Statistica & Applicazioni - Vol. XI, n. 1, 2013, pp. 79-102

APPLICATION OF ZENGA'S DISTRIBUTION TO A PANEL SURVEY ON HOUSEHOLD INCOMES OF EUROPEAN MEMBER STATES

Alberto Arcagni *
Michele Zenga*

SUMMARY

In this paper Zenga's distribution is applied to 114 household incomes distributions from a panel survey conducted by Eurostat. Previous works showed the good behaviour of the model to describe income distributions and analyzed the possibility to impose restrictions on the parametric space so that the fitted models comply with some characteristics of interest of the samples. This work is the first application of the model on a wide number of distributions, showing that it can be used to describe incomes distributions of several countries. Maximum likelihood method on grouped data and methods based on the minimization of three different goodness of fit indexes are used to estimate parameters. The restriction that imposes the equivalence between the sample mean and the expected value of the fitted model is also considered. It results that the restriction should be used and the changes in fitting are analyzed in order to suggest which estimation method use jointly to the restriction. A final section is devoted to the direct proof that Zenga's distribution has Paretian right-tail.

Keywords: *Zenga's Model, Income Distributions, Inequality, Paretian Right-tail.*

1. INTRODUCTION

In this paper, several applications of the density function proposed by Zenga (2010) for non-negative variables are presented. Zenga's distribution is characterized by three parameters: a scale parameter $\mu > 0$, which is equal to the expected value, and the shape parameters $\alpha > 0$ and $\theta > 0$. This new density is particularly suitable for modelling income distributions since it is positively skewed and it has Paretian right tail.

For $\theta > 1$, Zenga (2010) has obtained the expressions of the distribution function, the moments, the truncated moments, the mean deviation and Zenga's point inequality $A(x)$ at $x = \mu$ (see Zenga, 2007). Zenga, Pasquazzi, Polisicchio and Zenga (2011) have generalized the previous results to the case $\theta > 0$. First applications of the model to empirical income distributions (Italy, Swiss, US and UK) can be found in Zenga, Pasquazzi and Zenga (2012). They estimated the parameters through the method of moments, D'Addario's invariant method and by minimizing the Mortara's goodness of fit index (A_1). Arcagni and Porro (2012) applied the maximum likelihood method and made some further research on the method of moment, obtaining its analytical solution.

* Dipartimento di Statistica e Metodi Quantitativi - Università Milano-Bicocca - Edificio U7, Via Bicocca degli Arcimboldi, 8 - 20126 MILANO (e-mails: alberto.arcagni@unimib.it; michele zenga@unimib.it).

In this paper, Zenga's density function has been applied to 114 household income distributions taken from the European Community Household Panel survey (Eurostat, 2003). The parameters are estimated by maximum likelihood on grouped data and by minimizing three different goodness of fit measures (the Mortara index A_1 the quadratic K. Pearson index A_2 and the modified quadratic index A_2').

The paper is organized as follows. Zenga's model and its main properties are presented in Section 2. Section 3 is devoted to the illustration of the dataset and of the criterion for the division of each empirical distribution into 25 classes. The estimation methods and the goodness of fit indexes are defined in Section 4. In Section 5 the overall results of the application are provided. In the same section it is illustrated the use of Zenga's model to compare different income distributions which allow to make considerations about inequality measures such as the Gini index, the Zenga inequality index, the $I(p)$ curve and the Lorenz curve. Section 6 provides the direct proof of the Paretian right-tail of Zenga's distribution. Finally, Section 7 is devoted to conclusions.

2. DEFINITION OF THE MODEL

Zenga (2010) introduced a new three parameter density function for non-negative variables considering a mixture of Polisicchio's truncated Pareto densities (see Polisicchio, 2008) with Beta weights. The Polisicchio density is

$$f(x:\mu;k) = \begin{cases} \dfrac{\sqrt{\mu}}{2} k^{0.5}(1-k)^{-1}x^{-1.5}, & \mu k \le x \le \frac{\mu}{k}, \mu > 0, 0 < k < 1 \\ 0 & \textit{otherwise}, \end{cases} \tag{1}$$

and the mixture is obtained adopting the following Beta density as mixing distribution on the parameter k

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

where $B(\alpha;\theta)$ is the Beta function. Then, the density function of the Zenga's model resulted

$$\begin{aligned} f(x:\mu;\alpha;\theta) &= \\ &= \int_0^1 f(x:\mu;k)g(k:\alpha;\theta)dk = \\ &= \begin{cases} \frac{1}{2\mu B(\alpha;\theta)} \left(\frac{x}{\mu}\right)^{-1.5} \int_0^{\frac{x}{\mu}} k^{\alpha+0.5+1}(1-k)^{\theta-2}dk, & 0 < x < \mu \\ \frac{1}{2\mu B(\alpha;\theta)} \left(\frac{\mu}{x}\right)^{1.5} \int_0^{\frac{\mu}{x}} k^{\alpha+0.5+1}(1-k)^{\theta-2}dk, & \mu < x. \end{cases} \\ &= \begin{cases} \frac{1}{2\mu B(\alpha;\theta)} \left(\frac{x}{\mu}\right)^{-1.5} \sum_{i=1}^{\infty} IB\left(\frac{x}{\mu} : \alpha - 0.5 + i;\theta\right), & 0 < x < \mu \\ \frac{1}{2\mu B(\alpha;\theta)} \left(\frac{\mu}{x}\right)^{1.5} \sum_{i=1}^{\infty} IB\left(\frac{\mu}{x} : \alpha - 0.5 + i;\theta\right), & \mu < x, \end{cases} \end{aligned} \tag{2}$$

while the distribution function is

$$F(x:\mu;\alpha;\theta) = \begin{cases} \frac{1}{B(\alpha;\theta)}\Big\{\sum_{i=1}^{\infty} IB\Big(\frac{x}{\mu}:\alpha+i-1;\theta\Big)+ \\ -\Big(\frac{x}{\mu}\Big)^{-0.5}\sum_{i=1}^{\infty} IB\Big(\frac{x}{\mu}:\alpha+i-0.5;\theta\Big)\Big\}, & 0<x\leq\mu \\ \\ 1-\frac{1}{B(\alpha;\theta)}\Big\{\big(\frac{\mu}{x}\big)^{0.5}\sum_{i=1}^{\infty} IB\big(\frac{\mu}{x}:\alpha+i-0.5;\theta\big)+ \\ -\sum_{i=1}^{\infty} IB\big(\frac{\mu}{x}:\alpha+i;\theta\big)\Big\}, & \mu<x, \end{cases} \quad (3)$$

where $\mu > 0$, $\alpha > 0$ and $\theta > 0$ and $IB(x:\alpha;\theta)$ is the Incomplete Beta function (see Zenga, 2010; Zenga *at al.*, 2011).

Let X be a random variable following the density function (2). Zenga *et al.* (2011) determined that

$$E(X^r) = \frac{\mu^r}{(2r-1)B(\alpha;\theta)}\sum_{i=1}^{2r-1} B(\alpha-r+1;\theta) \qquad \textit{for} \qquad r<(\alpha+1) \quad (4)$$

which shows that the distribution has Paretian right tail. If $r = 1$, from expression (4) it follows that $E(X) = \mu$, which, indeed, is always finite. Moreover, μ is also scale parameter, since $f(x:\mu;\alpha;\theta) = \frac{1}{\mu}f\Big(\frac{x}{\mu}:1;\alpha;\theta\Big)$. The expected value is always greater than the median, because $F(\mu:\mu;\alpha;\theta) \geq 0.5$, therefore Zenga's model is positively skewed (see Zenga, 2010). The upper bound of the density function depends on the shape parameters' values. In detail

$$\lim_{x\to\mu} f(x:\mu;\alpha;\theta) = \begin{cases} \frac{B(\alpha+0.5;\theta-1)}{2\mu B(\alpha;\theta)}, & \theta>1 \\ \infty, & 0<\theta\leq 1; \end{cases} \quad (5)$$

$$\lim_{x\to 0} f(x:\mu;\alpha;\theta) = \begin{cases} 0, & \alpha>1 \\ \frac{1}{3}\frac{\theta}{\mu}, & \alpha=1 \\ \infty, & 0<\alpha<1. \end{cases} \quad (6)$$

From expression (5) it follows that θ determines if the density (2) is finite for $x = \mu$ or not. Expressions (4) and (6) show that α controls the behaviour of the tails of the density function. Porro (2011) and Arcagni and Porro (2012) determined that α is an inverse inequality indicator and that θ is a direct inequality indicator. This set of characteristics make Zenga's distribution a model of interest to describe income distributions. Differently from other models, such as Pareto and Dagum distributions, its expectation is always finite and equal to the parameter μ. Moreover its parameters govern separately the location and the inequality.

We conclude this section remembering that, if $\theta > 1$, the analytical expressions (2) and (3) are given by

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

$$F(x:\mu;\alpha;\theta)=\begin{cases}\frac{1}{B(\alpha;\theta)}\Big\{IB\Big(\frac{x}{\mu}:\alpha;\theta-1\Big)+\\ -\left(\frac{x}{\mu}\right)^{-0.5}IB\Big(\frac{x}{\mu}:\alpha+0.5;\theta-1\Big)\Big\}, & 0<x\le\mu\\ 1-\frac{1}{B(\alpha;\theta)}\Big\{\left(\frac{\mu}{x}\right)^{0.5}IB\big(\frac{\mu}{x}:\alpha+0.5;\theta-1\big)+\\ -IB\big(\frac{\mu}{x}:\alpha+1;\theta-1\big)\Big\}, & \mu<x.\end{cases} \quad (7)$$

This observation is particularly useful from the computational point of view, since, in almost all cases of practical interest, it is reasonable to assume that the value of $f(\mu:\mu;\alpha;\theta)$ is finite.

3. DATA AND CLASSES DEFINITION

We fitted Zenga's model to 114 household income distributions from the European Community Household Panel (ECHP), conducted by Eurostat (2003). The data of the ECHP were collected from 1994 to 2001 and regards 15 European Member States: Austria (since 1995), Belgium, Denmark, Finland (since 1996), France, Germany, Greece, Ireland, Italy, Luxembourg, Netherlands, Portugal, United Kingdom, Spain and Sweden (since 1997). In Table 1, for each sample distribution, are shown the sample size (n), the mean ($\bar{x}$), the ratio between median and mean ($me/\bar{x}$), the Gini index (G) and the Zenga inequality index (I) (see Zenga, 2007). The sample size and these statistics are presented in order to illustrate the main features of the income sample distributions, with particular emphasis on inequality.

Eurostat provides the micro-data expressed in National Currency. All the estimation procedures considered in this paper, require a preliminary grouping of the micro-data. We have grouped the observations into classes following the proposal of Zenga *et al.* (2012). For completeness, this grouping scheme is recalled below. The empirical distributions have different sample sizes n and different extreme values. Therefore, a general and uniform grouping scheme is obtained subdividing each distribution into $s=25$ classes starting from the relative frequency f'_j (for $j=1,...,s$). The chosen values of f'_j, and the related cumulative relative frequencies p'_j are reported in Table 2. Obviously, the quantities np'_j are not necessarily integers. Therefore, Zenga *et al*. (2012) defined p_j as the cumulative relative frequencies such that the np_j are the rounded values of np'_j.

Now, the class C_j, for $j=1,2,\ldots,s$, is defined as

$$C_j=\{x_{(np_{j-1}+1)},\ldots,x_{(np_j)}\}$$

TABLE 1. - *Sample size (n), sample mean ($\bar{x}$), ratio between sample median and sample mean ($me/\bar{x}$), Gini inequality index (G) and Zenga inequality index (I) of the 114 micro-data distributions*

Country		Year							
		1994	1995	1996	1997	1998	1999	2000	2001
Austria	n		3367	3281	3130	2952	2809	2637	2535
	$\bar{x}$		380824	383425	372720	374034	388136	397799	398877
	$me/\bar{x}$		0.8781	0.8903	0.8891	0.8964	0.8838	0.8942	0.8937
	G		0.3415	0.3299	0.3251	0.3181	0.3361	0.3238	0.3190
	I		0.6904	0.6787	0.6707	0.6609	0.6853	0.6688	0.6621
Belgium	n	3454	3343	3191	3013	2863	2691	2558	2342
	$\bar{x}$	1017973	1039514	1043165	1061628	1085580	1141939	1177792	1223753
	$me/\bar{x}$	0.8548	0.8589	0.8548	0.8617	0.8541	0.8346	0.8156	0.8394
	G	0.3518	0.3521	0.3461	0.3374	0.3391	0.3566	0.3634	0.3553
	I	0.7026	0.7025	0.6941	0.6819	0.6831	0.7000	0.7034	0.6958
Denmark	n	3478	3218	2951	2740	2505	2381	2273	2279
	$\bar{x}$	196699	205366	207620	216906	225742	233853	243061	248249
	$me/\bar{x}$	0.8845	0.8807	0.8759	0.8911	0.9023	0.9160	0.8971	0.9177
	G	0.3171	0.3182	0.3191	0.3060	0.3133	0.3116	0.3141	0.3149
	I	0.6549	0.6576	0.6593	0.6430	0.6556	0.6535	0.6583	0.6606
Finland	n			4138	4103	3917	3818	3101	3106
	$\bar{x}$			149293	146438	147498	151305	153805	160284
	$me/\bar{x}$			0.9206	0.9263	0.9244	0.9003	0.8853	0.8750
	G			0.2961	0.3057	0.3247	0.3339	0.3374	0.3487
	I			0.6350	0.6513	0.6787	0.6867	0.6883	0.7014
France	n	7108	6680	6556	6142	5849	5594	5332	5268
	$\bar{x}$	156187	159616	162640	163536	165895	170341	171196	175603
	$me/\bar{x}$	0.8074	0.8578	0.8562	0.8577	0.8664	0.8500	0.8582	0.8692
	G	0.3879	0.3461	0.3426	0.3416	0.3337	0.3432	0.3335	0.3287
	I	0.7367	0.6962	0.6896	0.6900	0.6807	0.6886	0.6771	0.6729
Germany	n	6196	6329	6252	6156	5955	5845	5687	5559
	$\bar{x}$	49992	48704	49782	51273	51404	53103	55371	56521
	$me/\bar{x}$	0.8991	0.9188	0.9163	0.9242	0.9304	0.9180	0.9201	0.9204
	G	0.3238	0.3083	0.2960	0.2866	0.2885	0.2931	0.2914	0.2944
	I	0.6766	0.6572	0.6374	0.6245	0.6271	0.6327	0.6294	0.6324
Greece	n	5480	5173	4851	4543	4171	3952	3893	3895
	$\bar{x}$	2935058	3262795	3550018	3856494	4246580	4410062	4648298	4804914
	$me/\bar{x}$	0.8120	0.8198	0.8184	0.8070	0.8183	0.8192	0.8232	0.8213
	G	0.4068	0.3982	0.3877	0.3985	0.4030	0.3954	0.3845	0.3828
	I	0.7625	0.7518	0.7419	0.7513	0.7562	0.7478	0.7358	0.7340
Holland	n	5139	5035	5097	5019	4922	4981	4976	4824
	$\bar{x}$	46397	47168	48358	50331	51757	54366	54942	57589
	$me/\bar{x}$	0.9347	0.9125	0.9071	0.9270	0.9334	0.9226	0.9341	0.9283
	G	0.2989	0.3179	0.3221	0.2922	0.2872	0.2927	0.2746	0.2894
	I	0.6417	0.6683	0.6725	0.6319	0.6254	0.6335	0.6065	0.6274
Ireland	n	4038	3569	3164	2935	2723	2372	1944	1757
	$\bar{x}$	15996	17068	17503	18545	20232	21016	22160	24636
	$me/\bar{x}$	0.8497	0.8445	0.8381	0.8410	0.8255	0.8441	0.8779	0.8766
	G	0.3573	0.3698	0.3706	0.3713	0.3917	0.3828	0.3726	0.3742
	I	0.7043	0.7178	0.7180	0.7180	0.7374	0.7323	0.7246	0.7264

follows

Country		Year							
		1994	1995	1996	1997	1998	1999	2000	2001
Italy	n	6915	7004	7026	6627	6478	6273	5989	5525
	$\bar{x}$	30981	33274	34195	34438	36347	37678	38851	40183
	$me/\bar{x}$	0.8527	0.8349	0.8527	0.8593	0.8585	0.8496	0.8590	0.8601
	G	0.3554	0.3548	0.3423	0.3393	0.3376	0.3338	0.3310	0.3381
	I	0.7121	0.7077	0.6939	0.6904	0.6869	0.6790	0.6768	0.6855
Luxembourg	n	1010	2976	2471	2651	2521	2550	2373	2428
	$\bar{x}$	1502184	1584630	1637555	1636614	1683883	1717866	1804012	1846854
	$me/\bar{x}$	0.8433	0.8732	0.8726	0.8670	0.8657	0.8458	0.8582	0.8635
	G	0.3397	0.3093	0.3007	0.3049	0.3010	0.3090	0.3018	0.2982
	I	0.6866	0.6412	0.6306	0.6352	0.6293	0.6395	0.6300	0.6250
Portugal	n	4787	4870	4807	4767	4666	4645	4606	4588
	$\bar{x}$	1664684	1761507	1886649	1998450	2114116	2255393	2353262	2577289
	$me/\bar{x}$	0.7638	0.7770	0.7918	0.8015	0.8239	0.8109	0.8190	0.8046
	G	0.4455	0.4295	0.4167	0.4120	0.4130	0.4074	0.4017	0.4038
	I	0.7921	0.7782	0.7666	0.7611	0.7637	0.7568	0.7526	0.7523
Spain	n	7142	6449	6133	5714	5439	5301	5048	4950
	$\bar{x}$	2198713	2316971	2442086	2510380	2625487	2814373	2967596	3115968
	$me/\bar{x}$	0.8225	0.8260	0.8216	0.8155	0.8137	0.8277	0.8198	0.8251
	G	0.3727	0.3639	0.3655	0.3736	0.3680	0.3643	0.3700	0.3679
	I	0.7219	0.7112	0.7127	0.7218	0.7151	0.7111	0.7167	0.7155
Sweden	n				5286	5208	5165	5116	5085
	$\bar{x}$				213864	222074	225094	234192	249974
	$me/\bar{x}$				0.9594	0.9366	0.9462	0.9332	0.9290
	G				0.2827	0.2956	0.2928	0.3024	0.3048
	I				0.6206	0.6373	0.6345	0.6481	0.6488
United Kingdom	n	5041	4999	4991	4958	4975	4935	4866	4779
	$\bar{x}$	13990	15230	15490	17011	18470	17964	19506	21006
	$me/\bar{x}$	0.8629	0.8461	0.8624	0.8655	0.8479	0.8532	0.8596	0.8676
	G	0.3625	0.3744	0.3576	0.3456	0.3636	0.3612	0.3569	0.3536
	I	0.7157	0.7261	0.7085	0.6939	0.7138	0.7113	0.7084	0.7036

TABLE 2. - *Prefixed relative frequencies f'_j and cumulative relative frequencies p'_j of the $s = 25$ intervals*

j	f'_j	p'_j	j	f'_j	p'_j	j	f'_j	p'_j	j	f'_j	p'_j	j	f'_j	p'_j
1	0.010	0.010	6	0.050	0.150	11	0.100	0.500	16	0.050	0.850	21	0.010	0.960
2	0.010	0.020	7	0.050	0.200	12	0.100	0.600	17	0.050	0.900	22	0.010	0.970
3	0.015	0.035	8	0.050	0.250	13	0.100	0.700	18	0.020	0.920	23	0.010	0.980
4	0.015	0.050	9	0.050	0.300	14	0.050	0.750	19	0.015	0.935	24	0.010	0.990
5	0.050	0.100	10	0.100	0.400	15	0.050	0.800	20	0.015	0.950	25	0.010	1.000

where $p_0 = 0$, $x_{(0)} = 0$ and $x_{(np_j)}$ is the p_j-quantile of the observed distribution.

In the study of income distributions the value of the arithmetic mean is very important. So it seems useful to have a bound equal to the sample mean $\bar{x}$ of the observed distribution. This goal is achieved substituting the $x_{(np_j)}$ nearest to $\bar{x}$ with $\bar{x}$, itself.

In conclusion, the (integer) frequencies n_j are given by

$$n_j = n(p_j - p_{j-1});\ j = 1, 2, \ldots, s;\ \text{with } p_0 = 0$$

and the true relative frequencies f_j of each class are

$$f_j = \frac{n_j}{n} = p_j - p_{j-1};\ j = 1, 2, \ldots, s.$$

For $j = 1, 2, \ldots, s-1$ observe that through this procedure, if $x_{(np_j)} = v$ is a repeated value of the sample, some observations equal to v are attributed to the class of index j and the remaining to the class of index $j+1$, so that the relative frequencies f_j are not far from the prefixed ones f'_j.

4. ESTIMATION METHODS AND GOODNESS OF FIT INDEXES

Some of the estimation methods considered in this paper, are based on well known goodness of fit indexes. These indexes are based on the evaluation of the discrepancy between the observed frequencies n_j and the theoretical frequencies

$$\hat{n}_j = \begin{cases} n \cdot F(x_1 : \mu; \alpha; \theta), & j = 1 \\ n\{F(x_j : \mu; \alpha; \theta) - F(x_{j-1} : \mu; \alpha; \theta)\}, & j = 2, 3, \ldots, s-1 \\ n\{1 - F(x_{j-1} : \mu; \alpha; \theta)\}, & j = s \end{cases} \tag{8}$$

where x_j is the upper bound of the j-th class, for $j = 1, 2, \ldots, s-1$, and $x_s = \infty$. We considered the Mortara index A_1, the quadratic K. Pearson index A_2 and the modified quadratic index A'_2, which are defined as follow

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

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

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

The parameters of Zenga's distribution can be estimated by numerical minimization of A_1, A_2 and A'_2:

- minimum A_1 estimates

$$(\hat{\mu}, \hat{\alpha}, \hat{\theta})_{A_1} = \arg\min_{\mu,\alpha,\theta} A_1(\mu, \alpha, \theta);$$

- minimum A_2 estimates

$$(\hat{\mu}, \hat{\alpha}, \hat{\theta})_{A_2} = \arg\min_{\mu,\alpha,\theta} A_2(\mu, \alpha, \theta);$$

- minimum A_2' estimates

$$(\hat{\mu}, \hat{\alpha}, \hat{\theta})_{A_2'} = \underset{\mu,\alpha,\theta}{\arg\min} A_2'(\mu, \alpha, \theta).$$

We also apply the maximum likelihood method on grouped data. In these case, the likelihood function is obtained from the multinomial distribution with parameters n and

$$\pi_j(\mu,\alpha,\theta) = \begin{cases} F(x_1 : \mu; \alpha; \theta), & j = 1 \\ F\left(x_j : \mu; \alpha; \theta\right) - F\left(x_{j-1} : \mu; \alpha; \theta\right), & j = 2, 3, \ldots, s-1 \\ 1 - F\left(x_{j-1} : \mu; \alpha; \theta\right), & j = s. \end{cases}$$

The likelihood and log-likelihood functions are

$$\mathcal{L}(\mu, \alpha, \theta; n_1, \ldots, n_s) = \frac{n!}{\prod_{j=1}^{s} n_j!} \prod_{j=1}^{s} \pi_j(\mu, \alpha, \theta)^{n_j}; \tag{9}$$

$$\log \mathcal{L}(\mu, \alpha, \theta; n_1, \ldots, n_s) = \log \frac{n!}{\prod_{j=1}^{s} n_j!} + \sum_{j=1}^{s} n_j \log \pi_j(\mu, \alpha, \theta);$$

and the maximum likelihood estimates are defined as follow

$$(\hat{\mu}, \hat{\alpha}, \hat{\theta})_{\mathcal{L}} = \underset{\mu,\alpha,\theta}{\arg\max} \log\, \mathcal{L}(\mu, \alpha, \theta; n_1, \ldots, n_s).$$

The considered estimation methods require the optimization of complicated functions and the analytical solution can not be obtained. Therefore the estimates are found by numerical procedures which need the definition of a starting point. Arcagni and Porro (2012) have obtained the analytical solution of method of moments, which do not perform well but it can be easily applied. We used this estimates as starting point in our numerical procedure. For completeness, here, we recall the expression of the method of moments estimates

$$\begin{cases} \hat{\mu} = \bar{x} \\ \hat{\theta} = \dfrac{-\left[\frac{1}{3}\frac{\bar{x}^2}{m_2} - 3\frac{\bar{x}\, m_2}{m_3} - 1\right] + \sqrt{\left[\frac{1}{3}\frac{\bar{x}^2}{m_2} - 3\frac{\bar{x}\, m_2}{m_3} - 1\right]^2 + 4\left[\frac{1}{3}\frac{\bar{x}^2}{m_2} - \frac{3}{5}\bar{x}\frac{m_2}{m_3}\right]\left[\frac{18}{5}\frac{\bar{x}\, m_2}{m_3} + 2\right]}}{2\left[\frac{1}{3}\frac{\bar{x}^2}{m_2} - \frac{3}{5}\bar{x}\frac{m_2}{m_3}\right]} \\ \hat{\alpha} = \dfrac{-(\hat{\theta} - 1) + \sqrt{\left(\hat{\theta} - 1\right)^2 + 4\left[\frac{1}{3}\frac{\bar{x}^2}{m_2}\hat{\theta}\left(\hat{\theta} + 1\right) + \hat{\theta}\right]}}{2} \end{cases}$$

where m_2 and m_3 are respectively the empirical variance and the empirical third central moment.

As aforementioned, the sample mean $\bar{x}$ is an important statistic when studying income distributions. For this reason we study also the fitting of the Zenga's model to the empirical data imposing the restriction $\mu = \bar{x}$.

The goodness of fit indexes A_1, A_2 and A_2' can be also used to evaluate the performance of the estimation methods. Defining the absolute relative frequency deviations

$$a_j = \frac{|n_j - \hat{n}_j|}{\hat{n}_j}$$
$$a_j' = \frac{|n_j - \hat{n}_j|}{n_j}$$

it can be observed that

$$A_1 = M_1(a_j; \hat{n}_j) = M_1(a_j'; n_j)$$
$$A_2 = M_2(a_j; \hat{n}_j)$$
$$A_2' = M_2(a_j'; n_j)$$

that is, Mortara index A_1 is both the arithmetic mean of a_j with weights $\hat{n}_j$ and the arithmetic mean of a_j' with weights n_j, A_2 index is the quadratic mean of a_j with weights $\hat{n}_j$ and A_2' index is the quadratic mean of a_j' with weights n_j. Therefore

$$A_1 \leq A_2$$
$$A_1 \leq A_2'$$

and the variance of a_j with weights $\hat{n}_j$ and the variance of a_j' with weights n_j can be obtained as follows

$$\begin{aligned} Var(a_j; \hat{n}_j) &= {A_2}^2 - {A_1}^2 \\ Var(a_j'; n_j) &= {A_2'}^2 - {A_1}^2. \end{aligned} \tag{10}$$

Small differences between A_1 index and quadratic indexes (A_2 or A_2') mean low variability of absolute relative frequency deviations, and a 'uniform' fitting of the model on the whole range of the empirical distribution.

To take into account the differences between the sample mean and the expected value of the r.v. with distribution function equal to the fitted model, we evaluate the relative absolute deviation of $\hat{\mu}$ from $\bar{x}$

$$\rho_E = \frac{|\hat{\mu} - \bar{x}|}{\bar{x}}. \tag{11}$$

For the restricted model, ρ_E is equal to zero.

5. APPLICATION

For each of the 114 empirical distributions we have obtained the values of the three parameters (μ;α;θ) by the four estimation methods introduced in Section 4. Note that the parameters have also been evaluated imposing the restriction $\mu = \bar{x}$. For each estimated model we have evaluated the three indexes of goodness of fit, A_1, A_2 and A_2'.

In Section 5.1 for each of the 15 countries and for each estimation method (with and without restriction on μ) we present only some ‘syntheses’ of the indexes of goodness of fit. Viceversa, in Section 5.2 for two empirical distributions (Germany 2001 and Greece 2001) the detailed results are given.

5.1 *Overall results*

The principal aim of this section is to analyze the variation of the fitting due to the restriction on μ. Thus, for each of the European Member State analyzed we provide the mean and the Mean Absolute Deviation (MAD) of the goodness of fit indexes. We also point out for which countries the income distribution is best described by the model.

Minimization of goodness of fit indexes

Here we consider the estimation methods based on the minimization of the goodness of fit indexes A_1, A_2 and A_2'. Table 3 shows the mean and the MAD (in brackets) of the objective function by estimation method (both without and with restriction) and country. As an example, the value 0.1151 related to ‘Italy’ and ‘$\min A_1/A_1$’ is obtained averaging the values of A_1 obtained fitting the unrestricted Zenga’s model using the method of the minimum A_1 on the 8 Italian income distributions from 1994 to 2001. In the last row are provided the global means. The means by country which are lower than the global one are marked in bold type in order to indicate the countries for which Zenga’s model fit best. It can be observed that, whatever the method, the best results are obtained for Austria, France, Germany, Greece, Luxembourg and Portugal.

Maximum likelihood method on grouped data

Indexes A_1, A_2 and A_2' are also used to evaluate the fitting of the maximum likelihood method. In Table 4 mean and MAD by countries are evaluated for all the three indexes. Therefore maximum likelihood method can be compared with all the methods based on the minimization of the goodness of fit indexes. Indexes A_1, A_2 and A_2' are not the objective function of the maximum likelihood method, thus the values of such indexes are not lower than the ones obtained with the methods that minimize them. By the global means of Tables 3 and 4 it can be observed that the increment is small.

As observed with the previous methods, the best fitting is obtained again for Austria, France, Germany, Greece, Luxembourg and Portugal, regardless the index.

Implications of the restriction on the expected value

We now analyze the changes of the goodness of fit due to the introduction of the restriction on the expected value. In Table 5 the values of, the relative deviation of the expected value from the arithmetic mean (see expression 11), ρ_E are reported for each

TABLE 3. - *Minimization methods, goodness of fit indexes: mean, Mean Absolute Deviation (in round brackets), overall mean (bold: values lower than the overall mean) and overall median*

Estimation method	$min\, A_1$	$rest.\, min\, A_1$	$min\, A_2$	$rest.\, min\, A_2$	$min\, A_2'$	$rest.\, minA_2'$
g.d.f. index	A_1	A_1	A_2	A_2	A_2'	A_2'
Austria						
mean	**0.1093**	**0.1379**	**0.1694**	**0.1846**	**0.1786**	**0.1831**
(MAD)	(0.0071)	(0.0091)	(0.0114)	(0.0131)	(0.0112)	(0.0117)
Belgium						
mean	0.1495	**0.1542**	0.2125	**0.2194**	**0.1970**	**0.2008**
(MAD)	(0.0086)	(0.0090)	(0.0314)	(0.0343)	(0.0108)	(0.0094)
Denmark						
mean	0.2144	0.2205	0.2801	0.2887	0.2758	0.2815
(MAD)	(0.0062)	(0.0118)	(0.0117)	(0.0176)	(0.0187)	(0.0166)
Finland						
mean	0.1393	0.1834	**0.2026**	0.2343	0.2191	0.2257
(MAD)	(0.0100)	(0.0113)	(0.0111)	(0.0221)	(0.0141)	(0.0156)
France						
mean	**0.0817**	**0.0913**	**0.1565**	**0.1636**	**0.1390**	**0.1408**
(MAD)	(0.0042)	(0.0083)	(0.0202)	(0.0219)	(0.0090)	(0.0097)
Germany						
mean	**0.0788**	**0.1183**	**0.1482**	**0.1718**	**0.1556**	**0.1636**
(MAD)	(0.0053)	(0.0066)	(0.0236)	(0.0284)	(0.0117)	(0.0131)
Greece						
mean	**0.1110**	**0.1350**	**0.1913**	**0.2071**	**0.1829**	**0.1870**
(MAD)	(0.0094)	(0.0095)	(0.0128)	(0.0127)	(0.0080)	(0.0087)
Holland						
mean	0.1440	0.1823	0.2268	0.2479	0.2237	0.2281
(MAD)	(0.0066)	(0.0085)	(0.0086)	(0.0110)	(0.0079)	(0.0080)
Ireland						
mean	0.1928	0.2338	0.3458	0.3477	0.2906	0.2941
(MAD)	(0.0203)	(0.0347)	(0.0340)	(0.0354)	(0.0383)	(0.0412)
Italy						
mean	**0.1151**	**0.1277**	0.2196	0.2346	**0.1787**	**0.1828**
(MAD)	(0.0161)	(0.0161)	(0.0383)	(0.0433)	(0.0193)	(0.0203)
Luxembourg						
mean	**0.0871**	**0.0881**	**0.1308**	**0.1321**	**0.1194**	**0.1203**
(MAD)	(0.0155)	(0.0154)	(0.0325)	(0.0336)	(0.0210)	(0.0206)
Portugal						
mean	**0.1206**	**0.1407**	**0.1730**	**0.1823**	**0.1796**	**0.1875**
(MAD)	(0.0109)	(0.0113)	(0.0128)	(0.0129)	(0.0112)	(0.0130)
Spain						
mean	**0.1206**	**0.1432**	0.2533	0.2555	0.2044	0.2087
(MAD)	(0.0137)	(0.0164)	(0.0295)	(0.0280)	(0.0221)	(0.0225)
Sweden						
mean	0.1829	0.2256	0.2502	0.2822	0.2671	0.2707
(MAD)	(0.0225)	(0.0183)	(0.0257)	(0.0281)	(0.0252)	(0.0231)
United Kingdom						
mean	0.1370	0.1653	**0.1900**	**0.2048**	0.1988	**0.2015**
(MAD)	(0.0072)	(0.0091)	(0.0073)	(0.0105)	(0.0096)	(0.0100)
Overall mean	0.1310	0.1544	0.2094	0.2224	0.1988	0.2032
Overall median	0.1286	0.1504	0.1993	0.2128	0.1901	0.1964

TABLE 4. - *Maximum likelihood method, goodness of fit indexes: mean, Mean Absolute Deviation (in round brackets), overall mean (bold: values lower than the overall mean) and overall median*

Estimation method	*ML*			*rest. ML*		
g.d.f. index	A_1	A_2	A'_2	A_1	A_2	A'_2
Austria						
mean	**0.1223**	**0.1712**	**0.1880**	**0.1428**	**0.1857**	**0.1873**
(MAD)	(0.0073)	(0.0120)	(0.0147)	(0.0077)	(0.0137)	(0.0130)
Belgium						
mean	0.1563	0.2229	0.2182	**0.1603**	0.2298	0.2214
(MAD)	(0.0097)	(0.0427)	(0.0245)	(0.0116)	(0.0454)	(0.0212)
Denmark						
mean	0.2217	0.2834	0.2937	0.2245	0.2897	0.2863
(MAD)	(0.0081)	(0.0121)	(0.0291)	(0.0115)	(0.0185)	(0.0192)
Finland						
mean	0.1539	**0.2044**	0.2317	0.1862	0.2349	0.2282
(MAD)	(0.0103)	(0.0114)	(0.0180)	(0.0110)	(0.0226)	(0.0163)
France						
mean	**0.0969**	**0.1634**	**0.1553**	**0.1030**	**0.1699**	**0.1543**
	(0.0064)	(0.0233)	(0.0152)	(0.0083)	(0.0251)	(0.0153)
Germany						
mean	**0.0938**	**0.1523**	**0.1705**	**0.1233**	**0.1759**	**0.1713**
(MAD)	(0.0088)	(0.0286)	(0.0191)	(0.0089)	(0.0340)	(0.0181)
Greece						
mean	**0.1190**	**0.1940**	**0.1917**	**0.1374**	**0.2087**	**0.1902**
(MAD)	(0.0106)	(0.0135)	(0.0089)	(0.0098)	(0.0133)	(0.0097)
Holland						
mean	0.1657	0.2332	0.2475	0.1884	0.2537	0.2434
(MAD)	(0.0100)	(0.0103)	(0.0115)	(0.0092)	(0.0127)	(0.0109)
Ireland						
mean	0.2378	0.3559	0.3105	0.2437	0.3570	0.3106
(MAD)	(0.0293)	(0.0346)	(0.0446)	(0.0358)	(0.0356)	(0.0450)
Italy						
mean	**0.1432**	0.2369	**0.2123**	**0.1515**	0.2522	**0.2123**
(MAD)	(0.0214)	(0.0481)	(0.0367)	(0.0243)	(0.0541)	(0.0355)
Luxembourg						
mean	**0.0933**	**0.1348**	**0.1271**	**0.0936**	**0.1367**	**0.1270**
	(0.0189)	(0.0380)	(0.0296)	(0.0189)	(0.0403)	(0.0286)
Portugal						
mean	**0.1303**	**0.1742**	**0.1862**	**0.1448**	**0.1831**	**0.1925**
(MAD)	(0.0128)	(0.0134)	(0.0129)	(0.0122)	(0.0133)	(0.0149)
Spain						
mean	**0.1407**	0.2661	0.2380	**0.1532**	0.2689	0.2410
(MAD)	(0.0162)	(0.0295)	(0.0260)	(0.0160)	(0.0262)	(0.0259)
Sweden						
mean	0.1958	0.2537	0.2918	0.2305	0.2838	0.2811
(MAD)	(0.0185)	(0.0272)	(0.0374)	(0.0182)	(0.0290)	(0.0300)
United Kingdom						
mean	0.1501	**0.1915**	**0.2081**	0.1673	**0.2056**	**0.2049**
(MAD)	(0.0085)	(0.0077)	(0.0125)	(0.0097)	(0.0108)	(0.0110)
Overall mean	0.1469	0.2155	0.2161	0.1614	0.2279	0.2152
Overall median	0.1445	0.2015	0.2088	0.1546	0.2148	0.2089

TABLE 5. - ρ_E: *mean, Mean Absolute Deviation (in round brackets), overall mean (bold: values lower than the overall mean) and overall median*

Estimation method	$minA_1$	$minA_2$	$minA'_2$	ML
Austria				
mean	0.0790	0.0523	**0.0222**	0.0403
(MAD)	(0.0071)	(0.0055)	(0.0072)	(0.0052)
Belgium				
mean	**0.0324**	**0.0373**	**0.0186**	**0.0244**
(MAD)	(0.0234)	(0.0155)	(0.0134)	(0.0145)
Denmark				
mean	**0.0370**	**0.0422**	0.0297	**0.0208**
(MAD)	(0.0296)	(0.0287)	(0.0064)	(0.0142)
Finland				
mean	0.0989	0.0760	0.0330	0.0637
(MAD)	(0.0255)	(0.0216)	(0.0078)	(0.0147)
France				
mean	**0.0420**	**0.0297**	**0.0123**	**0.0224**
(MAD)	(0.0140)	(0.0122)	(0.0055)	(0.0091)
Germany				
mean	**0.0644**	0.0503	0.0283	0.0441
(MAD)	(0.0052)	(0.0114)	(0.0052)	(0.0087)
Greece				
mean	0.1016	0.0687	0.0288	0.0530
(MAD)	(0.0152)	(0.0066)	(0.0066)	(0.0075)
Holland				
mean	0.1003	0.0580	**0.0237**	0.0446
(MAD)	(0.0149)	(0.0094)	(0.0032)	(0.0073)
Ireland				
mean	0.1492	**0.0284**	0.0250	**0.0250**
(MAD)	(0.0616)	(0.0194)	(0.0211)	(0.0218)
Italy				
mean	**0.0554**	0.0590	**0.0225**	0.0437
(MAD)	(0.0139)	(0.0163)	(0.0059)	(0.0153)
Luxembourg				
mean	**0.0097**	**0.0082**	**0.0049**	**0.0067**
(MAD)	(0.0075)	(0.0090)	(0.0028)	(0.0051)
Portugal				
mean	0.0898	0.0509	0.0470	0.0476
(MAD)	(0.0140)	(0.0104)	(0.0067)	(0.0092)
Spain				
mean	**0.0710**	**0.0202**	0.0292	**0.0251**
(MAD)	(0.0093)	(0.0095)	(0.0050)	(0.0035)
Sweden				
mean	0.1122	0.0829	**0.0220**	0.0641
(MAD)	(0.0201)	(0.0099)	(0.0091)	(0.0064)
United Kingdom				
mean	0.1156	0.0602	**0.0217**	0.0478
(MAD)	(0.0188)	(0.0105)	(0.0056)	(0.0080)
Overall mean	0.0759	0.0468	0.0245	0.0371
Overall median	0.0750	0.0482	0.0239	0.0393

estimation method without restriction on the expected value. By the global means we observe that the lowest deviation, on average, is obtained with the method of the minimum A_2', which is equal to 0.0245. The results reported in Table 3 allow to evaluate the implications on the fitting of the introduction of the restriction. Since we are observing the values of the objective function, the restriction of the parametric space increases the goodness of fit indexes, in other words, it reduces the fitting of the model to the empirical distributions.

It can be observed that, on average the restriction increases the objective function of the method of the minimum A_1 of 0.0234, but it reduces the average ρ_E from 0.0759 to zero. For the method of the minimum A_2 the increment of the objective function is 0.0130 and the reduction of ρ_E is 0.0468. For the method of the minimum A_2' the increment of the objective function is 0.0044 and the reduction of ρ_E is 0.0245. It can be therefore observed that the loss in terms of fitting is lower than the gain in terms of ρ_E for each estimation method based on the minimization of goodness of fit indexes. We also observe that the lowest loss of fitting is obtained with the method of the minimum A_2'.

The introduction of the restriction, on average, reduces the ρ_E of the maximum likelihood from 0.0371 to zero. In Table 4 it can be observed that the increment of the indexes A_1 and A_2 are still lower than the gain in terms of ρ_E. The indexes can also decrease with the introduction of the restriction. It can be observed by comparing the third and the sixth columns of Table 4 referred to the index A_2' of the unrestricted and restricted models.

TABLE 6. - *Comparison of the uniformity of the fitting (bold: most uniform fitting over the estimation methods)*

Uniformity measures	Restrictions							
	no restrictions				$\mu = \bar{x}$			
	Estimation method				**Estimation method**			
	$\min A_1$	$\min A_2$	$\min A_2'$	ML	$\min A_1$	$\min A_2$	$\min A_2'$	ML
average ${A_2}^2 - {A_1}^2$	0.0886	**0.0232**	0.0871	0.0271	0.0727	**0.0238**	0.0875	0.0281
average ${A_2'}^2 - {A_1}^2$	0.0472	0.0380	**0.0174**	0.0264	0.0223	0.0283	**0.0166**	0.0212

Final remarks about global results

We suggest to estimate the parameters with the restriction on μ because this restriction can be easily applied to the Zenga's model and, moreover, the mean is one of the most important statistics in the study of incomes (D'Addario, 1934, 1939; Dagum, 1977; McDonald, 1984; McDonald and Ransom, 2008). The results show that in general the higher values of the A_1, A_2 and A_2' indexes due to the restriction are lower than the reduction of ρ_E. Among the estimation methods considered the lowest loss in fitting are obtained with the method of the minimum A_2' and the maximum likelihood method.

Table 6 summarizes the average values of the variances defined in equation (10) to compare the estimation methods in terms of uniformity.

It can be observed that, depending on the uniformity measure used, on average the most uniform fitting is obtained with the methods of the minimum A_2 and of the minimum A_2'.

From these considerations it emerges that with the restriction the recommended estimation methods are the method of the minimum A_2' and the maximum likelihood method.

5.2 *Specific cases*

In this section two specific cases are presented to show the behaviour of the Zenga's distribution and the consequences of the introduction of the restriction on μ.

Figure 1 shows the histogram of the Germany 2001 income distribution and the density functions of the Zenga's model with parameters estimated by the method of the minimum A_2' both without restrictions and with restriction on μ. In Table 7, we provide the parameters estimates, and the values of A_1, A_2, A_2' and ρ_E.

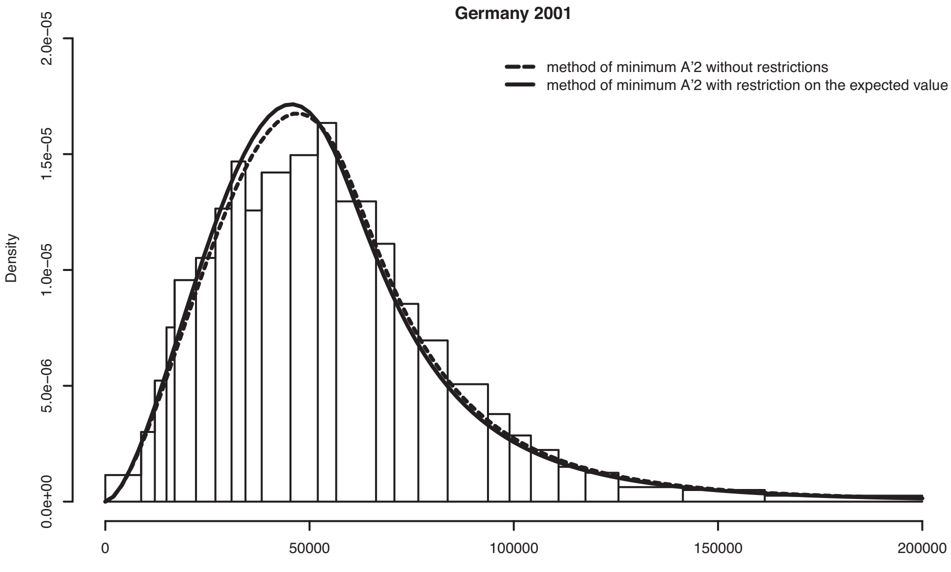

FIGURE 1. - *Histogram of Germany 2001 distribution and Zenga's model estimated by the method of minimum A_2' with and without restriction on μ*

TABLE 7. - *Germany 2001, parameters estimates and goodness of fit indexes obtained with the method of minimum A_2'*

Parameter estimates	$\hat{\mu}$	$\hat{\alpha}$	$\hat{\theta}$	A_1	A_2	A_2'	ρ_E
Unrestricted model	57857.65	2.6756	3.0430	0.1024	0.1318	0.1428	0.0236
Restricted model	56521.40	2.7189	3.1012	0.1147	0.1428	0.1490	

Imposing the restriction on μ we observe an increase of $\hat{\alpha}$ and $\hat{\theta}$. The variation of $\hat{\alpha}$ decreases the inequality by changing the tails, viceversa the variation of $\hat{\theta}$ increases the inequality. Therefore, to evaluate the consequences on the inequality due to the restriction on μ some specific statistics have to be evaluated. Table 8 shows the actual arithmetic mean, Gini index (G) and Zenga's inequality index (I) (see Zenga, 2007) compared with the corresponding indexes evaluated on the theoretical models with the previous parameter estimates. The inequality indexes of the estimated model are obtained through numerical procedures. The indexes G and I of the fitted models, are higher than the actual ones. Note that in this example the restriction also makes the theoretical values closer to the actual ones. The absolute differences between the inequality indexes of the restricted model and the empirical distribution are 0.0052 for I and 0.0074 for G.

To evaluate how the fitted models reproduce the empirical distribution, it can also be useful to compare the Zenga $I(p)$ (Zenga, 2007) and the Lorenz $L(p)$ curves. Figure 2 represents the two inequality curves. Differences between the inequality curves of the restricted and unrestricted models are not significant and observable. For this particular application there is only an intersection point between theoretical and actual curves. Both the curves show that the theoretical models underestimate the inequality for low incomes and overestimate the inequality for high incomes.

TABLE 8. - *Germany 2001, comparison between empirical and theoretical distributions*

Statistics	mean/expected value	G	I
Empirical distribution	*56521.40*	*0.2944*	*0.6324*
Unrestricted model	57857.65	0.3021	0.6382
Restricted model	56521.40	0.3018	0.6376

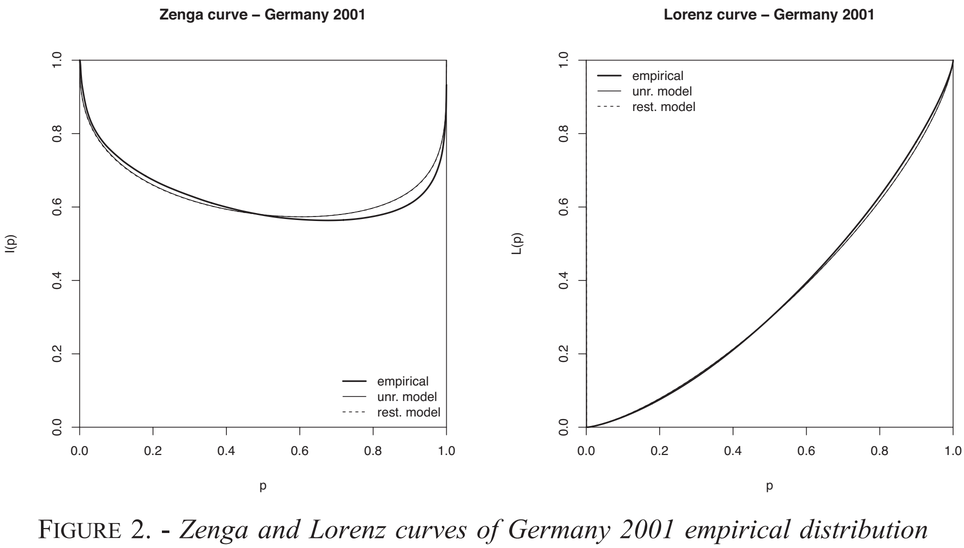

FIGURE 2. - *Zenga and Lorenz curves of Germany 2001 empirical distribution and of the correspondent estimated models*

Figure 3 shows the histogram of the Greece 2001 distribution and the density functions of the estimated models by the method of the minimum A_2' both with and without the restriction on μ. Figure 3 (and Figure 1 in relation to Germany) shows that the empirical distribution has more than one peak. Zenga's model and other traditional parametric models can not describe this characteristic therefore the fitting is reduced, even though it describes well the tails.

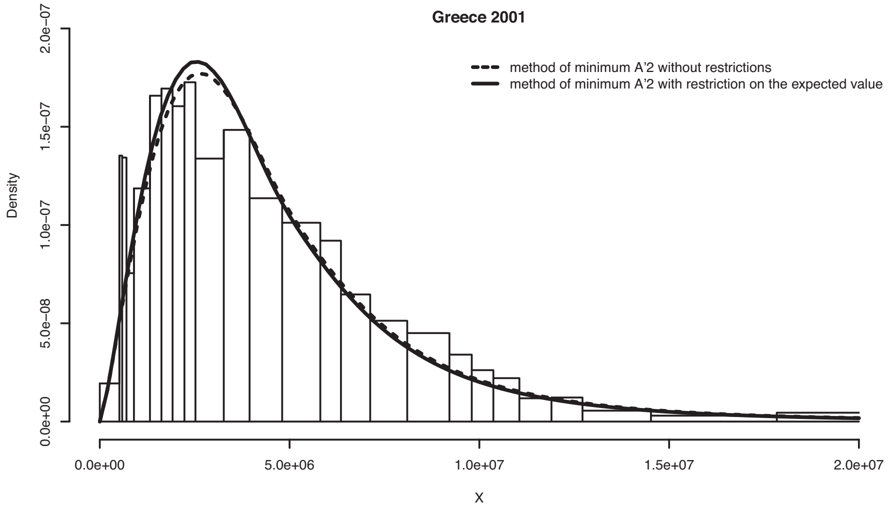

FIGURE 3. - *Histogram of Greece 2001 distribution and Zenga's model estimated by the method of minimum A_2' with and without restriction on μ.*

TABLE 9. - *Greece 2001, parameters estimates and goodness of fit indexes obtained with the method of minimum A_2'*

Parameter estimates	$\hat{\mu}$	$\hat{\alpha}$	$\hat{\theta}$	A_1	A_2	A_2'	ρ_E
Unrestricted model	4927569	2.3169	4.2245	0.1473	0.2136	0.1967	0.0255
Restricted model	4804914	2.3714	4.3440	0.1541	0.2181	0.1993	0

The estimated parameters, the goodness of fit indexes and the relative deviation of the expected value from the actual mean are provided in Table 9. The A_2' indexes are higher than the ones obtained for the Germany 2001 empirical distribution.

In Table 10 the actual arithmetic mean and inequality indexes are compared with the corresponding indexes evaluated on the theoretical models. As observed for the Germany 2001 distribution, the introduction of the restriction on μ reduces the difference between actual and theoretical inequality indexes. The absolute differences between the inequality indexes of the restricted model and the empirical distribution are 0.0058 for I and 0.0130 for G.

By Figure 4 it can be observed that the behaviour of the theoretical and the actual inequality curves is similar to the behaviour observed for Germany 2001.

TABLE 10. - *Greece 2001, comparison between empirical and theoretical distributions*

Statistics	mean/expected value	G	I
Empirical distribution	*4804914*	*0.3828*	*0.7340*
Unrestricted model	4927569	0.3962	0.7406
Restricted model	4804914	0.3958	0.7398

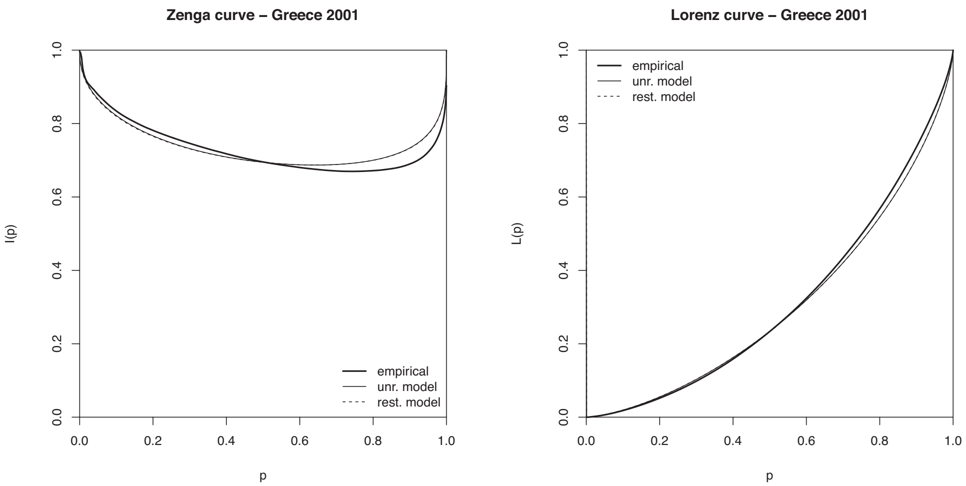

FIGURE 4. - *Zenga and Lorenz curves of Greece 2001 empirical distribution and of the correspondent estimated models*

Now we show how the model can be used to compare two different distributions. Let

$$X_1 \sim f(x : \hat{\mu}_1 = \bar{x}_1; \hat{\alpha}_1; \hat{\theta}_1)$$
$$X_2 \sim f(x : \hat{\mu}_2 = \bar{x}_2; \hat{\alpha}_2; \hat{\theta}_2)$$

be r.v. with Zenga's density function and parameters estimated by method of the minimum A_2' with the restriction on μ, respectively on the Germany 2001 and the Greece 2001 empirical distributions. The comparison between $\hat{\mu}_1$ and $\hat{\mu}_2$ allows to evaluate the difference in the 'magnitude' of the distributions. Eurostat (2003) states that incomes are expressed in national currency (NC) and that 'normally comparisons are made in equivalent units (purchasing power standards - PPS) taking into account differences in the NC's purchasing power. This is obtained by DIVIDING the NC amounts by the purchasing power parities (PPP)'. The PPP of Germany 2001 is equal to 1.9930 and the one of Greece 2001 is equal to 277.2001. Therefore we have to consider the variables

$$Y_1 = X_1/1.9930$$
$$Y_2 = X_2/277.2001.$$

TABLE 11. - *Comparison between Greece 2001 and Germany 2001 income distributions on the basis of fitted models parameters*

	Y_i	$\hat{\mu}_i/PPP_i$	$\hat{\alpha}_i$	$\hat{\theta}_i$	G_i	I_i
Germany 2001	Y_1	28359.76	2.7189	3.1012	0.3018	0.6376
Greece 2001	Y_2	17337.35	2.3714	4.3440	0.3958	0.7398

Since μ is scale parameter we can say that

$$Y_1 \sim f(y : \hat{\mu}_1/1.9930; \hat{\alpha}_1; \hat{\theta}_1)$$
$$Y_2 \sim f(y : \hat{\mu}_2/277.2001; \hat{\alpha}_2; \hat{\theta}_2)$$

are variables with Zenga's density function and that the inequality indexes do not change. We summarize in Table 11 the parameters and the inequality indexes of the variables Y_1 and Y_2. By Table 11 we can observe that on average the PPS in Germany 2001 is higher than the PPS in Greece 2001. Accordingly to both Gini and Zenga indexes, the inequality is higher in Greece than in Germany. The higher inequality of the Greece distribution is confirmed by the higher tails, $\hat{\alpha}_2 < \hat{\alpha}_1$, and by $\hat{\theta}_2 > \hat{\theta}_1$ (see Porro, 2011; Arcagni and Porro, 2012)

In order to compare the inequality of the income distributions in Germany and Greece, we show the Lorenz curve and Zenga curve (Zenga, 2007).

By Figure 5 it can be observed that inequality of Greece 2001 is always higher than inequality of Germany 2001, therefore the estimated model on Greece 2001 is larger than the one estimated on Germany 2001 in the Lorenz ordering and in the ordering based on the Zenga curve.

6. THE PARETIAN RIGHT-TAIL

Since the existence of the moments of Zenga's distribution depends on the value of the parameter α, it can be deduced that the model has a Paretian right-tail. Even if this consideration is enough to characterize the model, it is based on a consequence of the Paretian tail and it lacks a direct proof.

To provide a direct proof, we start from the definition of *Weak (asymptotic) Pareto Law* proposed by Mandelbrot (1960) and recalled by Esteban (1986) and Kleiber and Kotz (2003).

DEFINITION 1 (Weak Pareto Law)
Let X be a non-negative random variable with distribution function $F(x)$, then, the Weak Pareto Law (WPL) asserts that

$$\lim_{x\to\infty} \frac{1 - F(x)}{(x/x_0)^{-\gamma}} = 1, \qquad (12)$$

for some $\gamma > 0$ and $x_0 > 0$.

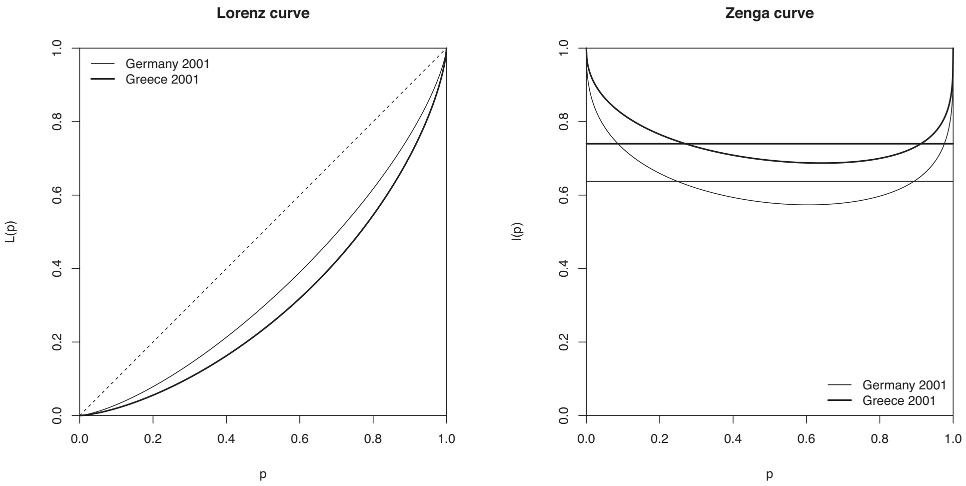

FIGURE 5. - *Comparison of the Zenga and Lorenz curves of the estimated models*

In other words, the WPL means that the *survival function* of the r.v. X ''behaves like'' (Mandelbrot, 1960) the *survival function* of the Pareto distribution.

Now we expose some considerations. Equation (12) can be written as follow

$$\lim_{x\to\infty} \frac{1-F(x)}{x^{-\gamma}} = x_0^{-\gamma},$$

so, it is necessary to find only γ and x_0 follows by the positive and finite result of the limit. It is necessary to proof the following Lemma.

LEMMA 1

Let $w(x; m; a; b)$ be the function of the real variables $x \geq 0$, $m > 0$, $a > 0$ and b, such that

$$w(x; m; a; b) = \frac{\int_0^{m/x} k^{a-1}(1-k)^{b-1} dk}{x^{-a}}$$

then

$$\lim_{x\to\infty} w(x; m; a; b) = \frac{m^a}{a}.$$

PROOF:

Observe that

$$\lim_{x\to\infty} \frac{\int_0^{m/x} k^{a-1}(1-k)^{b-1} dk}{x^{-a}} \tag{13}$$

leads to the indeterminate form $0/0$. Therefore we apply the De l'Hopital's rule. The numerator's derivative is

$$\frac{d}{dx}\int_0^{m/x} k^{a-1}(1-k)^{b-1}dk = \left(\frac{m}{x}\right)^{a-1}\left(1-\frac{m}{x}\right)^{b-1}\left(-\frac{m}{x^2}\right) = \\ = -m^a\left(\frac{1}{x}\right)^{a+1}\left(1-\frac{m}{x}\right)^{b-1}.$$

The denominator's derivative is

$$\frac{d}{dx}x^{-a} = -ax^{-a-1} = -a\left(\frac{1}{x}\right)^{a+1}.$$

For the De l'Hòpital's rule, the limit in equation (13) is equal to

$$\lim_{x\to\infty}\frac{-m^a\left(\frac{1}{x}\right)^{a+1}\left(1-\frac{m}{x}\right)^{b-1}}{-a\left(\frac{1}{x}\right)^{a+1}} = \lim_{x\to\infty}\frac{m^a}{a}\left(1-\frac{m}{x}\right)^{b-1} = \frac{m^a}{a}$$

for any real value of b.

Now, we can show that Zenga's distribution has Paretian right-tail.

THEOREM 1

Zenga's distribution satisfies the WPL.

PROOF:

We are interested to the right tail, therefore we recall the distribution function of Zenga's model for $x > \mu$

$$F(x:\mu;\alpha;\theta) = 1 - \frac{1}{B(\alpha;\theta)}\left\{\left(\frac{\mu}{x}\right)^{0.5}\int_0^{\mu/x} k^{\alpha-0.5}(1-k)^{\theta-2}dx + \\ - \int_0^{\mu/x} k^{\alpha}(1-k)^{\theta-2}dx\right\},$$

that for $\theta > 1$ it can be written as in equation (7), otherwise it requires the series expansion in equation (3).

For a given α, Zenga's distribution has the moment of order $\alpha + 1$, therefore it is reasonable to compare its right-tail with the power function decreasing with order $\gamma = \alpha + 1$. Consequentially to verify if Zenga's distribution satisfies the WPL we evaluate the following limit

$$\lim_{x\to\infty}\frac{1-F(x:\mu;\alpha;\theta)}{x^{-(\alpha+1)}} = \\ = \lim_{x\to\infty}\frac{1}{B(\alpha;\theta)}\left\{\left(\frac{\mu}{x}\right)^{0.5}\int_0^{\mu/x} k^{\alpha-0.5}(1-k)^{\theta-2}dx + \\ - \int_0^{\mu/x} k^{\alpha}(1-k)^{\theta-2}dx\right\}\frac{1}{x^{-(\alpha+1)}} =$$

$$= \lim_{x\to\infty} \frac{1}{B(\alpha;\theta)} \left\{ \mu^{0.5} \frac{\int_0^{\mu/x} k^{\alpha-0.5}(1-k)^{\theta-2}dx}{x^{+0.5}x^{-(\alpha+1)}} + \right.$$

$$\left. - \frac{\int_0^{\mu/x} k^{\alpha}(1-k)^{\theta-2}dx}{x^{-(\alpha+1)}} \right\} =$$

$$= \lim_{x\to\infty} \frac{1}{B(\alpha;\theta)} \left\{ \mu^{0.5} \frac{\int_0^{\mu/x} k^{(\alpha+0.5)-1}(1-k)^{(\theta-1)-1}dx}{x^{-(\alpha+0.5)}} + \right.$$

$$\left. - \frac{\int_0^{\mu/x} k^{(\alpha+1)-1}(1-k)^{(\theta-1)-1}dx}{x^{-(\alpha+1)}} \right\}.$$

Observe that the limit shows two times the function defined in Lemma 1 therefore

$$\lim_{x\to\infty} \frac{1-F(x:\mu;\alpha;\theta)}{x^{-(\alpha+1)}} =$$

$$= \lim_{x\to\infty} \frac{1}{B(\alpha;\theta)} \left\{ \mu^{0.5} w(x;\mu;\alpha+0.5;\theta-1) - w(x;\mu;\alpha+1;\theta-1) \right\} =$$

$$= \frac{1}{B(\alpha;\theta)} \left\{ \mu^{0.5} \frac{\mu^{\alpha+0.5}}{\alpha+0.5} - \frac{\mu^{\alpha+1}}{\alpha+1} \right\} =$$

$$= \frac{1/2}{B(\alpha;\theta)} \frac{\mu^{\alpha+1}}{(\alpha+0.5)(\alpha+1)}.$$

By this demonstration we shown that given a Zenga's distribution of parameters $\mu > 0$, $\alpha > 0$ and $\theta > 0$, its survival function decreases like the survival function of the Pareto distribution with shape parameter $\alpha + 1$ and $x_0 = \frac{1}{\mu}[2B(\alpha;\theta)(\alpha+ +0.5)(\alpha+1)]^{-1/(\alpha+1)}$.

7. CONCLUSIONS

This paper proposes an application of Zenga's model on a wide number of household income distributions. The estimation method used are: minimization of goodness of fit indexes (A_1, A_2 and A_2') and maximum likelihood on grouped data. Each method has been applied without restrictions and with restriction on μ.

We observed that, for each index and estimation method, the best results are obtained with the empirical distributions of the same countries: Austria, France, Germany, Greece, Luxembourg and Portugal. By the values of ρ_E, we observed that the introduction of the restriction on μ implies, on average, a loss of fitting lower than the gain in terms of relative distance of the expected value from the actual mean. In particular, the lowest values of ρ_E are observed with the method of minimum A_2' and the method of the maximum likelihood. With these methods the introduction of the restriction implies also the lowest variations of parameters estimates and the lowest loss in fitting. The method of minimum A_2' also maximize the uniformity of the fitting, therefore we suggest to use this method when the restriction on μ is imposed.

We chose two specific cases (Germany 2001 and Greece 2001) to show how Zenga's model can be used to compare different income distributions. By these two examples it can be observed that the restriction on μ reduces the difference between the actual inequality indexes (Gini's index and Zenga's index) and the ones related to the estimated model. We also compared the actual inequality curves (Lorenz curve and Zenga curve) and the ones of the estimated models. Therefore we observed that the inequality curves of the estimated models do not change significantly with the restriction on μ. However, empirical and theoretical inequality curves intersect only in one point and it is observable that Zenga's model underestimates the inequality for low incomes and overestimates the inequality for high incomes. These last considerations are limited to the two examples we have shown, but they suggest further research to study the effects of a restriction on a global inequality index.

Finally we proved that Zenga's distribution has Paretian right-tail, by showing that is satisfies the Weak Pareto Law.

REFERENCES

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

D'Addario R. (1934). *Sulla misura della concentrazione dei redditi*. Poligrafico 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*.

Dagum C. (1977). A new model of personal income distribution: specification and estimation. *Economie Appliquee*, **30**, 413-437. (Reprinted in: Duangkamon Chotikapanich Editor, ''Modeling Income istributions and Lorenz Curves'', Springer, 2008, pp. 3-25.).

Esteban J. (1986). Income-Share Elasticity and the Size Distribution of Income. *International Economic Review*, **27**(2), 439-444.

Eurostat (2003). *ECHP UDB - version of December 2003*.

Kleiber C., Kotz S. (2003). *Statistical Size Distributions in Economics and Actuarial Sciences*. John Wiley and Sons.

Mandelbrot B. (1960). The Pareto-Lévy Law and the Distribution of Income. *International Economic Review*, **1**(2), 79-106.

McDonald J.B. (1984). Some Generalized Functions for the Size Distribution of Income. *The*

Economic Society, **52**, 647-663. (Reprinted in: Duangkamon Chotikapanich Editor, "Modeling Income Distributions and Lorenz Curves", Springer, 2008, pp. 37-55.).

McDonald J.B., Ransom M. (2008). The Generalized Beta Distribution as a Model for the Distribution of Income: Estimation of Related Measures of Inequality. In D. Chotikapanich (Eds.), *Modeling Income Distributions and Lorenz Curves*, (pp. 147-166). Springer.

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

Porro F. (2011). Inequality order for Zenga distribution. Technical Report 215, Dipartimento di Metodi Quantitativi per le Scienze conomiche Aziendali, Università degli Studi di Milano - Bicocca.

Zenga M.M. (2007). Inequality curve and inequality index based on the ratios between lower and upper arithmetic means. *Statistica & Applicazioni*, **V**(1), 3-27.

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

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

Zenga M.M., Pasquazzi L., Zenga M. (2012). First Applications of a New Three Parameter Distribution for Non-Negative Variables. *Statistica & Applicazioni*, **X**(2), 131-147.