Rapporto n. 229

Titolo

On two possible generalizations of the Zenga distribution

Autori

Lucio De Capitani e Alessandro Zini

maggio 2012

Dipartimento di Metodi Quantitativi per le Scienze Economiche ed Aziendali
Università degli Studi di Milano Bicocca
Via Bicocca degli Arcimboldi , 8 - 20126 Milano - Italia
Tel +39/02/64483103- Fax +39/02/64483105
Segreteria di redazione: Andrea Bertolini

ON TWO POSSIBLE GENERALIZATIONS OF ZENGA DISTRIBUTION

Lucio De Capitani*
Alessandro Zini*

SUMMARY

We propose two generalizations of the three-parameters Zenga distribution obtaining two families of distributions with four and five parameters. The generalizations are done starting from two different generalized beta functions and using them as mixing distributions in place of the classical beta. We compare the flexibility of the resulting models with that of the Zenga distribution observing some improvements.

Keywords: *Polisicchio's truncated Pareto distribution, Zenga Distribution, Gauss Hypergeometric distribution, Confluent Hypergeometric distribution, income distribution*

1. INTRODUCTION

Recently, Zenga (2010) introduced a new family of distributions for non negative random variables built as a beta mixture of Polisicchio's truncated Pareto distributions (see Polisicchio, 2008). As it is shown in Zenga (2010) and in Zenga *et al.* (2011), the new family is quite flexible since it collects densities of very different shapes. For example, a Zenga Distribution (ZD) can be zero-modal, uni-modal or bi-modal and the value of the density near the origin ranges in $[0, \infty]$. Moreover, it has a Paretian right tail and it is positively skewed. In this paper, we generalize the ZD family building a mixture of Polisicchio's truncated Pareto distributions with Gauss Hypergeometric and Confluent Hypergeometric weights. The Gauss Hypergeometric distribution adopted in the first generalization was introduced in Armero and Bayarri (1994) and depends on 4 parameters. Then, the number of parameters which characterizes the first kind of Generalized Zenga Distribution (*Type-I GZD*) increases to five. The Confluent Hypergeometric distribution used in the second generalization was introduced in Gordy (1998) and depends on 3 parameters. Then, the second kind of Generalized Zenga Distribution (*Type-II GZD*) has 4 parameters.

As it will be seen in detail, the GZDs are more complex than the ZD from the analytical point of view. However, the additional parameters should further increase the flexibility of the ZD family. Then, a natural question follows: is the increase in the flexibility of the generalized models sufficient to compensate for the increase in the analytical complexity? We think it is important to give an answer to this question. In fact, the analytical complexity and a high number of parameters impact on the practical relevance of the model since they make the estimation procedures difficult and deteriorate the quality of the estimates.

*Dipartimento di Metodi Quantitativi per le Scienze Economiche ed Aziendali- Università degli Studi di Milano-Bicocca - via Bicocca degli Arcimboldi, 8 - 20126 MILANO
(e-mails: ✉ lucio.decapitani1@unimib.it; alessandro.zini@unimib.it).

We first recall the properties of the Polisicchio and Zenga distributions (Section 2) and, later, we introduce the *Type-I* and *Type-II* GZD (Sections 3 and 4, respectively). In Section 5 the flexibility of the two generalized models is compared with that of the ZD while in Section 6 some applications to income distributions are presented. Section 7 is devoted to the conclusions and to a brief description of the further possible lines of research.

2. RECALLING THE POLISICCHIO AND THE ZENGA DISTRIBUTION

The Polisicchio truncated Pareto distribution was introduced in Polisicchio (2008) as the distributional model providing a constant Zenga inequality curve $I(p)$ (see Zenga, 2007). Its density is given by

$$h(x;\mu,k)=\begin{cases}\dfrac{\sqrt{\mu}}{2}k^{\frac{1}{2}}(1-k)^{-1}x^{-\frac{3}{2}} & \mu k\leq x\leq \frac{\mu}{k}\\ 0 & \text{otherwise}\end{cases}, \tag{1}$$

where $\mu>0$ and $0<k<1$. The cumulative distribution function $H(\cdot;\mu,k)$ associated to $h(\cdot;\mu,k)$ is

$$H(x;\mu,k)=\begin{cases}0 & x<\mu\\ \dfrac{1}{1-k}\left[1-\left(\dfrac{x}{\mu}\right)^{-\frac{1}{2}}k^{\frac{1}{2}}\right] & \mu k\leq x\leq \frac{\mu}{k}\\ 1 & \text{otherwise}\end{cases}. \tag{2}$$

Let X be a random variable following the density (1). In Polisicchio (2008) it is shown that

$$E[X^r]=\frac{\mu^r k^{1-r}}{(2r-1)}\frac{1-k^{2r-1}}{1-k}\quad . \tag{3}$$

Following Zenga *et al.* (2011), it is useful to observe that (2) and (3) can be re-expressed as follows:

$$H(x;\mu,k)=\begin{cases}0 & x<\mu\\ \displaystyle\sum_{i=0}^{\infty}k^i-\left(\frac{x}{\mu}\right)^{-\frac{1}{2}}\sum_{i=0}^{\infty}k^{i+\frac{1}{2}} & \mu k\leq x\leq \frac{\mu}{k}\\ 1 & \text{otherwise}\end{cases}, \tag{4}$$

$$E[X^r]=\frac{\mu^r}{2r-1}\sum_{i=1}^{2r-1}k^{i-r}\quad . \tag{5}$$

Zenga (2010) introduced a beta mixture of Polisicchio's distributions integrating the density (1) with respect to k and assuming for this parameter the beta density

$$g(k;\alpha,\theta)=\begin{cases}\dfrac{k^{1-\alpha}(1-k)^{\theta-1}}{B(\alpha,\theta)} & 0<k<1\\ 0 & \text{otherwise}\end{cases}, \tag{6}$$

where $\alpha > 0$, $\theta > 0$, and $B(\alpha,\theta)$ denotes the Beta function. As shown in Zenga *et al.* (2011), the obtained density f and the corresponding distribution F are:

$$f(x;\alpha,\theta,\mu) = \begin{cases} \frac{1}{2\mu B(\alpha,\theta)} \left(\frac{\mu}{x}\right)^{\frac{3}{2}} \sum_{i=0}^{\infty} B\left(\frac{x}{\mu};\alpha+\frac{1}{2}+i,\theta\right) & 0 < x < \mu \\ \frac{1}{2\mu B(\alpha,\theta)} \left(\frac{\mu}{x}\right)^{\frac{3}{2}} \sum_{i=0}^{\infty} B\left(\frac{\mu}{x};\alpha+\frac{1}{2}+i,\theta\right) & \mu < x \end{cases}, \quad (7)$$

$$F(x;\alpha,\theta,\mu) = \begin{cases} \frac{1}{B(\alpha,\theta)} \sum_{i=0}^{\infty} D(x;\alpha+i,\theta,\mu) & 0 < x < \mu \\ 1+\frac{1}{B(\alpha,\theta)} \sum_{i=0}^{\infty} D(x;\alpha+i,\theta,\mu) & \mu < x \end{cases},$$

where $\mu > 0$, $\alpha >$, $\theta > 0$, $B(z;a,b)$ denotes the incomplete beta integral

$$B(z;a,b) = \int_0^z u^{a-1}(1-u)^{b-1}du \qquad z \leq 1, a > 0, b > 0\,,$$

and

$$D(x;\alpha,\theta,\mu) = \begin{cases} B\left(\frac{x}{\mu};\alpha,\theta\right) - \left(\frac{\mu}{x}\right)^{\frac{1}{2}} B\left(\frac{x}{\mu};\alpha+\frac{1}{2},\theta\right) & 0 < x < \mu \\ B\left(\frac{\mu}{x};\alpha+1,\theta\right) - \left(\frac{\mu}{x}\right)^{\frac{1}{2}} B\left(\frac{\mu}{x};\alpha+\frac{1}{2},\theta\right) & x > \mu \end{cases}. \quad (8)$$

When $\theta > 1$, the functions f and F admit a easier representations (see Zenga, 2010):

$$f(x;\alpha,\theta,\mu) = \begin{cases} \frac{1}{2\mu B(\alpha,\theta)} \left(\frac{x}{\mu}\right)^{-\frac{3}{2}} B\left(\frac{x}{\mu};\alpha+\frac{1}{2},\theta-1\right) & 0 < x < \mu \\ \frac{1}{2\mu B(\alpha,\theta)} \left(\frac{\mu}{x}\right)^{\frac{3}{2}} B\left(\frac{\mu}{x};\alpha+\frac{1}{2},\theta-1\right) & \mu < x \end{cases}, \quad (9)$$

$$F(x;\alpha,\theta,\mu) = \begin{cases} \frac{D(x;\alpha,\theta-1,\mu)}{B(\alpha,\theta)} & 0 < x < \mu \\ 1+\frac{D(x;\alpha,\theta-1,\mu)}{B(\alpha,\theta)} & \mu < x \end{cases}.$$

The behavior and the properties of the density (7) and (9) are extensively studied in Zenga (2010) and in Zenga *et al.* (2011). Here we recall that μ is a scale parameter and it coincides with the expectation of a random variable following the ZD. The parameters α and θ play a key role in determining the value assumed by f near the origin and around the value of μ:

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

$$\lim_{x\to\mu} f(x;\alpha,\theta,\mu) = \begin{cases} \dfrac{B\left(\alpha+\frac{1}{2},\theta-1\right)}{2\mu B(\alpha,\theta)} & \theta>1 \\ \infty & 0<\theta\leq 1 \end{cases} .$$

Moreover, denoting with X a random variable with density (7), we recall that

$$E[X^r] = \begin{cases} \dfrac{\mu^r}{2r-1}\displaystyle\sum_{i=1}^{2r-1}\dfrac{B(\alpha-r+i,\theta)}{B(\alpha,\theta)} & r<\alpha+1 \\ \infty & \text{otherwise} \end{cases} ,$$

and

$$E[|X-\mu|] = 2\mu[2F(\mu;\alpha,\theta,\mu)-1] \ . \tag{10}$$

From (10) it follows that the relative mean deviation P due to Pietra (1915) results $P = E[|X-\mu|]/2\mu = 2F(\mu;\alpha,\theta,\mu)-1$. In Zenga (2010) it is also provided the expression for the point inequality index $A(x)$ proposed in Zenga (2007) for the particular case $x=\mu$:

$$A(\mu) = 1-\frac{E[X|X\leq\mu]}{E[X|X>\mu]} = 1-\left(\frac{1-F(\mu;\alpha,\theta,\mu)}{F(\mu;\alpha,\theta,\mu)}\right)^2 \ . \tag{11}$$

Several interesting applications of the Zenga model in the context of income distribution are also given in Zenga *et al.* (2010a); while Zenga *et al.* (2010b) provides the analytical expression of the estimators for α, θ and μ obtained with the method of moments. Finally, Porro (2011) exploited the stochastic ordering induced by the parameters α and θ on the ZD obtaining that the latter is increasing in α and decreasing in θ in the sense of the convex stochastic ordering. He observed that, in the study of concentration this fact implies that α and θ are direct and indirect indicators of concentration, respectively. Finally, we recall that in Zenga *et al.* (2011) the Lower Partial Moment function of the ZD is derived in order to obtain the Lorenz curve and the Zenga's $I(p)$ curve.

3. GAUSS-HYPERGEOMETRIC MIXTURE OF POLISICCHIO'S DISTRIBUTION: TYPE-I GENERALIZED ZENGA DISTRIBUTION

3.1 *Definition of the Type-I GZD*

The procedure described in the previous section can be generalized by substituting to the beta density (6) the Gauss Hypergeometric distribution introduced in Armero and Bayarri (1994). It is given by

$$g(k;\alpha,\beta,\theta,\delta) = \begin{cases} \dfrac{Ck^{\alpha-1}(1-k)^{\theta-1}}{(1-k\delta)^{\beta}} & 0<k<1 \\ 0 & \text{otherwise} \end{cases} , \tag{12}$$

with $\alpha>0$, $\theta>0$, $\theta+\alpha>\beta>0$, and $|\delta|<1$. In (12), the constant C coincides with

$$C = \frac{1}{B(\alpha,\theta)\,{}_2F_1(\alpha,\beta;\theta+\alpha;\delta)} ,$$

where ${}_2F_1$ is the Gauss Hypergeometric series defined by

$$ {}_2F_1(a,b;c;d) = \sum_{h=0}^{\infty} \frac{(a)_h (b)_h}{(c)_h} \frac{d^h}{h!} \qquad \text{with} \qquad (s)_h = s(s+1)\dots(s+h-1) \ . $$

To clearly understand that C in (12) is a normalizing constant, it is sufficient to remember that, when $c > a > 0$, the Gauss Hypergeometric series admits the following *Euler's integral representation*:

$$ {}_2F_1(a,b;c;d) = \frac{1}{B(a,c-a)} \int_0^1 x^{a-1}(1-x)^{c-a-1}(1-xd)^{-b}\, dx \ . $$

Observing that ${}_2F_1(\alpha,\beta;\theta+\alpha;0) = 1$ and ${}_2F_1(\alpha,0;\theta+\alpha;\delta) = 1$ for all the admissible values of α, θ, β, and δ, it turns out that the densities (12) and (6) coincide if $\delta = 0$ or $\beta = 0$.

Now, a Gauss Hypergeometric mixture of Polisicchio's distributions can be introduced (*Type-I* GZD). In detail, let $\eta = (\alpha,\theta,\beta,\delta,\mu)$ and let $f^*(x;\eta)$ denote the density function of the *Type-I* GZD. We have that:

$$ \begin{aligned} f^*(x;\eta) &= \int_0^1 h(x;\mu,k) g(k;\eta)\, dk \\ &= \begin{cases} \dfrac{C\sqrt{\mu}}{2} x^{-\frac{3}{2}} \displaystyle\int_0^{x/\mu} \frac{k^{\alpha+\frac{1}{2}-1}(1-k)^{\theta-2}}{(1-k\delta)^{\beta}}\, dk & 0 < x < \mu \\ \dfrac{C\sqrt{\mu}}{2} x^{-\frac{3}{2}} \displaystyle\int_0^{\mu/x} \frac{k^{\alpha+\frac{1}{2}-1}(1-k)^{\theta-2}}{(1-k\delta)^{\beta}}\, dk & x > \mu \end{cases} . \end{aligned} $$

The cumulative distribution function of the *Type-I* GZD is, then:

$$ F^*(x;\eta) = \begin{cases} \dfrac{C\sqrt{\mu}}{2} \displaystyle\int_0^x \int_0^{y/\mu} y^{-\frac{3}{2}} \frac{k^{\alpha+\frac{1}{2}-1}(1-k)^{\theta-2}}{(1-k\delta)^{\beta}}\, dkdy & 0 < x < \mu \\ F^*(\mu;\eta) + \dfrac{C\sqrt{\mu}}{2} \displaystyle\int_\mu^x \int_0^{\mu/y} y^{-\frac{3}{2}} \frac{k^{\alpha+\frac{1}{2}-1}(1-k)^{\theta-2}}{(1-k\delta)^{\beta}}\, dkdy & x > \mu \end{cases} . $$

Obviously, a *Type-I* GZD with either $\beta = 0$ or $\delta = 0$ is a ZD.

In figures 1 and 2, we give some graphs in order to highlight the possible behavior of the density f^* for several parameters settings.

3.2 *Series representation of F^* and f^**

For computational purposes, it is useful to represent $f^*(\cdot,\eta)$ and $F^*(\cdot,\eta)$ as a series. To do this, as in Zenga *et al.* (2011), we remember that:

$$ (1-k)^{-1} = \sum_{i=0}^{\infty} k^i \quad \text{for all} \quad |k| < 1 \ . \tag{13} $$

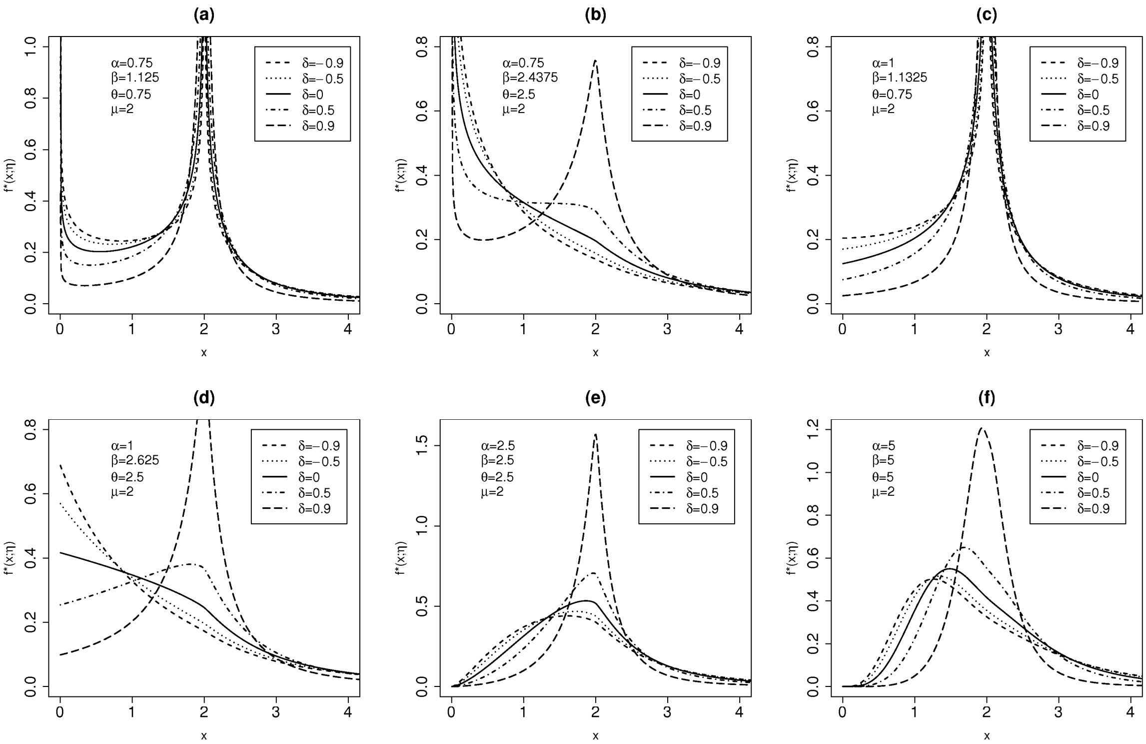

FIGURE 1: Behavior of the density of the *Type-I* GZD for various parameters settings. The plots highlight the impact of the parameter δ on the shape of the density. Note that the solid line corresponds to a ZD.

Moreover, we recall that

$$(1-k\delta)^{-\beta} = \sum_{i=0}^{\infty}(-1)^i \binom{-\beta}{i} \delta^i k^i \quad \text{for all} \quad |k|<1 \quad \text{and} \quad \beta \in \mathbb{R} \ , \tag{14}$$

with generalized binomial coefficient

$$\binom{-\beta}{i} = \prod_{j=1}^{i} \frac{-\beta - j + 1}{j} = \frac{-\beta(-\beta-1)\cdots(-\beta-i+1)}{i!} \ .$$

Observing that $\binom{-\beta}{i} = (-1)^i \binom{\beta+i-1}{i}$, expression (14) can be re-expressed as follows:

$$(1-k\delta)^{-\beta} = \sum_{i=0}^{\infty} \binom{\beta+i-1}{i} \delta^i k^i \quad \text{for all} \quad |k|<1 \quad \text{and} \quad \beta \in \mathbb{R} \ . \tag{15}$$

Thanks to (13) and (15) we obtain that:

$$f^*(x;\eta) = \begin{cases} \dfrac{C\sqrt{\mu}}{2} x^{-\frac{3}{2}} \displaystyle\sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \delta^j \binom{\beta+j-1}{j} B\left(\frac{x}{\mu};\alpha+i+j+\frac{1}{2},\theta\right) & 0<x<\mu \\ \dfrac{C\sqrt{\mu}}{2} x^{-\frac{3}{2}} \displaystyle\sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \delta^j \binom{\beta+j-1}{j} B\left(\frac{\mu}{x};\alpha+i+j+\frac{1}{2},\theta\right) & x>\mu \end{cases} . \tag{16}$$

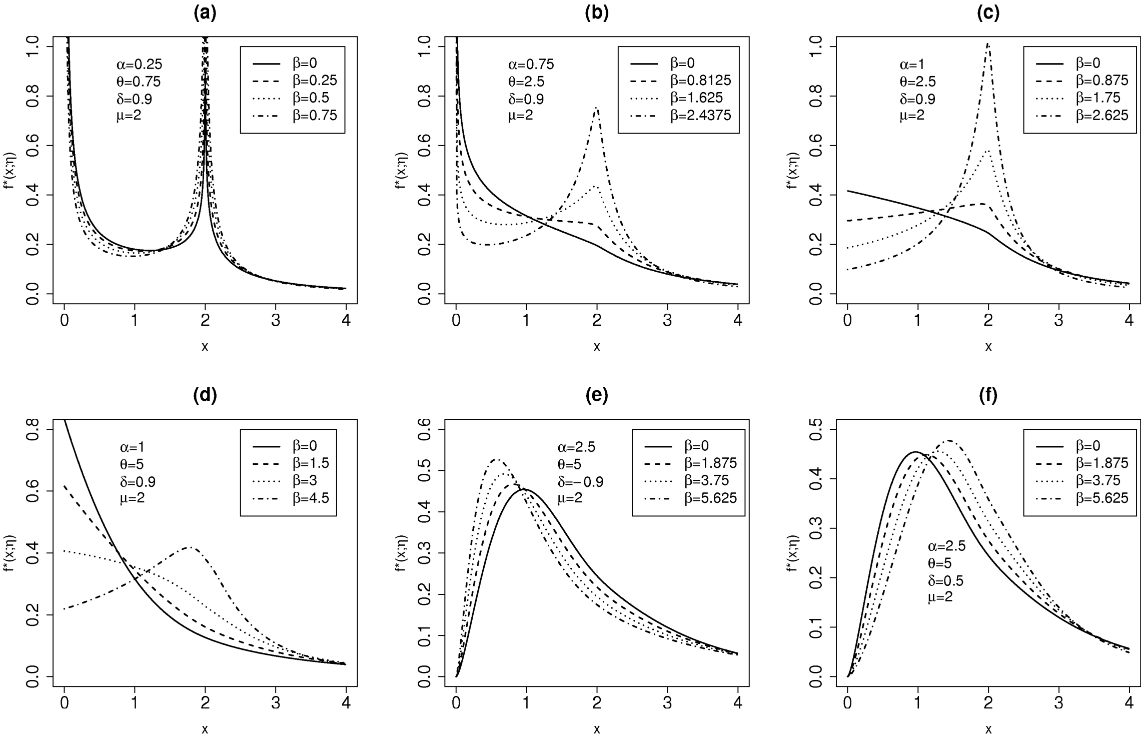

FIGURE 2: Behavior of the density of the *Type-I* GZD for various parameters settings. The plots highlight the impact of the parameter β on the shape of the density. Note that the solid line corresponds to a ZD.

In order to obtain a series representation of $F^*(\cdot;\eta)$ we observe that:

$$\begin{aligned} F^*(x;\eta) &= \begin{cases} \displaystyle\int_0^{x/\mu} H(x;\mu,k)g(k)\,dk & 0<x<\mu \\ \displaystyle\int_0^{\mu/x} H(x;\mu,k)g(k)\,dk + \int_{\mu/x}^{1} g(k)\,dk & x>\mu \end{cases} \\ &= \begin{cases} \displaystyle\int_0^{x/\mu} H(x;\mu,k)g(k)\,dk & 0<x<\mu \\ \displaystyle 1-\int_0^{\mu/x} g(k)\,dk + \int_0^{\mu/x} H(x;\mu,k)g(k)\,dk & x>\mu \end{cases}. \end{aligned}$$

Thanks to the expression for $H(\cdot;\mu,k)$ provided in (4) and to formula (15), the above expression can be rewritten as

$$F^*(x,\eta) = \begin{cases} \displaystyle C\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\binom{\beta+j-1}{j}\delta^j D(x;\alpha+i+j,\theta,\mu) & 0<x<\mu \\ \\ \displaystyle 1+C\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\binom{\beta+j-1}{j}\delta^j D(x;\alpha+i+j,\theta,\mu) & x>\mu \end{cases}. \tag{17}$$

As aforementioned, the series representation (16) and (17) are particularly useful from the computational point of view since they avoid the extensive use of numerical integration. However, they involve a double summation and, consequently, the computational time required to obtain a sufficiently precise value of f^* of F^* is high. From a practical point of view, the latter problem can be avoided observing that it is usually reasonable to assume that the value of the density in μ is finite and, then, $\theta > 1$. In particular, if $\theta > 1$, thanks to (15), we obtain the following "single series" representation of f^*:

$$f^*(x;\eta) = \begin{cases} \frac{C\sqrt{\mu}}{2} x^{-\frac{3}{2}} \sum_{j=0}^{\infty} \delta^j \binom{\beta+j-1}{j} B\left(\frac{x}{\mu}; \alpha+j+\frac{1}{2}, \theta-1\right) & 0 < x \leq \mu \\ \frac{C\sqrt{\mu}}{2} x^{-\frac{3}{2}} \sum_{j=0}^{\infty} \delta^j \binom{\beta+j-1}{j} B\left(\frac{\mu}{x}; \alpha+i+j+\frac{1}{2}, \theta-1\right) & x > \mu \end{cases} . \tag{18}$$

Moreover, from expressions (2) and (15), following a procedure similar to that applied to derive (17), it results that, if $\theta > 1$:

$$F^*(x,\eta) = \begin{cases} C \sum_{j=0}^{\infty} \binom{\beta+j-1}{j} \delta^j D(x; \alpha+j, \theta-1, \mu) & 0 < x < \mu \\ 1 + C \sum_{j=0}^{\infty} \binom{\beta+j-1}{j} \delta^j D(x; \alpha+j, \theta-1, \mu) & x > \mu \end{cases} . \tag{19}$$

3.3 *Some properties of the Type-I GZD*

The *Type-I* GZD shares several properties with the ZD. For example, as for the ZD, μ is a scale parameter and it coincides with the expectation of a random variable following the *Type-I* GZD. Moreover, also now, the parameters α and θ play a key role in determining the value f^* assumes for $x \to 0^+$ and $x \to \mu$, respectively:

$$\lim_{x\to 0^+} f^*(x;\eta) = \begin{cases} \infty & \alpha < 1 \\ \frac{\theta}{3\mu \, {}_2F_1(1,\beta;\theta+1;\delta)} & \alpha = 1 \\ 0 & \alpha > 1 \end{cases} ;$$

$$\lim_{x\to\mu} f^*(x;\eta) = \begin{cases} \frac{B\left(\alpha+\frac{1}{2}, \theta-1\right) \, {}_2F_1\left(\alpha+\frac{1}{2}, \beta; \theta+\alpha-\frac{1}{2}; \delta\right)}{2\mu B(\alpha,\theta) \, {}_2F_1(\alpha,\beta;\theta+\alpha;\delta)} & \theta > 1 \\ \infty & 0 < \theta \leq 1 \end{cases} .$$

The maximum order of the existing moments is determined by the parameter α. In fact, remembering that the r-th moment of a Polisicchio's distribution can be expressed as in (5), denoting with X a random variable following the *Type I*-GZD, we have that:

$$\begin{aligned} E[X^r] &= \int_0^1 \frac{C\mu^r}{2r-1}\left(\sum_{i=1}^{2r-1} k^{i-r}\right)\frac{k^{\alpha-1}(1-k)^{\theta-1}}{(1-k\delta)^{\beta}}\,dk \\ &= \begin{cases} \frac{C\mu^r}{2r-1}\sum_{i=1}^{2r-1} B(\alpha-r+i,\theta)\,{}_2F_1(\alpha-r+i,\beta;\theta+\alpha-r+i;\delta) & r<\alpha+1 \\ \infty & \text{otherwise} \end{cases} . \end{aligned}$$

It is worthwhile to note that the parameters β and δ added to the ZD do not have impact on the maximum order of the existing moments and on the finiteness/infiniteness of f^* around μ and in 0.

In order to obtain the MAD of X it is useful to note that, for a Polisicchio random variable Y, it can be easily expressed in terms of $H(\mu;\mu,k)$ as follows:

$$\begin{aligned} E[|Y-\mu|] &= 2\int_{\mu}^{\mu/k}[1-H(x;\mu,k)]dy \\ &= 2\mu[2H(\mu;\mu,k)-1]\ . \end{aligned} \tag{20}$$

Consequently, it results that

$$\begin{aligned} E[|X-\mu|] &= 2\mu C\int_0^1 (2H(\mu;\mu,k)-1)\frac{k^{\alpha-1}(1-k)^{\theta-1}}{(1-k\delta)^{\beta}}\,dk \\ &= 2\mu[2F^*(\mu;\eta)-1]\ . \end{aligned} \tag{21}$$

The Pietra relative mean deviation of a *Type-I* GZD is then given by $P=(2F^*(\mu;\eta)-1)$. The evident resemblance of expressions (10) and (21) stems from expression (20) and it emphasizes that the MAD of any mixture of Polisicchio random variables with common parameter μ can be expressed in terms of the mixture distribution function as in formula (21). A similar result holds also for the point inequality measure $A(\mu)$. In detail, note that, for a general positive random variable Z, with expectation μ and distribution function M, it results that

$$A(\mu)=1-\frac{2\mu M(\mu)-E[|Z-\mu|]}{2\mu(1-M(\mu))+E[|Z-\mu|]}\left(\frac{1-M(\mu)}{M(\mu)}\right)\ .$$

From the above expression, thanks to (21), we obtain that the index $A(\mu)$ for a *Type-I* GZD is

$$A(\mu)=1-\left(\frac{1-F^*(\mu;\eta)}{F^*(\mu;\eta)}\right)^2\ . \tag{22}$$

The most widespread and important tools to describe the inequality implicit in a distribution are undoubtedly the Lorenz curve and the Zenga $I(p)$ curve. These curves can be easily obtained from the Lower Partial Moment function (LPM) which can be derived using Lemma 4 and Lemma 5 on page 9 in Zenga *et al.* (2011). Also here, the LPM function is derived only for the particular case $\mu=1$ since μ is a scale parameter and, then, it does not influence the shape of the Lorenz and $I(p)$ curves. Using the aforementioned lemmas together with the series representation (16) we have that

$$LPM(x;\alpha,\theta,\beta,\delta,1) = \begin{cases} C\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\binom{\beta+j-1}{j}\delta^j\tilde{D}(x;\alpha+i+j,\theta) & 0<x<1 \\ 1+C\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\binom{\beta+j-1}{j}\delta^j\tilde{D}(x;\alpha+i+j,\theta) & x>1 \end{cases} ; \tag{23}$$

where

$$\tilde{D}(x;\alpha,\theta) = \begin{cases} \sqrt{x}B\left(x;\alpha+\frac{1}{2},\theta\right) - B\left(x;\alpha+1,\theta\right) & 0<x<1 \\ B\left(\frac{1}{x};\alpha,\theta\right) - \sqrt{x}B\left(\frac{1}{x};\alpha+\frac{1}{2},\theta\right) & x>1 \end{cases} . \tag{24}$$

It is worthwhile to note that if $\theta>1$, the "single series" representation (18) leads to the following easier expression for the LPM:

$$LPM(x;\alpha,\theta,\beta,\delta,1) = \begin{cases} C\sum_{j=0}^{\infty}\binom{\beta+j-1}{j}\delta^j\tilde{D}(x;\alpha+j,\theta-1) & 0<x<1 \\ 1+C\sum_{j=0}^{\infty}\binom{\beta+j-1}{j}\delta^j\tilde{D}(x;\alpha+j,\theta-1) & x>1 \end{cases} . \tag{25}$$

Thus, the Lorenz curve $L(p;\eta)$ of a *Type-I* GZD of parameters η is given by

$$L(p;\eta) = LPM(Q^*(p;\eta);\alpha,\theta,\beta,\delta,1) \qquad 0<p<1 ; \tag{26}$$

where $Q^*(\cdot;\eta)$ denotes the quantile function of the *Type-I* GZD. Now, the Zenga's inequality curve can be obtained using its relationship with the Lorenz curve (see Zenga, 2007, formula 5.15 on page 17):

$$I(p;\eta) = \frac{p-L(p;\eta)}{p[1-L(p;\eta)]} \qquad 0<p<1 . \tag{27}$$

A further interesting feature would be to investigate the stochastic ordering induced by the parameters of the *Type-I* GZD. This topic is particularly relevant for the meaning of the parameters, especially in applications related to income distribution. In order to study the role of the parameters α, β, θ, and δ, we introduce the following theorem, inspired by Theorem 1.A.6 on page 7 in Shaked and Shanthikumar (2007).

THEOREM 1 *Consider a family of distribution functions* $\{H_\gamma, \gamma\in\chi\}$ *where* χ *is a subset of the real line and let* $Y(\gamma)$ *denote a random variable with distribution* H_γ. *Let* K_1 *and* K_2 *be two random variables with supports in* χ *and distribution functions* G_1

and G_2, respectively. Let X_1 and X_2 be two random variables with distributions F_1 and F_2, where:

$$F_i(y) = \int_{\chi} H_\gamma(y) dG_i(\gamma), \qquad y \in \mathbb{R},\ i = 1,2\ .$$

If $Y(\gamma) \geq_{cx} Y(\gamma')$ whenever $\gamma \leq \gamma'$ and if $K_1 \leq_{st} K_2$, then $X_1 \geq_{cx} X_2$.

PROOF
The theorem can be proved following the scheme of proof of Theorem 4.A.18 on page 191 in Shaked and Shanthikumar (2007): we show that $E[\phi(X_1)] \geq E[\phi(X_2)]$ for all the convex functions $\phi : \mathbb{R} \to \mathbb{R}$ for which the last expectations exist (see the definition of convex stochastic order in Shaked and Shanthikumar, 2007, p. 109). Let ϕ be such a function and let us introduce the function $\psi(\gamma) = E[\phi(Y(\gamma))]$. Note that $E[\phi(X_i)] = E[\psi(K_i)]$, $i = 1,2$. Moreover, the function $\psi(\gamma)$ is decreasing in γ since, by hypotheses, $Y(\gamma) \geq_{cx} Y(\gamma')$ whenever $\gamma \leq \gamma'$ (see, again, the definition of convex stochastic order). Since ψ is increasing, the definition of the classical stochastic ordering (see Shaked and Shathikumar, 2007, page 4) assures that $E[\psi(K_1)] \geq E[\psi(K_2)]$ and the thesis then follows. $\diamond$

Now, we introduce the following proposition which describes a criterion very useful in order to establish the existence of the usual stochastic ordering between two random variables.

PROPOSITION 1 *Let Y_1 and Y_2 be two random variables with common support $\mathcal{S}$ and with continuous, strictly positive density g_1 and g_2, respectively. If $g_1(y)/g_2(y)$ is strictly decreasing in y for all $y \in \mathcal{S}$, then $Y_1 \leq_{st} Y_2$.*

PROOF
First, observe that if $g_1(y)/g_2(y)$ is strictly decreasing in y for all $y \in \mathcal{S}$ then there is one (and only one) point $y^* \in \mathcal{S}$ such that $g_1(y^*)/g_2(y^*) = 1$. Then we have that: $g_1(y) > g_2(y)$ for all $y < y^*$, $y \in \mathcal{S}$; $g_1(y) = g_2(y)$ if $y = y^*$; $g(y) < f(y)$ for all $y > y^*$, $y \in \mathcal{S}$. The proposition now follows from Theorem 1.A.12 on page 10 in Shaked and Shanthikumar (2007). $\diamond$

As aforementioned, Theorem 1 is useful in order to interpret the role of the parameters α, β, θ and δ in determining the shape of the *Type-I* GZD distribution. In particular we can state the following corollary.

COROLLARY 1 *Let $X(\mu, \alpha, \beta, \theta, \delta)$ be a random variable following the Type-I GZD with parameters μ, α, β, θ, and δ. We have that:*

1. *keeping fixed the values of μ, θ, and β the random variable X is stochastically decreasing in α and δ in the sense of the convex stochastic ordering:*

$$X(\mu,\alpha,\beta,\theta,\delta) \geq_{cx} X(\mu,\alpha',\beta,\theta,\delta') \quad \textit{whenever}\ \alpha \leq \alpha',\ \delta \leq \delta'.$$

2. *keeping fixed the values of μ, α, β, and δ, the random variable X is stochastically increasing in θ in the sense of the convex stochastic ordering:*

$$X(\mu,\alpha,\beta,\theta,\delta) \leq_{cx} X(\mu,\alpha,\beta,\theta',\delta) \quad \textit{whenever}\ \theta \leq \theta'.$$

3. *keeping fixed the values of μ, α, θ, and δ the random variable X is stochastically increasing (decreasing) in β if $\delta > 0$ ($\delta < 0$) in the sense of the convex stochastic ordering:*

$$X(\mu,\alpha,\beta,\theta,\delta) \geq_{cx} X(\mu,\alpha,\beta',\theta,\delta) \quad \textit{whenever} \quad \beta \leq \beta' \quad \textit{and} \quad \delta > 0\ ;$$
$$X(\mu,\alpha,\beta,\theta,\delta) \leq_{cx} X(\mu,\alpha,\beta',\theta,\delta) \quad \textit{whenever} \quad \beta \leq \beta' \quad \textit{and} \quad \delta < 0.$$

PROOF
Applying Theorem 3.A.44 on page 133 in Shaked and Shanthikumar (2007), it turns out that the Polisicchio random variables with the same value of μ are decreasing in k with respect to the convex stochastic order. Now, applying Proposition 1, we obtain that a Gauss Hypegeometric random variable is stochastically increasing in α and δ with respect to the usual stochastic ordering. To the contrary, a Gauss Hypegeometric random variable is stochastically decreasing in θ with respect to the usual stochastic ordering. Finally, we obtain that a Gauss Hypegeometric random variable is stochastically decreasing (increasing) in β if $\delta < 0$ ($\delta > 0$). Now, the thesis follows directly from Theorem 1. $\diamond$

Thanks to Corollary 1 we can assert that, for the Type-I GZD, the parameters α and δ are indirect inequality indicators while θ is a direct inequality indicator. The parameter β is a direct (indirect) inequality indicator if $\delta < 0$ ($\delta > 0$). Obviously the interpretation provided for the parameters α and θ is coherent with the one provided in Porro (2011) for the ZD.

4. CONFLUENT-HYPERGEOMETRIC MIXTURE OF POLISICCHIO'S DISTRIBUTION: GENERALIZED ZENGA DISTRIBUTION OF THE SECOND KIND

4.1 *Definition of the Type-II GZD*

A further generalization of the Zenga distribution can be obtained adopting the Confluent-Hypergemometric distribution (see Gordy, 1198) in place of the classical beta (6) or the Gauss Hypergeometric (12). The Confluent-hypergeometric distribution is given by:

$$g(k;\alpha,\theta,\gamma) = \begin{cases} C' k^{\alpha-1}(1-k)^{\theta-1}e^{-\gamma k} & 0<k<1 \\ 0 & \text{otherwise} \end{cases} \tag{28}$$

with $\alpha > 0$, $\theta > 0$, and $\gamma \in \mathbb{R}$. In (28), the constant C' coincides with:

$$C' = \frac{1}{B(\alpha,\theta)\,{}_1F_1(\alpha;\theta+\alpha;-\gamma)}\ ,$$

where ${}_1F_1$ is the Confluent Hypergeometric series defined by:

$${}_1F_1(a;b;c) = \sum_{h=0}^{\infty} \frac{(a)_h}{(b)_h}\frac{c^h}{h!} \qquad \text{with} \qquad (s)_h = s(s+1)\dots(s+h-1)\ .$$

Also in this case, to clearly see that C' in (28) is a normalizing constant, it is sufficient to remember the *Euler's integral representation* of the Confluent Hypergeometric series:

$$ {}_1F_1(a;b;c) = \frac{1}{B(a,b-a)} \int_0^1 x^{a-1}(1-x)^{b-a-1} e^{cx}\,dx \qquad \text{where} \qquad b > a > 0\,. $$

Observing that ${}_1F_1(\alpha;\theta+\alpha;0) = 1$ for $\alpha > 0$ and $\theta > 0$ it turns out that the densities (28) and (6) coincide if $\gamma = 0$.

The Confluent Hypergeometric mixture of Polisicchio's distributions (*Type-II* GZD) can, now, be introduced. Let $\eta' = (\alpha,\theta,\gamma,\mu)$ and let $f_e(x;\eta')$ denote the density ot the *Type-II* GZD. We have that:

$$ \begin{aligned} f_e(x;\eta') &= \int_0^1 h(x;\mu,k)g(k;\eta')\,dk \\ &= \begin{cases} \dfrac{C'\sqrt{\mu}}{2} x^{-\frac{3}{2}} \displaystyle\int_0^{x/\mu} k^{\alpha+\frac{1}{2}-1}(1-k)^{\theta-2} e^{-\gamma k}\,dk & 0 < x < \mu \\ \dfrac{C'\sqrt{\mu}}{2} x^{-\frac{3}{2}} \displaystyle\int_0^{\mu/x} k^{\alpha+\frac{1}{2}-1}(1-k)^{\theta-2} e^{-\gamma k}\,dk & x > \mu \end{cases}. \end{aligned} $$

The cumulative distribution function of the Type-II GZD is, then, obtained:

$$ F_e(x;\eta') = \begin{cases} \dfrac{C'\sqrt{\mu}}{2} \displaystyle\int_0^x \int_0^{y/\mu} y^{-\frac{3}{2}} k^{\alpha+\frac{1}{2}-1}(1-k)^{\theta-2} e^{-\gamma x}\,dkdy & 0 < x < \mu \\ F_e(\mu;\eta') + \dfrac{C'\sqrt{\mu}}{2} \displaystyle\int_\mu^x \int_0^{\mu/y} y^{-\frac{3}{2}} k^{\alpha+\frac{1}{2}-1}(1-k)^{\theta-2} e^{-\gamma x}\,dkdy & x > \mu \end{cases}. $$

Obviously, a *Type-II* GZD with $\gamma = 0$ is a ZD.

In Figure 3 we give some graphs in order to highlight the possible behavior of the density f^* for several parameters settings.

4.2 *Series representation of F_e and f_e*

For computational purposes, it is useful to represent also the density $f_e(\cdot,\eta')$ and the distribution $F_e(\cdot,\eta')$ as a series. Taking the series expansion of the exponential function and following the same procedure as in Section 3.2 we obtain that:

$$ f_e(x;\eta') = \begin{cases} \dfrac{C'\sqrt{\mu}}{2} x^{-\frac{3}{2}} \displaystyle\sum_{i=0}^{\infty}\sum_{j=0}^{\infty} (-1)^j \frac{\gamma^j}{j!} B\left(\frac{x}{\mu};\alpha+i+j+\frac{1}{2},\theta\right) & 0 < x \leq \mu \\ \dfrac{C'\sqrt{\mu}}{2} x^{-\frac{3}{2}} \displaystyle\sum_{i=0}^{\infty}\sum_{j=0}^{\infty} (-1)^j \frac{\gamma^j}{j!} B\left(\frac{\mu}{x};\alpha+i+j+\frac{1}{2},\theta\right) & x > \mu \end{cases} ; \tag{29} $$

$$ F_e(x,\eta') = \begin{cases} C' \displaystyle\sum_{i=0}^{\infty}\sum_{j=0}^{\infty} (-1)^j \frac{\gamma^j}{j!} D\left(x;\alpha+i+j,\theta,\mu\right) & 0 < x \leq \mu \\ 1 + C' \displaystyle\sum_{i=0}^{\infty}\sum_{j=0}^{\infty} (-1)^j \frac{\gamma^j}{j!} D\left(x;\alpha+i+j,\theta,\mu\right) & x > \mu \end{cases}. \tag{30} $$

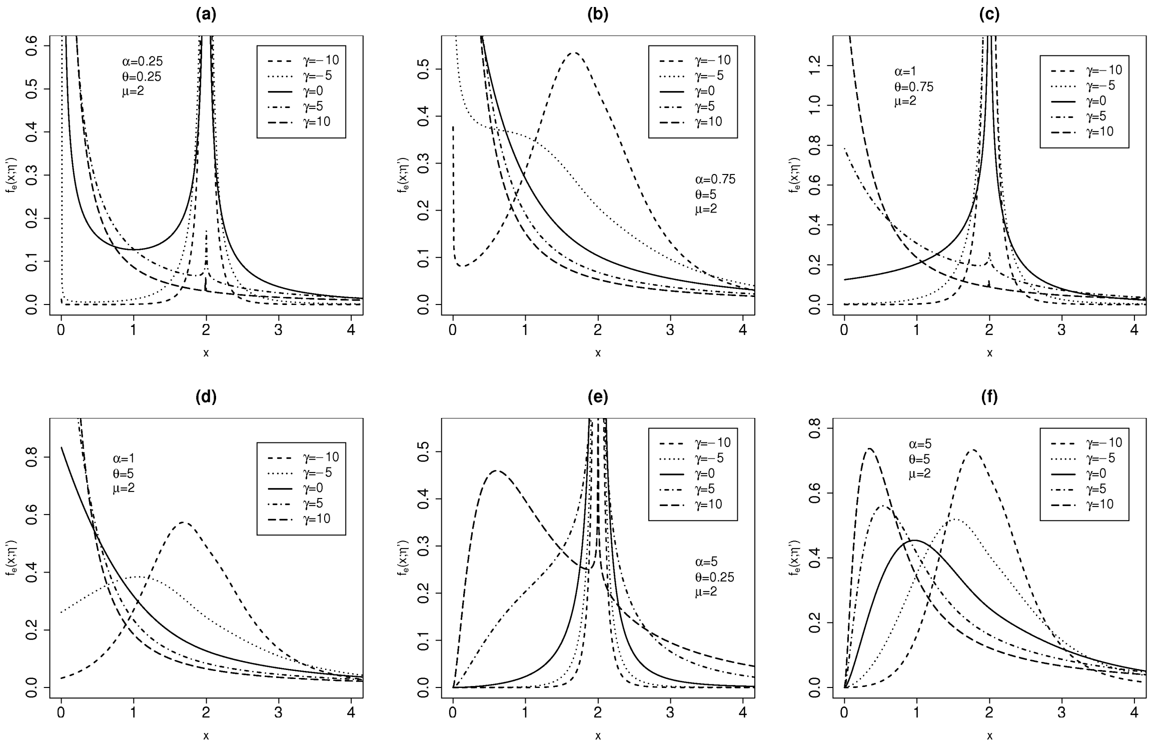

FIGURE 3: Behavior of the density of the *Type-II* GZD for various parameters settings. The plots highlight the impact of the parameter γ on the shape of the density. Note that the solid line correspond to a ZD.

If $\theta > 1$, the function f_e and F_e admit the following "single-series" representation:

$$f_e(x;\eta') = \begin{cases} \frac{C'\sqrt{\mu}}{2} x^{-\frac{3}{2}} \sum_{j=0}^{\infty} (-1)^j \frac{\gamma^j}{j!} B\left(\frac{x}{\mu}; \alpha + j + \frac{1}{2}, \theta - 1\right) & 0 < x \leq \mu \\ \frac{C'\sqrt{\mu}}{2} x^{-\frac{3}{2}} \sum_{j=0}^{\infty} (-1)^j \frac{\gamma^j}{j!} B\left(\frac{\mu}{x}; \alpha + j + \frac{1}{2}, \theta - 1\right) & x > \mu \end{cases} ; \tag{31}$$

$$F_e(x,\eta') = \begin{cases} C' \sum_{j=0}^{\infty} (-1)^j \frac{\gamma^j}{j!} D\left(x; \alpha + j, \theta - 1, \mu\right) & 0 < x \leq \mu \\ 1 + C' \sum_{j=0}^{\infty} (-1)^j \frac{\gamma^j}{j!} D\left(x; \alpha + j, \theta - 1, \mu\right) & x > \mu \end{cases} . \tag{32}$$

4.3 *Some properties of the Type-II GZD*

In analogy with the results remembered for the Zenga distribution and obtained for the *Type-I* GZD, we observe that μ is again a scale parameter and it coincides with the expectation of a random variable following the *Type-II* GZD. Moreover, as in the previous cases, the parameters α and θ play a key role in determining the value f_e

assumes for $x \to 0^+$ and $x \to \mu$, respectively:

$$\lim_{x\to 0^+} f_e(x;\eta') = \begin{cases} \infty & \alpha < 1 \\ \dfrac{\theta}{3\mu\, {}_1F_1(1;\theta+1;-\gamma)} & \alpha = 1 \\ 0 & \alpha > 1 \end{cases} ;$$

$$\lim_{x\to\mu} f_e(x;\eta') = \begin{cases} \dfrac{B\left(\alpha+\frac{1}{2},\theta-1\right)\, {}_1F_1\left(\alpha+\frac{1}{2};\theta+\alpha-\frac{1}{2};-\gamma\right)}{2\mu B(\alpha,\theta)\, {}_1F_1(\alpha;\theta+\alpha;-\gamma)} & \theta > 1 \\ \infty & 0 < \theta \le 1 \end{cases} .$$

Denoting with X a random variable following the *Type-II* GZD, it results that

$$E[X^r] = \begin{cases} \dfrac{C'\mu^r}{2r-1}\displaystyle\sum_{i=1}^{2r-1} B(\alpha-r+i,\theta)\, {}_1F_1(\alpha-r+i;\theta+\alpha;-\gamma) & r < \alpha+1 \\ \infty & \text{otherwise} \end{cases} .$$

Note that the parameter γ added to the ZD does not impact on the maximum order of the existing moments and on the finiteness/infiniteness of f_e around μ and 0. Moreover, from the observation made in Section 3.3 we deduce that

$$E[|X-\mu|] = 2\mu[2F_e(\mu;\eta')-1] , \qquad P = (2F_e(\mu;\eta')-1) , \qquad \text{and} \tag{33}$$

$$A(\mu) = 1 - \left(\frac{1-F_e(\mu;\eta')}{F_e(\mu;\eta')}\right)^2 . \tag{34}$$

Using again Lemma 4 and Lemma 5 on page 9 in Zenga *et al.* (2011) along with the series representation (29) we obtain that

$$LPM(x;\alpha,\theta,\gamma,1) = \begin{cases} C'\displaystyle\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}(-1)^j\frac{\gamma^j}{j!}\tilde{D}(x;\alpha+i+j,\theta) & 0 < x < 1 \\ 1 + C'\displaystyle\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}(-1)^j\frac{\gamma^j}{j!}\tilde{D}(x;\alpha+i+j,\theta) & x > 1 \end{cases} . \tag{35}$$

In addition, when $\theta > 1$, the "single series" representation (31) leads to:

$$LPM(x;\alpha,\theta,\gamma,1) = \begin{cases} C'\displaystyle\sum_{j=0}^{\infty}(-1)^j\frac{\gamma^j}{j!}\tilde{D}(x;\alpha+j,\theta-1) & 0 < x < 1 \\ 1 + C'\displaystyle\sum_{j=0}^{\infty}(-1)^j\frac{\gamma^j}{j!}\tilde{D}(x;\alpha+j,\theta-1) & x > 1 \end{cases} . \tag{36}$$

Thus, the Lorenz and Zenga's inequality curves of a *Type-II* GZD with parameters η' are given by

$$L(p;\eta') = LPM(Q_e(p;\eta');\alpha,\theta,\gamma,1) \qquad \text{and} \qquad I(p;\eta') = \frac{p - L(p;\eta')}{p[1 - L(p;\eta')]} \qquad 0 < p < 1\ ; \tag{37}$$

where $Q_e(\cdot;\eta')$ denotes the quantile function of the *Type-II* GZD.

Theorem 1 presented in Section 3.3 is also useful to investigate the stochastic ordering induced by the parameters of the *Type-II* GZD. In detail, it is easy to prove the following corollary.

COROLLARY 2 *Let $X(\mu,\alpha,\theta,\gamma)$ be a random variable following the Type-II GZD distribution with parameters μ, α, θ, and γ. We have that:*

1. *keeping fixed the values of μ, θ, and γ the random variable X is stochastically decreasing in α in the sense of the convex stochastic ordering:*

$$X(\mu,\alpha,\theta,\gamma) \geq_{cx} X(\mu,\alpha',\theta,\gamma) \quad \textit{whenever } \alpha \leq \alpha'.$$

2. *keeping fixed the values of μ and α the random variable X is stochastically increasing in θ and γ in the sense of the convex stochastic ordering:*

$$X(\mu,\alpha,\theta,\gamma) \leq_{cx} X(\mu,\alpha,\theta',\gamma') \quad \textit{whenever } \theta \leq \theta'\ ,\ \gamma \leq \gamma'.$$

Thanks to Corollary 2 we can assert that, for the *Type-II* GZD, the parameters α is an indirect inequality indicator while θ and γ are direct inequality indicators. Obviously the interpretations provided for the parameters α and θ are coherent with those provided in Porro (2011).

5. COMPARING THE FLEXIBILITY OF THE ZD AND THE GZDs

As mentioned in the introduction, it is now necessary to study the real flexibility improvement provided by the *Type-I* and *Type-II* GZD. So, we evaluate in a numerical exercise how dense the family of the ZDs is in the families of *Type-I* and *Type-II* GZDs. We consider the following values for the parameters: $\mu = 2$; $\alpha \in \{0.25, 0.75, 1, 2.5, 5\}$; $\theta \in \{0.25; 0.75, 2.5, 5\}$; $\delta \in \{-0.9, -0.5, 0.5, 0.9\}$; $\beta \in \left\{\frac{(\theta+\alpha)}{4}, \frac{(\theta+\alpha)}{2}, \frac{3\cdot(\theta+\alpha)}{4}\right\}$; $\gamma \in \{-10, -5, 5, 10\}$. Then, we take into consideration the *Type-I* GZDs obtained from all the $5\times4\times4\times3 = 240$ possible combinations of the aforementioned values of α, θ, β, and δ. For each possible parameters combination, we search for the ZD which is "less distant" to the particular *Type-I* GZD. Analogously, we consider the *Type-II* GZDs obtained from all the $5\times4\times4 = 60$ possible combinations of α, θ, and γ and we search for the ZD which is "less distant" to the particular *Type-II* GZD. The "distances" taken into consideration are:

- L_1-distance:
  $d_1(f,f^*) = \int_0^\infty |f(x) - f^*(x)|dx$ or $d_1(f,f_e) = \int_0^\infty |f(x) - f_e(x)|dx$

- Pearson's type "distance":
$$d_2(f, f^*) = \int_0^\infty \frac{(f(x) - f^*(x))^2}{f^*(x)} dx \quad \text{or} \quad d_2(f, f_e) = \int_0^\infty \frac{(f(x) - f_e(x))^2}{f_e(x)} dx$$

The desired values of the parameters of the ZD are the (numerical) solutions of the following optimization problems

$$\begin{cases} \min\limits_{\alpha',\theta',\mu'} d_i(f(\cdot;\alpha',\theta',\mu'), f^*(\cdot;\eta)) \\ sub \\ \alpha' > 0, \theta' > 0, \mu' > 0 \end{cases} \quad \text{and} \quad \begin{cases} \min\limits_{\alpha',\theta',\mu'} d_i(f(\cdot;\alpha',\theta',\mu'), f_e(\cdot;\eta')) \\ sub \\ \alpha' > 0, \theta' > 0, \mu' > 0 \end{cases}, \tag{38}$$

with $i = 1, 2$. The most interesting results obtained in the numerical exercise are reported in Table 1 and in Figures 4 and 5. We chose the distribution reported in these figures in order to cover all the relevant qualitative behaviors of the *Type-I* and *Type-II* GZDs. Observing Figures 4 it appears that the *Type-I* GZD assumes particular shapes that the ZD cannot reach. However, it emerges that the increase in flexibility is evident only when f^* assumes non-conventional shapes. On the contrary, as shown in Figure 4(h)-4(n), the densities of the *Type-I* GZD and the nearest ZDs become quite similar when analyzing the distributional shapes usually encountered in real applications (such as in the analysis of income distributions). In these cases, it emerges that the increase in flexibility provided by the two additional parameters is evident mainly around the mode of the distribution and near the origin. Similar observations can be made in the case of the *Type-II* GZD. However, we remark that this distribution shows a qualitative increase in flexibility (with respect to the ZD) very similar to that of the *Type-I* GZD, even if it is more parsimonious (4 parameters instead of 5).

Parameters of the *Type-I* GZD					Parameters of the ZD: minimum d_2				Parameters of the ZD: minimum d_1				
α	θ	β	δ	μ	α'	θ'	μ'	d_2	α'	θ'	μ'	d_1	
5.00	0.75	4.3125	0.9	2	5.1038	0.1050	2.0001	**0.3227**	5.0750	0.2489	2.0002	**0.2981**	Fig.4-(a)
5.00	2.50	5.6250	0.9	2	5.1258	0.4174	1.9954	**0.2330**	16.9745	1.5115	2.0015	**0.0642**	Fig.4-(b)
0.75	2.50	2.4375	0.9	2	0.8557	1.0119	2.0048	**0.0159**	0.9689	1.0135	2.0043	**0.0842**	Fig.4-(c)
1.00	2.50	2.6250	0.9	2	1.3748	1.0924	2.0042	**0.0145**	1.5524	1.0937	1.9973	**0.0794**	Fig.4-(d)
2.50	5.00	5.6250	0.9	2	4.1056	2.0068	1.9996	**0.0065**	5.6701	2.5618	2.0000	**0.0648**	Fig.4-(e)
1.00	5.00	4.5000	0.9	2	1.1216	1.7540	2.0079	**0.0042**	1.1968	1.7779	2.0036	**0.0441**	Fig.4-(f)
0.75	5.00	4.3125	0.9	2	0.7799	1.7068	2.0154	**0.0035**	0.8065	1.6904	2.0139	**0.0445**	Fig.4-(g)
5.00	5.00	5.0000	0.9	2	7.2101	2.5728	2.0000	**0.0023**	7.9696	2.7604	2.0000	**0.0301**	Fig.4-(h)
2.50	5.00	3.7500	0.9	2	2.8607	2.5076	2.0001	**0.0011**	3.0228	2.5921	1.9994	**0.0216**	Fig.4-(i)
2.50	2.50	3.7500	-0.9	2	2.1727	3.2444	1.9950	**0.0008**	2.0221	3.0617	1.9989	**0.0166**	Fig.4-(l)
2.50	5.00	5.6250	0.5	2	2.8069	3.5911	2.0017	**0.0005**	2.9651	3.7566	1.9998	**0.0131**	Fig.4-(m)
2.50	2.50	2.5000	-0.9	2	2.2460	2.9289	1.9977	**0.0004**	2.1370	2.8169	1.9997	**0.0111**	Fig.4-(n)

Parameters of the *Type-II* GZD				Parameters of the ZD: minimum d_2				Parameters of the ZD: minimum d_1				
α	θ	γ	μ	α'	θ'	μ'	d_2	α'	θ'	μ'	d_1	
2.50	0.25	5	2	1.5899	0.8802	2.0001	**0.0920**	0.8941	0.6086	2.0001	**0.1559**	Fig.5-(a)
0.25	2.50	-10	2	4.3799	1.7089	1.9999	**0.0675**	4.9415	1.8549	2.0002	**0.1106**	Fig.5-(b)
1.00	0.25	5	2	0.9225	2.8061	1.9687	**0.0365**	0.8559	2.6704	1.9508	**0.1231**	Fig.5-(c)
0.75	2.50	-5	2	1.5659	1.4586	2.0035	**0.0244**	1.9360	1.6512	1.9995	**0.0819**	Fig.5-(d)
0.75	5.00	-10	2	2.3033	2.3526	2.0039	**0.0232**	2.9413	2.8660	1.9989	**0.0753**	Fig.5-(e)
2.50	0.75	5	2	1.8141	1.9520	1.9963	**0.0181**	1.4591	1.7438	2.0009	**0.0786**	Fig.5-(f)
5.00	0.75	10	2	3.2118	3.1416	1.9968	**0.0176**	2.4330	2.5820	2.0020	**0.0757**	Fig.5-(g)
1.00	5.00	-10	2	2.8745	2.6154	2.0007	**0.0126**	3.5459	3.1114	1.9993	**0.0541**	Fig.5-(h)
1.00	2.50	-5	2	2.0566	1.6092	2.0010	**0.0125**	2.4570	1.7976	1.9997	**0.0566**	Fig.5-(i)
2.50	0.25	10	2	2.0262	5.1439	1.9678	**0.0110**	1.7637	4.6502	1.9795	**0.0631**	Fig.5-(l)
1.00	5.00	-5	2	1.2091	2.6646	2.0263	**0.0048**	1.3755	2.9658	2.0136	**0.0413**	Fig.5-(m)
5.00	2.50	10	2	3.9948	6.4916	1.9931	**0.0013**	3.5619	5.8505	1.9990	**0.0229**	Fig.5-(n)

TABLE 1: Parameters settings of the *Type-I* and *Type-II* GZD and parameters values of the ZD obtained solving the optimization problems (38) with $i = 1, 2$. In the last column the references to the subplots in Figures 4-5 are given.

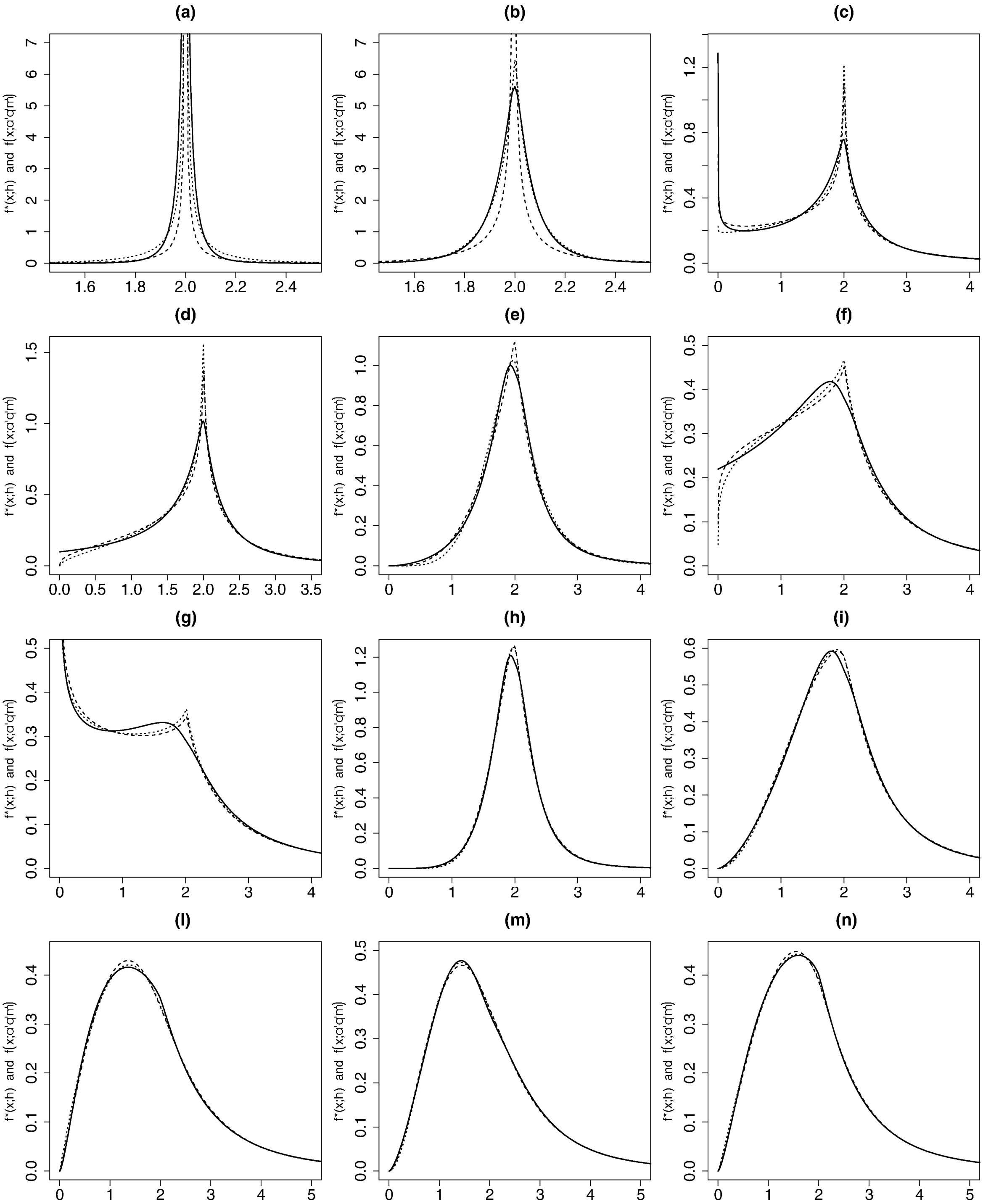

FIGURE 4: Plot of some of the *Type-I GZD* considered in the numerical example (solid line) with the corresponding plots of the ZDs obtained solving the optimization problem on the left of (38) with $i = 1$ (dashed line) and $i = 2$ (dotted line). The values of the parameters of the *Type-I GZD* and those of the ZDs can be found in Table 1.

6. APPLICATION TO REAL DATA: FITTING THE ITALY INCOME DISTRIBUTION

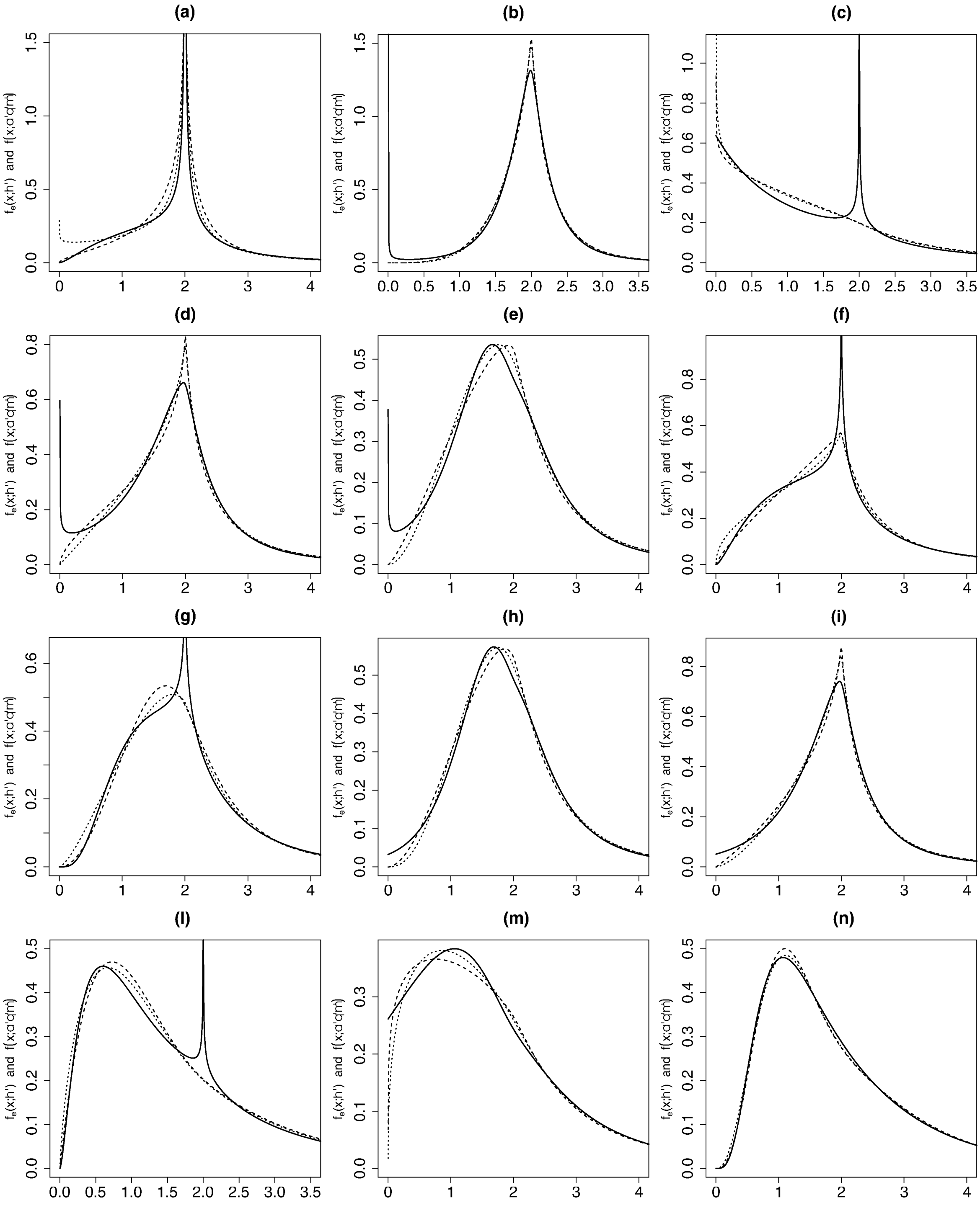

FIGURE 5: Plot of some of the *Type-II* GZD considered in the numerical example (solid line) with the corresponding plots of the ZDs obtained solving the optimization problem on the right of (38) with $i = 1$ (dashed line) and $i = 2$ (dotted line). The values of the parameters of the *Type-II* GZD and of the ZDs can be found in Table 1.

In this section we consider some income distributions in order to evaluate the increase in the goodness-of-fit obtained passing from the ZD to the generalized models. This analysis is performed to understand what is, in practice, the real improvement provided by the *Type-I* and *Type-II* GZDs. The real distributions taken into account are:

1. Italy 2006 Household Income Distribution: 7762 observations
2. Italy 2006 Individual Income Distribution: 13419 observations
3. Italy 2008 Household Income Distribution: 7958 observations
4. Italy 2008 Individual Income Distribution: 13616 observations

All these data are available on-line (see Banca d'Italia, 2008, 2010). The ZD and the GZD are fitted to the real distribution using the minimum Chi-square method and, as in Zenga *et al.* (2010a) the "minimum Mortara method". As it is well known, the minimum Chi-square estimates are obtained minimizing with respect to η (or η') the quantity:

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

where n denotes the sample size, s the number of intervals in which data are grouped, n_j and $\hat{n}_j(\eta)$ are the observed and the theoretical frequencies associated to the j-th interval, respectively. Similarly, the "minimum Mortara" estimates are obtained minimizing with respect to η (or η') the quantity:

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

The s class in which data are grouped and the theoretical frequencies $\hat{n}_j(\eta)$ are defined following a methodology similar to that adopted in Zenga *et al.* (2010a). In more detail, the data are grouped in 25 class. The end-points of each class coincide with the q_i-quantiles of the empirical distribution where $q_1 = q_2 = 0.01$, $q_3 = q_4 = 0.015$, $q_5 = ... = q_9 = 0.05$, $q_{10} = ... = q_{13} = 0.1$, $q_{14} = ... = q_{17} = 0.05$, $q_{18} = 0.02$, $q_{19} = q_{20} = 0.015$, and $q_{21} = q_{25} = 0.01$.

In addition to the minimum Chi Square and Mortara methods, we adopt, following Zenga *et al.* (2010a), also the "constrained" minimum Chi-square and Mortara methods obtained imposing the restriction $\mu = \bar{x}$ where $\bar{x}$ denotes the sample mean.

The estimated parameters of the ZD and GZDs obtained in the applications are provided in Table 2. In Table 3 we also provide the values of the Zenga inequality index I (see Zenga, 2007) and Gini concentration ratio R associated to the estimated and observed distributions (the observed values are given in bold at the top of the table). In Figure 6 we report the plots of the empirical and estimated distributions related to the 2008-Individual incomes along with the representations of the empirical and estimated Lorenz concentration curves and Zenga's inequality curves.

As it can be observed form Table 2, the GZDs provide some improvement in the goodness-of-fit. The improvement is more or less evident in relation to the features of the empirical distribution and to the estimation method adopted; it is higher when analyzing the Individual Income Distributions (which show a fatter left tail) and when the minimum Chi Square method is adopted. It is worthwhile to note that, the ZD and the GZDs provide an identical fitting when the Household Income Distribution related to 2008 is analyzed. The most evident difference is observed in the case of the 2008-Individual Income when the constrained minimum Chi-Square method is adopted (see Figure 6). Another interesting observation is that the GZD seems to have a goodness-of-fit which is more stable with respect to changes in the estimation method. It is important to note that the improvement in the goodness-of-fit of the *Type-I* GZD is almost the same of that of the *Type-II* even if the latter is more parsimonious. This fact confirms what observed in the numerical exercise performed in the previous section. From Table 3 it can be also observed that, in most cases, the values of the inequality indexes R and I provided by the GZDs tend to be slightly more adherent to the observed values with respect to those obtained from the ZD.

Estimates obtained via the Minimum Mortara method

Data	Model	$\hat{\alpha}$	$\hat{\theta}$	$\hat{\beta}$	$\hat{\delta}$	$\hat{\gamma}$	$\hat{\mu}$	A_1	A_2
HH-2006	ZD	4.0251	5.8149	–	–	–	31537	0.0659	0.1003
	Type-I	2.9163	22.7597	24.4341	0.8483	–	31328	0.0590	0.0841
	Type-II	2.3822	10.6939	–	–	-12.7451	31323	0.0572	0.0808
HH-2008	ZD	3.3408	4.9720	–	–	–	31953	0.0402	0.0659
	Type-I	3.3524	4.9399	4.2621	-0.0213	–	31951	0.0402	0.0657
	Type-II	3.3549	4.9323	–	–	0.1102	31951	0.0402	0.0659
Per-2006	ZD	1.8779	2.6314	–	–	–	18445	0.0833	0.1740
	Type-I	1.7464	3.3494	2.5908	0.5602	–	18655	0.0831	0.1644
	Type-II	1.6865	3.2161	–	–	-1.8695	18724	0.0826	0.1611
Per-2008	ZD	2.0536	2.7238	–	–	–	18845	0.0778	0.2427
	Type-I	1.6352	9.6895	8.6714	0.9361	–	18945	0.0680	0.1393
	Type-II	1.5057	3.6354	–	–	-3.3438	18971	0.0707	0.1308

Estimates obtained via the Minimum Chi Square method

Data	Model	$\hat{\alpha}$	$\hat{\theta}$	$\hat{\beta}$	$\hat{\delta}$	$\hat{\gamma}$	$\hat{\mu}$	A_1	A_2
HH-2006	ZD	3.4086	4.9917	–	–	–	31386	0.0680	0.0743
	Type-I	2.8994	8.4827	11.3821	0.5309	–	31325	0.0613	0.0707
	Type-II	2.7360	7.4051	–	–	-6.3086	31323	0.0606	0.0703
HH-2008	ZD	3.4206	5.0328	–	–	–	32266	0.0432	0.0612
	Type-I	3.3527	5.2667	7.3749	0.0813	–	32262	0.0431	0.0611
	Type-II	3.3497	5.2592	–	–	-0.6163	32262	0.0431	0.0611
Per-2006	ZD	1.6048	2.3682	–	–	–	18568	0.1014	0.1383
	Type-I	1.3078	10.7764	12.0842	0.8603	–	18514	0.0886	0.1143
	Type-II	1.1634	4.9020	–	–	-6.4495	18497	0.0874	0.1093
Per-2008	ZD	1.5818	2.2478	–	–	–	18984	0.1045	0.1444
	Type-I	1.2235	32.2709	33.4944	0.9608	–	19119	0.0735	0.0963
	Type-II	1.0584	5.0189	–	–	-7.2794	19082	0.0733	0.0913

Estimates obtained via the Minimum Mortara method with constraint $\mu = \bar{x}$

Data	Model	$\hat{\alpha}$	$\hat{\theta}$	$\hat{\beta}$	$\hat{\delta}$	$\hat{\gamma}$	$\hat{\mu}$	A_1	A_2
HH-2006	ZD	3.6754	5.2624	–	–	–	31829	0.0667	0.0909
	Type-I	2.6994	13.2925	14.9917	0.7637	–	31829	0.0619	0.0770
	Type-II	2.4098	8.8685	–	–	-9.8369	31829	0.0607	0.0765
HH-2008	ZD	3.4364	5.1461	–	–	–	32424	0.0424	0.0628
	Type-I	3.7594	4.3415	5.8127	-0.5258	–	32423	0.0418	0.0675
	Type-II	3.7076	4.1871	–	–	2.6577	32423	0.0419	0.0674
Per-2006	ZD	1.9005	2.6494	–	–	–	18412	0.0837	0.1798
	Type-I	1.8537	2.6833	3.6397	0.0542	–	18412	0.0835	0.1714
	Type-II	1.8532	2.6822	–	–	-0.2010	18412	0.0835	0.1713
Per-2008	ZD	2.0243	2.8554	–	–	–	18953	0.0787	0.2193
	Type-I	1.5985	4.7495	4.8766	0.7160	–	18953	0.0693	0.1369
	Type-II	1.5053	3.6083	–	–	-3.2924	18953	0.0707	0.1304

Estimates obtained via the Minimum Chi Square method with constraint $\mu = \bar{x}$

Data	Model	$\hat{\alpha}$	$\hat{\theta}$	$\hat{\beta}$	$\hat{\delta}$	$\hat{\gamma}$	$\hat{\mu}$	A_1	A_2
HH-2006	ZD	3.3759	4.9878	–	–	–	31829	0.0693	0.0772
	Type-I	2.9350	7.7160	10.6507	0.4679	–	31829	0.0642	0.0746
	Type-II	2.7982	7.0841	–	–	-5.4510	31829	0.0635	0.0743
HH-2008	ZD	3.4120	5.0356	–	–	–	32423	0.0448	0.0616
	Type-I	3.3610	5.2226	3.0396	0.1502	–	32423	0.0448	0.0616
	Type-II	3.3569	5.2125	–	–	-0.4804	32423	0.0448	0.0616
Per-2006	ZD	1.6095	2.3641	–	–	–	18412	0.1006	0.1390
	Type-I	1.3097	10.7049	12.0145	0.8596	–	18412	0.0895	0.1146
	Type-II	1.1627	4.9075	–	–	-6.4777	18412	0.0879	0.1095
Per-2008	ZD	1.5821	2.2461	–	–	–	18953	0.1043	0.1444
	Type-I	1.2270	28.2946	29.5215	0.9553	–	18953	0.0735	0.0973
	Type-II	1.0605	4.9892	–	–	-7.2395	18953	0.0743	0.0919

TABLE 2: Parameters estimates obtained with the estimation methods based on Mortara's and Chi-Square goodness-of-fit indexes

		Zenga's inequality index (I)				Gini concentration ratio (R)			
	Data	HH-2006	HH-2008	Per-2006	Per-2008	HH-2006	HH-2008	Per-2006	Per-2008
	Observed values	**0.6843**	**0.6784**	**0.7160**	**0.7119**	**0.3442**	**0.3397**	**0.3653**	**0.3599**
Est. Method	Model								
minimum A_1	ZD	0.6608	0.6772	0.7058	0.6882	0.3237	0.3371	0.3595	0.3426
	Type-I	0.6775	0.6771	0.7027	0.7037	0.3383	0.3370	0.3561	0.3568
	Type-II	0.6767	0.6771	0.7022	0.7012	0.3373	0.3370	0.3556	0.3540
"constrained"	ZD	0.6639	0.6768	0.7039	0.7010	0.3256	0.3370	0.3576	0.3554
minimum A_1	*Type-I*	0.6726	0.6801	0.7051	0.7027	0.3327	0.3401	0.3588	0.3558
	Type-II	0.6729	0.6800	0.7052	0.7013	0.3331	0.3400	0.3588	0.3542
minimum A_2	ZD	0.6728	0.6735	0.7283	0.7229	0.3331	0.3338	0.3822	0.3760
	Type-I	0.6725	0.6735	0.7231	0.7171	0.3328	0.3338	0.3760	0.3690
	Type-II	0.6726	0.6735	0.7218	0.7154	0.3328	0.3338	0.3744	0.3670
"constrained"	ZD	0.6751	0.6743	0.7272	0.7228	0.3352	0.3345	0.3810	0.3758
minimum A_2	*Type-I*	0.6751	0.6743	0.7224	0.7160	0.3352	0.3345	0.3753	0.3679
	Type-II	0.6752	0.6743	0.7212	0.7146	0.3352	0.3345	0.3738	0.3661

TABLE 3: Values of the Gini concentration ratio (R) and of the Zenga inequality index (I) associated to the observed distributions (in bold on the top of the table) and to the estimated distributions.

7. CONCLUSION

In this paper we propose two new distributional models generalizing the one recently proposed in Zenga (2010). The latter is defined as a Beta-mixture of Polisicchio's distributions and it has been extensively studied also in Zenga *et al.* (2011) and Zenga *et al.* (2010a, 2010b). The first new distribution (*Type-I* GZD) is characterized by 5 parameters and it is obtained considering a Gauss-Hypergeometric mixture of Polisicchio's distributions. The second new model (*Type-II* GZD) is a Confluent-hypergeometric mixture of Polisicchio's distributions and it has 4 parameters. The *Type-I* and *Type-II* GZDs shares several properties with the ZD. For example the parameter α regulates the maximum order of the existing moments and it influences the value assumed by the density near the origin of all the three distributions. The parameter θ regulates the finiteness/infiniteness of the density near the expected value μ. The additional parameters characterizing the GZDs are interpretable as inequality indicators (direct or indirect), and they provide an increase in the flexibility with respect to the ZD. We evaluate the flexibility increase both by a numerical exercise and by some applications. We observe that the ZD cannot assume some of the possible "qualitative" shapes of the GZDs. This fact is particularly evident when the *Type-II* GZD is considered even if the shapes the ZD cannot assume are, in general, "non-conventional". The numerical exercise points out that the flexibility improvement obtained with the *Type-I* GZD is almost the same as that obtained with the *Type-II* GZD, which is, indeed, more parsimonious. The applications corroborates the last observation since, in most cases, the *Type-II* GZD provides the best goodness-of-fit. From the applications it turns also out that the GZDs have a improved flexibility around the mode and near the origin. This is the reason why the goodness-of-fit of the generalized model is quite better than those provided by the ZD mainly when the individual income distributions are analyzed. In fact, these distributions exhibit a very fat left tail together with a very peaked mode. It is worthwhile to note that the analytical complexity of the GZDs is quite higher than that of the ZD, especially when the *Type-I* GZD is considered. In particular, we recall that if $\theta > 1$ (which is, indeed, the only case of practical interest) the density and the distribution function of the ZD coincide (roughly speaking) with the difference of two incomplete beta functions which can be immediately computed using , for example, R. On the contrary, in order to avoid the extensive use of numerical integration, the densities and the distribution functions of the *Type-I* and *Type-II* GZDs can be computed using the series representation provided in section 3.2 and 4.2. From a practical point of view, the use of these series representation requires a high computational time since, in order to reach a good approximation of the true densities and distribution functions, the series should be truncated at a high order (we consider the first 170 terms). For example, the procedure we write (in R code) in order to estimate the parameters takes a few seconds in the case of the ZD, a few minutes for the *Type-II* GZD and about *20* minutes for the *Type-I* GZD. For all the aforementioned reasons we advise using the *Type-II* GZD which, in some situations, provides a sensible improvement with respect to the ZD with (only) one additional parameter and with an acceptable increase in the analytical and computational complexity.

In further research, we want to fit the GZDs to several other income distributions in order to accumulate further evidence supporting the results outlined in the present paper. Moreover, we want to compare the results provided by these two new distributions with those provided by the ZD and by some other models usually employed in modeling incomes such as: the Dagum distribution, the Generalized Gamma and the recently introduced Beta-Dagum distribution (see Domma and Condino, 2010).

REFERENCES

Armero C., Bayarri M. J. (1994). Prior Assessments for Prediction in Queues. *Journal of the Royal Statistical Society. Series D (The Statistician)*, **43**, 1, 139-153.

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

Banca d'Italia (2010). *The 2008 Bank of Italy sample Survey on Household Income and Wealth*. Supplement to the Statistical Bulletin Sample Survey, **XX** (8). Available at `http://www.bancaditalia.it`.

Domma F., Condino F. (2010). The Beta-Dagum distribution. *XLV Conference of the Italian Statistical Society*, University of Padova, Padova, Italy, 15-18 June 2010. Available at `http://homes.stat.unipd.it/mgri/SIS2010/Program/contributedpaper/803-1194-1-RV.pdf`

J.L. Gastwirth, *Large Sample Theory of Some Measures of Income Inequality*. Econometrica, 42 (1974), 191-196.

Gordy M.B. (1998). A generalization of generalized beta distributions. *Finance and Eco-*

nomics Discussion Series, Paper No 1998-18, Board of Governors of the Federal Reserve System, US.

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

Porro F. (2008). Equivalence Between Partial Order Based on Curve $L(p)$ and Partial Order Based on Curve $I(p)$. *XLIV Conference of the Italian Statistical Society*, University of Calabria, Arcavacata di Rende, Italy, 25-27 June 2008. Available at http://www.sis-statistica.it/files/pdf/atti/rs08_spontanee_a_3_4.pdf

Porro F. (2011). Inequality order for Zenga distribution. *Tech. Rep. n. 215, Dipartimento di Metodi Qunatitativi per le Scienze Economiche ed Aziendali*, University of Milano-Bicocca, Milano, Italy.

Shaked M., Shanthikumar J.G. (2007). *Stochastic Orders*, Springer.

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., Polisicchi M., Ma. Zenga (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 Ma. (2010). First applications of a new three-parameter distribution for non-negative variables. *Tech. Rep. n. 188, Dipartimento di Metodi Qunatitativi per le Scienze Economiche ed Aziendali*, University of Milano-Bicocca, Milano, Italy.

Zenga M.M., Porro F., Arcagni A. (2010). Method of moments for Zenga's distribution, *Tech. Rep. n. 193, Dipartimento di Metodi Qunatitativi per le Scienze Economiche ed Aziendali*, University of Milano-Bicocca, Milano, Italy.

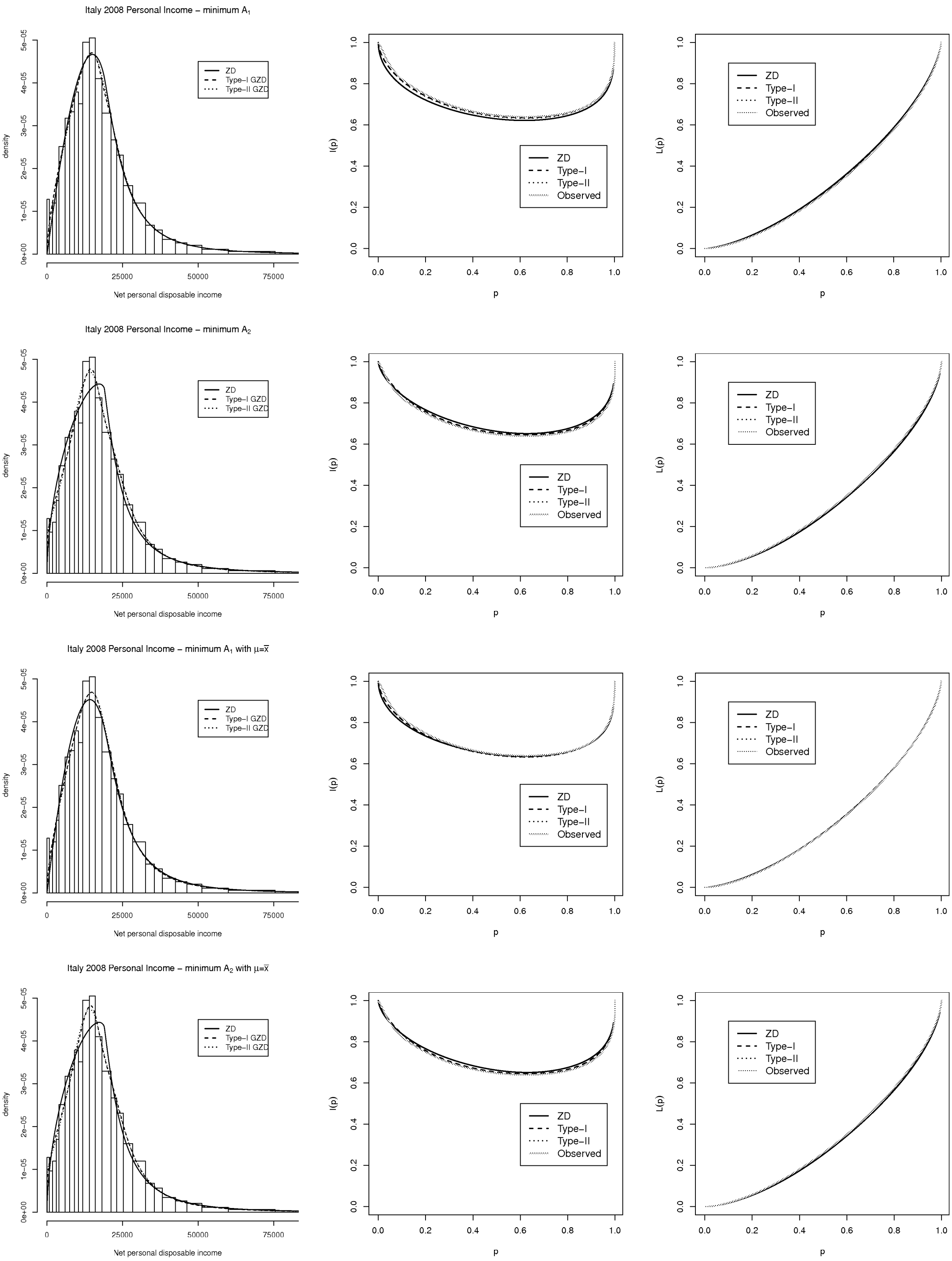

FIGURE 6: Fitting the Italy 2008 Individual Income Distribution