the review of income and wealth

Review of Income and Wealth
Series 63, Number 1, March 2017
DOI: 10.1111/roiw.12180

ON THE WORLD DISTRIBUTION OF INCOME

BY DAVIDE PROVENZANO*

Universita' degli Studi di Palermo

In this paper we demonstrate that the size distribution of the world income may be reasonably approximated by a log-normal distribution rather then by a power law, as has previously been believed. This result has been shown to be quite persistent as we move from 1985 to 2011.

JEL Codes: C13, C52, E25

Keywords: generalized gamma distribution, log normal, power law, world income

1. INTRODUCTION

With reference to the world income distribution, Di Guilmi *et al*. (2003) and Furceri (2008) have lent support to a power-law distribution linking the GDP per capita with the rank-size rule. In particular, in Di Guilmi *et al*. (2003), the GDP per capita of countries between the 30th and the 85th percentiles of the distribution is shown to follow a Pareto distribution over the period 1960–97. Furceri (2008), instead, provides empirical evidence of regularity in the GDP per capita data for 175 countries, from 1980 to 2004, collected by the International Monetary Fund (IMF) (World Economic Outlook (WEO), 2005). He shows that the "long-run" (average) world income distribution can be well approximated by Zipf's law.[1] Results in both papers are derived through the ordinary least squares (OLS) procedure.

As discussed in Clauset *et al*. (2009) and Urzúa (2011), the use of OLS for testing Zipf's law introduces systematic biases into the value of the exponent and returns unsatisfactory answers because it gives no indication of whether the data obey a power law at all. Indeed, several distributions, from the standard exponential to the log normal, may show linearity in their right tail on a Zipf plot, even if they are not power laws.

Note: The authors wish to thank Prof. Salvatore Piraino (Maths), Prof. Christian Kleiber, and Prof. Vito M.R. Muggeo for useful comments and discussions. We also want to thank an anonymous reviewer for the positive and helpful suggestions, which have improved the quality of this work.

*Correspondence to: Davide Provenzano, Dipartimento di Scienze Economiche, Aziendali e Statistiche, Ed. 13, Universita' degli Studi di Palermo, Viale delle Scienze (Parco D'Orleans), 90128, Palermo, Italy (davide.provenzano@unipa.it).

[1]The terms Pareto distribution and Zipf's law (the latter is also known as rank-size distribution or rank-size rule), after the two early researchers who championed their study, refer to cumulative distributions with a power-law form. Since power-law cumulative distributions imply a power-law form, Zipf's law and Pareto distribution are effectively synonymous with power-law distribution. Zipf's law and Pareto distribution differ from one another in the way the cumulative distribution is plotted: Zipf made his plots with x on the horizontal axis and $F(x)$ on the vertical one, Pareto did it the other way around (Newman, 2006, p. 4).

The issue is important because only a few distributions are consistent with Gibrat's law of Proportionate Effect (Gibrat, 1931), which provides the classical explanation for the empirical distribution of income. In particular, while the log-normal distribution can be derived as an outcome of stochastic growth models that rely on Gibrat's law (Sutton, 1997), the Pareto distribution is not consistent with it (Bottazzi, 2009).

This paper presents new evidence on the world income distribution. As in earlier studies, we show that the distribution of income is right skewed but, as a new result, we demonstrate that data distribute following a log-normal function rather than a power-law formulation. We demonstrate this point using a parametric method to estimate the income distribution for about 170 countries between 1985 and 2011 (IMF, WEO database, April 2013). For our analysis we use the three-parameter generalized gamma (GG) distribution, which includes the log-normal and the Pareto distributions as limiting cases (Kleiber and Kotz, 2003).

In the next section we present the database used for our analysis. The basic characteristics of the GG distribution and our estimation results are discussed in Section 3. Section 4 is devoted to assessment of the goodness of fit. Section 5 concludes.

2. Data Description and Basic Statistics

In our study the world distribution of income is measured in terms of GDP per capita based on purchasing-power-parity (PPP). The unbalanced panel we use has been taken from the IMF, WEO database.[2] It contains selected macroeconomic data series from the statistical appendix of the WEO report, which presents the IMF staff's analysis and projections of economic developments at the global level in major country groups and in many individual countries.

Most countries' macroeconomic data presented in the WEO conform broadly to the 1993 version of the System of National Accounts (SNA)[3] and reflect information from both national source agencies and international organizations. The WEO database is created during the biannual WEO exercise, which begins in January and June of each year, and results in a publication twice a year (April and September).

Our analysis refers to the April 2013 country-level database, which is made up of an average of 170 countries per year, over the time horizon 1985–2011. For every year in the sample, Table 1 shows the number of countries in the panel, some descriptive statistics, and the income concentration as measured by the Gini coefficient. Not surprisingly, the income data exhibit positive skewness. Moreover, for all the years in the panel, the empirical distribution of the world income looks leptocurtic. The Gini index shows a quite large income gap even though the inequality shows a slow tendency to decrease over the time horizon under consideration. In the same period, the median and mean have been increasing instead.

[2]http://www.imf.org/external/ns/cs.aspx?id=28 (accessed June 27, 2013).
[3]For details, see http://www.imf.org/external/pubs/ft/weo/2013/01/pdf/statapp.pdf.

TABLE 1
Basic Descriptive Statistics of Data

Year	No. countries	Min	Median	Mean	S.D.	Skew	Kur	Max	Gini
1985	138	176.70	2,873	5,687	6,854.35	2.196	9.258	39,680	0.574
1986	139	174.30	2,822	5,807	6,856.94	2.027	8.329	40,050	0.572
1987	139	203.80	3,011	6,069	7,081.24	1.895	7.378	39,900	0.571
1988	139	226.30	3,349	6,340	7,287.41	1.821	7.098	41,520	0.567
1989	139	248.40	3,357	6,749	7,844.09	1.793	6.732	43,670	0.572
1990	145	258.60	3,242	6,850	7,830.97	1.546	4.956	36,240	0.571
1991	146	280.00	3,384	7,083	8,073.54	1.548	5.000	37,660	0.570
1992	162	266.60	3,548	7,043	8,129.12	1.676	5.521	39,370	0.570
1993	164	281.80	3,727	7,259	8,411.52	1.692	5.540	39,430	0.570
1994	166	267.80	3,763	7,474	8,773.37	1.698	5.516	40,490	0.576
1995	168	266.40	3,920	7,768	9,018.78	1.652	5.311	42,160	0.573
1996	168	261.80	4,124	8,095	9,331.83	1.609	5.090	43,120	0.572
1997	170	246.80	4,424	8,461	9,746.16	1.658	5.446	47,390	0.570
1998	173	240.60	4,462	8,593	9,943.30	1.678	5.580	49,680	0.570
1999	174	229.20	4,784	8,888	10,307.36	1.693	5.647	50,680	0.570
2000	182	213.20	5,109	9,302	10,869.82	1.756	6.002	55,410	0.572
2001	182	213.40	5,483	9,624	11,165.46	1.754	6.039	57,450	0.569
2002	183	217.90	5,714	9,874	11,413.01	1.743	6.018	60,130	0.568
2003	185	228.50	5,807	10,250	11,808.88	1.734	5.942	61,630	0.567
2004	185	242.70	5,861	10,980	12,620.08	1.791	6.467	67,770	0.565
2005	185	261.70	6,197	11,660	13,186.24	1.718	6.075	70,460	0.561
2006	185	276.90	6,685	12,520	14,091.08	1.740	6.299	75,170	0.559
2007	185	294.00	7,625	13,360	14,844.16	1.705	6.216	81,100	0.555
2008	185	309.70	7,961	13,820	15,046.98	1.646	5.950	80,860	0.548
2009	185	311.90	7,908	13,390	14,417.90	1.685	6.221	77,450	0.542
2010	185	328.90	8,289	13,920	15,083.78	1.802	7.077	88,310	0.541
2011	185	348.50	8,604	14,440	15,753.80	1.899	7.874	97,990	0.542

3. The GG and the World Distribution of Income

A positive random variable X is said to follow a GG(a, β, p) if its probability density function (*pdf*) is given by

$$f(x)=\frac{ax^{ap-1}e^{-(x/\beta)^a}}{\beta^{ap}\Gamma(p)}, \quad x>0. \tag{1}$$

Here $\Gamma(\cdot)$ is the standard gamma function, $\beta^a = \sigma^2 a^2$ is a scale parameter, whereas $a > 0$ and $p=\frac{a\mu+1}{\beta^a}$ are shape parameters[4] (McDonald, 1984; McDonald and Xu, 1995; Kleiber and Kotz, 2003). This parameterization was first introduced by Amoroso (1925) as the family of GG distributions.

The GG is a fairly flexible family of distributions that includes many distributions supported on the positive halfline as special or limiting cases (McDonald, 1984; McDonald and Xu, 1995; Kleiber and Kotz, 2003). In particular, as $a \to 0$, $p \to \infty$, $\beta \to 0$, but $a^2p \to 1/\sigma^2$ and $\beta p^{1/a} \to e^{\mu}$, the distribution tends to a log normal LN(μ, σ)[5]; for $a \to \infty$, $p = r/a \to 0$, with $r > 0$, the distribution tends to a power

[4]It is sometimes convenient to allow for $a < 0$. One then simply replaces a by $|a|$ in the numerator of (1).

[5]Kleiber and Kotz (2003, p. 149) inadvertently indicate that $\beta \to \infty$, $a^2 \to 1/\sigma^2$ and $\beta p^{1/a} \to \mu$, as $a \to 0$.

function distribution PF(r, β)[6] (McDonald *et al.*, 2013). Since the power function distribution is the inverse Pareto distribution (Kleiber and Kotz, 2003, p. 73), for $a \to \infty$, $p \to 0$, with $ap = -r$, the Pareto distribution Par(β, r) is a limiting case as well. This makes the GG useful for discriminating among these models.

The GG was used by Kloek and Van Dijk (1978), Taillie (1981), McDonald (1984), Atoda *et al.* (1988), and Bordley *et al.* (1996) to study the income distribution in a number of different countries. In this paper, we use the GG to model the size distribution of the world income. The fitted model is given by $X \sim \mathrm{GG}(\hat{a}, \hat{\beta}, \hat{p})$, where $\hat{a}$, $\hat{\beta}$, $\hat{p}$ are maximum likelihood estimates of a, β, and p. For the estimation of the three unknown parameters, we use a generalization of the Rigby and Stasinopoulos (1996a, 1996b) algorithm, which does not require accurate starting values to ensure convergence.[7]

Results of fitting the GG to our dataset are reported in Table 2. We give the estimates for the whole time window to analyze the evolution of the world income distribution over time. Values of $\hat{a}$, $\hat{\beta}$, and $\hat{p}$ suggest that the log-normal distribution may be a good model for our data. Indeed, $\hat{\beta}$ and $\hat{p}$ respectively tend to zero and infinity, namely to the limit value for the GG to converge to the log-normal function. $\hat{a}$ is very close to zero, though it tends to increase over the time period under consideration, suggesting that the agreement with the log normal may deteriorate over time.

The relative fits of the log normal and GG with each other and with the data will be further explored in the next section. For the case under study, an increase in a confirms that the world income distribution has become less concentrated over time.

4. The Goodness of Fit

In order to get a formal test of the appropriateness of the log-normal hypothesis versus the GG model, we perform a likelihood ratio (LR) test, which provides the basis for comparing the goodness-of-fit (GoF) of nested models.

The LR test statistic is

$$LR = -2\ln\left(\frac{L(\Theta_r)}{L(\Theta_g)}\right) = 2[\ln L(\Theta_g) - \ln L(\Theta_r)], \tag{2}$$

where $L(\cdot)$ is the likelihood function, and Θ_g and Θ_r denote maximum likelihood estimators of the general and restricted model, respectively. The null hypothesis that the distribution with fewer parameters fits the data as well as the more general model is rejected if the value of the test is greater than a chi-squared percentile with k degrees of freedom, where the percentile corresponds to the chosen confidence level. In fact, the test statistic is asymptotically distributed as a χ^2 random variable with degrees of freedom equal to the difference in the number of free parameters between the two models. Based on the LR test results shown in Table 3, the hypothesis that the GG is observationally equivalent to the log-normal function

[6]Kleiber and Kotz (2003, p. 149) inadvertently indicate that the PF is a limiting case of the GG as $a \to 0$; the correct limit is $a \to \infty$.

[7]R users may use the gamlss package.

TABLE 2

ESTIMATES FOR THE GG DISTRIBUTION

Year	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998
$\hat{a}$	0.068	0.086	0.090	0.111	0.084	0.049	0.058	0.055	0.058	0.020	0.046	0.064	0.083	0.095
$\hat{\beta}$	0.000	0.000	0.000	0.000	0.000	0,000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$\hat{p}$	27,371	22,138	22,008	19,348	25,714	44,531	39,100	42,840	41,871	116,478	54,994	41,605	34,620	30,712

Year	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
$\hat{a}$	0.102	0.135	0.147	0.169	0.189	0.253	0.276	0.311	0.359	0.427	0.440	0.440	0.425
$\hat{\beta}$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.010	0.044	0.050	0.050	0.039
$\hat{p}$	29,943	24,063	23,552	21,235	20,160	16,967	17,134	16,926	16,607	15,860	15,872	16,580	17,593

TABLE 3

LR STATISTICS OF THE LOG-NORMAL MODEL AGAINST THE GG

Year	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998
LR	0.06	0.11	0.11	0.16	0.09	0.01	0.04	0.04	0.05	0.01	0.03	0.06	0.13	0.17
	(0.81)	(0.74)	(0.74)	(0.69)	(0.76)	(0.92)	(0.84)	(0.84)	(0.82)	(0.92)	(0.86)	(0.74)	(0.72)	(0.68)

Year	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
LR	0.21	0.41	0.49	0.66	0.86	1.59	1.82	2.30	2.94	4.05	4.21	4.29	4.02
	(0.65)	(0.52)	(0.48)	(0.42)	(0.35)	(0.20)	(0.18)	(0.13)	(0.09)	(0.044)	(0.040)	(0.038)	(0.045)

p-Values in parentheses.

TABLE 4

p-VALUES FOR THE FIT TO THE LOG-NORMAL MODEL

Year	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998
KS	0.12	0.17	0.16	0.13	0.17	0.24	0.24	0.43	0.33	0.34	0.32	0.26	0.23	0.37
AD	0.92	0.92	0.92	0.89	0.90	0.89	0.89	0.92	0.92	0.91	0.90	0.90	0.92	0.93
CvM	0.35	0.28	0.27	0.25	0.28	0.26	0.24	0.33	0.32	0.31	0.29	0.30	0.33	0.36

Year	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
KS	0.44	0.51	0.44	0.44	0.51	0.57	0.53	0.51	0.53	0.42	0.39	0.43	0.46
AD	0.93	0.93	0.93	0.93	0.94	0.93	0.93	0.93	0.90	0.90	0.89	0.90	0.90
CvM	0.40	0.40	0.40	0.35	0.39	0.42	0.38	0.34	0.29	0.26	0.26	0.23	0.26

can not be rejected at conventional levels of significance, except for the end of the time period being considered, which is consistent with the estimated values of a increasing.

Finally, our log-normal hypothesis is investigated using three GoF tests: the Kolmogorov–Smirnov (KS), the Anderson–Darling (AD), and the Cramér–von Mises (CvM) tests.

For a hypothesized cumulative distribution function (CDF), $F(x)$, the KS statistic is

$$D_n = sup_x |F_n(x) - F(x)|, \tag{3}$$

where $F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I_{X_i \le x}$ is the empirical cumulative distribution function (ECDF) ($I_{X_i \le x}$ is the indicator function that is equal to 1 if $X_i \le x$ and 0 otherwise). The null hypothesis that the sample comes from $F(x)$ is rejected at level α if $\sqrt{n}D_n \ge K_\alpha$, where K_α is calculated by $P(K \le K_\alpha) = 1 - \alpha$, according to the Kolmogorov distribution (D'Agostino and Stephens, 1986).

Rather than computing the largest vertical distance between the two compared CDFs, the AD statistic is calculated over the whole range of data. This gives more weight to outliers than KS and makes the AD GoF test more powerful than KS for detecting departures in the tails from the hypothesized distribution.

The AD test statistics is

$$AD = -n - S, \tag{4}$$

with

$$S = \sum_{i=1}^{n} \frac{2i-1}{n}[\ln(F(x_i)) - \ln(1 - F(x_{n-i+1}))]. \tag{5}$$

Differently from KS, the AD test makes use of the specific distribution for calculating critical values. This is an advantage for sensitivity, but it requires that critical values be calculated for every single distribution tested. Several tabulated values can be found in D'Agostino and Stephens (1986).

An alternative to KS is the CvM statistic given by

$$CvM = \frac{1}{12n} + \sum_{i=1}^{n}\left[\frac{2i-1}{2n} - F(x_i)\right]^2. \tag{6}$$

Critical values for this test can be found in D'Agostino and Stephens (1986) and Anderson (1962). When KS and CvM tests give opposite results, the second one should be trusted (Anderson, 1962).

When, as in our study, parameters of the tested distribution are not known in advance but have to be estimated from the sample itself, simulation experiments or other methods are required to correct the test statistic and the critical values.

We have determined critical values following a parametric bootstrap procedure. We have first fit a log-normal model to our observed data to calculate parameters $\hat{\mu}$ and $\hat{\sigma}$, and then the test statistic for the observed data has been compared against the test value calculated for a large number (2500) of synthetic data sets, as large as the real data, drawn from a log-normal distribution with mean $\hat{\mu}$ and standard deviation $\hat{\sigma}$. The fraction of synthetic measurements larger than the empirical one defines our p-value.

Table 4 shows, for every year in the panel, the statistics for the fit to the log-normal model. p-Values are not statistically significant for every year under study and GoF test statistic, indicating that our dataset is consistent with a log-normal distribution.

5. Conclusions

In this paper we have analyzed the size distribution of the world income measured in terms of GDP per capita (PPP), using data taken from the IMF, WEO database (April 2013), for the period 1985–2011.

We have shown that the world income has a right-skewed distribution with an inequality Gini coefficient ranging between 0.541 and 0.576. For all the years in the panel we have also provided a fit to a three-parameter GG function, which includes the log-normal and the Pareto distributions as limiting cases.

The estimated values of parameters suggest that the world income distribution may be reasonably approximated by a log-normal model rather than by a power law. This result has been shown to be quite persistent as we move from 1985 to 2011. Goodness of fit tests support this finding.

The classic explanation for the empirical distribution of income is Gibrat's law of Proportionate Effect (Gibrat, 1931). In fact, the log-normal distribution can be derived as an outcome of stochastic growth models that rely on that famous empirical law (Sutton, 1997). This being the case, the possibility of Pareto's law shaping the size distribution of the world income has to be excluded because, as demonstrated in Bottazzi (2009), the law of Pareto and the law of Gibrat can not, in any respect, be reconciled.[8]

References

Amoroso, L., "Ricerche intorno alla curva dei redditi," *Annali Di Matematica Pura Ed Applicata*, 2, 123–59, 1925.

Anderson, T. W., "On the Distribution of the Two-Sample Cramer–von Mises Criterion," *Annals of Mathematical Statistics*, 33, 1148–59, 1962.

Atoda, N., T. Suruga, and T. Tachibanaki, "Statistical Inference of Functional Form for Income Distribution," *Economic Studies Quarterly*, 39, 14–40, 1988.

Bordley, R. F., J. B. MacDonald, and A. Mantrala, "Something New, Something Old: Parametric Models for the Size Distribution of Income," *Journal of Income Distribution*, 6, 91–103, 1996.

Bottazzi, G., "On the Irreconcilability of Pareto and Gibrat Laws," *Physica A: Statistical Mechanics and its Applications*, 388, 1133–6, 2009.

Clauset, A., C. R. Shalizi, and M. E. J. Newman, "Power-Law Distributions in Empirical Data," *SIAM Review*, 51, 661–703, 2009.

D'Agostino, R. B. and M. A. Stephens, *Goodness-of-fit Techniques*, Marcel Dekker, New York, 1986.

Di Guilmi, C., M. Gallegati, and E. Gaffeo, "Power Law Scaling in the World Income Distribution," *Economics Bulletin*, 15(6), 1–7, 2003.

Furceri, D., "Zipf's Law and World Income Distribution," *Applied Economics Letters*, 15, 921–3, 2008.

Gibrat, R., *Les Inégalités Economiques, Libraire du Recueil Sirey*, Paris, 1931.

International Monetary Fund, World Economic Outlook Database, April 2005 (http://www.imf.org/external/pubs/ft/weo/2005/01/data/).

[8]In this regard, conclusions of Di Guilmi *et al.* (2003) and Furceri (2008) should be reformulated.

Kleiber, C. and S. Kotz, *Statistical Size Distributions in Economics and Actuarial Sciences*, John Wiley & Sons, Hoboken, NJ, 2003.
Kloek, T. and H. K. van Dijk, "Efficient Estimation of Income Distribution Parameters," *Journal of Econometrics*, 8, 61–74, 1978.
McDonald, J. B., "Some Generalized Functions for the Size Distribution of Income," *Econometrica*, 52, 647–63, 1984.
McDonald, J. B. and Y. J. Xu, "A Generalization of the Beta Distribution with Applications," *Journal of Econometrics*, 66, 133–52, 1995.
McDonald, J. B., J. Sorensen, and P. A. Turkley, "Skeness and Kurtosis Properties of Income Distribution Models," *Review of Income and Wealth*, 59, 360–74, 2013.
Newman, M. E. J., "Power Laws, Pareto Distributions and Zipfs Law," (http://arxiv.org/pdf/cond-mat/0412004.pdf; accessed July 3, 2013), 2006.
Rigby, R. A. and D. M. Stasinopoulos, "A Semi-Parametric Additive Model for Variance Heterogeneity," *Statistics and Computing*, 6, 57–65, 1996a.
———, "Mean and Dispersion Additive Models," in W. Härdle and M. G. Schimek (eds), *Statistical Theory and Computational Aspects of Smoothing*, Physica-Verlag, HD, 215–30, 1996b.
Sutton, J., "Gibrat's Legacy," *Journal of Economic Literature*, 35, 40–59, 1997.
Taillie, C., "Lorenz Ordering Within the Generalized Gamma Family of Income Distributions," in C. Taillie, G. P. Patil and B. Baldessari (eds), *Statistical Distributions in Scientific Work*, 6, Reidel, Boston, 181–92, 1981.
Urzúa, C. M., "Testing for Zipf's Law: A Common Pitfall," *Economics Letters*, 112, 254–5, 2011.