Estimating the three parameters of Zenga's distribution for income by size

Zenga, Michele
Department of Quantitative Methods for Economics and Business (DIMEQUANT), University of Milano-Bicocca
Edificio U7, Via Bicocca degli Arcimboldi 8
Milan, 20126, Italy
E-mail: michele.zenga@unimib.it

Arcagni, Alberto
Department of Quantitative Methods for Economics and Business (DIMEQUANT), University of Milano-Bicocca
Edificio U7, Via Bicocca degli Arcimboldi 8
Milan, 20126, Italy
E-mail: alberto.arcagni@unimib.it

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

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

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

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

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

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

Since μ is the expectation and scale parameter of each component $h(x : \mu; k)$ of the mixture, it is the expectation and scale parameter of the mixture model as well. The mean of the Zenga's model is thus always finite. It may be shown that the Zenga's model has Paretian right tail, i.e.

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

and the asymmetry is always positive.

Expressions for the moments about zero, the cumulative distribution function and some inequality measures are derived and reported in Zenga (2010a) and Zenga *et al.* (2010a).

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 $\theta > 1$ the cumulative distribution function $F(x : \mu; \alpha; \theta)$ is given by

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

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

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

for $x > \mu$, where

$$IB(x : \alpha; \theta) = \int_0^x t^{\alpha-1}(1 - t)^{\theta-1}\, dt, \quad 0 < x < 1$$

is the incomplete beta integral.

The cumulative distribution function for $\theta > 0$ can be expressed by series.

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

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

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

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

Zenga *et al.* (2010c) showed that, in case of the expectation of conditioned distributions is constant, the central moments of a mixture can be obtained in a similar way to the moments about zero, then they derived the variance and the third central moment of the Zenga's model as follow

$$Var(X) = \frac{\mu^2}{3} \cdot \frac{\theta(\theta + 1)}{(\alpha - 1)(\alpha + \theta)}, \quad \text{if } \alpha > 1;$$

$$E(X - \mu)^3 = Var(X) \cdot \frac{3}{5} \cdot \mu \frac{(\theta + 3)(\theta + 2)}{(\alpha + \theta + 1)(\alpha - 2)}, \quad \text{if } \alpha > 2.$$

Empirical distributions and application

Our goal is to provide a fitted model which closely follows the histogram and which preserves the values of some important descriptive indexes like the mean, the median and some inequality measure. In order to do this, *household income* distributions from the European Community Household Panel (ECHP, Eurostat 2003) have been selected. The duration of the ECHP was from 1994 to 2001 and covered 15 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). By the combination of each year and each state, we obtained 114 distributions.

The estimation methods we considered are the method of moments, methods based on minimization of goodness-of-fit indexes and the maximum likelihood method. To evaluate the accuracy of the fitted model have been used goodness-of-fit indexes and relative deviations between empirical indexes and model ones.

Method of Moments

Zenga, *et al.* (2010c) obtained the analytical definitions of Method of Moments estimators for the Zenga's model parameters, by solving the following system

$$\begin{cases} \mu = \bar{x} \\ \frac{\bar{x}^2}{3} \cdot \frac{\theta(\theta+1)}{(\alpha-1)(\alpha+\theta)} = m_2 \\ m_2 \cdot \frac{3}{5} \cdot \frac{(\theta+3)(\theta+2)}{(\alpha+\theta+1)(\alpha-2)} = m_3 \end{cases}$$

where $\bar{x}$, m_2 and m_3 are respectively the arithmetic mean, the variance and the third central moment of the sample. In order to the third equation is defined the third moment has to be finite, then α has to be greater than 2.

The parameters estimates are thus given by

$$\hat{\mu} = \bar{x};$$

$$\hat{\theta} = \frac{-\left[\frac{1}{3}\frac{\bar{x}^2}{m_2} - 3\frac{\bar{x}m_2}{m_3} - 1\right] + \sqrt{\left[\frac{1}{3}\frac{\bar{x}^2}{m_2} - 3\frac{\bar{x}m_2}{m_3} - 1\right]^2 + 4\left[\frac{1}{3}\frac{\bar{x}^2}{m_2} - \frac{3}{5}\bar{x}\frac{m_2}{m_3}\right]\left[\frac{18}{5}\frac{\bar{x}m_2}{m_3} + 2\right]}}{2\left[\frac{1}{3}\frac{\bar{x}^2}{m_2} - \frac{3}{5}\bar{x}\frac{m_2}{m_3}\right]};$$

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

As above remarked, the reported analytical solutions make sense only under the restrictions $\hat{\alpha} > 2$, and $\hat{\theta} > 0$.

Goodness-of-fit indexes

Each empirical distribution, of n observations, has been grouped into s intervals with bounds x'_j for $j = 1, \ldots, s - 1$, $x'_0 = 0$ and $x'_s = \infty$. Let n_j be the empirical frequencies of the s intervals and

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

be the estimated model frequencies.

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.

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

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

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

Minimization of goodness-of-fit indexes

Goodness-of-fit indexes above defined are functions of model parameters through the model frequencies, then estimates can be found by numerical minimization of A_1 (or either one of A_2 and A'_2).

Minimization methods have been implemented both without restrictions and with restrictions on statistics that are considered of interest to describe income distributions. The main interest in the income distribution description is the location, then the estimation with one restriction on the arithmetic mean has been furnished. For the new model it is of easy realization because

$$E(X) = \bar{x} \Leftrightarrow \hat{\mu} = \bar{x}.$$

Moreover an inequality index can be of interest. We realized estimations with the following two restrictions:

- $E(X) = \bar{x}$ and $P(\bar{x}; \alpha; \theta) = \hat{P}$;

- $E(X) = \bar{x}$ and $A(\bar{x}; \alpha; \theta) = \hat{A}(\bar{x})$;

where $P(\bar{x}; \alpha; \theta)$ is the Pietra index of the model as defined in equation (1), $\hat{P}$ is the empirical Pietra index, $A(\bar{x}; \alpha; \theta)$ is the Zenga's point inequality index of the model at $x = \hat{\mu} = \bar{x}$ as defined in equation (2) and $\hat{A}(\bar{x})$ the empirical one. For the Zenga's model, both $P(\bar{x}; \alpha; \theta)$ and $A(\bar{x}; \alpha; \theta)$ are functions of the distribution function in correspondence of the parameter μ, then the restrictions can be easily imposed. The differences between parameter estimates are due to the empirical indexes.

Maximum likelihood method

The maximum likelihood estimates have been provided with and without restrictions.

Generation of random values from the Zenga's model

The cumulative distribution function can not be inverted analytically and the numerical inversion can be time expansive if used in Monte Carlo techniques that require a large number of random values. A solution can be to use the mixture definition and generate the values x from the Zenga's distribution of parameters μ, α and θ by a two-step sampling.

First a value k of the conditioning parameter is generated from the beta distribution of parameters α and θ. The second step is to generate the value x from the truncated Pareto distribution assuming μ equal to the Zenga's distribution one and k equal to the value obtained in the first step.

The major statistical software have optimized algorithms to generate values from the beta distribution. The generation of values from the truncated Pareto distribution can be obtained through the *inverse transformation method* because the distribution function of the the truncated Pareto,

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

is invertible as follow

$$H^{-1}(p: \mu; k) = \mu k \left[1 - (1-k)p\right]^{-2} \quad \text{for } 0 \le p \le 1.$$

Distributions of estimators and goodness-of-fit indexes

The distribution of estimators has been analyzed by parametric Bootstrap resampling, in order to study the bias and the variability. The distribution of goodness-of-fit indexes have been also analyzed, in order to define a bound to compare with values of the goodness-of-fit indexes obtained on the real distributions.

REFERENCES (RÉFERENCES)

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

Dagum, C. (1977), A New Model of Personal Income Distribution: Specification and Estimation, Economie Appliquee, 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, 147-166. Springer, New York.

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

Polisicchio, M. (2008), The continuous random variable with uniform point inequality measure I(p), Statistica & Applicazioni, VI, 2, 137151.

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 Polisicchios Truncated Pareto distributions with beta weights, Statistica & Applicazioni, VIII(1), 3-25.

Zenga, M.M., Pasquazzi L., Zenga Ma. (2010a), First Applications of a New Three Parameter Distribution for Non-Negative Variables, Rapporto di Ricerca N. 188, Dip.to Metodi Quantitativi per le Scienze Economiche ed Aziendali - Università degli Studi di Milano - Bicocca.

Zenga, M.M., Polisicchio M., Zenga Ma., Pasquazzi L. (2010b), More on a new three-parameter distribution for non-negative variables, Rapporto di Ricerca N. 187, Dip.to Metodi Quantitativi per le Scienze Economiche ed Aziendali - Università degli Studi di Milano - Bicocca.

Zenga, M.M., Porro F., Arcagni A. (2010c), Method of momets for Zengas distribution, Rapporto di Ricerca N. 193, Dip.to Metodi Quantitativi per le Scienze Economiche ed Aziendali - Università degli Studi di Milano Bicocca.

RÉSUMÉ (ABSTRACT)

M.M. 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. One of the parameters is equal to the expected value and the other two are inequality indicators. In this work we considered several methods of fitting the new density to empirical distributions and we studied the properties of the resulting estimators. The methods we considered are the method of moments, methods based on minimization of goodness-of-fit indexes and the maximum likelihood method. These indexes are obtained as different types of means of relative distances between observed and theoretical frequencies which depend on the model parameters. Minimization methods have been implemented both without restrictions and with restrictions on statistics that are considered of interest to describe income distributions. These statistics are the arithmetic mean and inequality indexes such as Pietra index and Zenga's point inequality index (Zenga 2007). The analysis of the distributions of the estimators and goodness-of-fit indexes has been conducted by Monte Carlo methods.