Physica A xx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Non-stationary individual and household income of poor, rich and middle classes in Mexico

P. Soriano-Hernández [a], M. del Castillo-Mussot [a,*], O. Córdoba-Rodríguez [b], R. Mansilla-Corona [c]

[a] Instituto de Física, Universidad Nacional Autónoma de México, 04510 Mexico City, Mexico

[b] Facultad de Economía, Universidad Nacional Autónoma de México, 04510 Mexico City, Mexico

[c] Centro de Investigaciones Interdisciplinarias en Ciencias y Humanidades, Universidad Nacional Autónoma de México, 04510 Mexico City, Mexico

HIGHLIGHTS

- Economic crisis on household income distributions in Mexico for years 1992–2008.
- Definition of three-class structure; poor, rich (Pareto) and middle classes.
- Adjusting income distributions to exponential, Log-normal or Gamma functions.
- Not very visible poor households or agents due to very low cut-off values.

ARTICLE INFO

Article history:
Received 23 February 2016
Received in revised form 2 July 2016
Available online xxxx

Keywords:
Income distribution
Pareto
Econophysics

ABSTRACT

Despite Mexican peso crisis in 1994 followed by a severe economic recession, individual and household income distributions in the period 1992–2008 always exhibit a two-class structure; a highly fluctuating high-income class adjusted to a Pareto power-law distribution, and a low-income class (including poor and middle classes) adjusted to either Log-normal or Gamma distributions, where poor agents are defined as those with income below the maximum of the uni-modal distribution. Then the effects of crisis on the income distributions of the three classes are briefly analysed.

1. Introduction

Inequality in social systems has been a universal and robust phenomenon, not bound by either time or geography, but fortunately for scholars, it has a few statistical regularities, as in the case of income and wealth distributions over a wide range of societies and time periods [1–5]. For example, Pareto [6] made extensive studies in Europe and found that wealth distribution follows a power-law tail for the richer sections of society. In general, the upper end (the richer) of the income and wealth distributions is believed to be described by a power-law, as Pareto [6] argued over 100 years ago. Years later, Gibrat [7] found that not all income intervals follow a Pareto distribution, so he proposed a multiplicative stochastic process where the proportional rate of growth of a firm is independent of its absolute size, yielding a log-normal probability distribution function (PDF). Silva and Yakovenko [8] found that the data analysis of income distribution in the USA reveals coexistence

* Corresponding author.
E-mail addresses: pavelsoriano@fisica.unam.mx (P. Soriano-Hernández), mussot@fisica.unam.mx (M. del Castillo-Mussot), ocr@ciencias.unam.mx (O. Córdoba-Rodríguez), mansy@unam.mx (R. Mansilla-Corona).

http://dx.doi.org/10.1016/j.physa.2016.08.042
0378-4371/

Please cite this article in press as: P. Soriano-Hernández, et al., Non-stationary individual and household income of poor, rich and middle classes in Mexico, Physica A (2016), http://dx.doi.org/10.1016/j.physa.2016.08.042

of two social classes; the large lower class is characterized by the exponential Boltzmann–Gibbs (B–G) distribution, and the very small upper class exhibits the power-law Pareto distribution with characteristic fat tails. They found that the exponential distribution in the lower class is very stable in time for all years, since they collapse to a single curve after being adjusted to inflation, whereas the power-law distribution of the upper class is volatile. They claimed that the local conservation of money (energy) leads to the robust exponential (B–G) distribution in the large USA lower class. Although this conservation is just an approximation, the B–G analogy works very well in the low-income region, since all the yearly distributions collapse to a single curve after being adjusted to inflation. Since the (B–G) analogy allows transfers of all energy Q3 (or money) from an agent to another, then Chakrabarti and Chakraborti [9] extended the model to account for restrictions on transferred energy (savings), because we know that in real world it is unlikely that agents put into play or risk most of their money on a single economic negotiation, unlike what happens in collisions between particles when a particle can elastically yield all its kinetic energy to another. Within this context the Gamma function is a good fit for a model of constant saving factor fraction [9,10]. Incidentally, the Gamma function also appears in a utility maximization model [10] within a standard exchange-model employing Cobb–Douglas utility function. In summary as mentioned in Ref. [9], in general, the bulk of the low range of the distribution of both income and wealth in many societies seems to be well fitted by both log-normal and gamma distributions, for which the exponential term dominates for large values. Economists usually prefer the log-normal distribution [11], whereas statisticians and, more recently, physicists [2,3,9,12] tend to rely more on alternate forms such as the Gamma distribution for the probability density (PDF) or the Gibbs/exponential distribution for the complementary cumulative distribution (CCDF) [9]. An important recent study on US income distribution [13] claimed that the B–G and Pareto mixtures fit can be improved upon further by adding a Log-normal component.

On the other hand it is important to investigate less stable emergent economies, especially in times of economic crisis. In Mexico there was a strong depreciation of the currency at the end of 1994, which was followed by severe economic recession. Here, we analyse statistical information on income distribution of population provided by the national survey ENIGH of the national institute INEGI for the period 1992–2008. The layout of this work is the following. Section 2 is devoted to theory and empirical evidence of exponential-like (Gamma and Log-normal) and Pareto income distributions. In Section 3, for all years we estimate the demarcation income value for the exponential-like and Pareto mixture's fit distributions, and then we present the adjustments employing individual and household data. If we define poor (middle-class) agents as those with income below (above) the maximum of the uni-modal distribution, then the middle class population lies between that maximum and the Pareto rich class. Section 4 is devoted to conclusions.

2. Pareto and exponential-like (Log-normal and Gamma) distributions

Empirical income distribution in countries such as US [8], European Union [14] UK [15] and Canada [16], among others, presents a two-class structure, in the sense that most of the population of these countries or regions belong to a low class characterized by a distribution similar to an exponential function, while the highest follows a Pareto distribution (as was found by Pareto many years ago for a number countries). As mentioned in Ref. [9] in general, the low-income bulk of the distribution of both income and wealth seems to be fitted by either Log-normal or Gamma distributions. Similar results were found in US subgroups categorized by gender and race (Whites and African-Americans), confirming for the subgroups the same qualitative two-class income distribution [17]. In all these cases, the upper Pareto tail changes much more in time than the lower regions. Creation and destruction of money in complex processes (through investments, credit, financial derivatives, big stock market crisis, etcetera) are much more clearly related to the Pareto tail. For reviews on income and wealth distributions, see Refs. [3,9]. Let us review in some detail a particular case, the analysis income distribution in USA in the period 1983–2001 by Silva and Yakovenko [8]. The exponential distribution in the lower class was modelled by analogy to a thermodynamic phenomenon, by considering that conservation of money in economic interactions is similar to the conservation of kinetic energy in elastic collisions representing trade for goods and services. Since the exponential distribution in the lower class was shown to be very stable in time after inflation adjustment, whereas the power-law distribution of the upper class is highly dynamic and volatile, it was concluded that the lower class is in thermal equilibrium, and the upper class is out of it [8]. From the time evolution of the integrated wealth of people in the upper income Pareto tail in the period 1983 and 2001, they observed that this total Pareto wealth tail is more or less in phase with the stock market index S&P 500 divided by inflation. The S&P 500 is a capitalization-weighted index published by Standard & Poor's of the prices of 500 large market capitalization common stocks actively traded in the United States. That is, the Pareto wealth tail is correlated with the rise and fall of the stock market since it swells and shrinks following the stock market. This empirical fact shows that some of the mechanisms to make money followed by richest people are different from the majority of the population, which mostly exchange goods and services.

2.1. Power law or Pareto distributions

More than a century ago, the economist Vilfredo Pareto (1848–1923) found that wealth distribution follows a power law tail for the richer sections of society [6] following a probability distribution form, known now as the Pareto law [9]:

$$P(m) \propto m^{-\alpha} \tag{1}$$

where m is money and α is a constant parameter of the distribution known as the Pareto exponent or scaling parameter, found to vary slightly for different economies between 2 and 3 [8,18], since typically less than 10% of the population of any country possesses around 40% of the wealth and follows the above power-law [9]. In practice, few empirical phenomena obey in general power-laws for all values of m. More often, in many distributions for different fields of knowledge, the power-law applies only for values greater than some critical values [19]. In these cases, it is said that the tail of the distribution follows a power-law. For a tutorial on different ways of plotting related mathematical descriptions (Pareto, power and Zipf' laws), see Ref. [20]. It is remarkable that the Pareto distribution shows up in the tail of wide arrays of complex systems. Certain systems develop power-law distributions at special 'critical' points in their parameter space because of the divergence of some characteristic scale. A mechanism for generating power-laws is that of critical phenomena [21]. For more information about these systems, see Section 3.5 of Ref. [21]. Usually the circumstances under which the divergence takes place are very specific ones. The parameters of the system have to be tuned very precisely to produce the power-law behaviour [21]. The precise point at which the length-scale in a system diverges is called a critical point or a phase transition. Processes that happen in the vicinity of continuous phase transitions are known as critical phenomena, of which power-law distributions are one example. As first proposed by Per Bak [22] et al., it is possible that some dynamical systems actually arrange themselves so that they always reach to critical points, regardless of initial states (self-organized systems).

2.2. Log-normal distribution

In 1931, Robert Gibrat [15] clarified that Pareto's law is valid only for the high-income range, whereas for the small and middle-income ranges he suggested that the income distribution is described by a log-normal probability density and proposed a law of proportionate effect, which states that a small change in a quantity is independent of the quantity itself. Thus the distribution of a quantity $dz = dm/m$ should be Gaussian, and hence, x is log-normal, yielding Gibrat's law:

$$P(m) \propto \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\log^2(m/m_o)}{2\sigma^2}\right) \tag{2}$$

where $\log(m_0) = \langle \log(m) \rangle = \mu$ is the mean value of the logarithmic variable and $\sigma^2 = \langle [\log(m) - \log(m_0)]^2 \rangle$ is the corresponding variance. The factor $\beta = 1/\sqrt{2\sigma^2}$, also known as the Gibrat index, measures the equality of the distribution and empirically, is known to lie between 2 and 3 [9].

2.3. Gamma distribution

The main argument in favour of the kinetic exchange models, according to Refs. [2,9], is that millions of small transactions that take place in a very short span of time can generate the essential stochastic features of the kinetic exchange models and the corresponding distributions. Thus a particle gas interpretation of economic interactions, motivated by local conservation of money or energy and random collisions (or interactions) between economic agents, allows the use of statistical description used in physics if money works as a proxy for income/wealth. Since the distributions derived for money compare extremely well with the empirical data of income/wealth, these models provide important insights for such distributions. Gallegati [5] noted that "in industrialized capitalist economies, income is most definitely not conserved", which is certainly true. But while income and wealth in an economy grows over time, their growth can be studied as a time-series phenomenon with data taken at a single instance or within a very short period of time in order to get statistical averages over longer periods of time. The proposal of Dragulescu and Yakovenko [15] to study a conservative economic system as a particle gas has the advantage that the resulting distribution, a B–G energy distribution is very robust under perturbations, because the exponential distribution is the one that maximizes the "economic entropy".

However, in any trading, savings come naturally [23]. A saving propensity factor λ was introduced in the random exchange model by Chakraborti and Chakrabarti [9], in which each trader at time t saves a fraction λ of its money $m_i(t)$ and trades randomly with another:

$$\begin{aligned} m_i(t+1) &= \lambda m_i(t) + \epsilon_{ij}[(1-\lambda)(m_i(t)+m_j(t))] \\ m_j(t+1) &= \lambda m_j(t) + (1-\epsilon_{ij})[(1-\lambda)(m_i(t)+m_j(t))] \end{aligned} \tag{3}$$

where

$$\Delta m = (1-\lambda)[\epsilon_{ij}\{m_i(t)+m_j(t)\} - m_i(t)]. \tag{4}$$

Here ϵ_{ij} is a random fraction. Although there is no analytic expression in the case of fixed constant λ, a gamma distribution has been used as a good approximation of the resulting distribution [9]:

$$\begin{aligned} f_n(m) &= a_n m^{n-1} \exp\left(-\frac{nm}{\langle m \rangle}\right) \\ &\frac{1}{\Gamma(n)}\left(\frac{n}{\langle m \rangle}\right)^n \end{aligned} \tag{5}$$

where n is defined in terms of saving factor λ as:

$$n(\lambda) = 1 + \frac{3\lambda}{1-\lambda}. \tag{6}$$

The exponential Boltzmann–Gibbs distribution (Dragulescu and Yakovenko model, Ref. [3]),

$$P(m) = Ce^{-m/S} \tag{7}$$

is the special case for $n = 1$, i.e., $\lambda = 0$ or not saving at all (elastic collision).

3. Fitting income distributions in Mexico

In Mexico, the National Institute INEGI periodically presents results of the National Survey of Income and Household Expenditure Survey [24]. The data source with which we worked is the current per capita quarterly income obtained from the National Survey of Income and Expenditure Survey (ENIGH) that raises the INEGI biennially. For instance, for the third quarter of 2004 the design of this survey nationwide coverage allowed the breakdown of the information by location of 2500 or more inhabitants and less than 2500 inhabitants and for five strata according to their level of marginalization, where the sample size was 25,115 households [24]. The present study was realized over the years 1992, 1994, 1996, 1998, 2000, 2002, 2004 and 2005, 2006 and 2008.

3.1. Estimation of low-high class demarcation parameter and Pareto exponent

To determine the minimum value of income from which there is power-law behaviour, the Pareto exponent and the "goodness of fit test" by power-law distribution, a recent programming code [19] was used. This function implements both the discrete and continuous maximum likelihood estimators for fitting the power-law distribution to data, along with the goodness-of-fit based approach to estimating the lower cut-off for the scaling region. The fundamental idea behind this method is simple and involves choosing the value of $m_{\min}$ that makes the probability distributions of the measured data and the best-fit power-law model as similar as possible above this value. There are a variety of measures for quantifying the distance between two probability distributions, but for non-normal data the commonest and hence, the one used in this code, is the Kolmogorov–Smirnov or KS statistics [19,21], which is simply the maximum distance between the cumulative distribution functions (CDF's) of the data and the fitted model:

$$D = \max|O(x) - P(x)| \tag{8}$$

where $O(x)$ is the CDF of the data for the observations with value at least $m_{\min}$, and $P(x)$ is the CDF for the power-law model that best fits the data in the region. Our minimum income estimate is then the value of $m_{\min}$ that minimizes D. Moreover, find the exponent α correctly requires a value for the lower bound $m_{\min}$ of power-law behaviour in the data. Assuming that our data are drawn from a distribution that follows a power law exactly for $m \geq m_{\min}$, the program derives the maximum likelihood estimators (MLE's) of the scaling parameter for both the discrete and continuous cases. Details of the derivations are given in Appendix B of Ref. [21]. The MLE for the continuous case is:

$$\alpha = 1 + n\left[\sum_{i=1}^{n} \ln \frac{m_i}{m_{\min}}\right]^{-1} \tag{9}$$

where $m_i, i = 1 \ldots n$ are the observed values of m such that $m_i \geq m_{\min}$.

3.2. Household cumulative income distributions

Thanks to the accessibility of the ENIGH database, we found that it is possible to determine which individual income belongs to which family since each data of individual income includes a folio number that determines the family to which the agent belongs. We observed families that depend on up to 36 individual incomes. For different years, Eq. (8) is employed to find the separation parameter $m_{\min}$ and Eq. (9) for the Pareto exponent α, which are shown in Table 2, together with other parameters such as an update inflation factor, which sets the family income to the value of Mexican peso in 2012. This factor is obtained by dividing the recent national consumer price index (INPC for its acronym in Spanish) by the INPC for the respective year. The fraction of high-income (rich) households and the respective percentage of total income is included there. The fitting of high-income class data [12,14] to Eq. (1) yields an average correlation of $R^2 = 0.9775$. It is important to Q4 notice that fluctuations in the Pareto parameters are much more pronounced than in those characterizing the low-income zone (see Table 1).

From correlation coefficients in Table 2, the gamma distribution Eq. (5) is the one with slightly better fit to the empirical data provided by the ENIGH survey, except for 1996, 2000 and 2002 where the Log-normal distribution gives a higher correlation. The saving factor varies from year to year, but not drastically, to yield an average saving from 1992 to 2008 of %14.42 of family income, which in the model is constant for all households. In general, it is useful to use the complementary

Table 1
Parameters obtained for the high-income interval of the households distribution.

Year	Updated inflation	α	$m_{\min}$ ($)	% of "rich" households	% of total income	R^2
1992	6.29	2.653	79,977.35	8.48	44.17	0.9880
1994	5.44	2.999	81,542.88	6.68	33.5	0.9779
1996	2.80	3.430	124,152.00	1.45	14.62	0.9736
1998	2.04	3.029	57,501.48	8.48	37.36	0.9730
2000	1.67	3.122	86,422.50	5.38	29.34	0.9736
2002	1.51	3.701	143,200.85	1.55	12.20	0.9772
2004	1.38	3.274	115,023.00	4.20	24.63	0.9783
2005	1.337	3.118	72,304.75	9.70	38.9	0.9694
2006	1.285	3.325	107,473.39	4.71	24.9	0.9762
2008	1.1627	3.062	84,370.56	8.84	37.4	0.9878

Table 2
Parameters obtained for the low-income range households distributions (Gamma and Log-normal).

Year	Gamma				Log-normal		
	$\langle m \rangle$	n	λ	R^2	σ (log)	μ (log)	R^2
1992	22,503.12	1.60	0.17	0.9993	0.8687	9.711	0.9981
1994	22,275.19	1.60	0.17	0.9994	0.8753	9.694	0.9979
1996	19,746.24	1.20	0.06	0.9956	0.9100	9.510	0.9996
1998	17,136.02	1.57	0.16	0.9997	0.8658	9.443	0.9976
2000	21,673.52	1.48	0.14	0.9988	0.8493	9.673	0.9991
2002	25,644.82	1.33	0.09	0.9978	0.9190	9.778	0.9989
2004	28,118.81	1.47	0.13	0.9992	0.9150	9.897	0.9976
2005	23,148.61	1.74	0.20	0.9998	0.8817	9.743	0.9955
2006	27,587.63	1.49	0.14	0.9994	0.9040	9.888	0.9977
2008	26,555.85	1.64	0.18	0.9998	0.8657	9.882	0.9964

cumulative distribution (also known as survival function, CCDF). CCDF is defined as $C(m) = \int_m^\infty P(m')\mathrm{d}m' = 1 - CDF$, where m is money and P is the probability distribution function (PDF). A useful property is that the CCDF of a power law PDF is also a power law, but with a different exponent. Thus on vertical and horizontal logarithmic scales we get for both a straight line, but with a shallower (or less negative) slope. Similarly, the CCDF of an exponential PDF is also an exponential with the same slope. The CCDF's for data from 1992 to 2008 with inflation adjustment are shown in log–log scale in Fig. 1, showing both low and high income regions. The values of low-income slopes S (which are good approximations for the average quarterly income per household reported in the second column of Table 2) are obtained as the slopes of the data adjusted to a simple B–G exponential in log-linear scales. Notice how after adjusting to inflation all curves collapse in the low-income zone with small changes in average slope (except from 1994 to 1996 and 2002 to 2004), denoting more stability than in the high-income region.

Fig. 2 shows a similar structure for the household income distribution for year 2008 with at least 6 members up to 36 who contribute to the total income. Q5

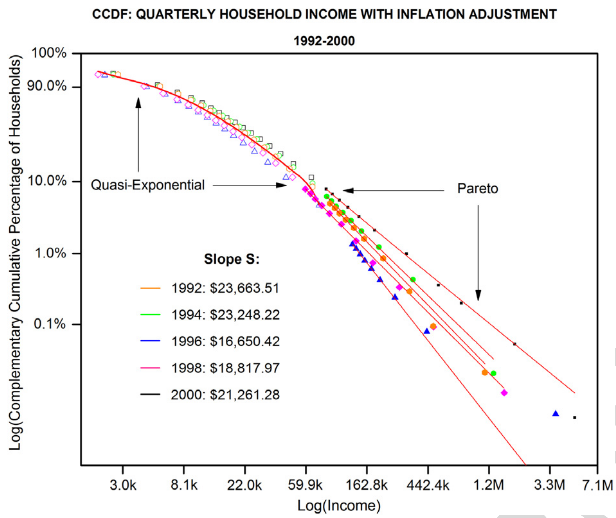
Fig. 3(a) shows changes in the annual average slope S in low-income region and Mexican gross domestic product (GDP) per capita adjusted to the value of Mexican Peso in 2015. Since S, (which is directly related to the income of the majority of households) has increased less than the GDP per person in the last years (2004, 2006 and 2008), then it seems that the growing GDP of the Mexican economy in those years did not reflect as a better income for the majority of its population. Fig. 3(b) signals agent transfers from high-income to low-income class when both α and $m_{\min}$ grow together due to high correlation between them. An important plot is given in 3(c) which yields the parity or number of pesos per US dollar (Nominal Exchange Rate). The Mexican peso crisis (also known as the Tequila crisis or December mistake) was a currency crisis sparked by the Mexican government's sudden devaluation of the peso against the US dollar in December 1994, followed by a severe recession in the Mexican economy. Q6

3.3. *Individual cumulative income distributions*

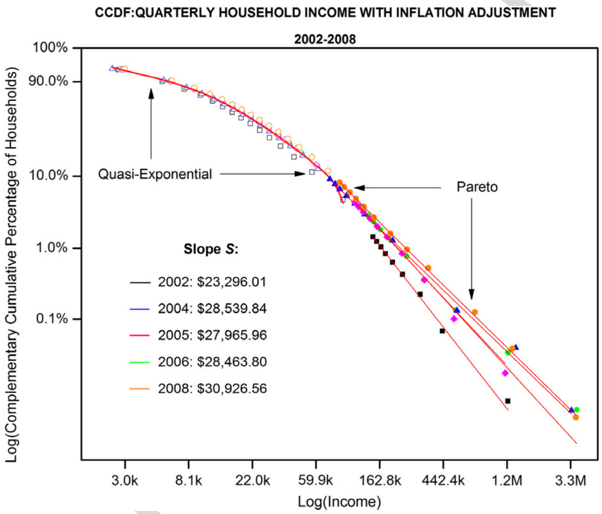
Following the same methodology as before but now for individual income with the same inflation adjustment, we show in Fig. 4 for simplicity only the results of the individual cumulative income distribution for quarterly individual income for year 2000, since the other plots for the remaining years are similar.

Fig. 4 exhibits the same overall behaviour as in Figs. 1–2. Similarly, in the years between 1992–2005 a similar collapse of all quasi-exponential curves is observed with inflation adjustment. The characteristic parameters of the distributions for different years are shown in Table 3. The average $R^2 = 0.9846$ for all years in the low-income area was best fitted by Eq. (7), i.e. which means there is no saving propensity; the average R^2 using Eq. (2) is 0.9849, while the average R^2 for high-income area (adjusted by Eq. (1)), was 0.9548.

Please cite this article in press as: P. Soriano-Hernández, et al., Non-stationary individual and household income of poor, rich and middle classes in Mexico, Physica A (2016), http://dx.doi.org/10.1016/j.physa.2016.08.042



(a) 1992–2000.



(b) 2002–2008.

Fig. 1. CCDF of quarterly household income, plotted in log–log scale for different years considering inflation adjustment, set to the value of Mexican Peso in 2012. The two-class structure is shown, together with the values of the slopes S in the low-income region.

Table 3
Parameters obtained for quarterly income per person with inflation adjustment of Table 1.

Year	α	$m_{\min}$ ($)	S	R^2 (Pareto)	R^2 (Exponential)	R^2 (Log-normal)
1992	2.664	24,342.3	5802.26	0.9639	0.9894	0.9824
1994	2.871	35,904.0	8004.96	0.9566	0.9946	0.9801
1996	3.082	54,600.0	5951.27	0.9587	0.9908	0.9889
1998	3.031	62,118.0	5541.67	0.9668	0.9849	0.9881
2000	2.863	34,141.5	5997.81	0.9613	0.9855	0.9820
2002	2.837	16,308.0	3904.70	0.9476	0.9762	0.9906
2004	3.051	57,657.8	6992.63	0.9401	0.9792	0.9835
2005	3.162	54,929.5	6436.91	0.9435	0.9759	0.9842

3.4. Poor and middle classes in the low-income interval

It is challenging to investigate further class subdivisions, particularly in the low-income zone, which could be small compared to the full income interval. Then the effects of a crisis could be investigated in more detail. After calculating in Section 3.1 the demarcation parameters $m_{\min}$ between low-income and high-income classes, let us divide the low-income interval into two parts. The first one, labelled as the "poor class", ranges from zero income to the peak of the uni modal PDF,

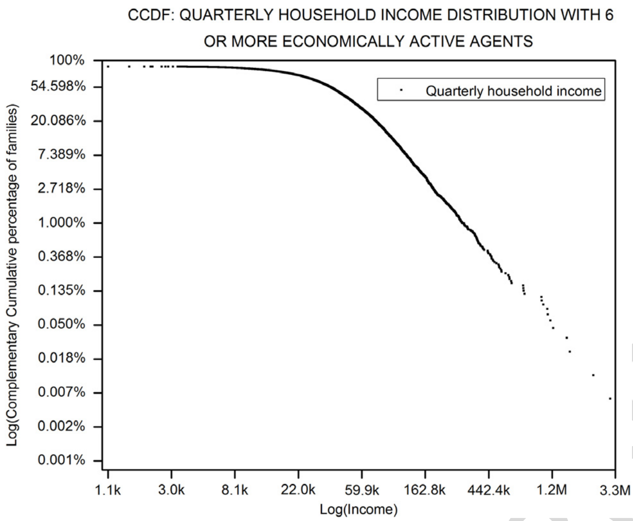
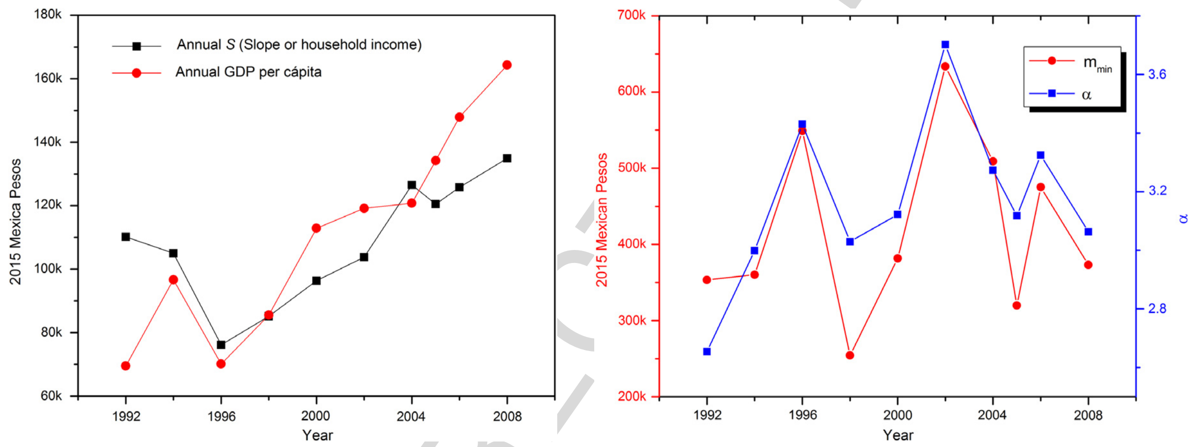
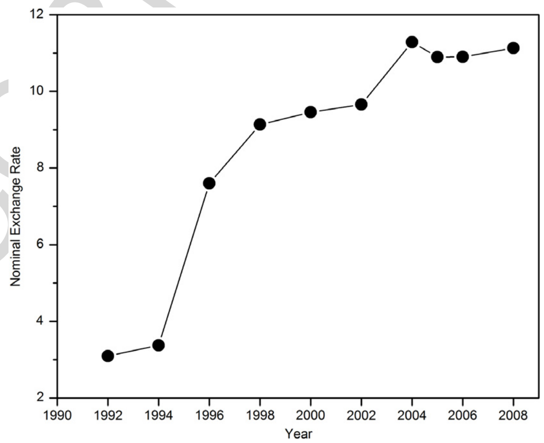


Fig. 2. Same as Fig. 1 for year 2008, but with households including at least 6 economically members up to 36.



(a) Average slope or annual family income and GDP per capita.

(b) α and $m_{\min}$.



(c) Nominal exchange rate.

Fig. 3. Temporal evolution of important parameters: (a) Average slope or quarterly family income (red curve) and GDP per capita (black curve), (b) α (blue curve) and $m_{\min}$ (red curve), in the 1992–2008 period, assuming that quarterly income is preserved for the rest of the year, set to the value of Mexican peso in 2015 and (c) Nominal Exchange rate from 1992 to 2008. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

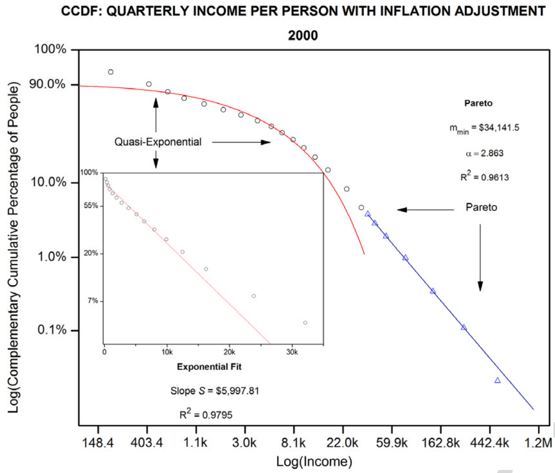


Fig. 4. Same as Fig. 1, but now for individual quarterly income for year 2000 considering inflation adjustment. S is adjusted to a semi-log exponential plot in the inset.

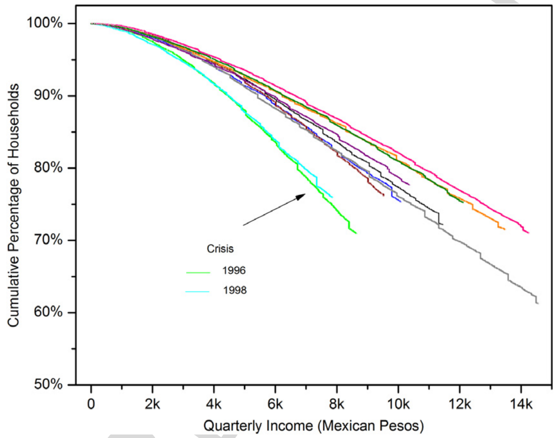


Fig. 5. CCDF of poorest class from a zero income value up to PDF peak values.

Table 4
Parameters of household distributions for different periods, adjusted by inflation.

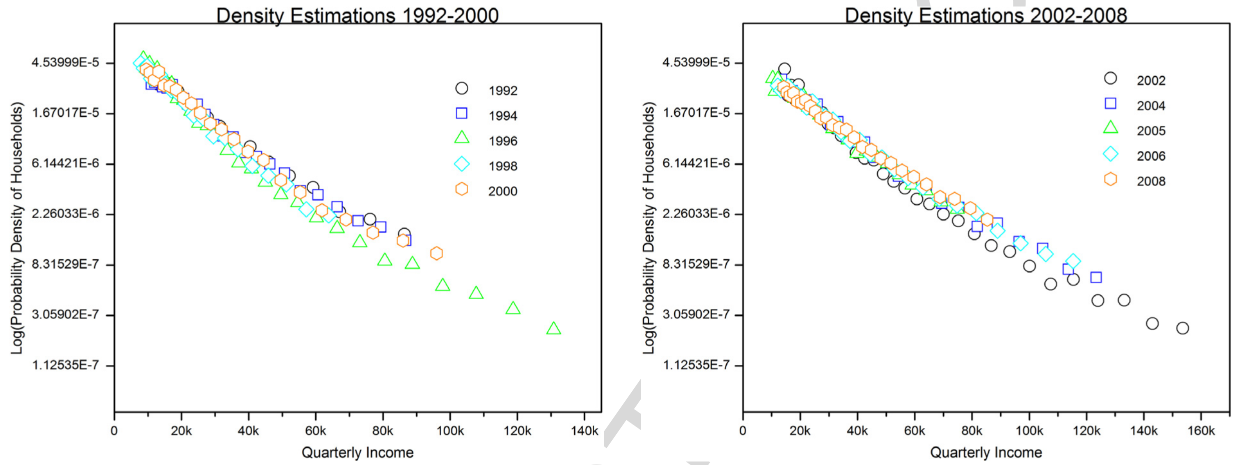
Period	$\langle S \rangle$	m_{peak}	$\langle m_{\min} \rangle$
Pre-crisis (1992–1994)	20,515.50	10,779.67	80,760.12
Crisis (1996–1998)	15,803.05	8,249.36	90,826.74
Post-crisis (2000–2008)	22,732.10	12,348.60	101,467.96

denoted as m_{peak}, and the second, labelled as "middle class", from this peak value to $m_{\min}$. Notice that in the case the PDF takes its maximum at value of zero income, then this definition does not make sense because the fraction of poor households is also zero. Although these definitions are somehow arbitrary, these class definitions are useful to further analyse their time evolution, particularly in epochs of economic crisis, as reflected by the parity or the Mexican peso (see Fig. 3(c)). Fig. 5 shows the poor-class quarterly income CCDF curves plotted up to the PDF peak value for different years. The curves labelled as "crisis" after devaluation of the Mexican currency (years 1996 and 1998) fall faster than in other ("pre-crisis" and "post-crisis") years. Also, in the period 1996–1998 the values of S and m_{peak} are the lowest ones, as shown in 4.

The middle class curves plotted from m_p to $m_{\min}$ are surprisingly well fitted by simple exponential curves $P(m) = Ce^{-m/S}$, as shown in Table 5, and illustrated for all years in Fig. 6 via density estimations (a histogram is constructed and the frequency of each partition is divided by the total number of data, and then divided by the width of the bin). The parameter S and the average income per household $\langle m \rangle$ are also shown in such table. Then, the middle class was also adjusted by Gamma distribution (average correlation $R^2 = 0.9832$) and Log-normal ($R^2 = 0.9895$).

Table 5
Parameters and correlations obtained for quarterly income per household with inflation adjustment obtained via density estimators set to simple exponential function $Ce^{-m/S}$.

Year	S	$\langle m \rangle$	R^2
1992	20,459.80	29,297.37	0.9865
1994	20,571.20	28,089.70	0.9759
1996	14,707.40	25,776.32	0.9901
1998	16,898.70	21,544.76	0.9868
2000	19,111.40	26,953.60	0.9877
2002	18,378.60	36,828.13	0.9559
2004	23,800.00	36,692.07	0.9856
2005	24,146.10	28,748.31	0.9728
2006	25,406.90	34,773.12	0.9865
2008	25,549.60	35,054.22	0.9919



(a) 1992–2000. (b) 2002–2008.

Fig. 6. Graphs of density estimators constructed from histograms for years (a) 1990–2000 and (b) 2002–2008 plotted in log-linear scale.

3.5. Modelling household income distribution with two providers from individual incomes

Notice from Table 3 that the best fits to the individual low-income region are obtained with a pure exponential distribution, whereas from Table 2 the best fits to the households with several members are obtained via Gamma distribution Eq. (5). This qualitative difference between household and individual income distributions was already discussed in Sec. III of Ref. [25], as well as in Ref. [3].

To construct feasible household income distributions with more than one member from the added individual incomes, it would be desirable to have information on how they group together in a single household. The simplest grouping model is to assume that they do it in an uncorrelated random way. In this case, for the low-income region when the individual distributions are exponential functions $Ce^{-m/S}$, in Refs. [3,25], then the simple analytical expression $\int_r^\infty Cr' e^{-Sr'} dr'$ was obtained for two-provider households as a convolution, where $r = m_1 + m_2$ is the added income. This procedure is labelled as analytical model (AM), since the corresponding CCDF is also analytical. In the general case of any distribution functions, one can employ the same idea in a numerical simulation selecting randomly two incomes, labelled as random model (RM).

For instance, for the low-income region, for year 2008 in Fig. 7 we compare the empirical two-provider household income CCDF obtained from the two models, AM and RM, of uncorrelated individual agents. The correlations between the two-income empirical or observed distribution with each of the two distributions constructed from the individual distributions are not very different, since $R^2 = 0.9889$ in the case of RM, and $R^2 = 0.9928$ for the AM.

4. Conclusions

From individual and household data in Mexico every two years in the period 1992–2008, the distributions always exhibited a high-income class Pareto power-law distribution and a lower income class characterized by a quasi-exponential distribution, as also observed in many countries [8,9,18], despite strong changes in the Mexican economy that started in 1994. If inflation is taken into account by multiplying the Mexican peso by the corresponding adjustment factor in each year, then the slope of the middle-income class distributions gets closer, while the parameters of the high-income class Pareto remain strongly fluctuating over time, specially in the 1990s decade. In all years the low-income curves were successfully

Please cite this article in press as: P. Soriano-Hernández, et al., Non-stationary individual and household income of poor, rich and middle classes in Mexico, Physica A (2016), http://dx.doi.org/10.1016/j.physa.2016.08.042

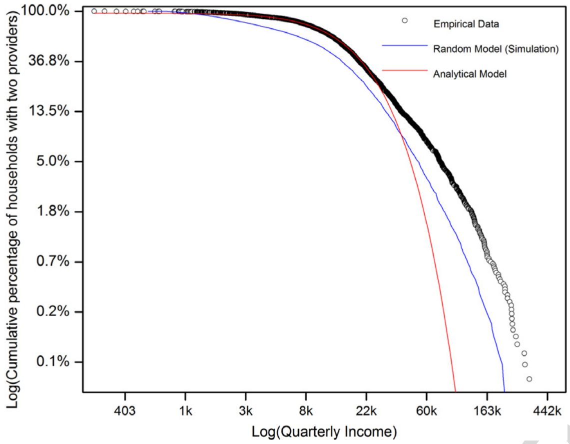


Fig. 7. Empirical CCDF of households for year 2008 with two active members compared to two models constructed from 2008 one-individual households individual data; RM in blue and AM in red. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

adjusted by either Log-normal or Gamma distributions, whereas the high-income range was successfully fitted by the Pareto power-law distribution. Also, taking into account the differences between individual and family income distributions, we successfully constructed for the low-income region a two-provider income distribution with individuals randomly selected. The Log-normal function arises from the Gibrat model and the Gamma function arises from two different contexts; from a constant saving factor fraction model in trade exchanges, and from a utility maximization model in a standard exchange-model. These low-income results even hold for income distribution of households or families with 6 or more economically active members (up to 36 members for year 2008). In order to better investigate the time-evolution of the distributions, the income of the lower class below the Pareto region was further subdivided into "poor" and "middle-class" zones, separated by the income value at the maximum of the uni-modal income probability distribution function (PDF). Since at the end of the year 1994 the Mexican currency suffered a strong depreciation, the subsequent recession clearly impacted the distributions in different ways, being the middle class less affected than the rich class, even in the worst years, 1996, 1998 and 2002. The most striking changes as a whole occurred in 1994–1996, and to a lesser extent, also in 2002. However, since the Mexican ENIGH data is obtained from personal surveys, and not from information obtained indirectly from taxes or consumption, then it is possible that data may not reflect real incomes. Usually income data of poor and rich agents are not very reliable, particularly if obtained by personal interviews and in times of crisis. Thus, the observed resilience of the middle class income may be related to a kind of an averaging effect that compensates errors. On the other hand, in the more stable years at the end of the 2000s, the GDP per capita of the Mexican economy grew faster than the income of the majority of its population. The average population over the years of the three classes in Mexico are 25.89%, 68.01% and 6.09% for poor, middle and rich classes, respectively, in the decade of the 1990s, and in the 2000s are 28.65%, 65.55% and 5.80%. Poor households account for an important chunk of the population. Thus, it was important to study here the poor class, because the higher limit of its household quarterly income (around 11 Kpesos) is relatively small when compared to the corresponding highest limit of the middle class (around 100 Kpesos), without mentioning the higher Pareto incomes. This fact explains why global distribution functions may not capture the low visibility of the poorest agents when it is used in the whole range of income values.

Acknowledgements

We acknowledge the partial financial support given by DGAPA-UNAM, México, through Grant Nos. IN102911 and IN105814, and thank G. Soriano, Diego Juárez, O.F. Favela, I. González, D. Marín and P. Rodríguez for their technical help in gathering and manipulation of data.

References

[1] R.N. Mantegna, H.E. Stanley, An Introduction to Econophysics, Cambridge University Press, 2000.
[2] A. Chaterjee, S. Yarlagadda, B.K. Chaterjee, Econophysics of Wealth Distributions, Springer-Verlag, 2005.
[3] V. Yakovenko, J.B. Rosser, Colloquium: Statistical mechanics of money, wealth and income, Rev. Modern Phys. 81 (2009) 1703.
[4] A. Chakraborti, I.M. Toke, M. Patriarca, F. Abergel, Econophysics review: I. Empirical facts, Quant. Finance 11 (2011) 991–1012.
[5] Gallegati, Mauro, S. Keen, T. Lux, P. Ormerod, Worrying trends in Econophysics, Physica A 370 (2006) 1–6.
[6] V. Pareto, Cours d'Economie Politique, Rouge, Lausanne, Switzerland, 1896.
[7] R. Gibrat, Les Inégalités Economiques. Paris, France, 1931.
[8] C.A. Silva, V.M. Yakovenko, Temporal evolution of the "thermal" and "superthermal" income classes in the USA during 1983–2001, Europhys. Lett. 69 (2) (2005) 304–310.
[9] B.K. Chakrabarti, A. Chakraborti, S.R. Chakravarty, A. Chatterjee, Econophysics of Income and Wealth Distributions, Cambridge University Press, 2013.
[10] A.S. Chakrabarti, B.K. Chakrabarti, Microeconomics of the ideal gas like market models, Physica A 388 (2009) 4151–4158.

[11] C. Gini, Measurement of inequality of incomes, Econom. J. 31 (1921) 124–126.
[12] E.W. Montroll, M.F. Shlesinger, On 1/f noise and other distributions with long tails, Proc. Natl. Acad. Sci. USA (1982) 3380–3383.
[13] M.P. Schneider, Revisiting the thermal and superthermal two-class distribution of incomes, Eur. Phys. J. B 88 (1) (2015) 1–10.
[14] M. Jagielski, R. Kutner, Modelling of income distribution in the European Union with the Fokker–Planck equation, Physica A 392 (2013) 2130–2138. http://dx.doi.org/10.1016/j.physa.2013.01.028.
[15] A.A. Dragulescu, V.M. Yakovenko, Exponential and power-law probability distributions of wealth and income in the United Kingdom and the United States, Physica A 299 (2001) 213–221.
[16] T. Ogwang, Power laws in top wealth distributions: evidence from Canada, Empir. Econom. 41 (2) (2011) 473–486.
[17] A. Shaikh, N. Papanikolaou, N. Wiener, Race, gender and the econophysics of income distribution in the USA, Physica A 415 (2014) 54–60.
[18] M. Brzezinski, Do wealth distributions follow power laws? Evidence from 'rich lists', Physica A 406 (2014) 55–162.
[19] A. Clauset, C.R. Shalizi, M.E.J. Newman, Power-law distributions in empirical data, SIAM Rev. 51 (4) (2009) 661–703. http://tuvalu.santafe.edu/~aaronc/powerlaws/ as consulted in February 21, 2015.
[20] L.A. Adamic, Zipf, Power-laws, and Pareto: a ranking tutorial. Xerox PARC, 2000. http://ginger.hpl.hp.com/shl/papers/ranking/ranking.html.
[21] M.E.J. Newman, Power laws, Pareto distributions and Zipf's law, Contemp. Phys. (2007).
[22] P. Bak, C. Tang, K. Wiesenfeld, Self-organized criticality: An explanation of the 1/f noise, Phys. Rev. Lett. 59 (1987) 381.
[23] P.A. Samuelson, W. Nordhaus, Economics, McGraw-Hill, New York, 1985.
[24] INEGI, Encuesta Nacional de Ingresos y Gastos de los Hogares 2004 (ENIGH-2004) biennial publication. First edition, 2004. http://www.inegi.org.mx/est/contenidos/Proyectos/encuestas/hogares/regulares/enigh/.
[25] A. Dragulescu, V.M. Yakovenko, Evidence for the exponential distribution of income in the USA, Eur. Phys. J. B 20 (2001) 585–589.