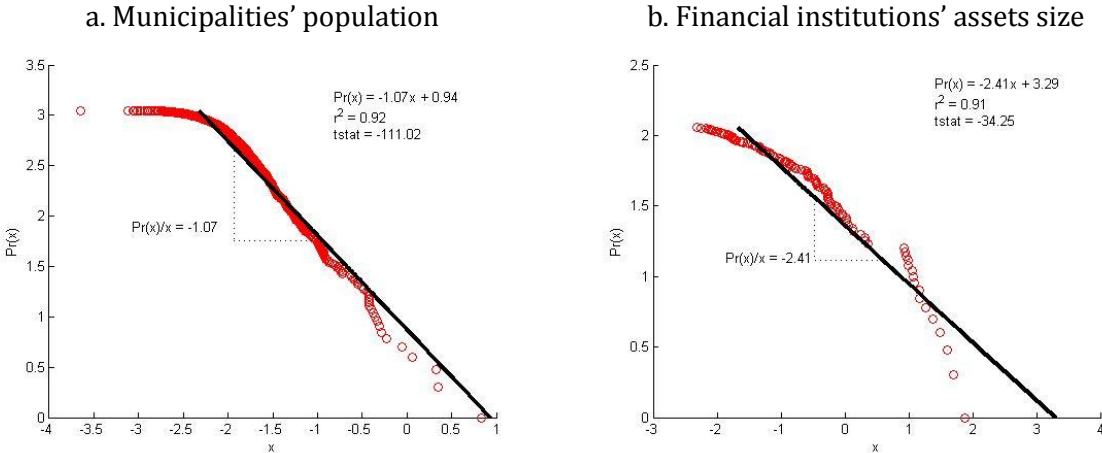
Source: author's calculations with data from DANE and Financial Superintendence of Colombia.

This slope of the straight line in Figure 2a is commonly used in urban economics as a *metropolization index* (Lanaspa et al., 2004), where the higher the absolute value of the slope the more egalitarian the distribution of population among cities. The estimated slope ($\hat{\gamma} = -1.07$) approximates to the traditional Zipf's exponent ($\gamma = -1.00$) and to the slope of metropolitan areas of the United States reported by Krugman (1996) and Simon (1955), and concurs with several estimations by Sánchez and España (2012) for the same Colombian data set; moreover, the fit of the OLS is fairly good (i.e. $r^2 = 0.92$ and $tstat = -111.02$ for the slope).

---

[4] The Monte Carlo simulation procedure implemented consisted of simulating 20,000 random numbers based on the estimated mean and standard deviation of the observed data. Since negative values are unfeasible, the absolute value of the Gaussian random numbers was used.

Remarkably, Zipf's law ($\gamma \sim -1.00$) is not documented for city sizes only: many other phenomena are accepted to follow the same regularity, such as the distribution of word frequencies (i.e. human vocabulary) in several languages, the size of avalanches, highway traffic and the water level of rivers (as reported in Bak, 1996; Simon, 1955; Mandelbrot and Hudson, 2004). Regarding the size of financial institutions, Fiaschi et al. (2013) find that the assets of financial firms in the United States also comply with Zipf's law; in the Colombian case, however, despite the fit is –again- fairly good (i.e. $r^2 = 0.91$ and $tstat = -34.25$ for the slope), Zipf's law does not hold according to Figure 2b ($\hat{\gamma} \sim -2.41$).

Mathematically, a straight line on a double logarithmic plot is called a "power-law" (Bak, 1996). In the case in hand this means that the probability ($\Pr(\cdot)$) of finding a municipality or financial institution of size x can be expressed as some power of x, as in [§1], where estimating the order of the power is usually done by means of the logarithmic transformation in [§2]. Under this mathematical representation the slope is positive by construction, where Zipf's law corresponds to $\hat{\gamma} \sim 1.00$; thus, hereafter the minus sign vanishes when estimating γ.

$$\Pr(x) = x^{-\gamma} \qquad [§1]$$

where

$$\log \Pr(x) = -\gamma \log x \qquad [§2]$$

Finding that a straight line fits the observed data on a double logarithmic scale has another interpretation. Since the straight line is the same everywhere, *where there are no features at some scale that makes that particular scale stand out* (Bak, 1996), the distribution of such data is called scale-free or scale-invariant.

However, Zipf's law is a particular case of a power-law. Other magnitudes for $\hat{\gamma}$ are feasible, and have been documented for various phenomena. The most well-known case of estimating the slope of a straight line that fits observed data on a double logarithmic scale dates back to Pareto's work on the distribution of wealth: in the wake of the twentieth century, based on tax record data from Basel (Switzerland) and Augsburg (Germany), rental income from Paris, personal income from Britain, Prussia, Saxony, Ireland, Italy and Perú, Vilfredo Pareto found that the straight line that fitted plotting income against the number of people had a particular slope, $\hat{\gamma} \sim 3/2$, which was consistent with much wealth concentrated in very few hands (Mandelbrot and Hudson, 2004).

Another well-known phenomenon that has been documented as approximating a power-law with $\hat{\gamma} \sim 3/2$ pertains to geology, and is related to the magnitude of earthquakes and their frequency. As in the case of wealth and city sizes, most earthquakes are of low –almost imperceptible- magnitude, whereas a few are devastating. Such distribution of the energy released by earthquakes is characterized by the Gutenberg-Richter scale (i.e. the Richter scale), which states that the probability of finding an earthquake releasing an e amount of energy is $\Pr(e) \sim e^{-1.5}$.

Mandelbrot is credited for introducing the double logarithmic plot to determine the scaling properties of financial time-series.[5] Using a century of daily U.S. cotton prices Mandelbrot (1963) found $\hat{\gamma} \sim 1.7$, a result verified by Fama (1963 & 1965). Based on his findings Mandelbrot would contend the Brownian motion assumption with its generalized version (i.e. fractional Brownian motion), and would suggest changing from the Gaussian hypothesis for price changes to the stable Paretian or stable Lévy hypothesis, in which the exponent that determines the height of the tails (γ) is the most important for comparing the goodness-of-fit against the traditional Gaussian hypothesis (Fama, 1963).[6]

2   Advances in the estimation of power-law exponents

Despite its simplicity and informational content, using the double logarithmic plot or the analogous OLS regression to determine the scaling properties of data is problematic. Several authors (Clauset et al., 2009; Sinha et al., 2011; Stumpf and Porter, 2012) have documented the main problems behind the traditional OLS method for estimating and confirming (or rejecting) a distribution approximating a power-law. Moreover, as in Clauset et al. (2009), new estimation techniques have rejected many of those data sets long considered as approximating a power law (e.g. earthquake intensity, net worth, links to web sites), whereas some others have been confirmed (e.g. words frequencies, city sizes, co-citation in scientific papers, scientific papers authored) or revealed (e.g. solar flare intensity, intensity of wars, terrorist attack severity, power blackouts).

Based on Clauset et al. (2009), the three main drawbacks of traditional estimation of power-law exponents are the following:

a. *Few empirical phenomena obey power laws for all values of x*; in such cases the tail of the distribution follows a power-law, and –thus- the estimation should be confined to the tail.
b. *The ordinary formula for the calculation of the standard error on the slope of the regression is correct when the assumptions of linear regression hold, which include*

---

independent, Gaussian noise; using a logarithmic transformation turns the noise into non-Gaussian, which may turn the estimation of error unreliable.[7]

c. Distributions that are nothing like a power-law can appear to follow a power-law for small samples and some, like the log-normal, can approximate a power-law closely over many orders of magnitude, resulting in high values of r^2; then, traditional OLS goodness-of-fit tests are non-informative since *the probability of successfully detecting a violation of the power-law assumption is low*.

Therefore, Clauset et al. introduce a maximum likelihood estimation (MLE) procedure for estimating the power-law exponent for observations pertaining to the tail of the data. Let $i = 1 \cdots n$ be the observed values of x such that $x_i \geq x_{min}$, in which x_{min} is the threshold for defining the part of the data that will be considered for fitting the power-law model (i.e. the tail), the MLE is defined as in [§3].[8] This MLE is known as the Hill estimator, which is documented to yield asymptotically normal and consistent estimates of γ from random samples of a distribution with an asymptotic power-law form (Clauset et al., 2009; Sinha et al., 2011; Dowd, 2005).[9]

$$\hat{\gamma} = 1 + n \left[\sum_{i=1}^{n} \ln \frac{x_i}{x_{min}} \right]^{-1} \qquad \text{[§3]}$$

Fitting a Pareto-type (i.e. power-law) parametric distribution cannot accommodate at the same time the features of the bulk of the data and of the tail (Carmona, 2014); hence, defining x_{min} is key for the procedure. If x_{min} is set too low (i.e. not discarding too much data) the estimated exponent will be biased due to fitting the model to non-power-law data. If x_{min} is set too high the estimation will be biased due to discarding potentially useful data, and statistical error will emerge from finite size effects.

As discussed by Clauset et al. (2009), several methods are available for defining x_{min}, including the visual inspection of $\hat{\gamma}$ for different values of x_{min}. Yet, Clauset et al. choose to employ the Kolmogorov-Smirnov standard non-parametric test for equality of probability distributions, which calculates the maximum distance (D) between the cumulative distribution function from the data and the fitted model.[10]

Let $\mathcal{P}(x)$ be the cumulative probability function (CDF) for the power-law model that best fits the data in the $x_i \geq x_{min}$ sample, and $\mathcal{F}(x)$ the CDF for the same sample, x_{min} corresponds to the value that minimizes D in [§4].

---

[7] Some authors prefer to use the cumulative density function (CDF) instead of the probability density function (PDF), where the latter is the one used in Figure 2. When using the CDF not only the errors are non-Gaussian, but may be dependent due to their cumulative nature.

[8] Clauset et al. (2009) defines separate methods for the continuous and discrete data; since the data considered in this document is continuous, the latter is not discussed. As usual, all power-law exponents correspond to estimations from data, therefore the usage of $\hat{\gamma}$ instead of γ; the true value of the power-law exponent is unknown.

[9] Some drawbacks of the Hill estimator are documented in Dowd (2005).

[10] Sinha et al., (2011) suggest an alternative based on the minimization of the estimation error in [§3] by the subsample bootstrap method.

$$D = \max_{x_i \geq x_{min}} |\mathcal{F}(x) - \mathcal{P}(x)| \hspace{4cm} [\S4]$$

However, the MLE in [§3] will only yield the best fit to the power-law functional form under the choice of x_{min}. Since any data can be fitted to any theoretical distributional form (e.g. Gaussian, exponential, power-law), a goodness of fit test is required in order to assess if the data significantly deviates from the theoretical target.

Again Clauset et al. (2009) rely on the Kolmogorov-Smirnov standard non-parametric test for equality of probability distributions. The proposed test consists of comparing the empirical data set with a significant number of samples of synthetic data sets from a true power-law distribution. If empirical data deviates *too much* from the synthetic samples, it is possible to state that the data is not drawn from a power-law distribution.

As designed by Clauset et al., the *p-value* (p) resulting from this goodness-of-fit-test is the fraction of the time the distance from the synthetic samples and their best fit to the power-law functional form is larger than the distance from the empirical data and its best fit.[11] Thus, the higher p, the better the fit to the power-law functional form. According to Clauset et al., a conservative threshold for ruling out the power-law hypothesis is $p \leq 0.10$; this threshold corresponds to those cases in which 10% or less of the synthetic power-law data and their corresponding best fit resulted in larger deviations (i.e. poorer fits) than that of empirical data and its best power-law fit.

There are two main limitations of the estimation and testing procedures by Clauset et al. (2009). First, attaining a high p does not necessarily mean that the power-law form is the correct distribution for the data; other similar (i.e. skewed) distributions may provide an equal or better fit. Second, if n (i.e. the number of observations beyond x_{min}) is small ($n \lesssim 100$), p may become inaccurate. However, compared to the old-fashioned OLS estimation method and its said downsides, the MLE method and the goodness-of-fit-test are a major and patent enhancement to the fitting of power-laws in empirical data.

Based on the method presented in this section, Figure 3 presents the double logarithmic plot corresponding to the two data sets used so far. The circles (in red) represent the observed data, whereas the triangles (in green) correspond to a Gaussian Monte Carlo simulation based on the mean and standard deviation estimated from observed data[12], in which each solid line represents the best power-law fit attained with the MLE procedure; unlike Figure 2, the vertical axis uses a log cumulative distribution scale.

---

[11] As in Clauset et al. (2009), the procedure is as follows: (i) based on [§3] and the choice of x_{min}, calculate $\hat{\gamma}$; (ii) calculate the K-S test (i.e. the distance) for the empirical data and the best fit; (iii) generate a large number of power-law distributed synthetic data sets with $\hat{\gamma}$ and x_{min}; (iv) calculate the K-S test for the synthetic data and their corresponding individual best fit; (v) count the fraction of the time the K-S test for synthetic data is larger than for the empirical data.

[12] All Monte Carlo simulations hereafter consisted of generating Gaussian distributed random numbers, where the quantity of random numbers equals three times the size of the original data set. As usual, the corresponding process uses the estimated mean and standard deviation.

Figure 3
Distribution of Colombian municipalities and financial institutions
(MLE on double logarithmic scale)*

a. Municipalities' population

b. Financial institutions' assets size**
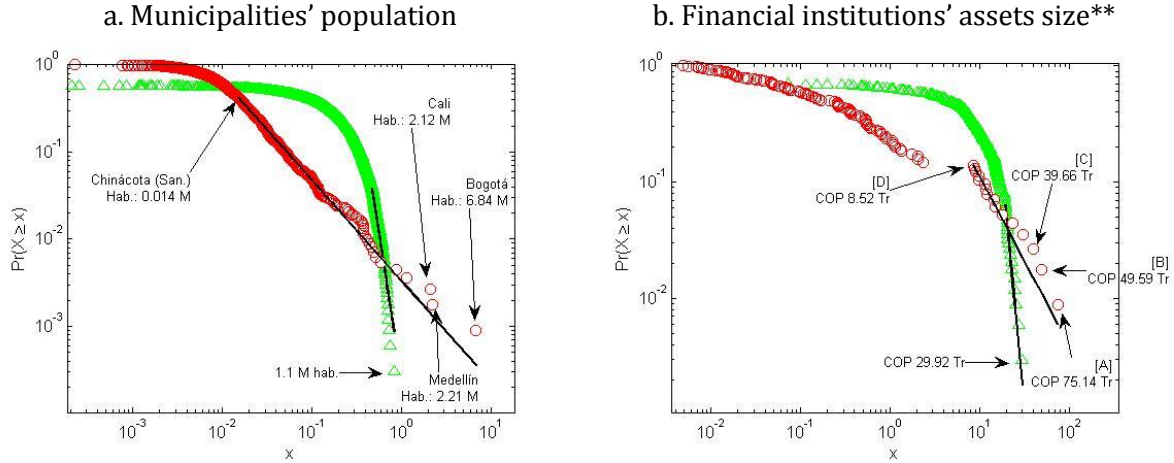
(*) Circles in red correspond to observed data; triangles in green to Gaussian synthetic data.
(**) Bancolombia [A]; Banco de Bogotá [B]; Banco Davivienda [C]; Citibank Colombia [D].
Source: authors' calculations.

Two main remarks may be extracted from Figure 3a. First, the difference between the observed data (circles in red) and the Gaussian synthetic data (triangles in green) is evident; as expected, Gaussian distributed data result in a large horizontal line with a short fast-decaying (i.e. vertical) tail, whereas the observed data has a short horizontal section and a large slow-decaying tail. Second, the line that corresponds to the power-law fit for the Gaussian synthetic data is steeper, almost vertical, whereas the fit for the observed data exhibits a moderate slope.

Regarding Figure 3b, the most striking remark arises from the "interruption" in the observed data, which appears to break the size of financial institutions in two clear groups, each one with a different distributional form.[13] Other remarks come in the form of the starting point of the MLE fit, in which x_{min} corresponds to the first financial institution of the "big financial institutions" group, and –again- the rapid decay of the Gaussian synthetic data.

Table 1 presents the numerical results. Based on the MLE method presented and the corresponding goodness-of-fit-test, both data sets follow a power-law distributional form. Yet, the value of the exponent for the municipalities' population is much higher than that estimated with OLS (i.e. almost twice as much); therefore, Zipf's law does not hold under the MLE estimation method for the size of Colombian municipalities, and the best fit for the power-law is for municipalities equal or larger than 14,784 habitants. The MLE estimated power-law for the financial institutions' assets size is not very different from that of the OLS, but the best fit is limited to financial institutions equal or larger than COP 8.52 trillion.

---

[13] Fiaschi et al. (2013) document a similar "interruption" for United States' financial institutions as well.

Stats	Population	Assets size
Observations	1,119	114
Mean	0.04	3.34
Standard deviation	0.23	10.11
Kurtosis	661.06	28.80
Skewness	23.77	4.74
$\hat{\gamma}_{OLS}$	1.07	2.41
r^2	0.92	0.91
$\hat{\gamma}_{MLE}$	2.15	2.45
p	0.30	0.79
x_{min}	0.014	8.52

(*) Mean, standard deviation and x_{min} are in the original units: population in habitants (millions) and asset size in COP (trillions).
Source: author's calculations.

3   Estimating power-law tail exponents on Colombian financial time-series

Cont (2001) documents that the distribution of financial returns display power-law or Pareto-like tails exponents in the $2 < \hat{\gamma} \leq 5$ range.[14] Sinha et al. (2011) and Gabaix et al. (2003a,b) document that the tails of the cumulative return distribution of several stock indexes (e.g. S&P 500, FTSE, DAX, IPC, WIG) and some individual stocks actually follow a power-law with an exponent $\hat{\gamma} \approx 3$, also known as the *inverse cubic law*. Likewise, Gopikrishnan et al. (1998) confirm the inverse cubic law for the three major US stock markets (i.e. NYSE, NASDAQ and AMEX), and find that the right and left tail display different power-law exponents (i.e. $\hat{\gamma}_{right} = 3.10 \pm 0.03$ and $\hat{\gamma}_{left} = 2.84 \pm 0.12$). On a wide sample of 202 stock markets, Eryigit et al. (2009) finds that the inverse cubic law holds under some assumptions (i.e. $x_{min} = 1$ standard deviation), but that other distributions could better fit the data as well. When estimating x_{min} by means of K-S test (as in Clauset et al., (2009)) Eryigit et al. found power-law exponents close to 4 (i.e. $\hat{\gamma}_{right} = 3.93$ and $\hat{\gamma}_{left} = 3.97$).

Three Colombian financial time-series are used to test the fit of a power-law exponent to their positive and negative tails based on the method developed by Clauset et al. (2009). The first series correspond to the official Colombian Peso – United States Dollar exchange rate, known in the local market as TRM (*Tasa Representativa del Mercado*); the second corresponds to IGBC (*Índice General de la Bolsa de Valores de Colombia*), the main index of the Colombian stock exchange; the third corresponds to IDXTES, a total-return index for local sovereign

---

[14] Such tail regime ($2 < \gamma \leq 5$) excludes infinite variance regimes ($\gamma < 2$) and the Normal distribution ($\gamma = 2$).

securities developed by Reveiz and León (2010).[15] The Standard and Poor's 500 (S&P500) U.S. stock market index is presented for comparison purposes.

The main statistics of these four time-series are presented in Table 2. Standard Normality tests (i.e. Jarque-Bera and Kolmogorov Smirnov) rejected the null hypothesis of normality of returns at the 5% significance level.

Table 2

Main statistics of selected time-series

| Statistic	TRM	IGBC	IDXTES	S&P 500
Observations	3,316	3,387	3,311	8,544
Mean	8.8×10^{-6}	7.6×10^{-4}	5.3×10^{-4}	3.3×10^{-4}
Standard deviation	6.4×10^{-3}	1.3×10^{-2}	3.1×10^{-3}	1.1×10^{-2}
Skewness	1.5×10^{-2}	4.6×10^{-3}	6.7×10^{-1}	-1.16
Kurtosis	12.49	16.90	14.47	29.57
Source: authors' calculations				

Time-series were transformed to individual vectors in a unitary form, namely by subtracting to each record its corresponding mean value and normalizing it to its standard deviation. Not only this transformation allows for comparisons between time-series, but it also allows to work x and x_{min} in a terms of standard deviations.

The substantial departure of observed data (crosses in blue) from the straight line in the Q-Q plots (Figure 4) evidences substantial excess kurtosis as a common feature of the four time-series. Both tails, left and right, corresponding to price decreases and increases, respectively, display thicker tails than the Gaussian hypothesis assumes.

---

[15] Based on spot transactions between foreign exchange market intermediaries, TRM is calculated by the Financial Superintendence of Colombia. As of April 15 2013 IGBC was replaced by COLCAP, but IGBC was calculated until the end of November 2013; since the new index (COLCAP) has limited data (i.e. from January 2008), the IGBC is preferred. The three local time-series comprise daily data from January 2000 to November 2013.

Figure 4
Q-Q plots of selected time-series
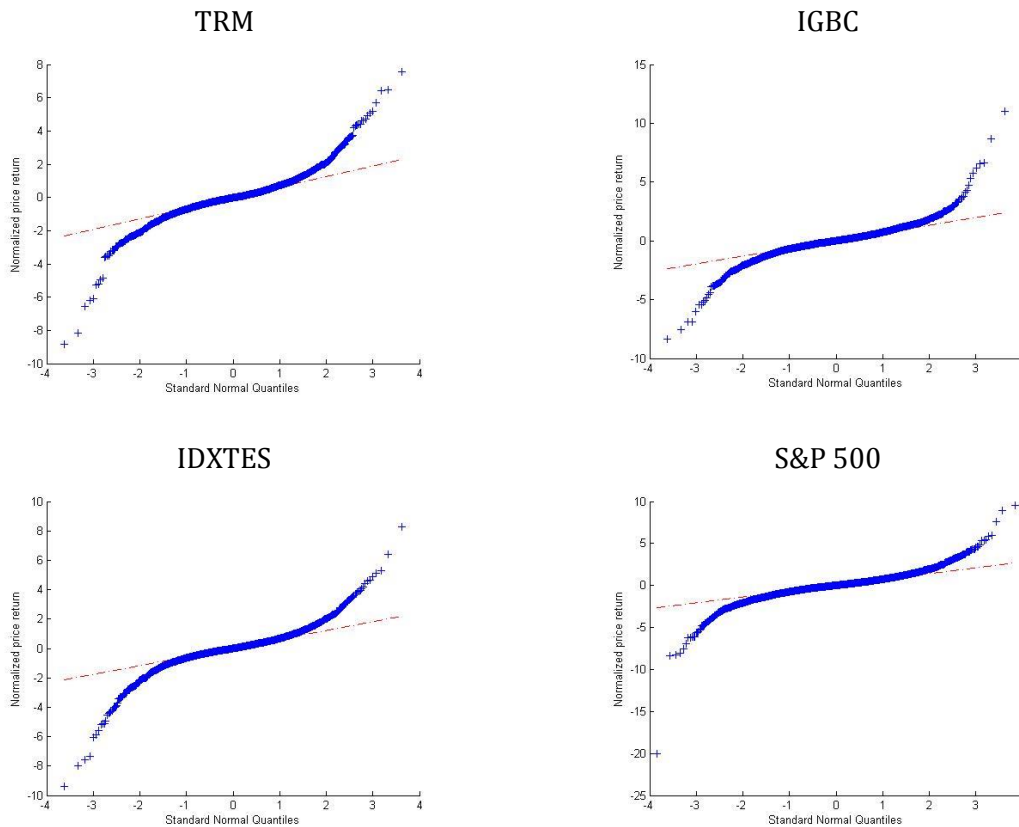
Source: author's calculations.

Figure 5 displays the distribution of the left and right parts of the distribution of the four selected time-series. As before, circles (in red) correspond to observed data, whereas triangles (in green) result from Gaussian synthetic data. Unlike the plots for Colombian municipalities' population and financial institutions' asset size, the departure from the Gaussian hypothesis is less evident: observed and synthetic data share a large horizontal line with a fast-decaying tail, in which the Gaussian decays faster (i.e. it is steeper).

However, despite the graphical similarity, the differences are quantitatively unmistakable. For instance, in the TRM case several observed returns exceed the largest Gaussian, with the largest negative (positive) change corresponding to about -8.8 (7.54) standard deviations, more than twice the largest change under the Gaussian simulation (± 3.53); in the S&P 500 case, the largest negative change, corresponding to the -20 standard deviation drop in October 19 1987, is about 5.8 times the largest under Gaussian assumptions (± 3.42).

Graphical comparison between time-series exhibits different degrees of deviation of the observed data with respect to the best power-law fit. It seems that the fit for the IDXTES series is the less appropriate, in which the most extreme price changes, positive and negative, do not follow the straight line corresponding to the attained power-law form. On the other hand, the best power-law fit for TRM, IGBC and S&P500 seems to be appropriate.

11

Figure 5
Distribution of time-series
(MLE on double logarithmic scale)*
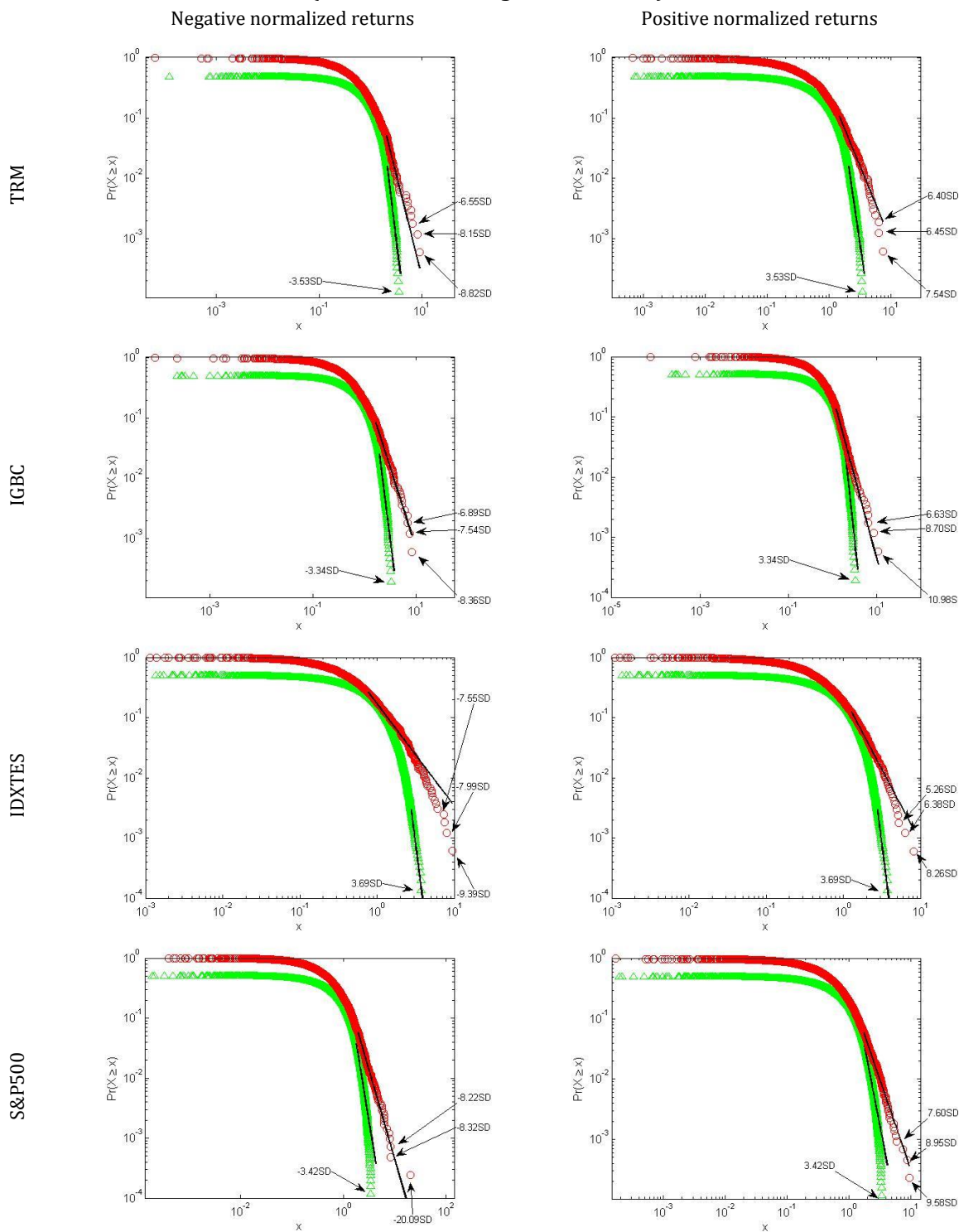
(*) Circles in red correspond to observed data; triangles in green to Gaussian synthetic data.
Source: authors' calculations

Table 3 summarizes the quantitative results from fitting the power-law to the four selected time-series. Three immediate remarks arise. First, all OLS estimations verify Zipf's law ($\hat{\gamma}_{OLS} \sim 1.00$) with high goodness-of-fit levels ($0.62 \le r^2 \le 0.70$). Second, all MLE estimations invalidate Zipf's law ($2.7 \le \hat{\gamma}_{MLE} \le 4.42$), and concur with the asymptotic behavior reported by Cont (2001). Third, the positive and negative returns of TRM and IGBC, and the negative returns of the S&P 500 are consistent with their tail approximating a power-law hypothesis, with $p > 0.10$, in which $\hat{\gamma}_{MLE} \sim 4$, as also reported by Eryigit et al. (2009); on the other hand, the power-law hypothesis is incompatible with IDXTES and the positive returns of S&P 500.

<div align="center">

Table 3

Power-law fit for selected time-series*

</div>

Stats	TRM $-\mid+$	IGBC $-\mid+$	IDXTES $-\mid+$	S&P 500 $-\mid+$
Observations (m)	**1,694 | 1,622**	**1,667 | 1,720**	1,636 | 1,675	**4,160** | 4,384
$\hat{\gamma}_{OLS}$	**1.02 | 1.08**	**1.03 | 0.98**	1.04 | 1.01	**0.99** | 0.96
r^2	**0.63 | 0.65**	**0.65 | 0.63**	0.70 | 0.65	**0.64** | 0.62
$\hat{\gamma}_{MLE}$	**4.42 | 3.48**	**3.67 | 3.71**	2.70 | 3.43	**4.01** | 3.97
p	**0.44 | 0.17**	**0.65 | 0.90**	0.00 | 0.04	**0.47** | 0.05
x_{min}	**2.01 | 1.48**	**1.67 | 1.22**	0.77 | 1.27	**1.93** | 1.76
n	**87 | 177**	**140 | 233**	428 | 211	**237** | 259
n/m	**0.05 | 0.11**	**0.08 | 0.14**	0.26 | 0.13	**0.06** | 0.06

(*) + and – correspond to positive and negative returns, respectively; x_{min} is in a standard deviation scale; since n corresponds to the number of x observations such that $x \ge x_{min}$ and m to the number of positive or negative returns, n/m is the percentage size of the tail regime according to the MLE procedure. Source: author's calculations.

The strongest cases for a power-law hypothesis ($p \gg 0.10$) are those of IGBC, negative returns of S&P 500 and negative returns of TRM; positive returns of TRM being consistent with a power-law are of moderate plausibility. For these cases the tail regime is in the $1.22 \le x_{min} \le 2.01$ standard deviations range, in which the number of data as a percentage of the corresponding (i.e. negative or positive) returns is in the range $0.05 \le n/m \le 0.14$.[16]

Since a lower exponent corresponds to a slower decay, the right tail of TRM appears to be fatter, which may be linked to an observation made by Rebonato (1999) and Derman (2008): volatility smiles tend to exhibit a pronounced asymmetry in the emerging markets' exchange rates case, where the higher volatility corresponds to the depreciation of the local currency.

---

[16] Christofferesen (2003) documents that using 5% of the data for samples around 1000 observations is a good rule of thumb for this type of estimations. Results in Table 3 suggest that this rule would be fair for the positive tail of TRM, in which the number of observations in each tail (n) is 5.34% of the data set. However, as expected from this type of rules, other tails are larger (i.e. IGBC positive tail, 6.9%) or smaller (i.e. TRM negative tail, 2.6%; S&P 500 negative tail, 2.8%; IGBC negative tail, 4.1%.

On the other hand, IGBC's negative tail having a lower power-law exponent may be linked to the well-known volatility smirk for stock markets, in which extreme negative price changes are overpriced with respect to extreme positive changes (Hull, 2003; Geman, 2005), in what has been called "crashophobia". Despite the difference between the power-law exponents in the IGBC case is small enough to be questionable, it agrees with findings by Gopikrishnan et al. (1998) and Eryigit et al. (2009).

4   Final remarks

Some findings are worth emphasizing. First, most of the empirical data analyzed here appears to follow Zipf's law ($\gamma \sim 1$) when implementing typical –yet questionable- OLS regression-based methods, in which standard goodness-of-fit statistics performed rather well ($r^2 \geq 0.62$). However, when implementing enhanced estimation methods Zipf's law resulted an invalid functional form; this is true even for the renowned and well-documented distribution of cities or municipalities.

These results support critiques regarding the flaws of OLS-based estimations of power-laws and their potential for misleading analysis. As Clauset et al. (2009) conclude: *the common practice of identifying and quantifying power-law distributions by the approximately straight-line behavior on a double logarithmic plot should not be trusted: it is a necessary but by no means a sufficient condition for true power-law behavior.*

Second, the enhanced estimation method and goodness-of-fit statistic designed by Clauset (2009) verify that the size of Colombian municipalities and financial institutions, the local exchange rate index (TRM), the local stock market index (IGBC) and the negative returns of the S&P 500 U.S. stock index approximate a power-law tail distribution. On the other hand, despite the evidence of fat positive and negative tails, a power-law tail distribution is not a plausible distribution for the local sovereign securities index (IDXTES) and the positive returns of S&P 500; other heavy tail distributions may provide a better fit.

Third, regarding TRM, IGBC and the negative returns of S&P 500, results concur with the stylized facts of financial returns reported by Cont (2001), in which the power-law tail exponents are in the $2 < \hat{\gamma} \leq 5$ range. Results coincide with those of Eryigit et al. (2009), who find that for a 202 stock market indexes the power-law exponents are close to 4; thus, results do not approximate the inverse cubic law suggested by Sinha et al. (2011), Gabaix et al. (2003a,b) and Gopikrishnan et al. (1998).

Fourth, based on the results, it seems reasonable to suggest that TRM, IGBC and the negative returns of S&P 500 may be described as following a *gradually truncated Lévy flight* (Gupta and Campanha, 1999).[17] This type of model combines a Lévy flight distribution model ($0 < \gamma \leq 2$)

---

[17] Using a Lévy stable distribution or *Lévy flight* alone is problematic since in most of the cases variance is infinite; the only case in which variance is finite is when $\gamma = 2$, corresponding to a Gaussian distribution. For instance, findings by Mandelbrot (1963) and Fama (1963 & 1965) converge to a Lévy stable distribution for asset prices

for the bulk of the distribution (e.g. $x < x_{min}$) and a gradual cut-off outside the Lévy flight; in our case, the gradual cut-off may be provided by the tail power-law estimated exponent, which is outside the Lévy flight (i.e. $\hat{\gamma} > 2$).

To the best knowledge of the author, fitting any sort of a gradually truncated Lévy flight has not been attempted for the Colombian indexes here considered. However, the shape of the double logarithmic plots in Figure 5, in which the bulk of the data ($x < x_{min}$) is mostly straight but noticeable flatter than the tail regime ($x \geq x_{min}$), suggests that if a power-law is in place its exponent would be higher than zero but lower than the herein estimated tail's exponent, possibly in the Lévy regime ($0 < \gamma \leq 2$). Verifying the fit of a gradually truncated Lévy flight to local financial time-series is a pending task.

Fifth, there is an obvious closeness of the contents and results of this research document with extreme value theory (EVT) basics. For some strange reason, in financial applications, Pareto-like heavy tail distributions are parameterized by $\xi = 1/\gamma$ which is called the shape parameter of the distribution (Carmona, 2014). Within the EVT framework, all attained exponents concur with the Fréchet distribution ($\hat{\xi} > 0$), with most of them resulting in the $0 < \hat{\xi} < 0.35$ range, which is the typical range of heavy-tailed financial returns (Dowd, 2005).[18]

Sixth, finding that a distribution is heavy-tailed and that the tail approximates a power-law should not be an end in itself; despite identifying the presence of fat tails is relevant for risk management, finding that empirical financial data fits a power-law is a first step towards the understanding of financial markets. Several authors, in different scientific realms, agree on power-law distributions being characteristic of *self-organizing systems* (Andriani and McKelvey, 2009; Strogatz, 2003; Barabási, 2003; Barabási and Albert, 1999; Bak, 1996; Krugman, 1996). Under the self-organized systems framework the assumptions of homogeneity, linearity and equilibrium are absent, and large fluctuations (e.g. outside the Gaussian) may occur from small frequent events, and not *only if many random events accidentally pull in the same direction, which is prohibitively unlikely* (Bak, 1996).

Several authors have already linked the divergence of financial returns and other economic data from Gaussian distributed returns to the self-organizing properties of financial markets and the economy as a whole. For instance, Marsili (2003) points out that the main stylized facts of financial fluctuations (i.e. fat tails, scaling, long-range volatility correlations) and their considerable deviation from Gaussian statistics provide empirical evidence of financial markets as complex self-organizing critical systems; this is, markets typically behave in a Gaussian manner, but when approaching a critical point or *phase transition* they behave according to the stylized facts. Likewise, Sinha et al. (2011) points out that the apparent

---

with $\gamma \sim 1.7$, which results in lack of convergence of distributions towards a Gaussian at longer time-scales. Mantegna and Stanley (1994) introduced the truncated Lévy flight, in which the Lévy flight is abruptly cut to zero at a certain critical threshold.

[18] The only time-series not complying with the $0 < \hat{\xi} < 0.35$ range is the negative returns of IDXTES. Besides, according to the goodness-of-fit tests here implemented, a power-law is not a plausible distribution for its tails.

universality behind power-law tails may indicate that different markets self-organize to an almost identical non-equilibrium steady state.

If financial markets display some sort of hierarchical architecture resulting from its self-organization[19], such architecture may help explain the behavior of prices as well. For instance, as suggested by Gabaix et al. (2003a), large movements in stock markets result from large transactions from large financial institutions, with the size of transactions and financial institutions following a power-law as well[20] (the size of Colombian financial institutions following a power-law is also verified in this document). In this sense, heterogeneity would be a key factor behind the behavior and dynamics of financial markets, against typical economic models based on the homogeneity of financial institutions and their linkages (as in Allen and Gale (2000) and Freixas et al. (2000)).

Under the same self-organization concept, Krugman (1996) describes the economy as a self-organizing system, in which economic cycles describe a *punctuated equilibrium*, with *long periods of relative quiescence divided by short periods of rapid change,* in which *sudden changes come when a previously state of equilibrium becomes unstable, setting the system adrift while it searches for a new equilibrium*. Within an economic framework, this could help to understand why some of the most extreme price changes (e.g. the Great Depression) occurred in the absence of any obvious cause, or why a modest linkage between two economies could be the conduit for a large cascade (Krugman, 1996).

Finally, some challenges arise from this document: (i) increasing the size of the data sets, with intraday data sets being a typical solution to this issue; (ii) fitting a gradual truncated Lévy flight model; (iii) testing the results for individual stocks and sovereign securities; (iv) testing scale-free nature of financial markets' transactions volumes; (v) further interpreting the power-law exponent, which is a long-lived issue.

---

[19] As recently suggested by León and Berndsen (2013) for the Colombian case.
[20] Gabaix et al. (2003) reveals some regularities in financial fluctuations: the cubic law of returns, the half cubic law of volumes and the approximate cubic law of number of trades. They link these regularities to economic optimization by heterogeneous agents.

5   References

Allen, F.; Gale, D., "Financial contagion", *Journal of Political Economy*, Vol.108, No.1, 2000.

Andriani, P.; McKelvey, B., "From Gaussian to Paretian thinking: causes and implications of power laws in organizations", *Organization Science*, No.6, Vol.20, 2009.

Bak, P., *How Nature Works*, Copernicus, 1996.

Barabási, A.-L., *Linked*, Plume, 2003.

Barabási, A.-L.; Albert, R., "Emergence of scaling in random networks", *Science*, Vol.286, October, 1999.

Carmona, R., *Statistical Analysis of Financial Data in R*, Springer, 2014.

Christoffersen, P.F., *Elements of Financial Risk Management*, Academic Press, 2003.

Clauset, A.; Shalizi, C.R.; Newman, M.E.J., "Power-law distributions in empirical data", *SIAM Review*, No.4, Vol.51, 2009.

Cont, R., "Empirical properties of asset returns: stylized facts and statistical issues", *Quantitative Finance*, Vol1, 2001.

Derman, E., 2008, Laughter in the dark: an introduction to the volatility smile – lecture notes, <http://www.ederman.com/new/docs/laughter.html>.

Dowd, K., *Measuring Market Risk*, John Wiley & Sons, 2005.

Eryigit, M.; Cukur, S.; Eryigit, R., "Tail distribution of index fluctuations in World markets", *Physica A*, No.388, 2009.

Fama, E.F, "Mandelbrot and the stable Paretian hypothesis", *The Journal of Business*, Vol.36, No.4, 1963.

Fama, E.F., "The behavior of stock-market prices", *The Journal of Business*, Vol.38, No.1, 1965.

Fiaschi, D.; Kondor, I.; Marisli, M., "The interrupted power law and the size of shadow banking", *mimeo*, 2013.

Freixas, X.; Parigi, B.M.; Rochet J-C., "Systemic risk, interbank relations, and liquidity provision by the central bank", *Journal of Money, Credit and Banking*, Vol. 32, No. 3, 2000.

Gabaix, X.; Gopikrishnan, P.; Plerou, V.; Stanley, H.E., "A theory of power-law distributions in financial market fluctuations", *Nature*, Vol.423, 15 May, 2003a.

Gabaix, X.; Gopikrishnan, P.; Plerou, V.; Stanley, H.E., "Understanding the cubic and half-cubic laws of financial fluctuations", *Physica* A, No.324, 2003b.

Geman, H., *Commodities and commodity derivatives*, John Wiley & Sons, 2005.

Gopikrishnan, P.; Meyer, M.; Nunes, L.A.; Stanley. E., "Inverse Cubic Law for the Distribution of Stock Price Variations", *European Physics Journal B*, No.3, 1998.

Gupta, H.M.; Campanha, J.R., "The gradually truncated Lévy flight for systems with power-law distributions", Physica A, No.268, 1999.

Hull, J., *Options futures and other derivatives*, Prentice Hall, 2003.

Krugman, P., *Self-Organizing Economy*, Blackwell, 1996.

Lanaspa, L.; Perdiguero, A. M.; Sanz, F., "La distribución del tamaño de las ciudades en España, 1900-1999", *Revista de Economía Aplicada*, No.34, Vol. 12, 2004.

León, C.; Berndsen, R., "Modular scale-free architecture of Colombian financial networks: Evidence and challenges with financial stability in view", *Borradores de Economía*, No.799, Banco de la República, 2013.

Mandelbrot, B., "The variation of certain speculative prices", *The Journal of Business*, No.4, 1963.

Mandelbrot, B.; Hudson, R., *The (Mis)Behavior of Markets*, Basic Books, 2004.

Mantegna, R.N.; Stanley, H.E., "Stochastic process with ultraslow convergence to a Gaussian: the truncated Lévy flight", *Physical Review Letters*, Vol.73, No.22, 1994.

Marsili, M., "Scale invariance and criticality in financial markets", *Physica A*, No.324, 2003.

Mitchell, W.C., "The making and using of index numbers", *Introduction to index numbers and wholesale prices in the United States and foreign countries*, Bulletin No. 173, U.S. Bureau of Labor Statistics, 1915.

Rebonato, R., *Volatility and correlation: In the pricing of equity, FX and interest-rate options*, John Wiley & Sons, 1999.

Reveiz, A.; León, C., "Índice Representativo del Mercado de Deuda Pública Interna: IDXTES", *Pensiones y Portafolio: La Construcción de una Política Pública (Eds. Laserna, J.M. y Gómez, M.C.)*, Banco de la República y Universidad Externado de Colombia, 2010.

Sánchez, F.; España, I., "Urbanización, desarrollo económico y pobreza en el sistema de ciudades colombianas 1951-2005", *Documentos CEDE*, Universidad de los Andes, No.13, July, 2012.

Simon, H.A., "On a class of skew distribution functions", *Biometrika*, Vol.42, No.3/4, 1955.

Sinha, S.; Charrerjee, A.; Chakraborti, A.; Chakrabarti, B.K., *Econophysics: an Introduction*, Wiley-VCH, 2011.

Stanley, H.E.; Amaral, L.A.N.; Buldyrev, S.V.; Gopikrishnan; Plerou, V.; Salinger, M.A., "Self-organized complexity in economics and finance", *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*, Vol.99, Suppl.1, Feb. 19, 2002.

Strogatz, S., SYNC: *How Order Emerges from Chaos in the Universe, Nature and Daily Life*, Hyperion Books, 2003.

Stumpf, M.P.H.; Porter, M. A., "Critical truths about power laws", *Science,* Vol.335, February 10, 2012.