Statistica & Applicazioni - Vol. X, n. 2, 2012, pp. 131-147

FIRST APPLICATIONS OF A NEW THREE-PARAMETER DISTRIBUTION FOR NON-NEGATIVE VARIABLES

Michele M. Zenga*
Leo Pasquazzi*
Mariangela Zenga*

SUMMARY

Zenga (2010a) recently proposed a new three-parameter family of density functions for non-negative variables. Its properties resemble those of economic size distributions: it has positive asymmetry, Paretian right tail and it may be zeromodal, unimodal or even bimodal. In this paper we explore some methods for fitting the new density to empirical income distributions. We will see that D'Addario's invariants method clearly outperforms Pearson's moments method, which does not seem to work well with heavy tailed distributions. Further, we propose some new methods based on the minimization of a measure for the goodness of fit, imposing restrictions on the parameter space to preserve some features of the empirical distribution in the fitted model. We will see that these methods provide very satisfactory results with income distributions from Italy, Swiss, US and UK.

Keywords: *Income Distribution, Zenga's Distribution, Goodness of Fit, Moments Method, Invariants Method.*

1. INTRODUCTION

Recently Zenga (2010a) proposed a new three-parameter density function for non-negative variables. The density is obtained as a mixture of Polisicchio's (Polisicchio, 2008) following truncated Pareto densities

$$h(x:\mu;k)=\begin{cases}\frac{\sqrt{\mu}}{2}k^{0.5}(1-k)^{-1}x^{-1.5}, & \text{if } \mu k\le x\le \frac{\mu}{k};\\ 0, & \text{otherwise,}\end{cases}$$

with $\mu > 0$ fixed, and $0 < k < 1$. The weights of the mixture are put on the parameter k and are given by the beta density

$$g(k:\alpha;\theta)=\begin{cases}\frac{k^{\alpha-1}(1-k)^{\theta-1}}{B(\alpha;\theta)}, & \text{if } 0<k<1;\\ 0, & \text{otherwise,}\end{cases}$$

* Dipartimento di Statistica e Metodi Quantitativi - Università Milano-Bicocca - Via Bicocca degli Arcimboldi 8 - 20126 MILANO (e-mail: michele.zenga@unimib.it; ✉ leo.pasquazzi@unimib.it; mariangela.zenga@unimib.it).

where $B(\alpha;\theta)$, with α and θ positive, is the beta function. The new density is thus given by

$$f(x:\mu;\alpha;\theta) = \begin{cases} \frac{1}{2\mu B(\alpha;\theta)}\left(\frac{x}{\mu}\right)^{-1.5} IB\left(\frac{x}{\mu}:\alpha+0.5,\theta-1\right), & \text{if } 0<x<\mu \\ \frac{1}{2\mu B(\alpha;\theta)}\left(\frac{\mu}{x}\right)^{1.5} IB\left(\frac{\mu}{x}:\alpha+0.5,\theta-1\right), & \text{if } \mu<x. \end{cases}$$

where

$$IB(x:\alpha,\theta) = \int_0^x k^{\alpha-1}(1-k)^{\theta-1}dk, \qquad 0 \le x \le 1.$$

is the incomplete beta function. It is well defined only if α and θ are both positive. However, in this work we will use it also for negative values of θ, in which case $IB(x:\alpha,\theta)$ is not defined at $x=1$, but

$$\lim_{x\to 1} IB(x:\alpha,\theta) = \infty.$$

Note that:

$$\lim_{x\to 0} f(x:\mu;\alpha;\theta) = \begin{cases} 0, & \text{for } \alpha > 1 \\ \frac{1}{3}\frac{\theta}{\mu}, & \text{for } \alpha = 1 \\ \infty, & \text{for } 0 < \alpha < 1; \end{cases}$$

$$\lim_{x\to \mu} f(x:\mu;\alpha;\theta) = \begin{cases} \frac{B(\alpha+0.5;\theta-1)}{2\mu B(\alpha;\theta)}, & \text{if } \theta > 1 \\ \infty, & \text{if } 0 < \theta \le 1. \end{cases}$$

It is easily seen that μ is the expectation as well as the scale parameter for the new family of density functions. Moreover, it may be shown that the right tail is Paretian, i.e.

$$E(X^r) < \infty \quad \Leftrightarrow \quad r < \alpha + 1,$$

and that the asymmetry is always positive, i.e.

$$F(\mu:\mu;\alpha;\theta) \ge \frac{1}{2},$$

where $F(x:\mu;\alpha;\theta)$ is the cumulative distribution function. Thus, the median m cannot be larger than the expectation μ.

Expressions for the moments, the cumulative distribution function and some inequality measures are derived and reported in Zenga (2010a) and Zenga, Polisicchio, Zenga and Pasquazzi (2011).

If $r < \alpha + 1$, we have

$$E(X^r) = \frac{\mu^r}{2r-1}\frac{1}{B(\alpha;\theta)}\sum_{i=1}^{2r-1} B(\alpha - r + i;\theta).$$

For $\theta > 1$ this formula simplifies to

$$E(X^r) = \frac{\mu^r}{2r-1}\frac{1}{B(\alpha;\theta)}[B(\alpha - r + 1; \theta - 1) - B(\alpha + r; \theta - 1)].$$

The cumulative distribution function $F(x : \mu; \alpha; \theta)$ is given by

$$\frac{1}{B(\alpha;\theta)}\sum_{i=1}^{\infty}\left[IB\left(\frac{x}{\mu}; \alpha + i - 1; \theta\right) - \left(\frac{\mu}{x}\right)^{0.5} IB\left(\frac{x}{\mu}; \alpha + i - 0.5; \theta\right)\right], \qquad (1)$$

for $0 < x \leq \mu$, and by

$$1 - \frac{1}{B(\alpha;\theta)}\sum_{i=1}^{\infty}\left[\left(\frac{\mu}{x}\right)^{0.5} IB\left(\frac{\mu}{x}; \alpha + i - 0.5; \theta\right) - IB\left(\frac{\mu}{x}; \alpha + i; \theta\right)\right], \qquad (2)$$

for $x > \mu$. In the case $\theta > 1$, the formulae in (1) and (2) reduce to

$$\frac{1}{B(\alpha;\theta)}\left[IB\left(\frac{x}{\mu}; \alpha; \theta - 1\right) - \left(\frac{\mu}{x}\right)^{0.5} IB\left(\frac{x}{\mu}; \alpha + 0.5; \theta - 1\right)\right]$$

and

$$1 - \frac{1}{B(\alpha;\theta)}\left[\left(\frac{\mu}{x}\right)^{0.5} IB\left(\frac{\mu}{x}; \alpha + 0.5; \theta - 1\right) - IB\left(\frac{\mu}{x}; \alpha + 1; \theta - 1\right)\right],$$

respectively.

In this paper we will also need the Pietra relative mean deviation

$$P = \frac{E\{|X - \mu|\}}{2\mu} = 2F(1 : 1; \alpha; \theta) - 1$$

and Zenga's (2007) point inequality $A(x)$ at $x = \mu$, which is given by

$$A(\mu) = 1 - \frac{E\{X|X \leq \mu\}}{E\{X|X > \mu\}} = 1 - \left\{\frac{1 - F(1 : 1; \alpha; \theta)}{F(1 : 1; \alpha; \theta)}\right\}^2.$$

According to the graphs of $f(x : \mu; \alpha; \theta)$ reported in Zenga (2010a) and in Zenga *et al.* (2011), the new family of density functions allows for a wider variety of shapes than traditional three-parameter models for income distributions, as for example the Dagum model (Dagum, 1977). In this work we fit the new three-parameter density to several empirical income distributions, using the classical method of moments, D'Addario's method of invariants (D'Addario, 1939) and three other methods based on the minimization of Mortara's measure for the goodness of fit. To assess the performance of the methods for fitting the model, we rely on weighted averages of the absolute deviations between observed and theoretical class frequencies (see next section). The approaches for fitting the new model proposed in the present work have been compared with the maximum likelihood method in Zenga and Arcagni (2012), where the new model has been fitted to 114 empirical distributions from European countries. Comparisons with other models, in particular, with the Dagum, the Singh-Maddala and generalized log-normal models, have been analyzed by Porro and Arcagni (2012).

To deal with the numerical optimization problems that arise when we shall fit the new model to empirical distributions, we used the R implementation of the Nelder-Mead algorithm (Nelder and Mead, 1965) in the present work.

2. INDIVIDUAL OBSERVATIONS AND THEIR GROUPED FORMAT

From each empirical income distribution we have n individual values

$$0 \leq x_1 \leq x_2 \leq \cdots \leq x_i \leq \cdots \leq x_n,$$

which are grouped into $s = 25$ intervals as follows: we first approximate the values np'_j obtained from Table 1 by their nearest integers np_j to determine the (integer) frequencies n_j of the s intervals. This yields

$$n_j = np_j - np_{j-1}; \quad j = 1, 2, ..., s; \quad p_0 = 0.$$

The endpoints of the s intervals are then given by

$$x'_0 = 0; \quad x'_j = x_{np_j}, \quad j = 1, 2, ..., s-1; \quad x'_s = \infty.$$

Note that $x'_j = x_{np_j}$ is the np_j-th order statistic of the n individual observations and that the actual relative frequencies f_j of the intervals are given by

$$f_j = \frac{n_j}{n} \simeq p'_j - p'_{j-1}, \quad j = 1, 2, ..., s; \quad p'_0 = 0.$$

Finally, since the empirical value of the arithmetic mean

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i,$$

TABLE 1. - *Prefixed values of the relative frequencies f'_j and of the cumulative relative frequencies p'_j of $s = 25$ intervals*

j	f'_j	p'_j	j	f'_j	p'_j
1	0,010	0,010	14	0,050	0,750
2	0,010	0,020	15	0,050	0,800
3	0,015	0,035	16	0,050	0,850
4	0,015	0,050	17	0,050	0,900
5	0,050	0,100	18	0,020	0,920
6	0,050	0,150	19	0,015	0,935
7	0,050	0,200	20	0,015	0,950
8	0,050	0,250	21	0,010	0,960
9	0,050	0,300	22	0,010	0,970
10	0,100	0,400	23	0,010	0,980
11	0,100	0,500	24	0,010	0,990
12	0,100	0,600	25	0,010	1,000
13	0,100	0,700			

is very important in analyzing income distributions, it seems good practice to preserve this information when presenting the data in grouped form. We therefore want to have an endpoint equal to $\bar{x}$, which is achieved substituting the closest endpoint x'_j with $\bar{x}$.

Example 1

Applying the above procedure, we obtained from the $n = 13401$ individual incomes from the Bank of Italy 2006 Survey on Household Income and Wealth (Bank of Italy, 2008) the following interval distribution.

Now, let $\widehat{\mu}, \widehat{\alpha}, \widehat{\theta}$ denote the parameter values of the fitted model. The modelfrequencies $\widehat{n}_j$ of the $s = 25$ intervals are then given by

$$\widehat{n}_j = n\Big[F(x'_j : \widehat{\mu}; \widehat{\alpha}; \widehat{\theta}) - F(x'_{j-1} : \widehat{\mu}; \widehat{\alpha}; \widehat{\theta})\Big], \quad j = 1, 2, \ldots s.$$

Using the $\widehat{n}_j$'s, we may assess the goodness of fit by the following indexes: the Mortara index A_1, the quadratic K. Pearson index A_2 and the modified quadratic index A'_2.

$$A_1 = \frac{1}{n}\sum_{j=1}^{s}\frac{|n_j - \widehat{n}_j|}{\widehat{n}_j}\widehat{n}_j = \frac{1}{n}\sum_{j=1}^{s}\frac{|n_j - \widehat{n}_j|}{n_j}n_j = \frac{1}{n}\sum_{j=1}^{s}|n_j - \widehat{n}_j|$$

$$A_2 = \left\{\frac{1}{n}\sum_{j=1}^{s}\left|\frac{n_j - \widehat{n}_j}{\widehat{n}_j}\right|^2 \widehat{n}_j\right\}^{1/2} = \left\{\frac{1}{n}\sum_{j=1}^{s}\frac{|n_j - \widehat{n}_j|^2}{\widehat{n}_j}\right\}^{1/2}$$

$$A'_2 = \left\{\frac{1}{n}\sum_{j=1}^{s}\left|\frac{n_j - \widehat{n}_j}{n_j}\right|^2 n_j\right\}^{1/2} = \left\{\frac{1}{n}\sum_{j=1}^{s}\frac{|n_j - \widehat{n}_j|^2}{n_j}\right\}^{1/2}$$

TABLE 2. - *Bank of Italy 2006 Survey on Household Income and Wealth (Bank of Italy, 2008)*

j	np_j	x'_j	n_j	p_j	j	np_j	x'_j	n_j	p_j
1	134	801	134	0,010	14	10051	22545	670	0,750
2	268	1950	134	0,020	15	10721	24790	670	0,800
3	469	3000	201	0,035	16	11391	27696	670	0,850
4	670	3900	201	0,050	17	12061	32023	670	0,900
5	1340	5590	670	0,100	18	12329	35031	268	0,920
6	2010	7000	670	0,150	19	12530	37702	201	0,935
7	2680	8551	670	0,200	20	12731	41269	201	0,950
8	3350	10000	670	0,250	21	12865	44739	134	0,960
9	4020	11232	670	0,300	22	12999	49193	134	0,970
10	5360	13450	1340	0,400	23	13233	57063	134	0,980
11	6700	15450	1340	0,500	24	13267	72913	134	0,990
12	8413	18503	1713	0,628	25	13401	∞	134	1,000
13	9381	20586	968	0,700					

Note that A_1 is the weighted arithmetic mean of the absolute values of the relative frequency deviations $|\rho_j| = \frac{|n_j - \widehat{n_j}|}{\widehat{n_j}}$, with weights $\widehat{n}_j$, or the weighted arithmetic mean of the absolute values of the modified relative frequency deviations $|\rho'_j| = \frac{|n_j - \widehat{n_j}|}{n_j}$, with weights n_j. A_2 is the quadratic mean of the ρ_j's with weights $\widehat{n}_j$, while A'_2 is the quadratic mean of the ρ'_j's with weights n_j.

3. METHOD OF MOMENTS

In this section we fit the new three-parameter density to 18 empirical income distributions using the classical method of moments. The data have been obtained from National Public Statistical Offices. Empirical evidence suggests that the density is bounded in a neighborhood of μ. Therefore, we will assume $\theta > 1$.

Let

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad m_2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2, \quad m'_3 = \frac{1}{n}\sum_{i=1}^{n} x_i^3$$

be the mean, the variance and the third moment of an empirical income distribution, respectively. Having assumed $\theta > 1$, we may use the following formulae for their model counterparts:

$$\begin{cases} E(X) & = \mu \\ Var(X) & = \frac{\mu^2}{3}\frac{\theta(\theta+1)}{(\alpha-1)(\alpha+\theta)} \\ E(X^3) & = \frac{\mu^3}{5}\frac{1}{B(\alpha;\theta)}\{B(\alpha-2;\theta-1) - B(\alpha+3;\theta-1)\}. \end{cases}$$

According to the method of moments, we fit the new density solving the system

$$\begin{cases} \bar{x} & = \mu \\ m_2 & = \frac{\mu^2}{3}\frac{\theta(\theta+1)}{(\alpha-1)(\alpha+\theta)} \\ m'_3 & = \frac{\mu^3}{5}\frac{1}{B(\alpha;\theta)}\{B(\alpha-2;\theta-1) - B(\alpha+3;\theta-1)\} \end{cases} \tag{3}$$

We immediately see that $\widehat{\mu} = \bar{x}$ and that the parameters α and θ are determined by the equations

$$\begin{cases} 3\frac{m_2}{\bar{x}^2} & = \frac{\theta(\theta+1)}{(\alpha-1)(\alpha+\theta)} \\ 5\frac{m'_3}{\bar{x}^3} & = \frac{1}{B(\alpha;\theta)}\{B(\alpha-2;\theta-1) - B(\alpha+3;\theta-1)\}. \end{cases} \tag{4}$$

From the first equation in (4) we get

$$\alpha^2 + \alpha(\theta-1) - \left[\theta + \frac{1}{3}\frac{\bar{x}^2}{m_2}\theta(\theta+1)\right] = 0.$$

Thus,

$$\alpha(\theta) = \frac{-(\theta-1) \pm \sqrt{(\theta-1)^2 + 4\left[\theta + \frac{1}{3}\frac{\bar{x}^2}{m_2}\theta(\theta+1)\right]}}{2}$$
$$= \frac{-(\theta-1) \pm \sqrt{(\theta+1)^2\left[1 + \frac{4}{3}\frac{\bar{x}^2}{m_2}\frac{\theta}{\theta+1}\right]}}{2}$$
$$= \frac{-(\theta+1) + 2 \pm (\theta+1)\sqrt{1 + \frac{4}{3}\frac{\bar{x}^2}{m_2}\frac{\theta}{\theta+1}}}{2}$$
$$= \frac{(\theta+1)\left\{-1 \pm \sqrt{1 + \frac{4}{3}\frac{1}{cv^2}\frac{\theta}{\theta+1}}\right\}}{2} + 1,$$

where $cv^2 = \frac{m_2}{\bar{x}^2}$. The solution

$$\alpha^*(\theta) = \frac{(\theta+1)\left\{-1 - \sqrt{1 + \frac{4}{3}\frac{\bar{x}^2}{m_2}\frac{\theta}{\theta+1}}\right\}}{2} + 1 < 0$$

is not admissible because for $Var(X)$ to be finite, $\widehat{\alpha}$ has to be greater than 1. We are thus left with the solution

$$\widetilde{\alpha}(\theta) = \frac{(\theta+1)\left\{-1 + \sqrt{1 + \frac{4}{3}\frac{1}{cv^2}\frac{\theta}{\theta+1}}\right\}}{2} + 1 > 1. \tag{5}$$

Note that $\widetilde{\alpha}(\theta)$ is an increasing function of θ. Since the empirical value of the third moment m_3' is necessarily finite, we will be interested in the set of values of θ, such that $E(X^3) < \infty$, or equivalently, such that $\widetilde{\alpha}(\theta) > 2$. From (5) we get

$$\widetilde{\alpha}(\theta) = 2 \quad \Leftrightarrow \quad (\theta+1)\left\{-1 + \sqrt{1 + \frac{4}{3}\frac{1}{cv^2}\frac{\theta}{\theta+1}}\right\} = 2.$$

Substituting $\theta + 1 = y$, the above equation becomes $y^2 - y[1 + 3cv^2] - 3cv^2 = 0$. It follows that

$$y = \frac{[1 + 3cv^2] + \sqrt{[1 + 3cv^2]^2 + 12cv^2}}{2}$$

and

$$\theta = \frac{[1 + 3cv^2] + \sqrt{[1 + 3cv^2]^2 + 12cv^2}}{2} - 1. \tag{6}$$

Thus we have $\widetilde{\alpha} > 2$ if and only if θ is larger than the right hand side in (6, and $\widehat{\theta}$ is therefore a solution to the systems (3) and (4) only if it satisfies this constraint.

Now, we will exploit the above theory to define a procedure for finding the moment estimates $\widehat{\alpha}$ and $\widehat{\theta}$. Let $T(\alpha;\theta)$ be the right hand side of the second equation in (4) as a function of α and θ:

$$T(\alpha;\theta) = \frac{1}{B(\alpha;\theta)}\{B(\alpha-2;\theta-1) - B(\alpha+3;\theta-1)\} \tag{7}$$

Using some numerical algorithm we may first find a numerical solution $\widehat{\theta}$ to the equation

$$T(\widetilde{\alpha}(\theta);\theta) = 5\frac{m_3'}{\bar{x}^3},$$

which we then plug into $\widetilde{\alpha}(\theta)$ on the right hand side in (30) to get $\widehat{\alpha} = \widetilde{\alpha}(\widehat{\theta})$. The following example illustrates how this procedure works in practice.

Example 2

From the individual income distribution of the 2006 Bank of Italy Survey on Household Income and Wealth (Banca d'Italia, 2008), we obtain the following statistics

$$m = 13401; \qquad \bar{x} = 18502.7 \qquad m_2 = 341808922$$

$$m_3' = 1{,}163341\times 10^{14} \qquad cv = 0.9992 \qquad 5\frac{m_3'}{\bar{x}^3} = 91.8273.$$

Thus we have

$$\widetilde{\alpha}(\theta) = (\theta+1)\frac{-1+\sqrt{1+1.3355\frac{\theta}{\theta+1}}}{2} + 1;$$

Figure 1 reports the graphs of $\widetilde{\alpha}(\theta)$ and $T(\widetilde{\alpha}(\theta);\theta)$ for the empirical distribution examined in this example. Notice that the scales on the right and left vertical axes are different. The right vertical axis refers to range of the function $T(\widetilde{\alpha}(\theta);\theta)$. The equation

$$T(\widetilde{\alpha}(\theta);\theta) = 5\frac{m_3'}{\bar{x}^3}$$

has a solution at $\widehat{\theta} = 4.6497$. Plugging into $\widetilde{\alpha}(\theta)$, whose range is represented on the left ordinate axis, we get

$$\widehat{\alpha} = \widetilde{\alpha}(4.6497) = 2.2678.$$

Table 3 reports some results obtained applying the moments method for fitting Zenga's new density to empirical income distributions. Notice that the values of the measures for the goodness of fit A_1, A_2 and A_2' in Table 3 are rather large ($A_1 > 0.09777$). This suggests that the moments method is not adequate for fitting the new density. The bad performance may in part be explained by the fact that with

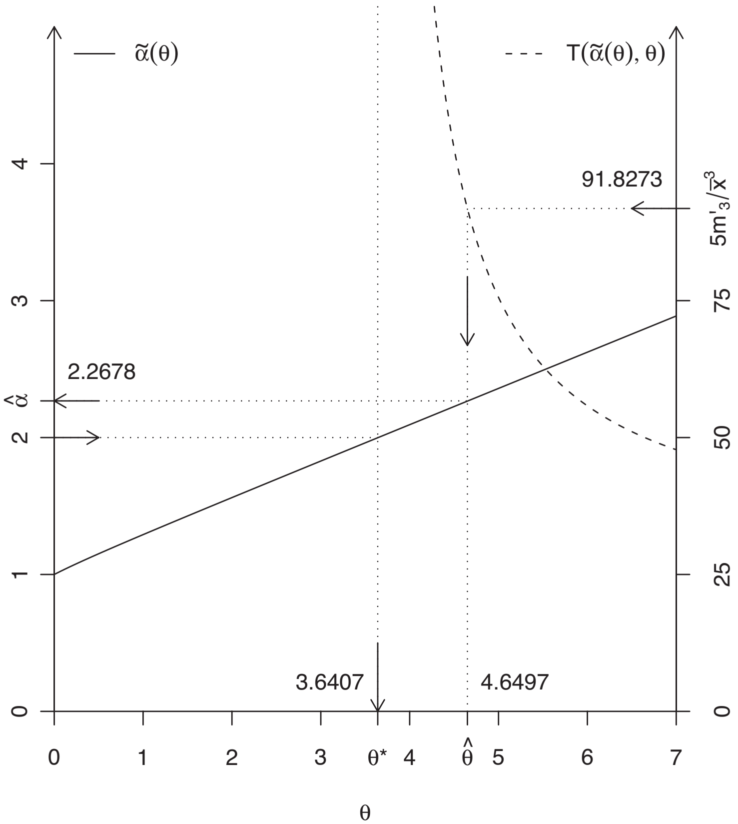

FIGURE 1. - *Graphs of* $\widetilde{\alpha}(\theta)$ *and* $T(\widetilde{\alpha}(\theta); \theta)$ *for* $cv = 0.9992$

this technique we can only yield values of α larger than 2. Indeed, $E(X^3) < \infty$ if and only if $\alpha > 2$.

TABLE 3. - *Moments method: estimated values of* $\widehat{\mu}$, $\widehat{\alpha}$, $\widehat{\theta}$ *and indexes of goodness of fit* A_1, A_2 *and* A_2' *for 35 income distributions*

dataset	n	$\widehat{\mu}$	$\widehat{\alpha}$	$\widehat{\theta}$	A_1	A_2	A_2'
CH 00 hh	3628	6764	2.2639	3.3526	0.18814	0.23478	0.31359
CH 01 hh	3729	6706	2.4945	2.4638	0.14328	0.17825	0.17144
CH 02 hh	3709	6915	2.8680	3.1782	0.10322	0.13110	0.12862
CH 03 hh	3454	6786	4.8198	5.7336	0.15884	0.24503	0.19554
CH 04 hh	3249	6560	3.9073	4.4095	0.09777	0.18888	0.14259
CH 05 hh	3071	6784	4.0210	4.8195	0.13758	0.17274	0.17007
I 06 hh eq	7762	19121	2.1939	3.6394	0.24793	0.29701	0.41231
I 06 hh	7762	31919	2.4447	4.0653	0.12317	0.16592	0.20295
I CE 06 hh eq	1574	21543	2.1880	5.9061	0.65516	0.73107	1.22580
I CE 06 hh	1574	35596	2.3365	4.2729	0.21665	0.28156	0.41298
I NW 06 hh eq	1987	21572	2.7120	3.0018	0.15727	0.19938	0.23182

(follows)

dataset	n	$\widehat{\mu}$	$\widehat{\alpha}$	$\widehat{\theta}$	A_1	A_2	A'_2
I NW 06 hh	1987	34693	2.5007	3.6521	0.13266	0.18971	0.24103
I S 06 hh eq	1618	13666	2.6086	3.6902	0.16516	0.23363	0.29830
I S 06 hh	1618	24550	2.4047	4.2064	0.15111	0.20619	0.25150
I 06 ind	13401	18503	2.2678	4.6497	0.25900	0.32080	0.29740
US 00 hh	5199184	56453	3.7480	11.0614	0.27891	0.68294	0.33490
US 08 hh	2899458	82460	3.9922	10.7071	0.26766	0.59313	0.32365
UK 99/00 hh	7086	446	2.8227	5.6448	0.13566	0.19394	0.24924

Countries: CH = Switzerland; *I* = Italy; *US* = United States; *UK* = United Kingdom. *Statistical unit: hh* = household; *ind* = individual. For Italy we consider also three macroregions (*CE* = Center; *NW* = Northwest; *S* = South) and we distinguish between actual household income (without label) and equivalent (labeled by *eq*) household income. The equivalent incomes have been obtained applying the modified Organization for Economic Cooperation and Development (OECD) equivalence scale, which gives weight 1 to the household head, 0.5 to the other adult members of the household, and 0.3 to the members under 14 years of age. Thus, the acronym *CH 00 hh* in the *dataset* column, for example, indicates the household income distribution from Switzerland in the year 2000.

4. D'ADDARIO'S INVARIANTS METHOD

The use of K. Pearson's moments method for determining the parameters of the Pareto model, was criticized by D'Addario (1934, 1939) who proposed the alternative invariants method. According to the latter, the parameters of a fitted model should be determined by matching some model characteristics – which are functions of the parameters – with their counterparts in the empirical distribution. D'Addario proposed using those characteristics that are actually employed to describe income distributions like the arithmetic mean, the median, the Gini inequality ratio, the Pietra index, etc.

In this section we will explore the performance of this method when applied to the new model. In choosing the characteristics that should be preserved in the fitted model we considered D'Addario's point of view as well as some peculiarities of the new density. In first place, we put $\widehat{\mu}$ equal to $\bar{x}$, since this seems the most natural choice. Then we determine $\widehat{\alpha}$ and $\widehat{\theta}$ by one of the following couples of equations:

i)

$$F(\widehat{m} : \bar{x}; \alpha; \theta) = 0.5 \quad \text{and} \quad F(\bar{x} : \bar{x}; \alpha; \theta) = \widehat{F}(\bar{x}), \tag{8}$$

where $\widehat{m}$ is the empirical median and $\widehat{F}(\bar{x})$ is the empirical cumulative distribution function at $\bar{x}$.

ii)

$$F(\widehat{m} : \bar{x}; \alpha; \theta) = 0.5 \quad \text{and} \quad P(\bar{x}; \alpha; \theta) = \widehat{P}, \tag{9}$$

where

$$\widehat{P} = \frac{1}{n} \sum_{i=1}^{n} |x_i - \bar{x}|$$

is the empirical value of the Pietra index and $P(\mu; \alpha; \theta)$ is its model counterpart;

iii)

$$F(\widehat{m} : \bar{x}; \alpha; \theta) = 0.5 \quad \text{and} \quad Var(\bar{x}; \alpha; \theta) = m_2, \tag{10}$$

where $Var(\mu; \alpha; \theta)$ is the variance in the new model;

iv)

$$F(\bar{x} : \bar{x}; \alpha; \theta) = \widehat{F}(\bar{x}) \quad \text{and} \quad Var(\bar{x}; \alpha; \theta) = m_2; \tag{11}$$

v)

$$F(\widehat{m} : \bar{x}; \alpha; \theta) = 0.5 \quad \text{and} \quad A(\bar{x}; \alpha; \theta) = \widehat{A}(\bar{x}), \tag{12}$$

where

$$\widehat{A}(\bar{x}) = 1 - \frac{\sum_{x_i \le \bar{x}} x_i}{\widehat{F}(\bar{x})} \times \frac{1 - \widehat{F}(\bar{x})}{\sum_{x_i > \bar{x}} x_i}$$

is the empirical value of Zenga's point inequality at $\bar{x}$, and $A(\mu; \alpha; \theta)$ is its counterpart in the new model.

We applied these five couples of constraints to 9 of the 18 empirical distributions considered with the method of moments. Table 4 reports the parameter values $\widehat{\mu}$, $\widehat{\alpha}$, $\widehat{\theta}$ of the fitted density functions and the corresponding goodness of fit measures A_1, A_2 and A_2'. Comparing the values of A_1, A_2 and A_2' obtained with the moments method with those achieved with the present method we notice a considerable improvement when α and θ are determined by the equations in (9) or in (12).

TABLE 4. - *Invariants method: estimated values of $\widehat{\mu}$, $\widehat{\alpha}$, $\widehat{\theta}$ and indexes of goodness of fit A_1, A_2 and $A_{2'}$ for 9 income distributions*

dataset	n	$\widehat{\mu}$	$\widehat{\alpha}$	$\widehat{\theta}$	A_1	A_2	A_2'
I 06 ind							
median and $F(\mu)$	13401	18503	1.9250	2.8360	0.0743	0.1725	0.1160
median and P	13401	18503	1.9697	2.8967	0.0731	0.1827	0.1168
median and A	13401	18503	1.9694	2.8962	0.0731	0.1826	0.1167
median and Var	13401	18503	1.7018	2.5213	0.0885	0.1379	0.1246
$F(\mu)$and Var	13401	18503	1.6576	2.3559	0.0970	0.1371	0.1242
I 06 hh							
median and $F(\mu)$	7762	31919	4.7267	7.0863	0.07081	0.15587	0.12241
median and P	7762	31919	3.3995	5.1556	0.06632	0.07682	0.07911
median and A	7762	31919	3.3959	5.1500	0.06637	0.07682	0.07911
median and Var	7762	31919	2.2337	3.4494	0.12058	0.16372	0.19876
$F(\mu)$and Var	7762	31919	1.9029	2.4884	0.18859	0.22609	0.24999
I NW 06 hh							
median and $F(\mu)$	1987	34693	3.0440	4.1997	0.10048	0.13964	0.15525
median and P	1987	34693	3.4224	4.6924	0.09269	0.12511	0.13680
median and A	1987	34693	3.4211	4.6905	0.09271	0.12514	0.13683
median and Var	1987	34693	2.3673	3.3136	0.13535	0.19237	0.24055
$F(\mu)$and Var	1987	34693	2.1843	2.8496	0.16504	0.21039	0.25062
I S 06 hh							
median and $F(\mu)$	1618	24550	1.0963	1.6909	0.27926	0.36422	0.57460
median and P	1618	24550	2.7254	4.0529	0.11824	0.14485	0.14809

(follows)

dataset	n	$\widehat{\mu}$	$\widehat{\alpha}$	$\widehat{\theta}$	A_1	A_2	A'_2
median and A	1618	24550	2.6904	4.0035	0.11834	0.14499	0.14995
median and *Var*	1618	24550	2.0392	3.0755	0.13421	0.18511	0.22138
$F(\mu)$and *Var*	1618	24550	2.7803	5.3712	0.20212	0.25482	0.29866
CH 00 hh							
median and $F(\mu)$	3628	6764	3.4245	3.9794	0.06466	0.12942	0.11786
median and *P*	3628	6764	2.8514	3.3605	0.06574	0.11195	0.11924
median and *A*	3628	6764	2.8507	3.3596	0.06576	0.11196	0.11928
median and *Var*	3628	6764	1.8819	2.2978	0.14285	0.19716	0.26134
$F(\mu)$and *Var*	3628	6764	1.6573	1.6823	0.22681	0.26590	0.30463
CH 05 hh							
median and $F(\mu)$	3071	6784	26.9006	23.6923	0.36900	499.82150	0.49010
median and *P*	3071	6784	1.8587	1.9291	0.21440	0.26323	0.27810
median and *A*	3071	6784	1.8390	1.9112	0.21603	0.26568	0.28253
median and *Var*	3071	6784	2.7740	2.7453	0.18231	0.21239	0.20124
$F(\mu)$and *Var*	3071	6784	1.9380	1.3743	0.40119	0.46642	0.47263
US 00 hh							
median and $F(\mu)$	5199184	56453	3.0054	6.2819	0.1705	0.6051	0.2532
median and *P*	5199184	56453	1.3836	2.8887	0.0735	0.1050	0.1133
median and *A*	5199184	56453	1.3755	2.8712	0.0739	0.1055	0.1165
median and *Var*	5199184	56453	2.1262	4.4545	0.1086	0.2436	0.1665
$F(\mu)$and *Var*	5199184	56453	1.9729	3.8317	0.1142	0.2344	0.1690
US 08 hh							
median and $F(\mu)$	2899458	82460	3.2436	6.2115	0.16988	0.57195	0.25072
median and *P*	2899458	82460	1.4456	2.7929	0.07062	0.08628	0.08804
median and *A*	2899458	82460	1.4374	2.7767	0.07104	0.08726	0.08932
median and *Var*	2899458	82460	2.2187	4.2755	0.10257	0.21259	0.15500
$F(\mu)$and *Var*	2899458	82460	2.0297	3.5927	0.11488	0.20741	0.16200
UK 99/00 hh							
median and $F(\mu)$	7086	446	3.1787	5.7379	0.10624	0.16006	0.17013
median and *P*	7086	446	2.9344	5.3059	0.11272	0.17232	0.18890
median and *A*	7086	446	2.9335	5.3046	0.11275	0.17237	0.18899
median and *Var*	7086	446	2.4424	4.4355	0.13321	0.20778	0.25001
$F(\mu)$and *Var*	7086	446	2.2649	3.8722	0.15558	0.22879	0.25989

5. METHOD OF MINIMUM A_1 WITH TWO RESTRICTIONS

The goodness of fit may be improved with respect to D'Addario's invariants method, by dropping one of the two equations that determinine α and θ in (8), (9), (10), (11) or (12), and by minimizing A_1 (or either one of A_2 or A'_2) subject to the restriction given by the other equation. This method seems very appealing since it addresses an issue raised first by Dagum (1977), and later by Mc Donald (1984) and Mc Donald and Ransom (2008), who consider both frequencies and characteristics such as the mean, median and inequality indexes when assessing a fitted model.

In this section we will apply this new method to fit Zenga's new density to the same nine empirical income distributions considered with the method of invariants in the previous section. As before, we put $\mu = \bar{x}$ and determine α and θ by minimizing

A_1 subject to one of the following restrictions:

i) $F(\widehat{m} : \bar{x}; \alpha; \theta) = 0,5$

ii) $F(\bar{x} : \bar{x}; \alpha; \theta) = \widehat{F}(\bar{x})$

iii) $P(\bar{x}; \alpha; \theta) = \widehat{P}$

iv) $A(\bar{x} : \bar{x}; \alpha; \theta) = \widehat{A}(\bar{x})$

Table 5 reports the results of this method. The best fit is achieved when we preserve the Pietra or Zenga indexes.

TABLE 5. - *Minimum Mortara with two restrictions: estimated values of $\widehat{\mu}$, $\widehat{\alpha}$, $\widehat{\theta}$ and indexes of goodness of fit A_1, A_2 and $A_{2'}$ for 9 income distributions*

dataset	n	$\widehat{\mu}$	$\widehat{\alpha}$	$\widehat{\theta}$	A_1	A_2	A_2'
I 06 ind							
median	13401	18503	1.9712	2.8983	0.07304	0.18318	0.11677
$F(\mu)$	13401	18503	1.9352	2.8522	0.07416	0.17458	0.11638
P	13401	18503	1.9591	2.8778	0.07288	0.18038	0.11633
A	13401	18503	1.9591	2.8778	0.07288	0.18038	0.11633
I 06 hh							
median	7762	31919	3.8348	5.7891	0.06301	0.08836	0.08667
$F(\mu)$	7762	31919	3.7855	5.5456	0.06396	0.08991	0.08543
P	7762	31919	3.8696	5.9613	0.06325	0.09131	0.09143
A	7762	31919	3.8683	5.9595	0.06326	0.09125	0.09140
I NW 06 hh							
median	1987	34693	3.6676	5.0113	0.09011	0.12182	0.13258
$F(\mu)$	1987	34693	3.8168	5.4224	0.09639	0.12849	0.14386
P	1987	34693	3.7188	5.1515	0.09231	0.12321	0.13568
A	1987	34693	3.7189	5.1517	0.09231	0.12321	0.13569
I S 06 hh							
median	1618	24550	3.4548	5.0828	0.11566	0.18531	0.13880
$F(\mu)$	1618	24550	1.9868	3.6121	0.17750	0.24972	0.33716
P	1618	24550	3.7452	5.8176	0.11257	0.20704	0.14826
A	1618	24550	3.6756	5.7126	0.11278	0.19990	0.14723
CH 00 hh							
median	3628	6764	3.1548	3.6886	0.06144	0.11545	0.11367
$F(\mu)$	3628	6764	3.1814	3.6604	0.06197	0.11797	0.11211
P	3628	6764	3.1087	3.7111	0.06256	0.11467	0.11826
A	3628	6764	3.1083	3.7107	0.06257	0.11467	0.11827
CH 05 hh							
median	3071	6784	2.7099	2.6892	0.18106	0.21209	0.20118
$F(\mu)$	3071	6784	5.3561	4.4032	0.29720	0.62936	0.35435
P	3071	6784	3.1243	3.5665	0.11974	0.15132	0.15119
A	3071	6784	3.0936	3.5399	0.11848	0.15102	0.15105
US 00 hh							
median	5199184	56453	1.5520	3.2471	0.07232	0.10800	0.10607

(follows)

dataset	n	$\widehat{\mu}$	$\widehat{\alpha}$	$\widehat{\theta}$	A_1	A_2	A_2'
$F(\mu)$	5199184	56453	1.9076	3.6782	0.11350	0.21944	0.16409
P	5199184	56453	1.5604	3.3666	0.06587	0.10177	0.10319
A	5199184	56453	1.5557	3.3590	0.06572	0.10122	0.10311
US 08 hh							
median	2899458	82460	1.5263	2.9493	0.06717	0.08066	0.08021
$F(\bar{x}) = \widehat{F}(\bar{x})$	2899458	82460	1.9966	3.5227	0.11461	0.20097	0.15979
P	2899458	82460	1.6763	3.3661	0.05993	0.07822	0.07578
A	2899458	82460	1.6713	3.3586	0.05965	0.07753	0.07528
UK 99/00 hh							
median	7086	446	3.4837	6.2750	0.10280	0.14971	0.15467
$F(\mu)$	7086	446	3.5395	6.4789	0.10432	0.14702	0.15497
P	7086	446	3.4700	6.4187	0.10644	0.14879	0.16017
A	7086	446	3.4699	6.4186	0.10644	0.14880	0.16017

6. METHOD OF MINIMUM A_1 WITH THE RESTRICTION $\mu = \bar{x}$

In this section we drop both restrictions that determined α and θ with the method of invariants and minimize A_1 preserving just the value of the mean. With respect to the method considered in Section 5, we achieve just a little decrease in the goodness of fit measures A_1, A_2 and A_2' (compare Tables 5 and 6).

TABLE 6. - *Minimum Mortara with restriction $\widehat{\mu} = \bar{x}$: estimated values of $\widehat{\mu}$, $\widehat{\alpha}$, $\widehat{\theta}$ and indexes of goodness of fit A_1, A_2 and $A_{2'}$ for 9 income distributions*

dataset	n	$\widehat{\mu}$	$\widehat{\alpha}$	$\widehat{\theta}$	A_1	A_2	A_2'
I 06 ind	13401	18503	2.0015	2.9215	0.07150	0.19150	0.11680
I 06 hh	7762	31919	3.8584	5.8840	0.06293	0.08975	0.08889
I NW 06 hh	1987	34693	3.6027	4.8350	0.08780	0.12168	0.13009
I S 06 hh	1618	24550	3.8224	5.9336	0.11234	0.21542	0.14959
CH 00 hh	3628	6764	3.1451	3.7117	0.06116	0.11502	0.11540
CH 05 hh	3071	6784	3.0513	3.5553	0.11453	0.14937	0.15098
US 00 hh	5199184	56453	1.4140	3.1376	0.06090	0.09760	0.11390
US 08 hh	2899458	82460	1.5375	3.1801	0.05213	0.07265	0.07462
UK 99/00 hh	7086	446	3.4562	6.1753	0.10225	0.15147	0.15492

7. METHOD OF MINIMUM A_1 WITHOUT RESTRICTIONS

Obviously, we may also minimize A_1 without restrictions on the parameter space. Sometimes this yields an appreciable improvement in terms of goodness of fit with respect to the other methods. However, these results are occasionally paid with a significant divergence between empirical and model mean (see Table 7).

TABLE 7. - *Minimum Mortara without restrictions: estimated values of* $\widehat{\mu}$, $\widehat{\alpha}$, $\widehat{\theta}$ *and indexes of goodness of fit* A_1, A_2 *and* $A_{2'}$ *for 9 income distributions*

dataset	n	$\widehat{\mu}$	$\widehat{\alpha}$	$\widehat{\theta}$	A_1	A_2	A'_2
I 06 ind	13401	18652	2.0349	2.9558	0.06980	0.20330	0.11750
I 06 hh	7762	31301	4.2575	6.3816	0.06083	0.10927	0.09841
I NW 06 hh	1987	33924	3.9262	5.1371	0.08633	0.12302	0.12476
I S 06 hh	1618	24207	3.8902	6.0044	0.11115	0.22039	0.15230
CH 00 hh	3628	6754	3.1849	3.7692	0.06103	0.11609	0.11609
CH 05 hh	3071	7155	2.8340	3.4984	0.10044	0.15147	0.18499
US 00 hh	5199184	59371	1.4497	3.3310	0.04140	0.09400	0.11280
US 08 hh	2899458	85925	1.5266	3.2534	0.03958	0.06912	0.07550
UK 99/00 hh	7086	446	3.4541	6.1752	0.10221	0.15182	0.15508

Figure 2 allows us to appreciate how the fit improves in passing from the invariants method proposed by D'Addario to the minimum A_1 methods. The histogram represents the empirical household income distribution from the US in 2008. The individual observations were grouped into classes according to the procedure described in Section 2.

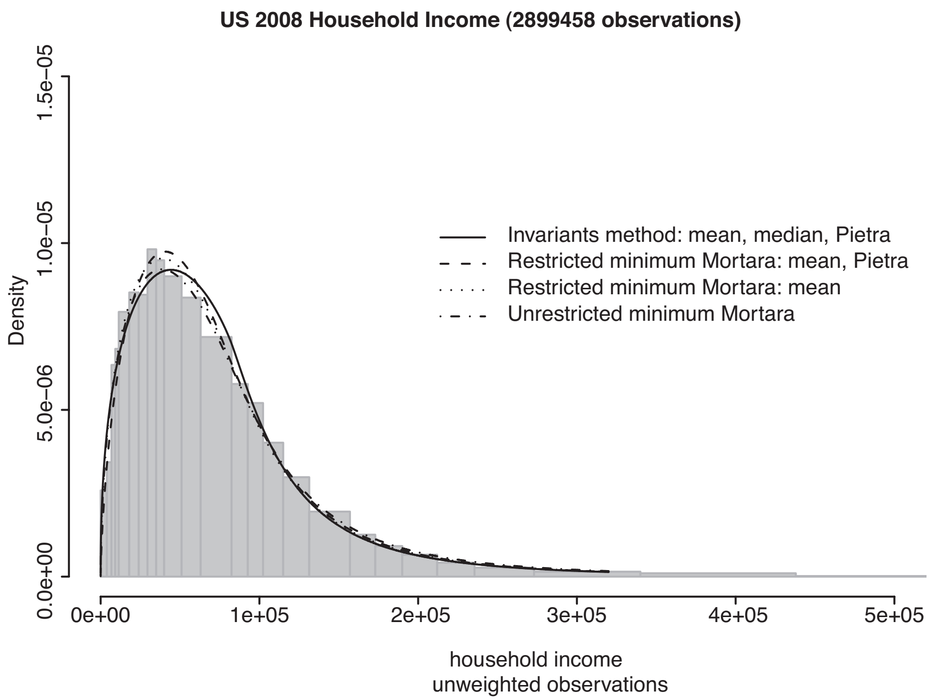

FIGURE 2. - *Zenga's new density fitted to empirical income distributions*

8. CONCLUSIONS

In this work we examined some methods for fitting Zenga's (2010a) new three-parameter model to empirical data. We considered 35 income distributions from Italy, Switzerland, United States and United Kingdom and five methods for fitting the new model: the classical method of moments, D'Addario's method of invariants and three other methods based on the minimization of Mortara's measure for the goodness of fit. We found that the moments method yields large values for the goodness of fit measures A_1, A_2 and A_2'. We therefore do not recommend its use. With D'Addario's method of invariants we achieved considerably better results. Pursuing lower values of the goodness of fit measures, we then proceeded to minimize Mortara's A_1 measure imposing suitable restrictions on the parameter space to preserve some important features of the empirical distribution in the fitted model. Finally, we also minimized A_1 without restrictions. Dropping the restrictions sometimes yields an appreciable improvement in terms of fit, but occasionally this is paid with a significant divergence between the empirical value of the mean and its model counterpart. Concerning the new model, we observe a good fit to the empirical income distributions from Switzerland, Italy and the US. Of course, further research is needed to confirm the results in this paper.

ACKNOWLEDGEMENTS

The authors wish to thank the Swiss Federal Statistical Office for kindly supplying them microdata from the Income and Consumption Survey and the UK Data Archive for providing through their webpage the computer file with survey data from Effects of Taxes and Benefits on Household Income, 1999-2000 conducted by the Office for National Statistics. The US family income data come from the US Census Bureau CPS Survey (2000, 2008) and are freely available on http://dataferrett.census.gov. Finally, we wish to thank Dr. Alberto Arcagni for his valuable help in processing the data.

REFERENCES

Banca d'Italia (2008). The 2006 Bank of Italy sample Survey on Household Income and Wealth. *Supplements to the Statistical Bulletin Sample Surveys*, **18**, 7. Available at http://www.bancaditalia.it.

D'Addario R. (1934). Sulla misura della concentrazione dei redditi. *Poligrafo dello Stato*, Roma.

D'Addario R. (1939). La curva dei redditi: sulla determinazione numerica dei parametri della seconda equazione paretiana. *Annali dell'Istituto di Statistica dell'Università di Bari*, **XII**(1939).

Dagum C. (1977). A New Model of Personal Income Distribution: Specification and Estimation. *Economie Appliqúee*, **30**, 413-437.

Mc Donald I.B., Ransom M. (2008). The Generalized Beta Distribution as a Model for the Distribution of Income: Estimation of Related Measures of Inequality. In Duangkamon Chotikapanich (Ed.), *Modelling Income Distributions and Lorenz Curves* (pp. 147-166). Springer, New York.

Mc Donald I.B. (1984). Some Generalized Functions for the Size Distribution of Income. *Econometrica*, **52**, 647-663.

Nelder J.A., Mead R. (1965). A simplex algorithm for function minimization. *Computer Journal*, **7**, 308-313.

Polisicchio M. (2008). The continuous random variable with uniform point inequality measure I(p). *Statistica & Applicazioni*, **VI**(2), 137-151.

Porro F., Arcagni A. (2012). On the parameters of Zenga distribution. To appear on *Statistical Methods & Applications*. DOI 10.1007/s10260-012-0219-y.

Zenga M.M. (2007). Inequality curve and inequality index based on the ratios between lower and upper arithmetic means. *Satistica & Applicazioni*, **V**(1), 3-28.

Zenga M.M. (2010a). Mixture of Polisicchio's Truncated Pareto distributions with beta weights. *Statistica & Applicazioni*, **XIII**(1), 3-25.

Zenga M.M., Arcagni A. (2012). *Application of Zenga's distribution to a panel survey on household incomes of European Member states*. Contributed paper at the RSS 2012 Conference.

Zenga M.M., Polisicchio M., Zenga M., Pasquazzi L. (2011). More on a new three-parameter distribution for non-negative variables. *Statistica & Applicazioni*, **IX**(1), 5-33.