Callealta-Barroso, F.J., García-Pérez, C., Prieto-Alaiz, M.: "Modelling income distribution using the log Student's t distribution: New evidence for European Union countries". Economic Modelling, 2019. DOI: https://doi.org/10.1016/j.econmod.2019.11.021

Modelling Income Distribution Using the Log Student's *t* Distribution: New Evidence for European Union Countries

Abstract

Income distribution remains a crucial topic in economic analysis, among other reasons, due to the increase in inequality in recent years, as one of the effects of the Great Recession. In this context, proposing parametric models that represent the full distribution through a small number of parameters arouses great interest as an instrument for economic analysis. This paper studies the ability of log Student's *t* distribution to model the size distribution of income due to its potential to reproduce the effect of a mode around low-incomes as well as its precision in capturing the degree of kurtosis of empirical distributions. These characteristics make the log-*t* an ideal analysis tool, for instance, for exploring the effects of anti-poverty policies. The model has been fitted to income data for the EU25 and for several years. The conclusion is that the log Student's *t* distribution offers the best fit in the vast majority of cases.

Keywords: Log Student's *t* distribution, parametric modelling, income distribution, EU countries

(*) We would like to thank the editor and the anonymous referees for the comments and suggestions on the first version of this paper.

1. Introduction

The personal distribution of income and wealth has been the subject of numerous recent studies, such as those of Piketty (2014, 2018) or Atkinson (2017), which, since the start of the Great Recession, have placed distributive aspects at the centre of the economic debate within the different international institutions (OECD, 2015). Indeed, in many countries, economic crisis has led to the increase of inequalities and polarisation, which raises the need to review the approaches of economic policies behind this situation and to seek new proposals of measures that will alleviate these effects (OECD, 2013).

One of the most used tools for the study of personal distribution of income and inequality has been the parametric modelling of the empirical distribution. This tool allows for the characterisation of the distribution using a scarce number of parameters, with the possibility of completely reconstructing the distribution of the variable of interest from limited information that is easy to manage.

The choice of a certain functional form, in general, is determined by the specific characteristics of the phenomenon to be analysed. In principle, any family of cumulative distribution functions could be used to model the personal distribution of income, so it seems necessary to restrict this large set to a subset of functions that satisfy certain properties. These properties may be based on standard characteristics of the observed income distribution (for example, the positive asymmetry or the heaviness of its right tail), desirable mathematical characteristics (for example, that it is twice differentiable), or economic properties (for example, that it is generated as a result of an underlying economic model).

This technique also offers certain advantages over the alternative use of non-parametric methods such as kernel estimates. One advantage is the possibility of providing theoretical justifications for the generation of the distribution and gauging its suitability to correct the usual problems derived from a lack of information in distribution tail surveys, which occur due to the under-declaration of high incomes and difficulties in accessing very low income and salary recipients (Pinkovsky and Sala-i-Martin, 2009). Parametric modelling is also used in numerous studies as a tool for general equilibrium microsimulation models (Boccanfuso, Richard and Savard, 2013), where the final result of the distribution is considered as a situation of equilibrium of the performance of economic agents (Clementi and Gallegati, 2016). In others works, it is used as a tool for measuring poverty or inequality (Pinkovsky and Sala-i-Martin, 2009; Vinh et al., 2010 and Chotikapanich et al., 2014).

In any case, among the proposed models of statistical distributions, some deficiencies are observed regarding the modelling of certain income intervals. Such is the case of negative or zero incomes, which, in situations of economic crisis, are more abundant in the distribution, due to indebtedness and the higher number of individuals who do not receive income after exhausting the subsidies for situations of long-term unemployment.

Thus, in this paper we try to answer the following questions: what is the best model to fit the empirical income distributions currently observed in different countries? and, additionally, which model is more capable of reproducing bimodal empirical distributions with an accumulation of incomes around zero? The question is relevant given that this model will be a fundamental tool for analysing policies that are dedicated to reducing inequality and poverty, especially related to the lower tail of the distribution.

In an effort to answer this question, this article studies and proposes the log Student's t distribution (log *t* distribution)[1] that contemplates this type of situation while enabling an acceptable modelling of other parts of the distribution, such as the upper tail, with a reduced number of parameters (three). This paper seeks to study in depth the theoretical and empirical properties that underlie this model and its applicability to income data based on its good fits and its advantages for economic interpretation. It also represents the existence of zero incomes.

To check the suitability of the log *t* distribution, EU Statistics on Income and Living Conditions (EU-SILC) data will be used for 25 European countries for the years 2006, 2011, and 2016, corresponding to different phases of the economic cycle before, during, and after the Great Recession, respectively.

The results obtained, using different goodness of fit measures, allow us to conclude definitively that the log t distribution produces the best fits compared to the most widely used models in the economic literature, especially in the first three quartiles of the distribution. This implies that such a distribution should be considered as a useful model to model the effects of policies aimed at improving the income distribution of countries.

The paper is structured as follows. There is a review of the economic literature on parametric modelling, followed by a discussion of the formal development of the log *t* model. The data used in the study is described and then the results are presented. The paper ends with a section that synthesises the main conclusions of the article.

2. Literature Review

Since Pareto (1896) presented the first law of personal income distribution, a large number of alternative probability density functions have been proposed as models of income distribution. Gibrat (1931) provided a theoretical basis (the proportional effect law) to obtain a lognormal distribution; this model was also studied by Aitchison and Brown (1957) and was the most widely used until Thurow (1970), and Salem and Mount (1974), proposed, respectively, the beta and gamma distribution as models of the distribution of income.

The results of goodness of fits were improved with the use of three-parameter models that were consolidated with the studies on the generalised gamma (Kloek and van Dijk 1978), the Singh-Maddala distribution (Singh and Maddala, 1976), and the Dagum type I distribution (Dagum 1977).

In the 1980s, models with more than three parameters were used, including the generalised beta distribution of first and second type. McDonald (1984) introduced two distributions of four parameters that nest most of the models of two and three

[1] The log t distribution has been shown to be appropriate for application to empirical income distributions (Bartels, 1977; Kloeck and Van Dijk, 1977, 1978; Pena et al., 1998),

parameters as special cases or limit distributions; their adequacy and validity has been demonstrated in recent studies (Prieto and García, 2009, García, Prieto and Simón, 2014).[2]

Subsequently, five-parameter models were proposed. Thus, McDonald and Xu (1995) introduced a generalised beta distribution with five parameters, and Reed and Wu (2008) introduced a generalised Double Pareto-lognormal distribution. Following the line of studies on distributions with a large number of parameters and the use of mixtures, Domma and Condino (2013) proposed the Beta-Dagum distribution.

In recent years, in addition to the proposals and applications of statistical models with a large number of parameters, studies on parametric distributions have been focusing on different topics: the modelling of specific parts of the distribution (Jenkins, 2017), comparisons of models (Oluyede et al., 2014, Tahir and Cordeiro, 2016), proposals for mixtures of distributions (Chotikapanich and Griffiths, 2008; Graf and Nedyalkova, 2017) or applications and theoretical developments of parametric models to obtain measures of poverty and inequality (Chotikapanich et al., 2013, Graf and Nedyalkova, 2014, Sarabia and Jordá, 2014, Sarabia et al., 2017), among other topics.

Among the recent contributions to the modelling of income distribution, the contributions to the econophysics literature by Dragulescu and Yakovenko (2001), Silva and Yakovenko (2004), and Yakovenko (2007) stand out. These authors decompose the distribution of income into labour and property incomes and provide empirical evidence that the bottom of the distribution of income is exponential, while the top is Pareto.

A recent line of research, also based on physics, refers to the probabilistic distributions generated from the agent-based models from which interesting studies on poverty measures have been developed (Clementi, et al., 2006, Chattopadhyay and Mallick, 2007; Chattopadhyay et al., 2010 and Chattopadhyay et al., 2017). In this framework, the authors start from a conception of the economy as a physical system where a multitude of heterogeneous agents interact. As a result of their aggregate behaviour, some models of income distribution can be generated, such as the κ-generalized models.

This work is based on the development and use of a three-parameter model, the log *t* distribution which, in addition to its good theoretical properties, has a microeconomic foundation and produces good fits in empirical income distributions throughout time and space.

3. The Log Student's *t* Model

3.1 Motivation

It is desirable that the functional form of any model that adequately represents personal income distribution has a certain microeconomic foundation that connects it with the generation and distribution of income processes.

---

[2] Comparisons between the models can be found in Kleiber (1996), Bordley et al., (1996), and Bandourian et al. (2003).

In this sense, the lognormal distribution is shown as a plausible model to justify the probabilistic distribution of personal incomes as the result of a process of growth produced by the accumulation of many small relative changes over time. These changes become additives on a logarithmic scale and, using the central limit theorem, can be approximated as a normal distribution (Gibrat, 1931). However, the lognormal distribution presents a poor fit in the right tail of the income distributions. As an alternative to positive model incomes, a mixture of the lognormal distribution with a Pareto distribution has been suggested. These two distributions are supported by economic laws that justify their use. In addition, the literature clearly identifies that the Pareto fits the right tail well (although not the left one), and that the lognormal fits the left tail adequately (although not the right one).

However, in the empirical distributions, the existence of an accumulation of null, and even negative, incomes that the aforementioned mixture would not adequately reproduce is often observed. The log t distribution, with only three parameters, is an alternative to approximate the behaviour empirically observed in the tails and the degree of kurtosis of income distributions, based on the same process for the generation and distribution of incomes underlying the lognormal model, but modifying it partially according to the following reasoning:

Let a variable, Y, distributed according to a lognormal distribution with parameters, $\mu = \log(\beta)$ and $\sigma = \alpha\sqrt{\eta}$, then:

$$\frac{\log(Y) - \mu}{\sigma} = \frac{\log(Y) - \log(\beta)}{\alpha\sqrt{\eta}} \sim N(0;1) \quad \Leftrightarrow \quad \log\left(\frac{Y}{\beta}\right)^{\frac{1}{\alpha\sqrt{\eta}}} \sim Z, \qquad (1)$$

where Z is a standard normal distribution.

On the other hand, a random variable Y is said to have a log t distribution with parameters $\alpha > 0$, $\beta > 0$ and $\eta > 0$ if $\frac{1}{\alpha}\ln\left(\frac{Y}{\beta}\right) \sim t_\eta$. So, when the variable Y is distributed according to a log t of parameters α, β, and η, then:

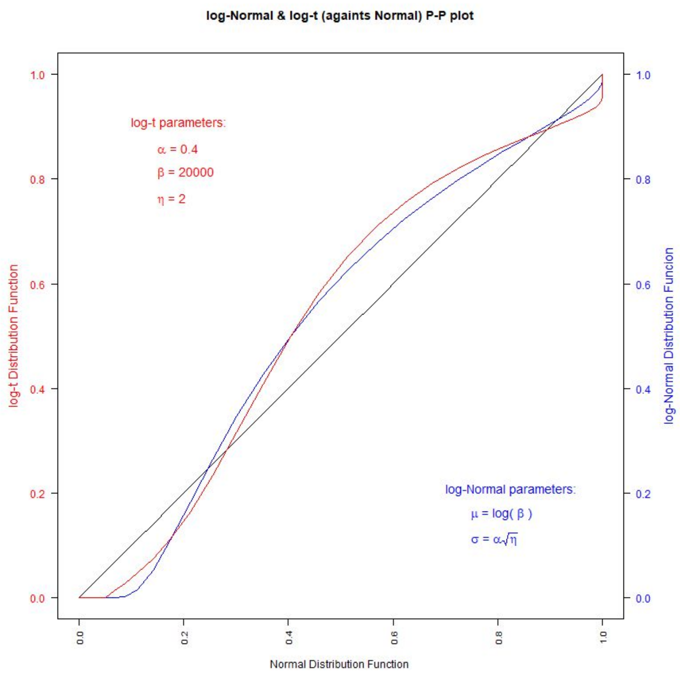
$$\frac{1}{\alpha}\log\left(\frac{Y}{\beta}\right) = \log\left(\frac{Y}{\beta}\right)^{\frac{1}{\alpha}} \sim t_\eta = \frac{Z}{\sqrt{D/\eta}} \quad \Leftrightarrow \quad \log\left(\frac{Y}{\beta}\right)^{\frac{1}{\alpha\sqrt{\eta}}} \sim \frac{Z}{\sqrt{D}}, \qquad (2)$$

where Z is a standard normal distribution and D is a χ^2_η distribution, both independent.

As can be deduced from the above expressions, in the process of income assignment of a distribution $log\ t(\alpha,\beta,\eta)$, we can initially identify the same process that is made in a distribution $log\ N(\log(\beta); \alpha\sqrt{\eta})$: the logarithm of income will be distributed following a normal distribution. Recall that the normal distribution emerges after applying the central limit theorem to the sum of many small relative changes over time. However, the log t corrects the initial normal variable with $\sqrt{D}$, the square root of the chi-squared random variable. When η is the integer, $\sqrt{D}$ could be interpreted as the Mahalanobis distance of a Gaussian random (η×1) vector to the centroid. The components of this Gaussian random vector are the different factors that control the individual heterogeneity, such as cognitive abilities, socio-economic background, or personality traits. Therefore, the parameter η would be the dimension of the factorial space of relevant determining factors of personal income distribution.

This correction procedure, using a measure of the heterogeneity of individuals ($\sqrt{D}$), generates a log t distribution that reproduces, quite well, three basic characteristics that are observed in a repetitive way in empirical income distributions: the heaviness of its right tail, a non-insignificant accumulation of individuals around zero and a high degree of kurtosis. This effect is shown in the following figure that presents the PP-plots (with respect to the normal distribution) of the log *t* distribution of parameters $\alpha = 0.4$, $\beta = 20000$, and $\eta = 2$, and the corresponding lognormal distribution that the log *t* modifies.

Figure 1. Lognormal and Log *t* PP-plots ($\alpha = 0.4$, $\beta = 20000$ and $\eta = 2$)



Additionally, as McDonald and Butler (1987) suggest, the log *t* distribution admits a mixture interpretation that allows a theoretical justification as a representation of income arising from a heterogeneous population. In fact, Hogg and Klugman (1983) show that the log *t* can be obtained as a compound lognormal distribution whose scale parameter follows an inverse gamma.

Finally, there are additional reasons that advise the in-depth study of this log *t* distribution from a purely instrumental perspective. When asymmetric variables (such as income, wages, or expenses) are considered as an endogenous variable in a general linear econometric model with normal errors, it is usually transformed by calculating its logarithms. This, implicitly, assumes that their distributions are lognormal, even though their tails are not adequately represented. In this or similar situations, some authors propose adapting these models considering non-normal errors; and some of them specifically considering Student's *t* errors (Zeckhauser & Thompson, 1970; Zellner, 1976; Burbano & Melo, 2015). When Student's *t* errors are considered, the distribution of the endogenous variable becomes a log *t*.

3.2. The model

The well-known Student's *t* distribution (or simply *t* distribution), t_η, is a symmetric distribution commonly used to model variables with tails heavier than those found in the normal distribution. It depends on one parameter, the degrees of freedom (η), which controls the dispersion and the kurtosis of the distribution (Johnson et al., 1995).

Similar to the way the lognormal comes from the normal distribution, a derived model from the *t* distribution is the log Student's *t* distribution (or simply log *t* distribution). Thus, a random variable *Y* is said to have a log *t* distribution with parameters $\alpha > 0$, $\beta > 0$, and $\eta > 0$ if $\frac{1}{\alpha}\ln\left(\frac{Y}{\beta}\right) \sim t_\eta$. The dispersion of $\ln\left(\frac{Y}{\beta}\right)$ depends on α and η, where β is the median of the distribution and η denotes the degrees of freedom. In the case where η=1, we have the log-Cauchy distribution, and when $\eta \to \infty$, the log *t* distribution tends to the lognormal distribution. Hence, when η decreases, the tails become heavier. The log *t* distribution could also be considered as a limiting case of the Generalized Beta of the second kind (McDonald, 1984 and Cummins et al., 1990).

The corresponding density function is given by:

$$f(y;\alpha,\beta,\eta) = \frac{1}{\alpha y} f_{t_\eta}\left(\frac{1}{\alpha}\ln\left(\frac{y}{\beta}\right)\right) = \frac{1}{\alpha y\sqrt{\eta}B\left(\frac{\eta}{2},\frac{1}{2}\right)}\left(1+\frac{\ln^2\left(\frac{y}{\beta}\right)}{\eta\alpha^2}\right)^{-\frac{1}{2}(\eta+1)}, y > 0 \quad (3)$$

where f_{t_η} is the density probability function of a *t* distribution with η degrees of freedom and *B* (.) is the beta function. The cumulative distribution function (cdf) cannot be expressed as a closed form, although it could be obtained using the cdf of the *t* distribution with η degrees of freedom, F_{t_η}; that is,

$$F(y;\ \alpha,\beta,\eta) = F_{t_\eta}\left(\frac{1}{\alpha}\ln\left(\frac{y}{\beta}\right)\right). \qquad (4)$$

Similarly, the quantiles could also be obtained by means of *t* distribution quantiles:

$$y_p = \beta\ e^{\alpha y_{t_\eta,p}} \qquad (5)$$

for $0 < p < 1$ and where y_p and $y_{t_{\eta,p}}$ are, respectively, the quantile of order p of the log t distribution and the quantile of order p of t distribution with η degrees of freedom. From the latter expression, it can be directly derived that β is the median of the log t distribution.

It is highly interesting to note that the log t distribution could accumulate an important probability around zero, since $f(y)\xrightarrow[y\to 0^+]{}+\infty$. Furthermore, if $0 < \alpha < \frac{\eta+1}{2\sqrt{\eta}}$, the log t has a peak (local maximum) at $y = \beta e^{-\frac{(\eta+1)-\sqrt{(\eta+1)^2-4\eta\alpha^2}}{2\alpha}}$.

The existence of a mode when y tends to zero, regardless of whether the distribution presents another within its range, is a specific advantage of this distribution compared to the most widely used and best models for fitting empirical income distributions[3]. In the section on parameter estimation and goodness of fit measures, we compare the log t distribution to the Singh-Maddala, the κ-generalized model, the Dagum type I, the lognormal distribution and the generalized beta of the second kind (GB2), which only present a unique mode, either at zero or inside its range (see, among others, Boccafuso et al, 2013, Bandourian et al., 2002, Callealta et al., 1997, Clementi and Gallegati, 2016 and Clementi et al., 2006).

On the other hand, the bimodal log t has an antimode at $y = \beta e^{-\frac{(\eta+1)+\sqrt{(\eta+1)^2-4\eta\alpha^2}}{2\alpha}}$.

It is worth noting that the parameters α and η control the probability around zero. This follows very simply from considering (2) at the antimode; that is,

$$F(y_{antimode}; \alpha,\beta,\eta) = F_{t_\eta}\left(-\frac{(\eta+1)+\sqrt{(\eta+1)^2-4\eta\alpha^2}}{2\alpha}\right). \qquad (6)$$

This probability tends to zero when $\alpha \to 0^+$, and tends to 0.5 when $\alpha \to \frac{\eta+1}{2\sqrt{\eta}}$ and $\eta \to 0^+$. Therefore, by conveniently modifying the parameters α and η, this probability can take values between 0 and ½.

As noted, the distribution is bimodal only for certain values of the parameters α and η. Thus, if α $\geq \frac{\eta+1}{2\sqrt{\eta}}$, the density probability function is decreasing, and thus the distribution is unimodal.

When attempting to find a mathematical expression for the moments with respect to the origin ($E[Y^r]$), we have obtained the following compact form making use of the Student's t distribution:

---

[3] This behaviour is not exclusive to this model. The Champernowne model can reproduce a similar performance for some parameter combinations. Other models, such as Dagum type II, need to incorporate an additional mixture parameter to specifically represent a mode at zero.

$$E[Y^r] = \int_0^\infty y^r \frac{1}{\alpha y} f_{t_\eta}\left(\frac{\log\left(\frac{y}{\beta}\right)}{\alpha}\right) dy = \beta^r E_{t_\eta}[e^{\alpha r T}] = \beta^r M_{t_\eta}[\alpha r], \qquad (7)$$

where $M_{t_\eta}(.)$ is the moment generating function of a t distribution. As $M_{t_\eta}(.)$ is not defined, the log t distribution does not have moments for any value of r (see Kleiber, and Kotz, 2003). This expression leads to the conclusion that the integrals do not converge and, consequently, the moments do not exist; the log t distribution has no moments, but when η tends to infinity and log t tends to the lognormal, it surprisingly has all moments.

This theoretical limitation is important, but it can be overcome from an applied perspective. From a practical point of view, when a value of some moments needs to be estimated for a finite population (the income mean or the variance), the corresponding truncated distribution at a realistic high income level can be used. [4]

Conversely, the non-existence of moments supposes that they cannot be used for statistical inference of the distributional parameters, while alternative measures can (such as e.g. quantiles).

3.3 Income elasticity of the cumulative distribution function

One of the properties of empirical income distributions pointed out by Dagum (1977) is that the income elasticity $\varepsilon(F,y)$ of the cdf F, is a decreasing and bounded function of F. In fact, the Dagum models are derived from this observed regularity of income. For the log t distribution, we have obtained the income elasticity of cdf and we have verified that:

$$\varepsilon(F,y) = \frac{1}{\alpha\sqrt{\eta}B\left(\frac{\eta}{2},\frac{1}{2}\right)}\left\{1+\frac{\ln^2\left(\frac{y}{\beta}\right)}{\eta\alpha^2}\right\}^{-\frac{\eta+1}{2}}\left\{F_{t_\eta}\left(\frac{1}{\alpha}\ln\left(\frac{y}{\beta}\right)\right)\right\}^{-1} > 0 \xrightarrow{y\to+\infty} 0. \qquad (8)$$

Thus, the log t distribution does not follow the Dagum requirement regarding income elasticity throughout the whole of the domain of the function. However, the income elasticity of the F decreases for incomes greater or around the income median; even for the incomes lower than the median if α or η are high enough.[5]

[4] To show the sensibility of moments to the truncation threshold, a graphical analysis of the effect on the average income of truncations in different high probability quantiles was conducted. In this analysis, we used different ranges of values of the parameters estimated for the income distributions of the countries of the EU considered in this paper. In this analysis it is verified that the truncations begin to be critical in some cases (small η and large α) if they are performed in quantiles of orders higher than 99.9%. In general, truncating around 99.5% seems to be a prudent decision. The different simulations and results of this sensitivity analysis are available to the reader upon request.

[5] This can be observed in Figure A.1, which shows the income elasticity of cdf for different parameter combinations. The former shows the sensitivity of elasticity when β changes, and the second and third when α and η vary, respectively. The chosen values for the parameters are similar to those found in the empirical application.

3.4 Income elasticity of the survival function

It is unanimously accepted that the weak law of Pareto (Mandelbrot, 1960) characterises the behaviour of the upper tail of income distributions. This law establishes that the income elasticity of the survival function, $\varepsilon(1-F,y)$, tends to a negative constant as income becomes larger. In the case of the log *t* distribution, the income elasticity of the survival function converges to zero as income becomes larger, as shown in the following expression:

$$\varepsilon(1-F,y) = -\frac{1}{\alpha\sqrt{\eta}B\left(\frac{\eta}{2},\frac{1}{2}\right)}\left\{1+\frac{\ln^2\left(\frac{y}{\beta}\right)}{\eta\alpha^2}\right\}^{-\frac{\eta+1}{2}}\left\{1-F_{t_\eta}\left(\frac{1}{\alpha}\ln\left(\frac{y}{\beta}\right)\right)\right\}^{-1} < 0 \xrightarrow{y\to+\infty} 0 \quad (9)$$

However, for the log *t* distribution, this convergence is too slow. In fact, as shown in Figure A.2 (which shows the sensitivity of the elasticity of the survival function to β, α, and η), this elasticity remains negative and rather stable for very high values of income in the range of values observed for the Pareto exponent in empirical studies (Atkinson et al., 2011).

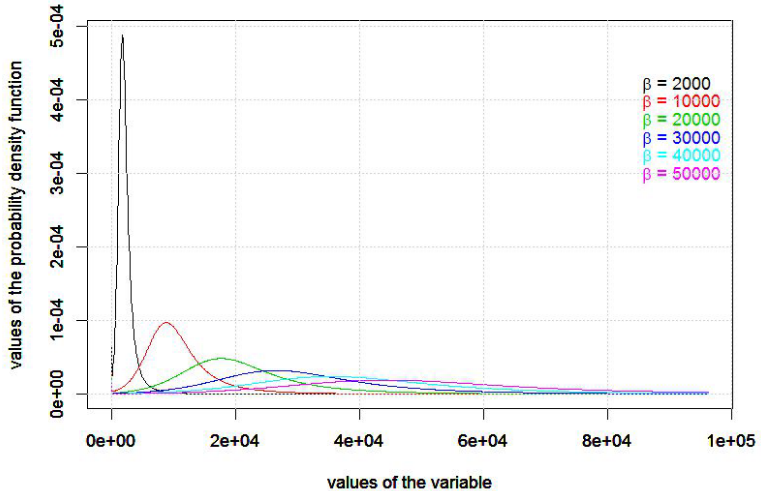
4. The Interpretation of the Parameters of the Log Student *t* Distribution

As indicated in the previous section, the log Student model has three parameters: α, β, and η. Parameter β is its median. Parameters α and η jointly control the probability close to zero as well as the shape and inequality of the distribution; in fact, increases of α imply an increase in inequality, while increases in parameter η, on the contrary, cause a decrease.

In order to interpret the parameters and their effects in greater depth, we have plotted different density functions and their corresponding Lorenz curves for habitual combinations of the parameter values, according to the distributions fitted in the section discussing parameter estimation and goodness of fit measures.

Thus, Figure 2 shows that, as parameter β grows, the entire distribution moves to the right, increasing the income levels. Inequality does not change, which can be seen in the Lorenz curves that remain unchanged when variations of the parameter occur. Therefore, an increase in parameter β, while the other parameters remain constant, produces a higher level of welfare based on an increase in the median and average income.

Figure 2. Effects of Parameter β on the Density Function (for α=0.4; η=3)



Parameter α is the shape parameter par excellence enabling the distribution to show an interior relative mode within it range (for low values of the parameter), or being unimodal at zero (when the parameter takes high values). As parameter α grows, individuals move from the centre of the distribution to their two tails, increasing the degree of inequality as demonstrated in the simulations of density functions and Lorenz curves in Figures 3 and 4, respectively. In turn, incomes decrease in the lower classes and increase in the upper classes.

Together with parameter η, parameter α allows representation of a probability close to zero, which grows when α grows and η tends to zero. Similarly, this probability can be made as small as desired when both parameters vary in the opposite direction (α tends to zero and η increases). The effect of the variations of α suggests that the parameter will be a good indicator of distribution inequality, a fact that is confirmed in the empirical results of this work, where a high correlation is observed between the estimates of parameter α and the values of the Gini index for the different income distributions of countries in the three years considered.

Figure 3. Effects of Parameter α on the Density Function (for β=20000; η=3)

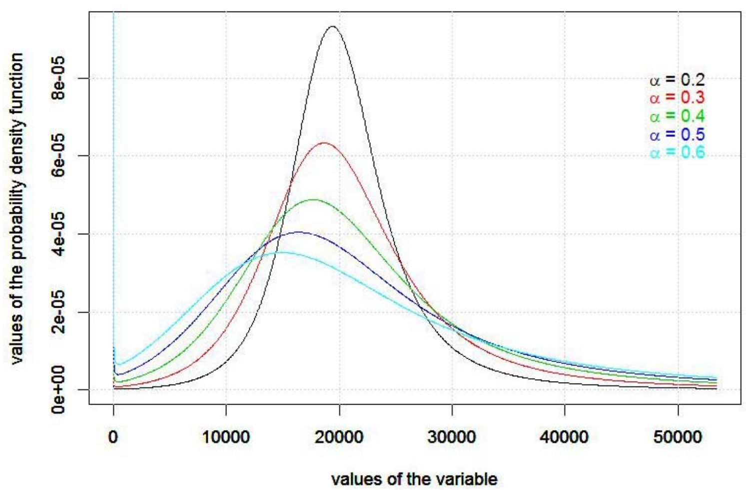
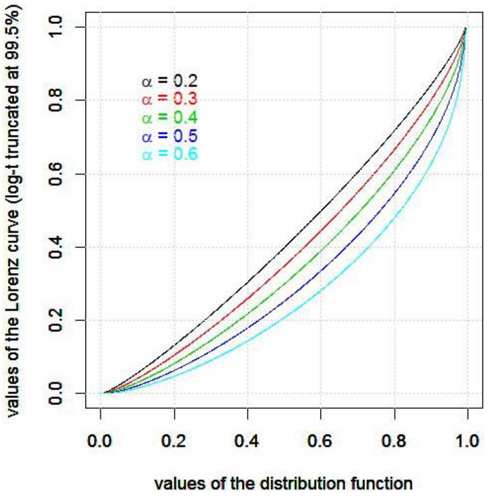


Figure 4. Effects of Parameter α on the Lorenz Curve (for β=20000; η=3)



Parameter η, like parameter α but in the opposite direction, allows the shape to be changed from a unimodal to a bimodal distribution. When η grows, individuals move from the two tails towards the centre of the distribution, thus decreasing inequality (Figures 5 and 6). This effect, however, is more moderate than that caused by the other shape parameter (α), since the sensitivity of the shape of the distribution to the parameter variations is much smaller, requiring larger relative variations of the parameter to produce significant effects in the distribution.

Figure 5. Effects of Parameter η on the Density Function (for α=0.4; β=20000)

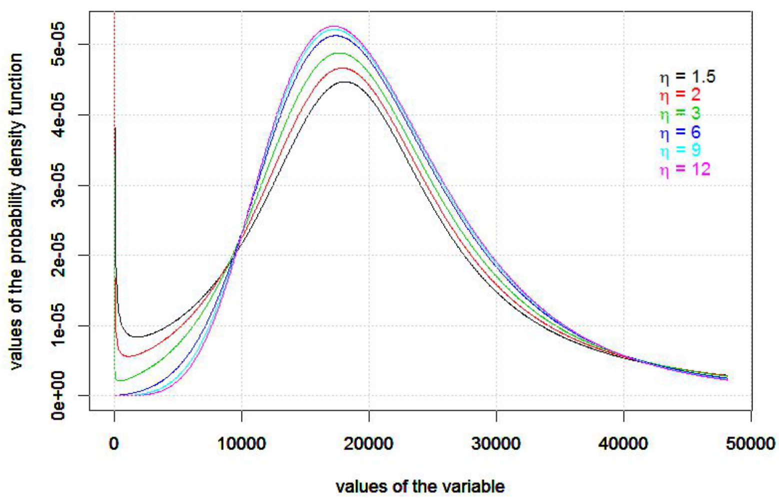
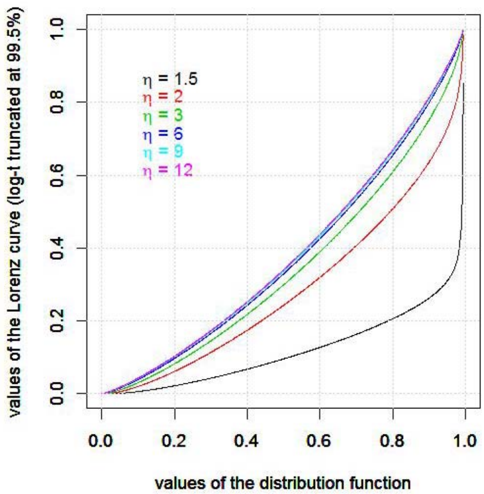


Figure 6. Effects of Parameter η on the Lorenz Curve (for α=0.4; β=20000)



Unlike parameter β, parameters α and η are not individually specialised in the explanation of separate effects. Their joint variation determines their effects on the distribution.

5. Parameter estimation and goodness of fit measures

Once an adequate statistical model has been selected to fit the empirical distribution, the next step is to estimate the vector of unknown parameters. A wide variety of estimators may be obtained; therefore, it is necessary to establish criteria to choose between them. Given that the finite properties of the estimators of the income distribution parameters are hardly known (since they are non-linear functions of the sample), we will focus on estimation procedures that provide estimators that are asymptotically efficient (see Rao 1973, p. 351 and Ghosh, 1994, p.4). Among these types of estimators are those obtained by the maximum likelihood method, which is used in this study. Therefore, the maximum likelihood principle has been applied to the individual income data weighted by the weights of the sample design and the size of each household, according to the data from the EU-SILC samples. To obtain the maximum likelihood estimates, we use the algorithms implemented in the *mle* and *mlexp* Stata commands (StataCorp, 2017).[6]

For comparisons between models, different goodness-of-fit criteria have been used. In this work, we have used the Kolmogorov-Smirnov, Anderson-Darling statistics and the sum of squared residuals (SSR)[7]. These statistics focus on different aspects of the distribution, and therefore provide complementary approaches to the global goodness of

[6] The parameters of the models have been estimated using a modified Newton-Raphson procedure (nr) and the Broyden-Fletcher-Goldfarb-Shanno (bfgs) algorithms, implemented in the *mle* and *mlexp* Stata commands (see StataCorp, 2017). To estimate the lognormal, Dagum, Singh-Maddala, and GB2 distributions we used Stata packages *lognfit* (Jenkins, 2007), *dagumfit* (Jenkins, 1999), *smfit* (Jenkins, 1999), and *gb2lfit* (Jenkins, 2014), which use the algorithms mentioned above. The combination of both algorithms allowed us to obtain the estimations in some countries for the κ-generalized distribution.

[7] A broader study has been conducted on the goodness of fit of the different models considering additional goodness of fit measures such as the log-likelihood values, the Akaike information criterion (AIC), and the Bayesian information criterion (BIC). The results of the above-mentioned statistics give results that conclusively favour the log-t as much as those presented in this paper. An appendix with these results is available to the interested reader upon request.

fit. Thus, the Anderson-Darling statistic (Anderson and Darling, 1954) weighs the deviations existing in the tails, while the Kolmogorov-Smirnov statistic focuses on the maximum distance between the cumulative distribution function of the model and the empirical one.

Goodness-of-fit statistics will be used only as descriptive measures; to perform a hypothesis test, a fully specified distribution function is required, as the parameters cannot be substituted by maximum likelihood estimates (obtained with the same sample) without altering the asymptotic distributions of the statistics. Therefore, in the case where the null hypothesis is simple, both the finite and the asymptotic distribution of the statistics are known and tabulated (see, for example, Stephens, 1986). However, in a case where the null hypothesis is composite, the distribution of the contrast statistics is, in general, unknown.[8]

To analyse the goodness of fit in the different percentiles of the income distributions, the probability plots (PP-plots) will also be constructed and presented, enabling the detection of deviations of a specific distribution from the observed income data. This plot will be approximately linear if the theoretical income distribution is the correct model. To complement this graphical information and in order to analyse the behaviour of each model for different parts of the distribution, the values of the measures of goodness of fit have also been calculated for different income ranges. In particular, we have calculated the sum of squares of the residuals separately for all quarters from the empirical distributions.

5. Data

In this article we consider the distributions of 25 European countries in three time periods – 2006, 2011, and 2016 – during different phases of the economic cycle – before the Great Recession, during the Great Recession, and in the subsequent recovery. These years were selected to try to validate the results in different situations, given the relationship of personal income with the macroeconomic variables that reflect the periods of growth and recession (Jenkins et al., 2012; Smeeding, 2012).

The income data are obtained from the EU-SILC and enable, wherever possible, comparisons using homogeneous income concepts obtained from surveys that follow a similar methodology. The countries analysed are those for which the empirical distributions are available in the three years selected.

The concept of 'income' is the disposable income of a household, which is determined by the sum of all monetary income received from any source by each member of the household (including labour income, capital income, and social benefits) and all other income, from which taxes and social contributions are deducted.

In order to reflect differences in the size and composition of households, this total income is divided by the number of 'equivalent adults' using the OECD-modified equivalence scale, which assigns a weight of 1.0 to the first adult in the household, a weight of 0.5 to the other members of the household over 14 years of age, and a weight

[8] Stephens (1986) indicates that this distribution depends on the distribution assumed under the null hypothesis, the estimated parameters, the estimation method and the sample size. Therefore, as Stephens (1986) and Gibbons and Chakrabarty (1992) warn, the critical values obtained considering that the null hypothesis is simple, can lead to false conclusions.

of 0.3 to members under 14 years of age. The resulting income is the equivalent disposable income, which is attributed to each member of the household.

The reference period of income is the previous year for all countries except the United Kingdom, whose reference period is the year of the survey, and Ireland, which conducts the survey continuously and collected data on income during the twelve months prior to the interview.

Table 1 shows the sample sizes, the median incomes, and the Gini index for each of the empirical income distributions obtained from the EU-SILC samples of the 25 countries considered in the waves of 2006, 2011, and 2016. Throughout the period of study, generalised increases in the median income can be observed between the different years, with the exception of the United Kingdom between 2006 and 2011, and Greece, Spain and Cyprus between 2011 and 2016. Between 2006 and 2016, Greece was the only economy that experienced a fall in nominal median income. Regarding the evolution of inequality, measured using the Gini index, different behaviour is observed between the countries. Thus, between 2006 and 2016, inequality grew in 13 countries and decreased in 12. Between 2006 and 2011, and between 2011 and 2016, 12 and 13 countries, respectively, experienced growth in the Gini index.

Table 1. Sample Sizes, Median Income and Gini Coefficient

	2006			2011			2016		
	N	Median	Gini	N	Median	Gini	N	Median	Gini
Austria	14883	17852.36	0.2533	13933	21462.56	0.2744	13049	23694.33	0.2694
Belgium	14329	17194.07	0.2724	14300	20007.89	0.2627	13773	22295.14	0.2603
Cyprus	11069	14532.30	0.2876	11443	16990.35	0.2917	11236	14020.00	0.3208
Czech R.	17830	4796.77	0.2530	20629	7451.35	0.2524	18964	7837.97	0.2503
Germany	31777	15616.67	0.2604	28644	19042.67	0.2896	26699	21358.00	0.2906
Denmark	14676	22662.79	0.2299	13151	26944.41	0.2656	13846	28664.87	0.2712
Estonia	15811	3637.91	0.3284	13426	5597.89	0.3192	15193	8644.68	0.3239
Greece	15190	9850.00	0.3368	15067	10984.78	0.3352	44094	7500.00	0.3375
Spain	34515	11433.50	0.3082	34756	13929.46	0.3404	36380	13680.87	0.3413
Finland	28039	18225.24	0.2579	23018	21826.19	0.2582	25983	23650.00	0.2540
France	24940	16186.88	0.2718	27071	20201.00	0.3083	26647	21710.00	0.2915
Hungary	19902	3847.19	0.3272	29474	4493.14	0.2693	18809	4767.76	0.2790
Ireland	14634	19679.36	0.3192	11005	19726.18	0.2980	13186	22407.48	0.2942
Italy	54512	14520.00	0.3164	47841	15970.50	0.3246	48316	16247.00	0.3221
Lithuania	12134	2532.12	0.3473	12492	3856.86	0.3301	10905	5644.92	0.3645
Luxembourg	10242	29480.38	0.2752	14891	32538.00	0.2725	10159	33818.27	0.3070
Latvia	10986	2533.76	0.3860	15891	4195.11	0.3506	13864	6364.63	0.3418
Netherlands	23096	17259.58	0.2555	25461	20309.64	0.2585	29559	22733.21	0.2651
Norway	14724	27784.92	0.2778	11730	36395.96	0.2291	16900	39572.88	0.2474
Poland	45122	3111.40	0.3319	36720	5025.36	0.3106	32609	5883.58	0.2966
Portugal	12071	7310.71	0.3767	14662	8409.51	0.3424	26565	8782.30	0.3389
Sweden	17149	17730.13	0.2308	16665	22506.42	0.2440	14072	25164.16	0.2735
Slovenia	31276	9318.38	0.2373	28747	11990.57	0.2383	25637	12320.00	0.2432
Slovakia	15147	3313.28	0.2801	15335	6305.94	0.2567	16507	6950.93	0.2403
United K.	23365	19307.05	0.3201	18670	17135.64	0.3304	22205	21150.00	0.3075

6. Results

As previously mentioned, for comparative purposes, six models have been fitted: the log *t* distribution, the lognormal distribution, the Dagum distribution, the Singh-Maddala distribution, the κ-generalized distribution, and the generalized beta of the second kind distribution (GB2).

After carrying out the estimation processes of the parameters, we obtained the values of statistics of goodness of fit for each fitted distribution presented in tables A.1, A.2, and A.3 of the annex. These results are summarised in Table 2, which identifies the distributions that provide the best fits according to the Kolmogorov-Smirnov (K-S), the Anderson-Darling (A-D), and the sum of squared residuals statistics (SSR). From this synthesis of results, it is observed that the log *t* distribution is the distribution that best fits the empirical distributions of European countries in the three years of the survey (86.7% of the 225 comparisons).

Indeed, in 2006, this distribution achieved the best fit in 20 countries, according to the Kolomogorov-Smirnov values, i.e. 22 countries according to the Anderson-Darling statistic, and in 20 countries according to the SSR statistic. In 2011, the log *t* distribution provides the best fit in 22 countries according to the Kolomogorov-Smirnov statistic, i.e. 21 countries according to the Anderson-Darling statistic, and in 22 countries according to the SSR statistic. In 2016, the results are similar since the log *t* distribution presents the best fit in 23 of the 25 countries according to the Kolomogorov-Smirnov statistic, in 22 countries according to the Anderson Darling statistic, and in 23 countries according to the SSR statistic.

Table 2. Best Fits According to the Goodness-of-Fit Statistics: Kolmogorov-Smirnov (K-S), Anderson-Darling (A-D) and the Sum of Squared Residual (SSR)

	2006				2011				2016					
	K-S	A-D	SSR	Average range	K-S	A-D	SSR	Average Range	K-S	A-D	SSR	Average range	Total	%
Lognormal	0	0	0	6.0	0	0	0	6.0	1	0	0	5.9	1	0.4
Dagum	1	0	0	4.2	0	0	1	4.2	0	0	0	4.4	2	0.9
Singh-Maddala	3	0	3	3.3	1	1	0	3.3	0	0	0	3.6	8	3.6
log *t*	20	22	20	1.5	22	21	22	1.5	23	22	23	1.2	195	86.7
κ-generalized	0	0	0	2.9	1	1	0	2.9	1	1	1	2.7	5	2.2
GB2	1	3	2	3.1	1	2	2	3.1	0	2	1	3.1	14	6.2

If the focus is on cases in which the goodness-of-fit measures indicate that the log *t* distribution presents worse fits, the first finding is that they represent a percentage of 13.3% (30 of the 225 comparisons). Among these cases, Slovenia is consistently the case in which the other models perform better than the log *t* distribution, which only provides the best fit according to one statistic. Portugal, Cyprus, and the Czech Republic also present worse fits using the log *t* in several cases. The remaining cases where the log *t* distribution is not the best model are isolated and refer to distributions in which the other models present better fits according to only one or two statistics. After the log t distribution, the GB2 distribution is better in fourteen cases, the Singh-Maddala distribution is better in eight cases, the κ-generalized distribution is better in five cases, the Dagum distribution in two cases, and the lognormal distribution in one case. If the average ranges in the different rankings of distributions, derived from goodness-of-fit statistics, are considered, the best adjustments, after log t distribution are, in this order, given by the κ-generalized distribution, the generalized beta of the second kind, the Singh-Maddala model, the Dagum distribution, and the lognormal, which is only the best model in one of the comparisons made.

In the comparisons in which the log t distribution presents the best fits (86,7%% of the cases), a clear superiority is also detected over the other models in terms of the difference between the values obtained for the goodness-of-fit statistics. Therefore, according to these three statistics we can confirm that of the distributions studied, the log t distribution clearly produces the best fits thanks to – among other factors – its flexibility to reproduce the lower tail of distribution, where zero and negative incomes are located, along with the kurtosis of income data.

The good performance of the log t for the different percentiles of the distribution can be observed in the probability plots (PP-plots) presented in the annex (Figures A.3, A.4, and A.5). The figures clearly show how the behaviour of this distribution is the best in the vast majority of cases, as indicated by the different goodness of fit measures. In particular, it is clear that the Dagum, the κ-generalized, and the Singh-Maddala distributions are not able to adequately fit the lower tail and, in general, the first two quartiles of the distribution, while the log t distribution can adapt to the lowest and middle incomes. The lognormal performs worst of all, both in the first quartile and in the middle and highest incomes.

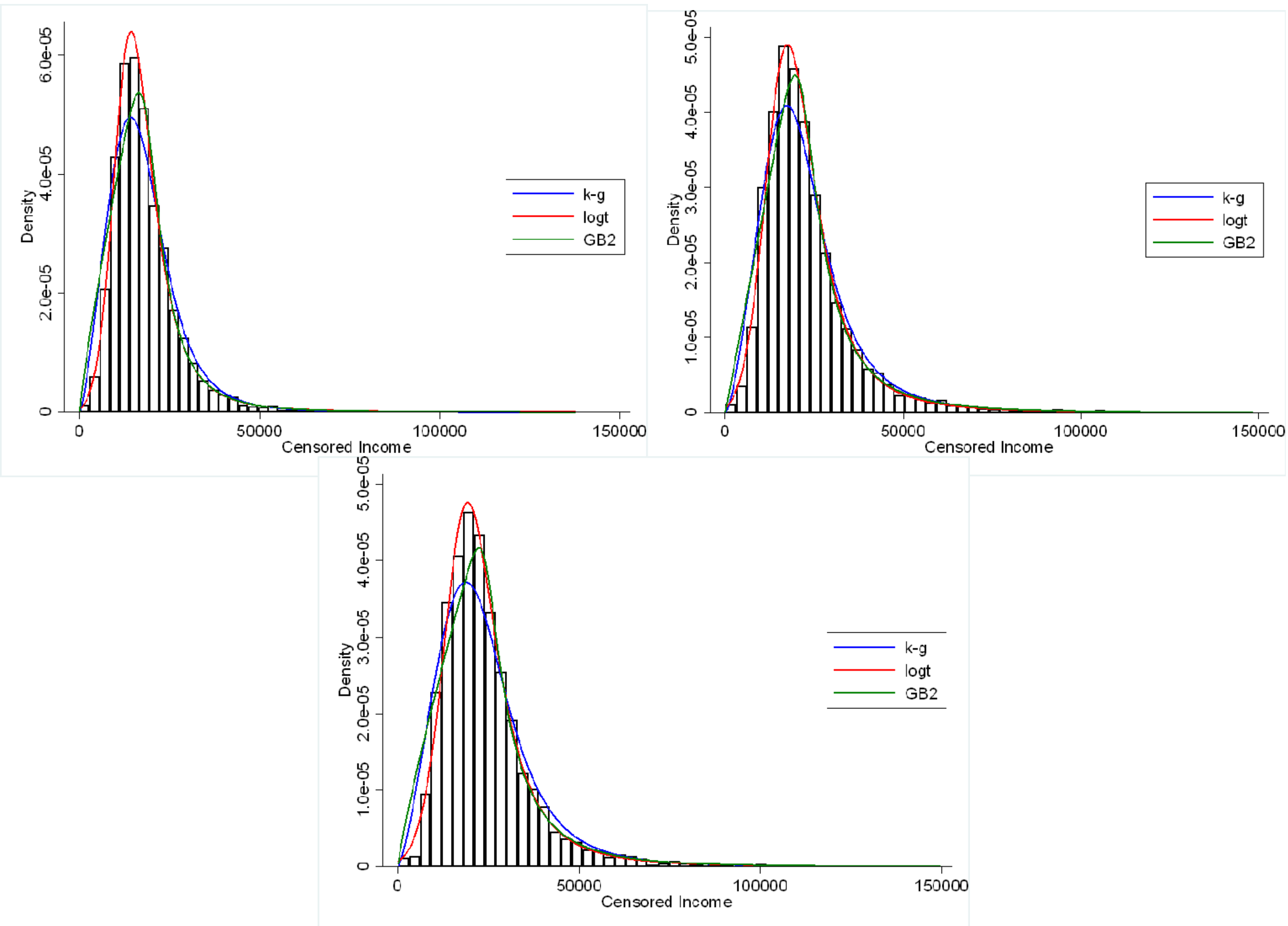
In order to further the appropriateness of the fits in the different parts of income distributions, goodness-of-fit statistics have also been calculated separately for each quartile. The SSR statistic, which can be obtained for the different income intervals, has been used. A summary of the results of the comparisons for each of the quartiles (Tables A.4–A.10), according to the goodness-of-fit statistics, is presented in Table 3. These results show that the log t distribution clearly produces the best fits in the first three quarters of the distribution, which is to be expected, given the good properties of the log t distribution to model low incomes and the kurtosis of the distribution. However, it should also be noted that, even though it does not satisfy the weak Pareto law, the slow convergence of the income elasticity of the survival function leads to the log t distribution presenting an acceptable fit in the upper quartile, similar to other distributions that satisfy the weak Pareto law. As a result, in the fourth quartile of the empirical income distributions, the log t presents the best fits in 9.3% of the comparisons, which is surpassed only by the Dagum distribution, which evidences better fits in nine cases (12% of the total) and by the GB2 distribution (with four parameters), which is the best model for the fourth quartile. Lognormal, Singh Maddala and κ-generalized distributions perform worse in the fourth quartile.

Table 3. Best Fits According to the Goodness-of-Fit Statistics Sum of Squared Residuals for the Distribution Quartiles

	2006				2011				2016			
Quartiles	*1*	*2*	*3*	*4*	*1*	*2*	*3*	*4*	*1*	*2*	*3*	*4*
Lognormal	0	0	0	0	0	0	0	0	0	0	1	0
Dagum	0	1	0	4	0	0	1	4	0	0	0	1
Singh-Maddala	1	2	3	2	0	2	0	2	0	1	0	1
log t	21	20	17	5	23	19	12	0	23	21	17	2
κ-generalized	0	1	0	1	1	0	0	1	1	1	2	3
GB2	3	1	5	13	1	4	12	18	1	2	4	18

For illustrative purposes, Figure 7 shows the histograms and the density functions of the three models that provide the three best fits to the income distributions of France in 2006, 2011, and 2016. In the three years considered, an excellent behaviour of the log *t* distribution is observed to describe income distribution, with particular precision in the lower tail. It also highlights its ability to reproduce the behaviour of incomes around the mode and the kurtosis of empirical income distributions. As in many other countries, the other distributions overestimate the number of recipients with low incomes, which causes underestimates of the average income.

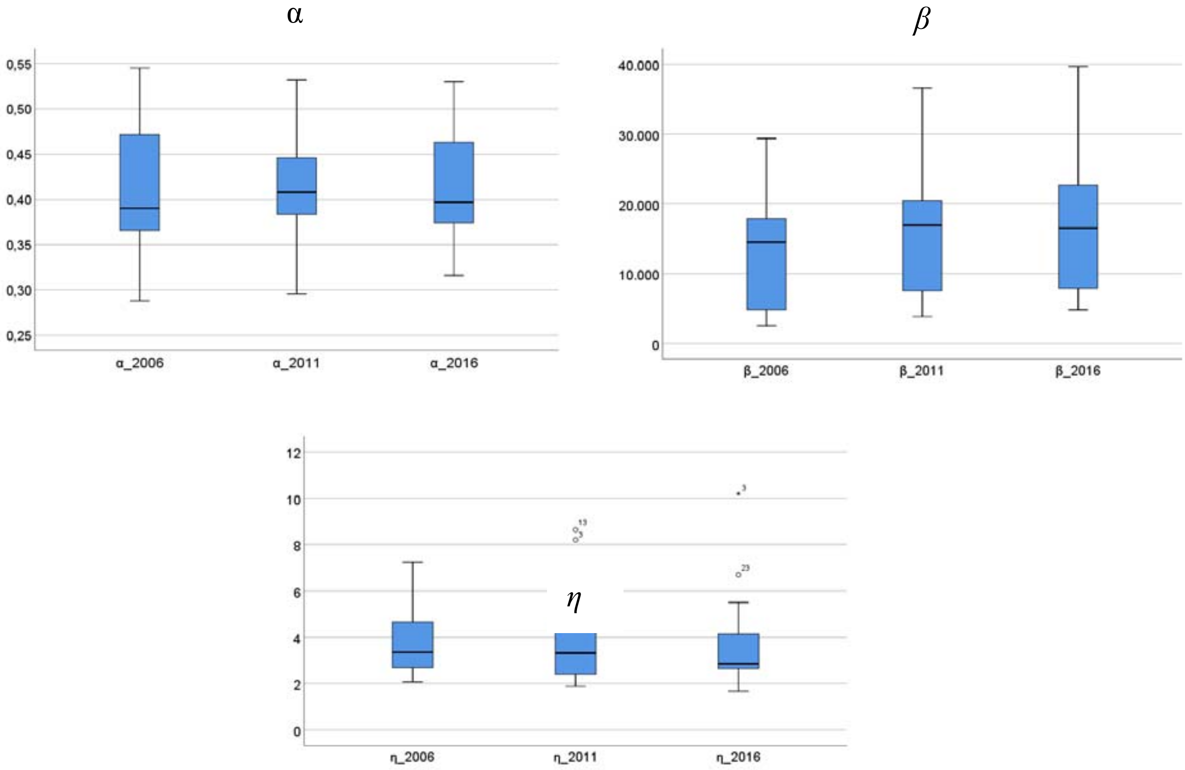
Figure 7. Histograms and Density Functions. France (2006, 2011, and 2016)



The parameter estimates of the log *t* distributions and their corresponding standard errors, for all countries and years, are presented in Table A.11. As indicated, the parameters of the log *t* can be interpreted as economic indicators of income level and distribution inequality. In Figure 8, which presents the box plots of the parameter estimates in 2006, 2011, and 2016, the increase in the nominal levels of income can be observed through the evolution of parameter β, which is the median of income distributions. The slight growth of β estimates only occurs in nominal terms. On the other hand, the box plots also show an increase in the dispersion of the β estimates, which indicates a growth in disparities between countries.

The estimates of parameter α, whose increases produced an increase in inequality, show a growth in mean and median in 2011, and a fall in 2016, returning to the levels of 2006. The estimates of the α parameter show similar variations to those experienced by the values of the Gini index (Table 1) obtained for the different countries analysed. The median of the estimates of parameter η remain stable in 2011 and decrease in 2016.

Figure 8. Box Plots of Estimates of Parameters α, β, and η (2006, 2011, and 2016)



7. Conclusions

This paper analyses the relevance of the log Student's *t* model to fit and describe income distributions in terms of their economic foundation, their theoretical properties, and the results obtained for a wide group of countries in the European Union in three time periods – 2006, 2011 and 2016 – during different phases of the economic cycle. Considering all of these aspects jointly makes it possible to conclude that log *t* distribution is an appropriate model for income distributions and therefore a useful tool for analysing the effect of income redistribution policy measures; in particular, those that focus on the left tail of distribution, related to poverty eradication, as well as others that focus on the middle classes.

With regard to its properties, the model's bimodal nature constitutes a clear advantage over other models, allowing the possibility of modelling incomes very close to zero or zero thanks to its mode at the origin. The existence of these incomes is a confirmed fact that is exacerbated in periods of crisis and in certain types of countries. A very important feature of log *t* distribution is its capacity to capture the mode and the kurtosis of the empirical income distributions, an advantage that is not highlighted in the studies on modelling the personal distribution of income, which are usually devoted to the study of the distribution tails.

An economic justification for the use of the log *t* model is also provided in this paper. Thus, the log *t* model arises from the process of income generation derived from a lognormal distribution that is corrected using a measure of the heterogeneity of individuals, which control the cognitive abilities, socio-economic background, or personality traits, among others individual factors. In addition, its parameters may be interpreted clearly, highlighting the capacity of parameter α as an indicator of inequality that, in practice, presents a high correlation with the Gini index.

In order to obtain the moments of the log *t* distribution, we have provided a compact form that makes use of the Student's t distribution. From these expressions, it has been verified that the integrals do not converge and, consequently, that the moments do not exist. Therefore, the log *t* distribution has no moments, but at the limit when $\eta \to \infty$ it has them all. This theoretical limitation can be overcome because the distribution can always be truncated using a realistic high value of income. Nonetheless, the non-existence of moments supposes that they cannot be used for statistical inference of the distributional parameters, although alternative measures based on quantiles could be used.

By analysing the elasticity of the survival function, it may be observed that the distribution does not obey the weak Pareto law, although the convergence of the elasticity of the survival function is very slow and a strictly negative value can be established for this elasticity, which is fairly stable for the highest incomes. The slow convergence of the Pareto exponent and its stabilisation in the values observed in practice causes the log *t* distribution to produce acceptable adjustments in the fourth quarter of the empirical income distributions; even better than some of the models that comply with the weak Pareto law. In any case, it should be noted that the performance of GB2 is better in the fourth quartile, although this distribution has one more parameter.

The comparative results obtained for the countries of the European Union clearly show the superiority of the log *t* model over the other four models, in particular, in 86.7% of the cases analysed. This is true regardless of the goodness-of-fit statistics used. In addition, the PP-plots obtained demonstrate the ability of the log t distribution to reproduce the lower tail and the kurtosis of the distributions, unlike the other distributions with which it is compared. The good fits are stable over time and also for the different countries, showing few cases in which, the log t distribution is not superior to the other models.

From this study, different practical implications are derived. The first one is the applicability of the log *t* distribution for the study of the left tail of the distribution and the kurtosis of income distributions. In relation to these advantages, the log t distribution will also be an appropriate model for the study of poverty measures and the simulation of the effects of economic policy measures directed at individuals in the left tail, as well as income intervals with a greater number of individuals, close to the income modes. Future research lines are also opened, such as the application of the log t distribution to developing countries, where numerous individuals have incomes close to zero, by performing comparative international analysis using harmonised information that is currently available in certain regional contexts. Finally, it would be interesting to examine the performance of mixtures that combines the log *t* distribution, for the first three quarters of distribution, and the Pareto distribution for the highest incomes.

References

Aitchinson, J. and Brown, J.A.C. (1957). The Lognormal Distribution, Cambridge University Press.

Anderson, T. W., and Darling, D. A. (1954). A test of goodness of fit. *Journal of the American statistical association*, *49*(268), 765-769.

Atkinson, A. B., Piketty, T., and Saez, E. (2011). Top incomes in the long run of history. *Journal of economic literature*, *49*(1), 3-71.

Atkinson, A. B. (2017). Pareto and the upper tail of the income distribution in the UK: 1799 to the present. *Economica*, *84*(334), 129-156.

Bandourian, R., McDonald, J. B. and Turley, R. S. (2003). Income Distributions: An Inter-temporal Comparison over Countries, *Estadistica,* 55, 135-152.

Bartels, C.P.A. (1977) Economic Aspects of Regional Welfare. *Martinus Nijhoff Social Sciences Division.*

Boccanfuso, D., Richard, P., and Savard, L. (2013). Parametric and nonparametric income distribution estimators in CGE micro-simulation modeling. *Economic Modelling*, *35*, 892-899.

Bordley, R. F., McDonald, J. B. and Mantrala, A. (1996). Something New, Something Old: Parametric Models for the Size Distribution of Income, *Journal of Income Distribution*, 6 (1), 91-104.

Brachmann, K., Stich, A. and Trede, M. (1996). Evaluating Parametric Income Distributions Models, *Allgemeines Statistisches Archiv*, 80, 285-298.

Callealta, J., Casas, J.M. and Núñez, J. J. (1996). Distribución de la renta per capita disponible en España: descripción, desigualdad y modelización en Pena B., J. Callealta, J.M. Casas, A. Merediz y J. J. Núñez (eds.), Capítulo 5. Editorial Pirámide.

Chattopadhyay, A. K., Ackland, G. J. & Mallick, S. K. (2010) Income and poverty in a developing economy. Euro physics Letters, 91 (5), 58003.

Chattopadhyay, A.K., Kumar, T.K., Mallick, S.K. (2017) Poverty index with time-varying consumption and income distributions, Physical Review E, 95 (3), art. no. 032109.

Chattopadhyay, A.K., Mallick, S.K. (2007) Income distribution dependence of poverty measure: A theoretical analysis, Physica A: Statistical Mechanics and its Applications, 377 (1): 241-252.

Chotikapanich, D., and Griffiths, W. E. (2008). Estimating income distributions using a mixture of gamma densities. In *Modeling Income Distributions and Lorenz Curves* (pp. 285-302). Springer, New York, NY.

Chotikapanich, D., G. Hajargasht, W. E. Griffiths and C. Xia (2014). Inequality and Poverty in Africa: Regional Updates and Estimation of a Panel of Income Distributions, Paper Prepared for the IARIW 33rd General Conference, Rotterdam, August 24-30, 2014.

Chotikapanich, D., W. Griffiths, W. Karunarathne, and D. S. Prasada Rao (2013). Calculating Poverty Measures from the Generalized Beta Income Distribution, *Economic Record*, 89, 48-66.

Clementi, F., and Gallegati, M. (2016). *Distribution of Income and Wealth*. Springer International Publishing.

Clementi, F. M. Gallegati, G. Kaniadakis, and S. landini (2006), κ-Generalized Models of Income and Wealth Distributions: A Survey, *The European Physical Journal Special Topics*, October 2016, Vol. 225, 10, pp 1959–1984.

Cummins, J. D., Dionne, G., McDonald, J. B., & Pritchett, B. M. (1990). Applications of the GB2 family of distributions in modeling insurance loss processes. *Insurance: Mathematics and Economics*, *9*(4), 257-272.

Dagum, C. (1977). A New Model of Personal Income Distribution: Specification and Estimation, *Economie Appliqueé*, 30, 413-436.

Dagum, C. (1980). Sistemas Generadores de la Distribución del Ingreso y la Ley de Pareto, *El Trimestre Económico*, 188, 877-917.

Dagum, C. (1990). Generation and Properties of Income Distribution Functions, *Studies in Contemporary Economics*, University of Ottawa.

Domma, F., and Condino, F. (2013). The beta-Dagum distribution: definition and properties. *Communications in Statistics-Theory and Methods*, *42*(22), 4070-4090.

Dragulescu, A. A., and Yakovenko, V. M. (2001). Evidence for the Exponential Distribution of Income in the USA. *The European Physical Journal B*, 20, 585–589.

García, C., Prieto, M., and Simón, H. (2014). La modelización paramétrica de las distribuciones salariales. *Revista de Economía Aplicada*, *22*(65).

Ghosh, J. K. (1994). Higher Order Asymptotics, Institute of Mathematical Statistics.

Gibbons, J. and Chakrabarty, S. (1992). Nonparametric Statistical Inference, Marcel Dekker.

Gibrat, R., (1931). Les Inegalites Economiques. Sirely, Paris.

Graf, M. and Nedyalkova, D. (2014). Modeling of Income and Indicators of Poverty and Social Exclusion Using the Generalized Beta Distribution of the Second Kind, *Review of Income and Wealth*, 821–42.

Graf, M., and Nedyalkova, D. (2017). Discretizing a compound distribution with application to categorical modelling. *Statistics*, *51*(3), 685-710.

Hogg, R. V., and Klugman. S.A. (1983). On the Estimation of Long-Tailed Skewed Distributions with Actuarial Data, *Journal of Econometrics*, 23, 91-102.

Jenkins, S. P. (2007). SMFIT: Stata module to fit a Singh-Maddala distribution by maximum likelihood.

Jenkins, S. P. (2012). GB2FIT: stata module to fit generalized beta of the second kind distribution by maximum likelihood.

Jenkins, S. P. (2013). LOGNFIT: stata module to fit lognormal distribution by maximum likelihood.

Jenkins, S. P. (2014). DAGUMFIT: Stata module to fit a Dagum distribution by maximum likelihood.

Jenkins, S. P. (2017). Pareto models, top incomes and recent trends in UK income inequality, *Economica*, *84*(334), 261-289.

Jenkins, S. P., Brandolini, A., Micklewright, J., and Nolan, B. (2012). *The great recession and the distribution of household income*. OUP Oxford.

Johnson, L. J., Kotz, S. K., & Balakrishnan, N. (1995). Limiting Distributions of Extremes. *Continuous univariate distributions*, *2*.

Kleiber, C. (1996). Dagum vs. Singh-Maddala expenditure distributions, *Economics Letters*, 53, 265-268.

Kleiber, C. and Kotz. S. (2003), Statistical Size Distributions in Economics and Actuarial Sciences, *Wiley Series in Probability and Statistics*.

Kloek, T., and H. K. Van Dijk (1978). Efficient estimation of income distribution parameters, *Journal of Econometrics*, 8, 61-74.

Kloek, T., and H. Van Dijk (1977). Further results on efficient estimation of income distribution parameters, *Economie Appliquée*, 30, 439-459.

Majumder, A. and Chakravarty, S. R. (1990). Distribution of Personal Income: Development of a New Model and its Application to US Income Data, *Journal of Applied Econometrics*, 5, 189-96.

McDonald, J. B. (1984). Some Generalized Function for the Size Distribution of Income. *Econometrica*, 52, pp. 647-664.

McDonald, J. B. and Butler, R. J. (1987). Some generalized mixture distributions with an application to unemployment duration. *The Review of Economics and Statistics*. 69, 232-240.

McDonald, J. B. and Xu, Y. J. (1995). A Generalization of the Beta Distribution with Applications, *Journal of Econometrics*, 66, pp. 133-152.

OECD (2015). *Final NAEC Synthesis*, Paris.

OECD (2013). *It's all about people: jobs, equality, and trust*, May 2013, Paris.

Oluyede, B. O., Huang, S., and Pararai, M. (2014). A new class of generalized Dagum distribution with applications to income and lifetime data. *Journal of Statistical and Econometric Methods*, *3*(2), 125-151.

Pareto, V. (1896). Ecrits sur la courbe de la répartition de la richesse, in *Oeuvres complètes de Vilfredo Pareto*, Libraire Droz, Ginebra, 1965.

Pena, J.B. et al (1998). Distribución personal de la renta en España. Pirámide (Madrid-Spain).

Piketty, T. (2014). Capital in the Twenty-First Century. Cambridge, MA: Harvard University Press.

Piketty, T. (2018). Capital in the 21st Century. In *Inequality in the 21st Century*. Routledge.

Pinkovskiy, M. and X. Sala-i-Martin (2009). Parametric Estimations of the World Distribution of Income, *NBER Working Papers* 15433, National Bureau of Economic Research.

Pinkovskiy, M. and Sala-i-Martin, X. (2016). Lights, Camera… Income! Illuminating the national accounts-household surveys debate. *The Quarterly Journal of Economics*, *131*(2), 579-631.

Prieto-Alaiz, M. and Victoria-Feser, M. P. (1996). Modelling Income Distribution in Spain: A Robust Parametric Approach, DARP discussion paper, 20.

Prieto-Alaiz, M. P., and García-Pérez, C. G. (2009). La distribución beta generalizada de segunda especie como modelo de la distribución personal de la renta en España. *Estadística española*, *51*(170), 33-62.

Rao, C. (1973), *Linear Statistical Inference and its Applications*, John Wiley & Sons.

Reed, W. J., and Wu, F. (2008). New four-and five-parameter models for income distributions. In *Modeling Income Distributions and Lorenz Curves* (pp. 211-223). Springer, New York, NY.

Sala-i-Martin, X. and Pinkovskiy, M. (2010). *African poverty is falling... much faster than you think!* (N. w15775). National Bureau of Economic Research.

Salem, A.B., and Mount, T.D. (1974) A convenient descriptive model of income distribution: the gamma density, *Econometrica*, 42 (1974), 1115-1127

Sarabia, J. M., and Jordá, V. (2014). Explicit Expressions of the Pietra Index for the Generalized Function for the Size Distribution of Income. *Physica* A 416, 582–89.

Sarabia, J. M., V. Jordá, and L. Remuzgo (2017). The Theil Indices in Parametric Families of Income Distributions. A Short Review, *Review of Income and Wealth*, 63: 867–80.

Silva, C. A., and Yakovenko, V. M. (2004). Temporal Evolution of the "Thermal" and "Superthermal" Income Classes in the USA during 1983–2001. *Europhysics Letters*, 3 31, 1–7.

Singh, S. K. and Maddala, G. S. (1976). A Function for the Size Distribution of Income, *Econometrica*, 44, 963-970.

Smeeding, T. (2012). Income, wealth, and debt and the Great Recession. *Report. New York, NY, The Russell Sage Foundation and the Stanford Center on Poverty and Inequality*.

StataCorp (2017). Stata: Release 15. Statistical Software. College Station, TX: StataCorp LLC.

Stephens, M.A. (1986). Test Based on EDF Statistics in D'Agostino, R.B. y Stephens, M.A. *Goodness of Fit Techniques*, chapter 4, Marcel Dekker.

Tahir, M. H., and Cordeiro, G. M. (2016). Compounding of distributions: a survey and new generalized classes. *Journal of Statistical Distributions and Applications*, 3(1), 13-26.

Tahir, M. H., Gauss M. Cordeiro, M. Mansoor, M. Zubair and Morad Alizadeh (2016) The Weibull–Dagum distribution: Properties and applications, *Communications in Statistics - Theory and Methods*, 45 (24), 7376-7398.

Thurow, L. C. (1970). Analyzing the American income distribution. *The American Economic Review*, *60*(2), 261-269.

Vinh, A., Griffiths, W. E., and Chotikapanich, D. (2010). Bivariate income distributions for assessing inequality and poverty under dependent samples. *Economic Modelling*, 27(6), 1473-1483.

Yakovenko, V. M. (2007). Statistical Mechanics Approach to Econophysics, arXiv:0709.3662 (November 2).