Sustainable Futures 2 (2020) 100008

Contents lists available at ScienceDirect

Sustainable Futures

journal homepage: www.elsevier.com/locate/sftr

A combined method based on kurtosis indexes for estimating p in non-linear L_p-norm regression

Massimiliano Giacalone

Department of Economics and Statistics, University of Naples "Federico II", Viale Cinthia, Monte S. Angelo, 80126 Naples, Italy.

ARTICLE INFO

Keywords:
Lp-norm estimators
Generalized error distribution
Adaptive procedures
Kurtosis indexes
Non-linear regression
Sustainability

ABSTRACT

The Generalized Error Distribution (G.E.D.) is a very flexible family of symmetric density distributions, which are characterized by their shape parameter p linked to the L_p*-norm* estimators. In fact, under this errors assumption in the regression model the G.E.D. parameter p coincides with the p exponent of the L_p*-norm*. In this paper, we examine the use of L_p*-norm* estimators in the framework of non-linear regression models assuming the G.E.D. as the errors distribution. More precisely, we introduce an exponential regression (Marković, & Borozan, 2015) and a new algorithm Lp_{med} consisting of two iterative procedures, one internal to estimate the regression parameters and another external for estimating p (the p exponent of the L_p*-norm*) based on two kurtosis indexes of the residuals distribution. In order to show the good results of the proposed method, an efficiency comparison of the new method, Lp_{med}, with other two well-known approaches as the maximum likelihood (Agrò, 1995) and the Money et al. (1982) method is performed. Our combined method shows better results asymptotically and, especially in presence of leptokurtic data, for the p parameter estimation. Finally an application on the Equitable and Sustainable Well-being (B.E.S) in the Italian context confirms the good properties of the proposed method.

1. Introduction

A central problem in regression analysis is the presence of errors in the detection of phenomena.

In this framework the assumption related to the form of the error distribution is very important and it is not necessarily Gaussian [36]. The approach of this paper highlights that, given the infinite ways of the accidental errors distribution; the assumption of normality is too restrictive and leads to less efficient estimates. The theoretical scheme of the Generalized Error Distributions (G.E.D.) represents an important generalization of the hypothesis of normality: indeed, by using this family of density functions, it is possible to describe unlimited symmetrical or asymmetrical forms of the accidental errors distribution [6].

This scheme constitutes an important generalization with respect to classical statistical theory based on the hypothesis of Gaussian normality. It could indeed be of great interest, where possible, to revisit the classical inferential procedures by placing the probabilistic assumption of normality of order p which has the great advantage of being less restrictive than the Gaussian one for describing various accidental errors distributions with symmetrical forms (or, in certain cases, asymmetrical).

In this perspective, the simulation studies and the applications performed in this paper are based on the generalization of the classical hypothesis of residual distribution normality.

As it is well known under the Gaussian hypothesis, the maximum likelihood estimators of the regression parameters are equivalent to the least squares estimators [20].

Similarly, the hypothesis of regression model assuming the G.E.D. residuals distribution (generalization of the classical hypothesis of the Gaussian normality) guarantees the identity between the maximum likelihood estimators and the L_p*-norm* estimators (with $p \neq 2$).

The objective of this generalization can be aimed at obtaining good estimates of the location (e.g. mean, median, etc.) and scale parameters (e.g. simple deviation from the median, variance, etc.) from a sample of measurements affected by accidental errors [32], or it can be to determine the most suitable distributional assumption to describe the regression model sample data.

The aim is to guarantee to parameters estimators some optimality properties, when the response variable is affected by not Gaussian random errors.

This problem has received a great deal of attention from statisticians and from various scientists who have considered the inductive approach to solving knowledge problems, and who have unanimously agreed that the best solution is to seek the exact distribution of the accidental errors, in order to limit theirs influence on the observed data [10].

E-mail address: massimiliano.giacalone@unina.it

https://doi.org/10.1016/j.sftr.2020.100008
Received 24 December 2019; Accepted 20 January 2020

This paper is structured as follows: after a brief introduction regarding the family of curves that is known as Generalized Error Distribution (also called Exponential Power Function or E.P.F), we consider the linear and non-linear regression models and the L_p-norm estimators.

Thus, we introduce Lp_{med}, a combined algorithm consisting of two iterative procedures for estimating the L_p-norm exponent p and the regression parameters. Moreover a comparative simulation study with two other procedures in order to prove empirically some important properties is performed.

In this paper a particular exponential regression model employed in many different areas such as biology, finance, sustainability, medicine, is investigated [29].

For all the three considered methods ([1, 33] and the combined method Lp_{med}) a Montecarlo simulation plan is carried out to estimate both the regression coefficients and the p parameter. Finally, to further test its validity, the new method is applied to a selection of indicators concerning the Italian “Equitable and Sustainable Well-being” named “B.E.S.” (Benessere Equo & Sostenibile) considered in the drafting of the D.E.F. (Economics and Finance Document).

In particular, a regression model containing four analyzed and predicted indicators defined by the M.E.F. (Ministry of Economics and Finance) in [12] is offered.

2. Generalized error distribution: some parameterizations

Starting from a partial modification of Gauss's law basic hypotheses, the mathematician Subbotin has formalized a law of errors' distribution, which is more flexible than Gaussian to describe most of the phenomena observable and applicable to the field of sustainability.

He considered the following axioms:

1. The probability of an error depends only on the size of the error itself and can be expressed by a function $\phi(z - x_1) = f(\varepsilon)$ having the first derivative continues in general (as Gauss);
2. The most probable value z of an amount of which direct measurements are known, x_1, does not depend on the unit of measurement used (according to Schiaparelli's axiom).

Subbotin [39] obtained the general formulation of the error distribution as follows:

$$f(\varepsilon) = \frac{mh}{2\Gamma(1/m)} exp\left\{-h^m|\varepsilon|^m\right\} \tag{1}$$

where $\Gamma(\lambda)$ is the complete Gamma function defined by the following integral:

$$\Gamma(\lambda) = \int_0^\infty t^{\lambda-1}e^{-t}dt$$

For particular values of h, m and ε, all the Generalized Error Distributions are precisely identified.

The Subbotin's formulation defines the G.E.D. [21] starting from the following frequency distribution:

$$f_s(x) = ae^{-b|x-c|^p} \tag{2}$$

With a being the constant, b is the scale parameter, c is the centrality parameter, p shape parameter, assuming $b > 0$ and $p > 1$. After a few steps on the integrals we get:

$$a = \frac{1}{2}\frac{(b)^{\frac{1}{p}}}{\Gamma\left(1+\frac{1}{p}\right)}$$

After obtaining the constant a, it is possible with a similar procedure to get a general formula of the moments of order k as a function of parameter p.

$$\mu_k = E|x-c|^k = b^{-\frac{k}{p}}\frac{\Gamma\left(\frac{k+1}{p}\right)}{\Gamma\left(\frac{1}{p}\right)} \tag{3}$$

This relation shows an interesting characteristic of the frequency distribution considered.

In fact, it is easy to verify that the relationships between centered moments of the same order are invariant with respect to the centrality and dispersion parameters and only depend on the shape parameter p.

From the previous formula, it is possible to obtain the theoretical relation relating to generalized kurtosis:

$$\beta_k = \frac{\mu_{2k}}{{\mu_k}^2} = \frac{\Gamma\left(\frac{1}{p}\right)\Gamma\left(\frac{2k+1}{p}\right)}{\left[\Gamma\left(\frac{k+1}{p}\right)\right]^2} \tag{4}$$

If $k = 2$ we obtain the Pearson classic kurtosis index of norm 2:

$$\beta_2 = \frac{\mu_4}{{\mu_2}^2} = \frac{\Gamma\left(\frac{1}{p}\right)\Gamma\left(\frac{5}{p}\right)}{\left[\Gamma\left(\frac{3}{p}\right)\right]^2} \tag{5}$$

If $k = 1$ we obtain the mean square deviation $\sigma_2 = \sqrt{\mu_2}$ and the average simple deviation $\sigma_1 = \mu_1$ which inverted provide the tails' length index introduced by Geary [15] also defined as the expected absolute value divided by the standard deviation and called norm one kurtosis.

$$I = \frac{\mu_1}{\sqrt{\mu_2}} = \frac{\Gamma\left(\frac{2}{p}\right)}{\sqrt{\Gamma\left(\frac{1}{p}\right)\Gamma\left(\frac{3}{p}\right)}} \tag{6}$$

Finally, if we put $k = p$ in (3) we get:

$$\mu_p = b^{-1}\frac{\Gamma\left(1+\frac{1}{p}\right)}{\Gamma\left(\frac{1}{p}\right)} = \frac{1}{bp}$$

The centered moment of p-th order is given by:

$$\sqrt[p]{\mu_p} = \left[\int_{-\infty}^{+\infty}\left|x-\mu_p\right|^p f(x)dx\right]^{\frac{1}{p}} = \sigma_p$$

This is also the p-th power of σ_p, the scale parameter of order p, which is the index of conditional variability. The value of b is obtained from the last two relations considered and the value of c is assumed equal to μ_p:

$$b = \frac{1}{p\sigma_p^p} \quad \text{and} \quad c = \mu_p$$

Therefore (2) can be written as follows:

$$f_s(x) = \frac{1}{2p^{\frac{1}{p}}\Gamma\left(1+\frac{1}{p}\right)\sigma_p} e^{-\frac{|x-\mu_p|^p}{p\sigma_p^p}} \tag{7}$$

Where $\mu_p = E(x)$ is the location parameter, $\sigma_p = [E(|x-\mu_p|^p)]^{\frac{1}{p}}$ is the scale parameter $p>0$ is the shape parameter and Γ as in (1). The (7) represents, in a more evident way than the (2), the family of the Generalized Error Distributions and in correspondence to each value of p provides particular errors distributions which in any case maintain the characteristic of symmetry. The (7) can be obtained furthermore from the (1) if it occurs in the latter $m = p,\ \varepsilon = x - \mu_p,\ h = 1/(\sigma_p p^{\frac{1}{p}})$.

The p parameter can be interpreted as a structure parameter characterizing each particular errors distribution and becomes decisive for distinguishing the particular mechanisms that determine them. These distributions therefore take infinite shapes with p varying from 0 to ∞ with respect to kurtosis, length of the tails and curvature.

In particular, if we use Pearson's kurtosis β_2 index to distinguish the different obtained distributions we note that:

(a) With $0 < p < 1$ we obtain double exponential distributions with characteristic cuspidate shapes, showing very long tails with $\beta_2 > 6$;

(b) With p=1, we have the first Laplace's law or double exponential distribution:

$$f_1(x) = (2\sigma_1)^{-1}\exp(-|x-\mu_1|/\sigma_1)$$

where σ_1 is the simple average deviation ed μ_1 is the median. It is also cuspidate with long tails and $\beta_2 = 6$;

(a) With $1 < p < 2$ leptokurtic distributions are obtained (with shapes that have increasingly higher values in the tails as p varies from 2 to 1). These curves have a curvature around the maximum, long tails and $3 < \beta_2 < 6$;
(b) With p=2 we get the second Laplace law or normal Gaussian distribution:

$$f_2(x) = \left(\sqrt{2\pi\sigma_2}\right)^{-1}\exp\left(-|x-\mu_2|^2/2\sigma_2^2\right)$$

(where σ_2^2 is the variance and μ_2 is the arithmetic mean). It is mesokurtic with $\beta_2 = 3$;

(a) With $p > 2$ we obtain distributions with values thickened around the central one. These curves are platikurtic with short tails and kurtosis and $\beta_2 < 3$;
(b) Finally, with $p \to \infty$ we have the rectangular or uniform distribution:

$$f_3(x) = \begin{cases} \dfrac{1}{b-a} & a \le x \le b \\ 0 & \textit{otherwise} \end{cases}$$

It has no tails, parameters a and b are equal to $a = \mu - \sigma_p$; $b = \mu + \sigma_p$ where μ is the location parameter and σ_p the enhanced deviation of order p and $\beta_2 \to 1.8$.

Ultimately, as the parameter p varies continuously from 0 to ∞, (7) assumes different characteristic shapes that vary from a degenerate form at the value $p = 0$, to a rectangular distribution when p tends to infinity. The dispersion parameter σ_p is also of considerable importance.

In fact, each errors distribution obtained in correspondence with a certain value of p can take different forms that are variable from the degenerate case of a single ordinate for $\sigma_p = 0$ to a uniform diffuse distribution for $\sigma_p \to \infty$.

Furthermore, it is possible to obtain from (7) also the family of the standardized Generalized Error Distributions by placing $z = (x-\mu_p)/\sigma_p$, which is given by:

$$f_p = \left(2^p\sqrt{p}\Gamma(1+1/p)\right)^{-1}\exp\left(-|z|^p/p\right) \tag{8}$$

Using the Nadarajah [34] notation, we obtain the probability density function for the Generalized Error Distribution that is characterized by three main parameters to be estimated [31] called location parameter (μ_p), scale parameter (σ_p) and shape parameter (p), which changing allows us to consider infinite type of symmetrical distributions (Fig. 1). Therefore, the probability density function (pdf) is the following:

$$f\left(x|\mu_p,\sigma_p,p\right) = \frac{p}{2\sigma_p}\Gamma\left(\frac{1}{p}\right)^{-1}\exp\left\{-\left|\frac{x-\mu_p}{\sigma_p}\right|^p\right\} \tag{9}$$

Where $\mu_p \in (-\infty,+\infty)$, σ_p is positive therefore $\sigma_p \in (0,+\infty)$ and also p, which is a measure of fatness of tails is $p \in (0,+\infty)$ and $x \in$ R. [44].

To integrate to one, the pdf needs to be specified as:

$$f(y|m,\theta,p) = \frac{p}{2\theta}\Gamma\left(\frac{1}{p}\right)^{-1} exp\left(-\left|\frac{y-m}{\theta}\right|^p\right) \tag{10}$$

Note that x, μ_p and σ_p are respectively replaced by y, m, and θ, simply because μ_p and $\sigma_p$are typically used to label the location parameter and the scale parameter of a random variable and x the explanatory variables in regression models. For the symmetric case, the mode m is the same as the mean. This is, however, not the case for the asymmetric G.E.D. pdf

Theodossiou [40], Savva & Theodossiou [37] considered the following asymmetric pdf

$$f(y|m,\theta,\lambda,p) = \frac{p}{2\theta}^{1-\frac{1}{p}}\Gamma\left(\frac{1}{k}\right)^{-1} exp\left(-\frac{1}{p}\left|\frac{y-m}{(1+sign(x-m)\lambda)\theta}\right|^p\right) \tag{11a}$$

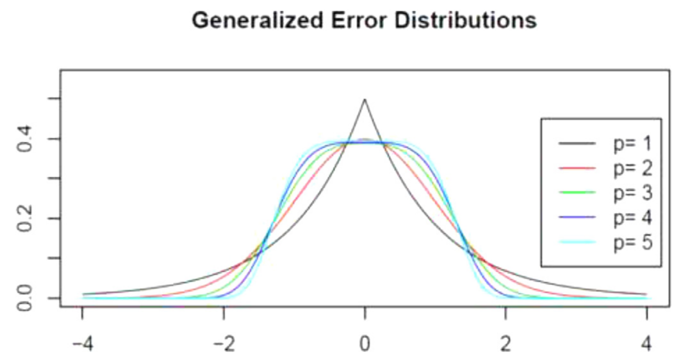

Fig. 1. Alternative G.E.D. distributions for different values of the shape parameter p.

where m is the mode of the random variable y, θ is a scaling constant related to the standard deviation of y, λ is a skewness parameter, k is a kurtosis parameter, *sign* is the sign function taking the value of -1 for $u < 0$ and 1 for $u > 0$.

The generalization (10) allows us to modify the classical hypothesis of residual distribution's normality usually employed in linear regression: if, for the classical hypothesis, the maximum likelihood estimators coincide with the least squares estimators, even for this more general hypothesis it is possible to identify a relationship between maximum likelihood estimators and the so-called L_p-norm estimators [43]. In particular this is true for values of $p \neq 2$, while assuming $p = 2$ we obtain the Gaussian distribution, so the L_p-norm estimators are equal to least squares estimators (L_2) [5].

Therefore, this generalization can be aimed to obtain good estimates of the location and scale parameters also from a sample whose measurements are affected by accidental errors, or it can be used to determine the most appropriate type of distribution to describe the data [14].

The last parameterization we propose in this paragraph concerns the probability density function for non-centered Skewed G.E.D. and can be defined as follows [40]:

$$f\left(z|\mu_p;\sigma_p;\lambda_p;p\right) = \frac{p\exp\left(-\frac{1}{p}\left|\frac{z-\mu_p+m}{v\sigma_p\left(1+\lambda_p sign\left(z-\mu_p+m\right)\right)}\right|^p\right)}{2v\sigma_p\Gamma\left(\frac{1}{p}\right)} \tag{11b}$$

Where $z \in$ R, μ_p is the location parameter, σ_p is the scale parameter, λ_p is the skewness parameter, p is the shape parameter, while Γ is as in (1). Function sign is the sign function which assumes value of -1 for negative values of its argument and 1 for positive ones. Moreover, m is defined as follow:

$$m = \frac{2^{\frac{2}{p}}v\sigma_p\lambda_p\Gamma\left(\frac{1}{2}+\frac{1}{p}\right)}{\sqrt{\pi}},$$

while v:

$$v = \frac{\pi\left(1+3\lambda_p^2\right)\Gamma\left(\frac{3}{p}\right)-16^{\frac{1}{p}}\lambda_p^2\Gamma\left(\frac{1}{2}+\frac{1}{p}\right)\Gamma\left(\frac{1}{p}\right)}{\pi\Gamma\left(\frac{1}{p}\right)}.$$

The shape parameter p controls the tails and the peak of the distribution; a small value of p means that the tails of the distribution become flat, with the center becoming largely peaked. The skewness parameter λ_p ranges between [−1, 1]; in the case of negative skewness ($\lambda_p < 0$) the density function is skewed to the left and vice versa for ($\lambda_p > 0$). Also the Skewed G.E.D. (S.G.E.D.) (11b) is a very special case of other distributions. For example, supposing $\lambda_p = 0$ (allowing p to change) we can obtain a wide family of non-skewed distributions. In particular, when $\lambda_p = 0$ we have the G.E.D.; $\lambda_p = 0$ and $p = 1$ means Laplace distribution; if $\lambda_p = 0$ and $p = 2$ we have Gaussian distribution; $\lambda_p = 0$ and $p = \infty$ means the Uniform distribution and $\lambda_p = 2$ and $p = 2$ is the skewed Gaussian.

However empirical evidence suggests the use of the univariate and multivariate Skewed G.E.D. in the context of financial market data especially to debate on volatility forecasting under non-normality hypothesis for asset returns. In particular the S.G.E.D. represents a suitable choice to model the empirical distribution of log-returns of financial asset for volatility prediction purposes [4, 8, 11].

3. L_p-norm estimators and G.E.D. moments derivation

This study deals with the construction of an adaptive estimation procedure for linear and non-linear regression models. In this context, the *Lp-norm* estimation methods are investigated for different values of p over a range of error distribution with varying kurtosis.

Considering the problem of regressing Y on X we assume a sample of n observed data (x_i, y_i) where y_i is the dependent variable and x_i the vector of k independent nonrandom predictors.

The general regression model is:

$$y_i = g(x_i, \theta) + \varepsilon_i$$

Where g is a derivable function, $\theta = (\theta_0, \theta_1, \dots, \theta_k)$ is the unknown real parameter vector to be estimated, and ε_i are the random, independent and identically errors distributed according to a G.E.D., with location parameter $\mu_p = 0$ and σ_p constant scale parameter. A common choice in order to obtain the L_p*-norm* estimators of the unknown parameter vector θ is to minimize the p-th power of the absolute deviations of the observed points from the regression function:

$$S(\theta) = \sum_{i=1}^{n} \left[y_i - g(x_i, \theta)\right]^p, \qquad \text{with } p \geq 1$$

Under the regular assumptions, the log-likelihood related to the sample is given by:

$$L(\theta, \sigma_p, p) = -n\log\left[2\, p^{1/p}\, \sigma_p\, \Gamma(1 + 1/p)\right] - \left[(p\sigma_p)^{1/p} \sum_{i=1}^{n} |y_i - g(x_i, \theta)|^p\right]$$

where we consider $z = y_i$ and $\mu_p = g(x_i, \theta)$. When p is known it is easy to calculate the first partial derivatives with respect to θ to get the following system composed by n nonlinear equations and k+*1* variables:

$$\frac{\partial L}{\partial \theta_i} = \sum_{i=1}^{n} |y_i - g(x_i, \theta)|^{p-1}\, sign\left(y_i - g(x_i, \theta)\right) \frac{\partial g}{\partial \theta_i} = 0$$

System solutions provide the maximum likelihood estimators for the regression parameter. The same equation are obtained by minimizing the sum of the p-th power of the absolute deviations of the observed points from the regression function, by applying the L_p*-norm* estimators:

$$\sum_{i=1}^{n} |y_i - g(x_i, \theta)|^p = \min, \qquad \text{with } p \geq 1$$

When the order p is specified, all the terms in the log-likelihood function, except for the last part containing the vector θ, are constants.

This result shows that the optimal exponent p is equal to the shape parameter of a G.E.D. assumed as underlying error distribution and it is very useful in connecting the L_p*-norm* estimators to the G.E.D. Therefore, maximum likelihood estimators are equivalent to L_p-norm estimators, either for the value of $p = 2$ or for any other value $p \neq 2$ [18].

However, if p is unknown, there are two important issues to solve: the estimation of a suitable exponent p based on the sample data [33, 35, 38] and the choice of the minimization algorithm to obtain the regression parameter estimation [31].

The moment's derivation which appears in the continuation of this paragraph [13] is very useful to underline their relations with the other G.E.D. parameters.

Assuming to consider a random variable X which follows a G.E.D. with the probability density function specified above and with shape parameter $p > 0$, it is possible to affirm [41] that the Moment Generation Function (m.g.f.) for a G.E.D. is given by:

$$M(t) = c_1 \int_{-\infty}^{\infty} e^{tx} e^{-c_2|x|^p} dx, \qquad -\infty < t < +\infty$$

Where:

$$c_1 = p\left[\vartheta 2^{1+\frac{1}{p}} \Gamma(1/p)\right]^{-1} \qquad \text{and} \qquad c_2 = (2\vartheta^p)^{-1}$$

Moreover, in order to simplify the notation, we considered the value:

$$\vartheta = 2^{-2/p}\left[\frac{\Gamma\left(\frac{1}{p}\right)}{\Gamma\left(\frac{3}{p}\right)}\right]^{\frac{1}{2}}$$

It is existing for any t when $p > 1$ and fail to exist when $0 < p < 1$. When $p = 1$, the m.g.f. exists for the interval $(-\sqrt{2}, \sqrt{2})$.

Now, supposing to have a random variable X which follows a G.E.D. with shape parameter $p > 1$, we consider the related standardized variable $Z = \frac{X-\mu_p}{\sigma_p}$.

So, if the pdf of the G.E.D. is defined as before, the standardized variable shows the following pdf, already obtained in another way in (8):

$$f_p(z) = \frac{p}{2\Gamma\left(1/p\right)} \exp\left\{-|z|^p\right\}$$

From this relationship we can easily define the k-th moment of Z given by:

$$E(Z^k) = \frac{1 + (-1)^k}{2\Gamma\left(1/p\right)} \Gamma\left(\frac{k+1}{p}\right)$$

Therefore, the n-th moment of the variable X can be obtained as follows:

$$E(X^n) = E\left[(\mu_p - \sigma_p Z)^n\right] =$$

$$= \sum_{k=0}^{n} \binom{n}{k} \mu_p^{n-k} \sigma_p^k E\left[Z^k\right] =$$

$$= \frac{\mu_p^n \sum_{k=0}^{n} \binom{n}{k} \left(\sigma_p/\mu_p\right)^k \left\{1 + (-1)^k\right\} \Gamma\left(\frac{k+1}{p}\right)}{2\Gamma\left(1/p\right)}$$

While the first four moments of the variable X can be determined as:

$$E[X] = \mu_p$$

$$E\left[X^2\right] = \mu_p^2 + \frac{\sigma_p^2 \Gamma\left(\frac{3}{p}\right)}{\Gamma\left(\frac{1}{p}\right)}$$

$$E\left[X^3\right] = \mu_p^3 + \frac{3\mu_p \sigma_p^2 \Gamma\left(\frac{3}{p}\right)}{\Gamma\left(\frac{1}{p}\right)}$$

and:

$$E\left[X^4\right] = \mu_p^4 + \frac{6\mu_p^2 \sigma_p^2 \Gamma\left(\frac{3}{p}\right)}{\Gamma\left(\frac{1}{p}\right)} + \frac{\sigma_p^4 \Gamma\left(\frac{5}{p}\right)}{\Gamma\left(\frac{1}{p}\right)}$$

Table 1
Values of β_2 and I for different theoretical values of p.

p	β_2	I	p	β_2	I	p	β_2	I
0.5	25.2000	0.5477	3.7	2.2406	0.8377	6.9	1.9546	0.8558
0.8	8.5651	0.6639	4.0	2.1884	0.8409	7.2	1.9440	0.8565
1.0	6.0000	0.7071	4.2	2.1558	0.8427	7.4	1.9376	0.8569
1.2	4.7434	0.7369	4.4	2.1327	0.8443	7.6	1.9316	0.8573
1.4	4.0178	0.7589	4.6	2.1009	0.8458	7.8	1.9260	0.8577
1.6	3.5537	0.7753	4.8	2.0887	0.8471	8.0	1.9208	0.8581
1.8	3.2323	0.7877	5.0	2.0701	0.8483	8.2	1.9159	0.8584
2.0	3.0000	0.7978	5.2	2.0454	0.8499	8.4	1.9113	0.8587
2.2	2.8247	0.8060	5.4	2.0309	0.8508	8.6	1.9069	0.8590
2.4	2.6884	0.8128	5.6	2.0177	0.8517	8.8	1.9028	0.8593
2.6	2.5797	0.8184	5.8	2.0056	0.8525	9.0	1.8990	0.8595
2.8	2.4914	0.8232	6.0	1.9945	0.8532	9.2	1.8953	0.8597
3.0	2.4184	0.8273	6.2	1.9844	0.8539	9.4	1.8936	0.8599
3.2	2.3571	0.8308	6.4	1.9750	0.8545	9.6	1.8903	0.8601
3.4	2.3082	0.8339	6.6	1.9664	0.8550	9.8	1.8871	0.8603
3.6	2.2606	0.8365	6.8	1.9584	0.8556	10.0	1.8841	0.8605

Now, starting from the following relationship:

$$E\left[\left(X-\mu_p\right)^n\right] = \sigma_p^n\left\{1+(-1)^k\right\} \int_{\mu_p}^{\infty}\left(\frac{X-\mu_p}{\sigma_p}\right)^n \frac{p\exp\left\{-\left(\frac{X-\mu_p}{\sigma_p}\right)^p\right\}}{2\sigma_p\Gamma\left(1/p\right)}dx \tag{12}$$

The n-th central moment can be derived as follows:

$$E\left[\left(X-\mu_p\right)^n\right] = \frac{p\sigma_p^n\left\{1+(-1)^k\right\}}{2\Gamma\left(1/p\right)}\int_0^{\infty} z^n \exp\left(-z^p\right)dz =$$

$$= \frac{\sigma_p^n\left\{1+(-1)^k\right\}}{2\Gamma\left(1/p\right)}\int_0^{\infty} y^{(n+1)/(p-1)} \exp(-y)dy =$$

$$= \frac{\sigma_p^n\left\{1+(-1)^k\right\}}{2\Gamma\left(1/p\right)}\Gamma\left(n+1/p\right)$$

So, from the (12) we can obtain the first four central moments that are:

$$E\left[\left(X-\mu_p\right)\right] = 0$$

$$E\left[\left(X-\mu_p\right)^2\right] = \frac{\sigma_p^2\Gamma\left(\frac{3}{p}\right)}{\Gamma\left(\frac{1}{p}\right)}$$

$$E\left[\left(X-\mu_p\right)^3\right] = 0$$

$$E\left[\left(X-\mu_p\right)^4\right] = \frac{\sigma_p^4\Gamma\left(\frac{5}{p}\right)}{\Gamma\left(\frac{1}{p}\right)}$$

The kurtosis itself plays a fundamental role in the combined estimation method we are presenting in the next section. In particular, the first four central moments have very nice properties from an algebraic point of view ([25] and [26]) and the kurtosis has important statistical applications especially, but not limited to financial data [23].

However, for the G.E.D. the skewness is zero because this distribution is symmetric (for the derivation for the asymmetric case see the Skewed G.E.D. defined in the previous paragraph) and the kurtosis that we call β_k is referred to k-th order. As we said before the so-called "Generalized Kurtosis" can be defined as in (4).

Indeed, replacing p =2 in (4) we obtain the Pearson kurtosis index that we call β_2 [2], already defined in (5). In the end, for $k = 1$ in (4), considering the square-root of the β_k reciprocal, we obtain the Geary Index (6), which measure the length of the tails.

4. The combined method based on the kurtosis indexes: Lp_{med}

Mineo [30] proposed an algorithm to estimate the shape parameter based on the errors distribution. This procedure is implemented by starting from a particular kurtosis index, called Generalized Kurtosis, whose formula is defined as (4).

However, some algorithms proposed in literature do not take into account all the properties of the p parameter conditioning the estimates of the shape parameter only to the kurtosis index β_2 as in (5). For example Sposito et al. [38] suggested the following relation to estimate p:

$$p_{sp} = \frac{6}{\hat{\beta}_2^2}, \qquad \text{for } 1 \le p < 2$$

We observe that the tails' distribution should also be taken into consideration as determinant factor for the shape parameter estimation, so we propose to consider the Geary index I, defined as (6).

For this reason we suggest to combine the Eqs. (5) and (6) jointly for the p estimation even if their behaviour differs according to the values assumed by p (Table 1).

The idea of combining estimators from two kurtosis indexes comes from the fact that the former (6) better approximates the value of the theoretical p for samples with many observations on the distribution tails, while the latter (5) is more suitable for the p estimation in samples with many values located around their centrality parameter.

In particular, these indexes essentially show two different aspects of the same phenomenon and for this reason, their common use is strongly indicated in order to get a combined estimator resuming both characteristics highlighted, because considering only the β_2 index to estimate p could be too restrictive, and the same we could say for the Geary index I.

In general, therefore for $p \to \infty$, the Geary index reaches the value of 0.866 while β_2 goes to 1.8.

For $p \to 0$ we have that $I \to 0$ and $\beta_2 \to \infty$. Finally the definition range is as follows, $I \epsilon [0; 0.866]$ and $\beta_2 \epsilon [1.8, +\infty]$. In order to prove it, remembering that as soon as $p \to \infty$, the G.E.D. p.d.f. converges to the uniform distribution, we can write:

$$\lim_{p\to\infty}\frac{1}{p}\Gamma\left(\frac{1}{p}\right) = \lim_{p\to\infty}\Gamma\left(1+\frac{1}{p}\right) = \Gamma(1) = 1$$

$$\lim_{p\to\infty} e^{-\left|\frac{x-\mu_p}{\sigma_p}\right|^p} = \begin{cases} 0 \; for \left|\frac{x-\mu_p}{\sigma_p}\right| > 1 \\ 1 \; for \left|\frac{x-\mu_p}{\sigma_p}\right| \le 1 \end{cases}$$

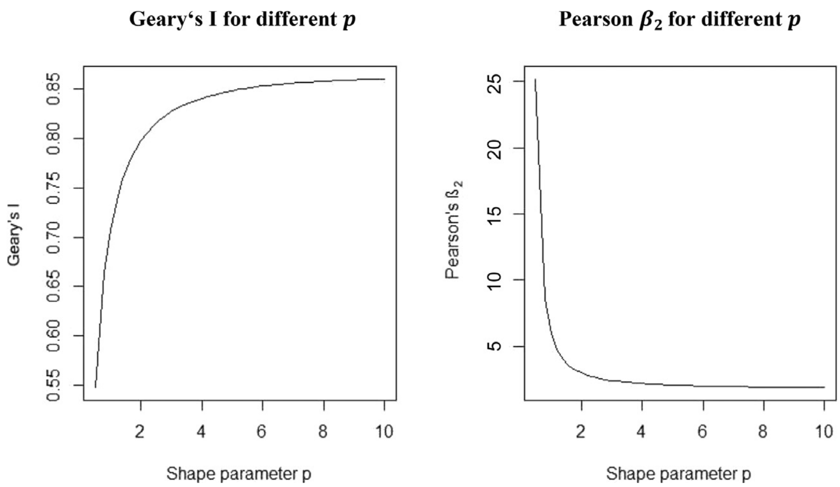

Fig. 2. Indexes I and β_2 for different values of shape parameter p.

and:

$$\lim_{p\to\infty} f(x) = \begin{cases} 0\ for \left|\frac{x-\mu_p}{\sigma_p}\right| > 1 \\ \frac{1}{2\sigma_p}\ for \left|\frac{x-\mu_p}{\sigma_p}\right| \le 1 \end{cases}$$

Since the k-th absolute moment is:

$$E[\left|x-\mu_p\right|]^k = \frac{1}{(k+1)}\sigma_p\ :$$

$$E\left[\left|x-\mu_p\right|\right] = \frac{1}{2}\sigma_p$$

$$E[\left|x-\mu_p\right|]^2 = \frac{1}{3}\sigma_p^2$$

Hence:

$$I = \frac{E\left[\left|x-\mu_p\right|\right]}{\sigma_p} = \frac{\sqrt{3}}{2} = 0.866$$

$$\beta_2 = \frac{E[\left|x-\mu_p\right|]^4}{\sigma_p^4} = \frac{9}{5} = 1.8$$

We can also note that the Geary index is less sensitive to changes in p in comparison to the Pearson index. For example, for p increasing from 2 to 6, it is possible to read from Table 1 that while β_2 decreases from 3.00 to 1.99459, the I index increases only from 0.79788 to 0.85324.

In particular, the behaviour of β_2 and I is opposite respect to the variation of p: if, in fact, with increasing p the Pearson kurtosis index decreases, the Geary index I increases, (Fig. 2).

This difference in sensitivity (and therefore also the greater "stability" of the index I) can largely be interpreted through the estimators of both indexes. Indeed, we have that:

$$\hat{\beta}_2 = \frac{n\ \sum_i (\varepsilon_i - \bar{\varepsilon})^4}{\left[\sum_i (\varepsilon_i - \bar{\varepsilon})^2\right]^2}$$

$$\hat{I} = \frac{\sum_i |\varepsilon_i - \bar{\varepsilon}|}{\sqrt{n}\sqrt{\sum_i |\varepsilon_i - \bar{\varepsilon}|^2}}$$

We can note that the estimator $\hat{\beta}_2$, due to the presence of the fourth moment is more influenced by the fluctuations of the extreme values on the tails, whereas the estimator $\hat{I}$ is more strongly affected by the errors in the values around the mean.

One of the problems in using these estimators is that both the sample values of β_2 and I are unbiased for small samples. In order to avoid these problems, we employed some changes by considering some correction factors. Gonin and Money [19] introduced a correction factor that depends on the sample size n. Obviously, this correction becomes less relevant increasing the sample size n. Calculating the estimators, it is immediate to identify Pearson's Kurtosis as a ratio of the previous estimators:

$$\hat{\beta}_2 = \frac{\widehat{\mu_4}}{\widehat{\mu_2}^2} \tag{13}$$

where:

$$\widehat{\mu_2}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\varepsilon_i - \bar{\varepsilon})^2$$

$$\widehat{\mu_4} = \frac{\left(n^2-2n+3\right)}{(n-1)(n-2)(n-3)}\sum_{i=1}^{n}(\varepsilon_i - \bar{\varepsilon})^4 - \frac{3(n-1)(2n-3)}{n(n-2)(n-3)}\widehat{\mu_2}^2$$

Where ε_i is the i-th residual of the estimated model and $\bar{\varepsilon}$ is the average of them. For Geary's index we decided to modify the estimator by adding a correction factor for the second moment, leaving the absolute first moment unchanged [22].

Therefore, the estimator of I becomes:

$$\hat{I} = \frac{\sum_i |\varepsilon_i - \bar{\varepsilon}|}{\sqrt{\sum_i |\varepsilon_i - \bar{\varepsilon}|^2}}\ \frac{\sqrt{n-1}}{n} \tag{14}$$

An useful algorithm that jointly uses both analyzed kurtosis indexes for the shape parameter estimation can be introduced after considering the following simple exponential regression model [29]:

$$y_i = \theta_2 \cdot e^{\theta_1 x_i} + \varepsilon_i \tag{15}$$

We apply the maximum likelihood method (since, in the hypothesis of p-known, it provides the Lp-norm estimators for the considered exponential regression model) and, in this scenario, we calculate the log-likelihood function:

$$L\left(\theta_1,\ \theta_2,\ \sigma_p,\ p\right) = \\ = -nlog\left[2\ p^{1/p}\ \sigma_p\ \Gamma(1+1/p)\right] - \left[\left(p\sigma_p^p\right)^{-1}\sum |y_i - \theta_2 e^{\theta_1 x_i}|^p\right]$$

Assuming that initially the value of p is known, the partial derivatives of the log-likelihood function can be calculated respect to $\theta_1, \theta_2, \sigma_p$ setting these derivatives equal to zero in order to solve the minimum problem. Indeed, we obtain in that context the following first order conditions:

$$\frac{\partial L}{\partial \theta_1} = \sum \left|y_i - \theta_2 e^{\theta_1 x_i}\right|^{p-1}\left(\theta_2 x_i e^{\theta_1 x_i}\right) sign\ \left(y_i - \theta_2 e^{\theta_1 x_i}\right) = 0$$

$$\frac{\partial L}{\partial \theta_2} = \sum \left|y_i - \theta_2 e^{\theta_1 x_i}\right|^{p-1}\left(e^{\theta_1 x_i}\right) sign\ \left(y_i - \theta_2 e^{\theta_1 x_i}\right) = 0$$

$$\frac{\partial L}{\partial \sigma_p} = \left(\frac{n}{p}\right) - \left(\frac{1}{\sigma_p^{p+1}}\right)\left(\sum \left|y_i - \theta_2 e^{\theta_1 x_i}\right|^p\right) = 0$$

Solving the partial derivative of the log-likelihood with respect to σ_p, we obtain S_p estimator of σ_p:

$$S_p = \left[\frac{\sum |y_i - \theta_2 e^{\theta_1 x_i}|^p}{n} \right]^{\frac{1}{p}}$$

On the other hand, if we assume that the value of p is unknown, we have to use an algorithm to compute its value. Considering the literature on this topic, we have selected two methods: the Agrò method (Lp_+) and the Money et al method (Lp_{gm}) and in the follow we report the equations used by the cited authors to get their p estimators.

The maximum likelihood equation to estimate p, proposed by Agrò [1] considered in the simulation study appearing in the next paragraph and indicated with p_+ is shown below:

$$\frac{\partial L}{\partial p_+} = (n/p^2)[\log(p) - 1 + zd(1+1/p)] + \left(1/p^2\sigma_p^p\right)\Big[\sum \left|y_i - \theta_2 e^{\theta_1 x_i}\right|^p - p\sum \left|y_i - \theta_2 e^{\theta_1 x_i}\right|^p \log\left|y_i - \theta_2 e^{\theta_1 x_i}\right| + p\sum \left|y_i - \theta_2 e^{\theta_1 x_i}\right|^p \log(\sigma_p)\Big] = 0 \quad (16)$$

where zd(1+1/p) indicates the digamma function or the derivative logarithm of the Gamma function.

However, as regards to the estimate of p, indicated with p_{gm}, obtained by Money et al. [33] on the basis of an extensive simulation study they carried out, the following formula is given:

$$p_{gm} = \frac{9}{\hat{\beta}_2^2} + 1, \qquad \text{for } 1 \le p < \infty \quad (17)$$

Finally considering the empirical and theoretical values of the Geary (6) and the Pearson (5) indexes, we obtain the p_{med} estimator proposed in this paper. We combine the two following nonlinear Eqs. (18) and (19) that give us two preliminary estimates:

$$I(p_1) - \hat{I} = 0 \quad (18)$$

$$\beta_2(p_2) - \hat{\beta}_2 = 0 \quad (19)$$

$$p = \lambda p_1 + (1-\lambda)\, p_2$$

Where the quantity $I(p_1)$ is defined as (6) and $\hat{I}$ is defined as (14). Moreover, the quantity $\beta_2(p_2)$ is defined as (5) and $\hat{\beta}_2$ is defined as (13).

Firstly the values of p_1 and p_2 are estimated from the former and the latter nonlinear equations. The last identity constitutes the final p_{med}, intended as the simple mean of p_1 and p_2. So, we obtain:

$$p_{med} = 0.5p_1 + 0.5p_2 \quad (20)$$

The weighting factor λ is set arbitrarily equal to 0.5 (so in our model λ= 0.5) because the logic of the model is that the characteristics of both kurtosis indexes are considered of equal importance. Indeed, we suppose that the characteristics of p_1, deriving from Geary's index (which better approximates the value of the theoretical p for samples with many observations on the distribution's tails) and the characteristics of p_2, deriving from Pearson's kurtosis (more suitable for the estimation of p in samples with many observations located around their central value.) are similarly important.

However, in the absence of a priori information, λ= 0.5 seems to be the right decision, even if arbitrary, because it seems to be the more logical one. The proposed algorithm is based on a two steps alternating procedure that firstly estimates the $\theta_1, \theta_2, \sigma_p$ parameters using the L_p-*norm* estimators and secondly estimates p with p_{med} in the following way:

Step 0: We set initially $i = 0$ and $p_o = 2$;
Step 1: We estimate the values of $\theta_1, \theta_2, \sigma_p$ using the L_p-*norm* estimators, calling these estimates $\theta_{1i}, \theta_{2i}, S_{pi}$;
Step 2: We compute the residuals $\varepsilon_i = y_i - \theta_{2i} \cdot e^{\theta_{1i} x_i}$, their average $\bar{\varepsilon}$ and insert these quantities in the nonlinear Eqs. (18) and (19);
Step 3: We compute p_1 and p_2 from the equations showed above.
Step 4: We compute p_{med}(20) as the average of p_1 and p_2 obtaining p_{i+1}, new estimate of p;
Step 5: We compare the last estimate of p obtained from the procedure (p_{i+1})with the previous p and if $|p_{i+1} - p_i| > 0,01$ we set $i = i + 1$ starting again from the step 1 until step 4; otherwise;
Step 6: We stop the algorithm assuming the values $\theta_{1i}, \theta_{2i}, S_{pi}$ as the L_p-*norm* estimates of the parameters $\theta_1, \theta_2, \sigma_p$ and the value $p = p_{med}$ as combined estimation of the exponent p.

5. The simulation study and the results

The performance of the above method was experimented by a Montecarlo simulation study to evaluate the unbiasedness and the asymptotic behaviour of the new estimation procedure for the exponential regression model (15).

Since the proposed method is of empirical nature because of the absence of prior information on the theoretical distribution of p_1 and p_2, it seems to show asymptotically very nice properties like asymptotic normality and a decreasing variance with increasing size n.

Starting from the simple exponential regression model showed in the previous section we have simulated 1000 samples of size n= 50, 100, 200 and 500 generated from a Generalized Error Distribution with six theoretical values (theoretical populations) of the shape parameter p (1.2, 1.5, 2.0, 2.5, 3.0, 3.5) [9].

The n-pairs of values (x_i, ε_i) were generated to get the y_i. The samples $x_i = (x_1, x_2, \ldots,\ x_n)$ where generated from a uniform distribution (0.5, 1.5).

In the model (15) we fixed the parameters $\theta_1' = 0.5$, $\theta_2' = 1$ and $\sigma_p = 1$ as theoretical values to be estimated with empirical frequency distributions. Obviously, we have indicated with θ_1, θ_2 and S_p the values of the sample estimates. The constants calculated on the sample estimates (for all the values of n) were mean and variance to evaluate their unbiasedness, different efficiency and asymptotic behaviour. The results of the simulation study are in the Table 2 showed below:

From the experimental results reported in Table 2 we can observe that, for any p, the parameter estimates of θ_1', θ_2' and σ_p are biased for n=50. Their variance decrease for increasing values of n.

This is true for all the parameters and for all the theoretical values of p and depends on the nonlinearity of the model that yields the unbiasedness of the estimates only for middle-large sample sizes (see n=200 and n=500).

As it is possible to note in Table 2, p_{med} shows a decreasing variance when the sample size increases too. Moreover, also its expected value is closer to its theoretical value in increasing of the sample size n, showing overall, as said before, interesting asymptotic properties. However, in order to show the better results of the proposed method, an efficiency comparison is needed.

Indeed, in the Tables 3, 4, 5 and 6 the Mean Square Error is calculated for three different methods to estimate the exponent p: the first proposed by Money et al. [33] (Lp_{gm}), the second (Lp_+) proposed by Agrò [1] and the last based on the combined algorithm we are proposing (Lp_{med}).

The Table 3 shows the results for small samples size n=50. In this case the method proposed by Money et al. [33] performs better than others especially when the theoretical value of p is equal to 2 (i.e. in the Gaussian normality situation).

However, for small samples the combined method based on kurtosis indexes (Lp_{med}) performs better than the one based on the maximum likelihood (16) for all the theoretical values of p we consider. Almost the same results obtained for sample size equal to n=100, are showed in the Table 4 below.

In this case in presence of an actual leptokurtic distribution (when $1 < p < 2$), the Lp_{med} method allows us to obtain better results for this samples size respect to both the other methods.

The most important result is related to the asymptotic behaviour of the p estimates in the cases n=200 and n=500 (Table 5 and 6). The methods Lp_+ and Lp_{med} seem to show a possible asymptotic convergence

Table 2
Results of simulation study with 1000 samples of size n=50, 100, 200 and 500 generated from a G.E.D., with different supposed values of p (Lp_{med} method).

p	$E[\theta_1]$	$V[\theta_1]$	$E[\theta_2]$	$V[\theta_2]$	$E[S_p]$	$V[S_p]$	$E[p_1]$	$V[p_1]$	$E[p_2]$	$V[p_2]$	$E[p_{med}]$	$V[p_{med}]$
n=50												
1.2	0.463	0.130	1.095	0.182	1.097	0.102	1.305	0.537	1.547	0.220	1.426	0.278
1.5	0.477	0.126	1.082	0.168	1.031	0.913	1.627	0.366	1.904	0.581	1.765	0.473
2.0	0.486	0.113	1.064	0.148	0.985	0.789	2.196	0.972	2.526	1.075	2.361	1.023
2.5	0.479	0.093	1.073	0.127	0.978	0.069	2.664	1.507	3.195	1.901	2.929	1.704
3.0	0.481	0.072	1.055	0.109	0.969	0.057	2.928	1.593	3.676	2.439	3.302	2.016
3.5	0.519	0.069	1.046	0.098	0.976	0.050	3.734	1.635	4.197	2.785	3.965	2.058
n=100												
1.2	0.483	0.070	1.056	0.112	1.043	0.062	1.251	0.186	1.403	0.086	1.327	0.151
1.5	0.479	0.066	1.049	0.101	1.017	0.058	1.561	0.242	1.697	0.187	1.629	0.274
2.0	0.480	0.054	1.042	0.087	1.187	0.048	2.115	0.362	2.290	0.363	2.202	0.362
2.5	0.479	0.487	1.036	0.098	0.959	0.041	2.539	0.586	2.771	0.556	2.655	0.771
3.0	0.486	0.413	1.048	0.086	0.975	0.036	3.152	1.255	3.382	1.028	3.267	1.141
3.5	0.485	0.038	1.044	0.068	0.942	0.031	3.457	1.739	3.979	1.401	3.718	1.570
n=200												
1.2	0.491	0.041	1.019	0.073	1.023	0.038	1.228	0.101	1.309	0.042	1.268	0.108
1.5	0.492	0.036	1.015	0.066	1.018	0.028	1.525	0.148	1.594	0.074	1.559	0.161
2.0	0.489	0.024	1.026	0.058	1.011	0.022	2.065	0.240	2.146	0.143	2.105	0.192
2.5	0.510	0.021	0.994	0.049	0.969	0.018	2.563	0.402	2.646	0.296	2.604	0.349
3.0	0.507	0.019	1.018	0.037	0.975	0.014	3.148	0.875	3.224	0.473	3.186	0.674
3.5	0.509	0.017	1.022	0.032	0.964	0.010	3.571	1.434	3.684	0.668	3.627	1.051
n=500												
1.2	0.498	0.038	1.008	0.047	1.011	0.025	1.208	0.064	1.209	0.032	1.208	0.048
1.5	0.496	0.033	0.997	0.034	1.008	0.021	1.485	0.046	1.514	0.059	1.499	0.059
2.0	0.497	0.022	1.008	0.026	1.007	0.019	2.055	0.130	2.063	0.110	2.059	0.121
2.5	0.503	0.019	0.098	0.022	0.998	0.013	2.473	0.302	2.555	0.166	2.514	0.219
3.0	0.502	0.017	1.005	0.016	0.995	0.011	3.042	0.785	3.137	0.313	3.089	0.589
3.5	0.501	0.016	1.013	0.012	0.993	0.008	3.501	1.304	3.651	0.496	3.576	0.846

Table 3
Expected value, variance and Mean Square Error of shape parameter p with sample size n=50 according to Money (Lp_{gm}), Agrò (Lp_+) and combined (Lp_{med}) methods.

p	1.2	1.5	2.0	2.5	3.0	3.5
n=50						
$E[p_{gm}]$	1.5467	1.8388	2.1884	2.7145	2.7593	2.9266
$V[p_{gm}]$	0.2103	0.3762	0.7445	0.8135	1.0966	1.3845
$MSE[p_{gm}]$	0.3346	0.4688	1.2265	1.3589	1.5503	1.9547
$E[p_+]$	1.6816	1.9458	2.4819	2.9446	3.4478	3.9522
$V[p_+]$	0.3477	0.4248	1.4501	1.8854	1.9669	2.3443
$MSE[p_+]$	0.5776	0.6439	1.9882	2.0255	2.1354	2.4453
$E[p_{med}]$	1.4268	1.7946	2.2692	2.7372	3.1987	3.8669
$V[p_{med}]$	0.1843	0.3488	0.8103	1.0235	1.0437	1.5687
$MSE[p_{med}]$	0.2295	0.4166	1.3245	1.3859	1.4366	1.6335

Table 4
Expected value, variance and Mean Square Error of shape parameter p with sample size n=100 according to Money (Lp_{gm}), Agrò (Lp_+) and combined (Lp_{med}) methods.

p	1.2	1.5	2.0	2.5	3.0	3.5
n=100						
$E[p_{gm}]$	1.4366	1.7728	2.0948	2.3342	2.4766	2.8469
$V[p_{gm}]$	0.0708	0.1175	0.1464	0.2257	0.3584	0.5397
$MSE[p_{gm}]$	0.1301	0.1684	0.1845	0.2846	0.4706	0.7865
$E[p_+]$	1.3876	1.7345	2.2055	2.6165	3.1066	3.5951
$V[p_+]$	0.0398	0.0891	0.2489	0.4561	0.6234	0.8789
$MSE[p_+]$	0.0865	0.1425	0.2768	0.4968	0.7582	0.9845
$E[p_{med}]$	1.3273	1.6292	2.1268	2.5889	3.1343	3.5736
$V[p_{med}]$	0.0579	0.1205	0.1942	0.3453	0.5708	0.8122
$MSE[p_{med}]$	0.0693	0.1369	0.2137	0.3912	0.7261	0.8906

to a Normal distribution, whilst the Lp_{gm} method shows an increasing biasedness even when the sample size increases (see the cases p =3 and p =3.5).

Looking at the simulation results it seems reasonable to distinguish two cases for n=200 and n=500.

When $1.2 \leq p < 2$ the Lp_{med} algorithm here proposed give us the best performance respect to the other methods. The case of $p \geq 2$ is well dealt using either Lp_+ or Lp_{med} . In fact, as it is well known, the maximum likelihood method generally achieves the best results when the sample size is higher.

Table 5
Expected value, variance and Mean Square Error of shape parameter p with sample size n=200 according to Money (Lp_{gm}), Agrò (Lp_+) and combined (Lp_{med}) methods.

p	1.2	1.5	2.0	2.5	3.0	3.5
n=200						
$E[p_{gm}]$	1.5078	1.6955	2.0472	2.3377	2.6108	2.9154
$V[p_{gm}]$	0.0248	0.0507	0.0775	0.1189	0.1814	0.2067
$MSE[p_{gm}]$	0.1141	0.1065	0.0910	0.1488	0.3478	0.7028
$E[p_+]$	1.2757	1.6304	2.0743	2.5277	2.9449	3.3867
$V[p_+]$	0.0495	0.0276	0.0724	0.1344	0.3165	0.4552
$MSE[p_+]$	0.0948	0.0806	0.0954	0.1458	0.4421	0.5187
$E[p_{med}]$	1.2585	1.5694	2.0605	2.5689	3.0504	3.5338
$V[p_{med}]$	0.0382	0.0439	0.0759	0.1749	0.3987	0.5844
$MSE[p_{med}]$	0.0618	0.0683	0.0944	0.2168	0.4787	0.5643

Table 6
Expected value, variance and Mean Square Error of shape parameter p with sample size n=500 according to Money (Lp_{gm}), Agrò (Lp_+) and combined (Lp_{med}) methods.

p	1.2	1.5	2.0	2.5	3.0	3.5
n=500						
$E[p_{gm}]$	1.2406	1.5337	2.0103	2.5189	2.8933	3.1207
$V[p_{gm}]$	0.0156	0.0268	0.0169	0.0118	0.0467	0.0645
$MSE[p_{gm}]$	0.0239	0.0436	0.0278	0.0467	0.1656	0.2934
$E[p_+]$	1.2279	1.5209	2.0096	2.5109	2.9754	3.4856
$V[p_+]$	0.0123	0.0234	0.0147	0.0145	0.1546	0.1578
$MSE[p_+]$	0.0202	0.0345	0.0229	0.0411	0.1821	0.2127
$E[p_{med}]$	1.2194	1.5148	2.0045	2.5236	3.0433	3.5298
$V[p_{med}]$	0.0086	0.0251	0.0142	0.0298	0.1758	0.1704
$MSE[p_{med}]$	0.0168	0.0303	0.0218	0.0767	0.2123	0.2678

6. Application to Italian well-being data

In order to show the application of the proposed method, we consider data from Italian National Statistical Institute (ISTAT) on Italian equitable and sustainable well-being. The awareness about the relevance of sustainable and equitable well-being (in Italian "Benessere Equo e Sostenibile", BES) indicators in terms of economic and financial planning has reached its pick recently with the declaration by the Italian Government to monitor the progress of some BES indicators considered relevant within the annual D.E.F. (Economics and Finance Document).

In particular, the Government together with the Committee for BES indicators, have enforced the monitoring of 12 indicators included in the dimensions of the BES [12].

Specifically, these are:

1. Available average income adjusted per capita;
2. Index of inequality of disposable income;
3. Index of absolute poverty;
4. Life expectancy in good health at birth;
5. Excess weight;
6. Early exit from the education and training system;
7. Rate of non-participation in the work, with relative breakdown by gender;
8. Ratio between the employment rate of women aged 25-49 with preschoolers and women without children;
9. Predatory crime index;
10. Index of efficiency of civil justice;
11. CO_2 emissions and other altering climate gases;
12. Index of illegal construction.

However, for the transitional phase 4 indicators have been selected from the 12 mentioned.

The selected indicators are:

1. Available average income adjusted per capita;
2. Index of inequality of disposable income;
3. Rate of non-participation in the work, with relative breakdown by gender;
4. Emissions of CO_2 and other altering climate gases.

In this section we want to study the impact of the 4 indicators included in the Italian DEF on the GDP growth rate of the Italian regions and on Italy as a country. In order to achieve this aim we run a pooled linear regression on Italian BES data from 2013 up to 2018 (Table 7) by using first the Ordinary Least Squares (O.L.S.) and, then, Lp-norm estimators with two different methods.

In particular, we estimate the shape parameter first considering the Lp_{min} algorithm [16] and then with the Lp_{med} procedure introduced in this paper based on (20). Hence, we estimate the following linear regression model:

$$Y_{i,t} = \beta_0 + \beta_1 X_{1i,t} + \beta_2 X_{2i,t} + \beta_3 X_{3i,t} + \beta_4 X_{4i,t} + \varepsilon_{i,t} \quad (21)$$

Where $Y_{i,\,t}$ is the "*GDP growth*", $X_{1i,\,t}$ is the "*Net average available income*", $X_{2i,\,t}$ is the "*Index of available income inequality*", $X_{3i,\,t}$ is the "*Work non participation rate*" and $X_{4i,\,t}$ is the "*Index of overall environmental conditions*".

Results are showed in the Table 8 below.

In terms of statistical significance, for all the introduced specifications we obtain the same results: net average available income, work non-participation rare and environmental condition affect regional GDP changes. Income inequality, instead, do not. Nevertheless, for evaluating the regional well-being, income inequality is one of the most important indicators to take into account. This means, in other words, that monitoring this variable is still more important on the light of these results. However, from statistical point of view several differences across models could be highlighted.

Firstly, the estimates of p are far from 2, indicating that OLS estimation probably is not the best choice. Instead, according to *Lp-norm* estimation with Lp_{min} method the value of p is equal to 1.62, while for the proposed Lp_{med} even it is less than 2, namely 1.48. Standard error associated to estimate regression parameters are then lowered for the *Lp-norm* estimation with Lp_{med} with respect to the other models. This means in a certain sense that estimates are more accurate with *Lp-norm* and, specifically, with Lp_{med}. To provide further results, we compared for all the three methods the p-th power of residuals (Table 9).

Table 7
Descriptive statistics of data.

Variables	Mean	St. Dev.	Min	Max
GDP growth rate	-0.0041268	0.199193	-0.0709512	0.0880617
Net average available income	17814.93	3492.195	11989.11	24623.2
Index of available income inequality	5.1730043	1.257491	3.6	10
Poverty risk	18.68783	9.94999	5.4	42.3
Life expectancy	82.42957	0.6963531	80.4	83.8
Overweight	44.65043	4.485403	35.7	53.4
Early exit from the education system	14.32	4.368291	6.7	25.4
Work non-participation rate	20.5887	10.72943	4.8	43
Woman employment ratio	79.23913	6.677938	62.4	95.4
Predatory crime index	96.48783	8.726257	76.8	113.2
Justice efficiency index	421.4252	228.6827	102.3	974
Index of overall environmental conditions	104.3165	7.259355	89.4	121.2
Index of illegal constructions	21.72609	20.4907	1.3	71.1

Table 8
Estimates from different pooled regressions.

Coefficients	L_2 estimates	Lp_{min} estimates	Lp_{med} estimates
β_0	-1.736e^{-01} *** (5.139673e^{-02})	-1.928e^{-01}*** (5.029461e^{-02})	-2.068e^{-01} *** (4.95823e^{-02})
β_1	4.494e^{-06}* (1.804327e^{-06})	6.836e^{-06}* (1.765637e^{-06})	7.349e^{-06} * (1.724633e^{-06})
β_2	-3.296e^{-03} (2.536642e^{-03})	-3.707e^{-03} (2.482248e^{-03})	-4.148e^{-03} (2.447298e^{-03})
β_3	1.830e^{-03} ** (6.862519e^{-04})	2.324e^{-03}** (6.715365e^{-04})	2.583e-03* (6.587729e-04)
β_4	6.596e^{-04} * (3.337399e^{-04})	4.855e^{-04} * (3.265835e^{-04})	4.041e^{-04} * (3.190983e^{-04})
p	**2**	**1.62**	**1.48**

Note: * is significant for 90%, ** for 95% and *** for 99%. Standard errors are reported in parentheses.

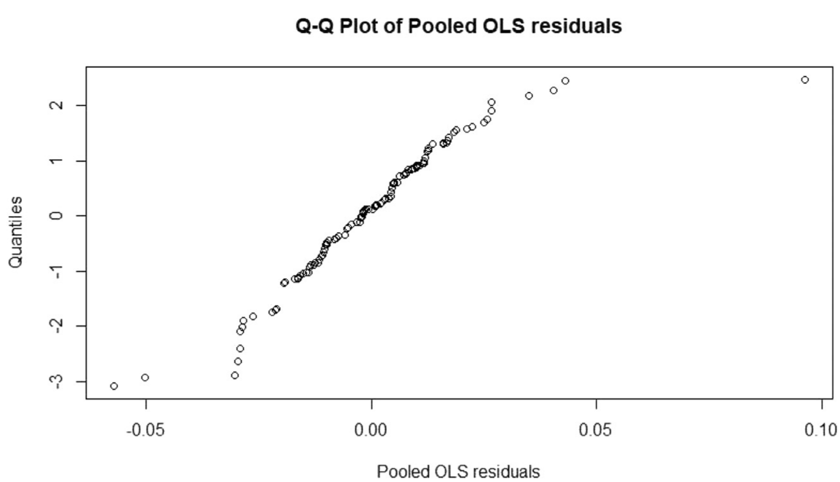

Fig. 3. Q-Q Plot of pooled OLS residuals.

Table 9
Sum of p-th power residuals.

Measure	Pooled OLS estimates	Lp_{min} estimates	Lp_{med} estimates
$\sum_{i=1}^{N} e_i^p$	0.04185381	0.04007809	**0.03830237**

Table 9 clearly shows the over-performance of the Lp_{med} algorithm respect to the alternatives.

In conclusions, residuals Q-Q plots are showed for all the three methods (Figs. 3, 4, 5).

Hence empirically Lp_{med} algorithm leads to more accurate estimates than Lp_{min} and Least Squares.

7. Final remarks

The Generalized Error Distribution is a very flexible family of distributions, where, by changing the value of the shape parameter p, it is possible to obtain several symmetric distributions.

The L_p-*norm* estimators can be derived assuming that the residuals of the regression model follow a G.E.D. and for p=2 the L_p-*norm* procedure gives exactly the same estimators of the Least Squares, while for $p \neq 2$ the L_p-*norm* estimators coincides with the maximum likelihood estimators when the p parameter is specified.

However, usually the value of the shape parameter needs to be estimated and, as a result, we have introduced a method called Lp_{med} which seems to show some advantages compared to those used by previous literature [1, 33] especially for medium-large samples and in presence of leptokurtic data, mainly because it considers more carefully the tails of the distribution by the joint computation of two G.E.D. kurtosis indexes.

This paper analyzes the computation of G.E.D. distribution moments too. Moreover, we also obtain, through the proposed algorithm, the estimated value of the regression parameters for a particular nonlinear exponential model.

Indeed, several papers [3, 7, 17, 24, 27, 28, 42] suggest the rejection of the normality assumption in data analysis and some of them underline the improvements achieved by using alternative and more flexible

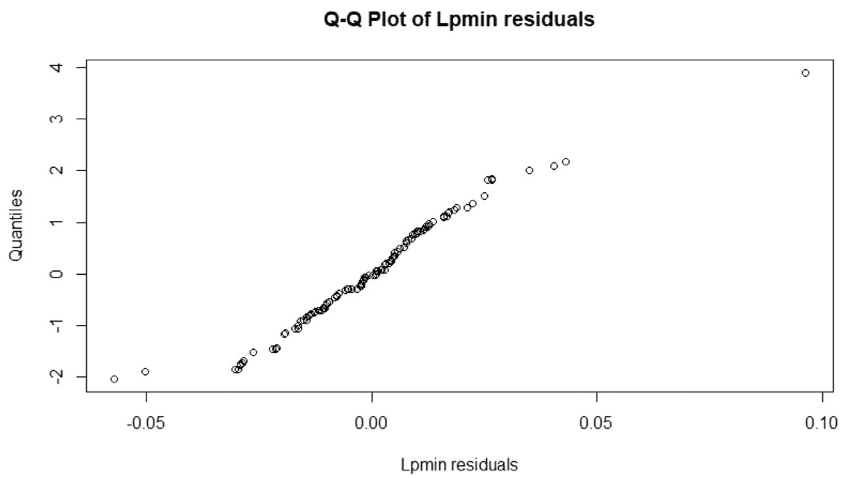

Fig. 4. Q-Q Plot of Lp_{min} residuals.

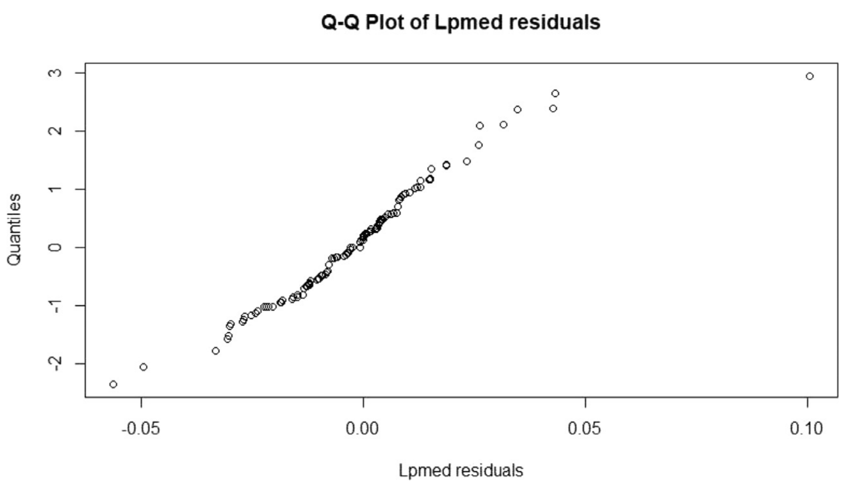

Fig. 5. Q-Q Plot Lp_{med} residuals.

distributional assumptions as these we have studied and analyzed in this paper. Very often, in order to minimize the *p*-th power considering the *Lp*-norm estimation rule, we have to find the value of *p* according to the distribution of the data contained in the sample. Among the most common approaches, in this paper we have proposed the combined use of two kurtosis indexes in order to obtain a robust *p* estimation as the characteristics of the samples change.

Looking at the simulation results the Lp_{med} findings show very good asymptotic properties for the parameters estimation and a very evaluable performance in terms of mean squared error of the exponent *p* especially in the case of leptokurtic residual distributions.

Finally a real data application about Italian Equitable and Sustainable Well-being (Benessere Equo e Sostenibile, B.E.S.) confirms the good performance of the proposed estimation method.

For all these reasons this method could be suggested for future research applied on both social sciences and financial theories.

Acknowledgements

The work this paper represents was partially financed by the Department of Economics and Statistics - University of Naples “Federico II”.

IN LOVING MEMORY OF MY MOM

In memory of my mother who costantly encouraged and supported me in the fulfilment of my studies and research activities sharing this great passion with me.

References

[1] G. Agrò, Maximum likelihood estimation for the exponential power function parameters, Commun. Stat. Simul. Comput. 24 (2) (1995) 523–536.
[2] F.J. Anscombe, W.J. Glynn, Distribution of the kurtosis statistic β_2 for normal samples, Biometrika 70 (1) (1983) 227–234.
[3] A. Azzalini, The Skew-Normal and Related Families, 3, Cambridge University Press, 2013.
[4] A. Azzalini, A. Capitanio, Statistical applications of the multivariate skew normal distribution, J. R. Stat. Soc.: Series B (Stat. Methodol.) 61 (3) (1999) 579–602.
[5] G. Bottazzi, A. Secchi, A new class of asymmetric exponential power densities with applications to economics and finance, Ind. Corp. Change 20 (4) (2011) 991–1030.
[6] G.E.P. Box, G. Tiao, Bayesian Inference in Statistical Analysis, John Wiley & Sons, New York, 1973.
[7] R. Cerqueti, M. Giacalone, D. Panarello, A Generalized Error Distribution Copula-based method for portfolios risk assessment, Physica A: Stat. Mech. Appl. 524 (2019) 687–695.
[8] R. Cerqueti, M. Giacalone, R. Mattera, Skewed non-Gaussian GARCH models for cryptocurrencies volatility modelling, Inf. Sci. (2020), doi:10.1016/j.ins.2020.03.075.
[9] M. Chiodi, Generation of pseudo random variates from a normal distribution of order *p*, Italian J. Appl. Stat. 7 (4) (1995) 401–416.
[10] M. Chiodi, Le curve normali di ordine *p* nell’ambito delle distribuzioni di errori accidentali: una rassegna dei risultati e problemi aperti per il caso univariato e per quello multivariato, Atti della XL Riunione Scientifica della SIS (2000) 289–300.
[11] D. Coin, A goodness-of-fit test for generalized error distribution, Commun. Stat.-Theory Methods 46 (23) (2017) 11485–11499.
[12] DEF, Documento di Economia e Finanza, Ministero dell’Economia e delle Finanze, 2017.
[13] A. Dytso, R. Bustin, H.V. Poor, S. Shamai, Analytical properties of generalized Gaussian distributions, J. Stat. Distrib. Appl. 5 (1) (2018) 6.
[14] B.V. Frosini, A general quantile-based estimation method with application to the normal distribution of order p, Italian J. Appl. Stat. 21 (3-4) (2011) 317–336.

[15] R.C. Geary, Moments of the ratio of the mean deviation to the standard deviation for normal samples, Biometrika 28 (1936) 295–307.
[16] M. Giacalone, D. Panarello, R. Mattera, Multicollinearity in regression: an efficiency comparison between Lp-norm and least squares estimators, Q. Quant. 52 (4) (2018) 1831–1859.
[17] M. Giacalone, R. Mattera, E. Nissi, Economic indicators forecasting in presence of seasonal patterns: time series revision and prediction accuracy, Q. Quant. 54 (1) (2020) 67–84.
[18] R. Gonin, A.H. Money, Non-linear Lp-norm estimation: part I-on the choice of the exponent, *p*, where the errors are additive, Commun. Stat.-Theory Methods 14 (4) (1985) 827–840.
[19] Ed. R. Gonin, A.H. Money, A review of computational methods for solving the non–linear L_1-norm estimation problem, in: Y. Dodge (Ed.), Statistical Data Analysis Based on the L_1-Norm and Related Methods North Holland, 1987. Ed..
[20] R. Gonin, A.H. Money, Non-linear L_p-Norm Estimation, Marcel Dekker Inc., New York, 1989.
[21] N.L. Johnson, S. Kotz, N. Balakrishnan, Continuous Univariate Distributions, 2, 2nd ed., John Wiley, New York, 1995.
[22] M. Kendall, A. Stuart, The Advanced Theory of Statistics. Vol. 1: Distribution theory, 4th Ed., Griffin, London, 1977.
[23] S. Kotz, E. Seier, An analysis of quantile measures of kurtosis: center and tails, Stat. Pap. 50 (3) (2009) 553–568.
[24] R. Li, S. Nadarajah, The true maximum-likelihood estimators for the generalized Gaussian distribution with p= 3, 4, 5, Commun. Stat.-Theory Methods 46 (18) (2017) 8821–8835.
[25] N. Loperfido, A note on the fourth cumulant of a finite mixture distribution, J. Multivariate Anal. 123 (2014) 386–394.
[26] N. Loperfido, A new kurtosis matrix, with statistical applications, Linear Algebra Appl. 512 (2017) 1–17.
[27] J. Ma, Multiplicative algorithms for maximum penalized likelihood inversion with non-negative constraints and generalized error distributions, Commun. Stat.—Theory Methods 35 (5) (2006) 831–848.
[28] J.M. Marín, G. Sucarrat, Modelling the skewed exponential power distribution in finance, in: Mathematical and Statistical Methods for Actuarial Sciences and Finance, Springer, Milano, 2012, pp. 279–286.
[29] D. Marković, L. Borozan, On parameter estimation by nonlinear least squares in some special two-parameter exponential type models, Appl. Math. Inf. Sci. 9 (6) (2015) 2925.
[30] A. Mineo, The norm-p estimation of location, scale and simple linear regression parameters, in: Statistical Modelling, Springer, New York, NY, 1989, pp. 222–233.
[31] A.M. Mineo, M. Ruggieri, A software tool for the exponential power distribution: the normalp package, J. Stat. Softw. 12 (2005) 1–24.
[32] A.M. Mineo, On the estimation of the structure parameter of a normal distribution of order *p*, Statistica 63 (1) (2007) 109–122.
[33] A.H. Money, J.F. Affleck-Graves, M.L. Hart, G.D.I. Barr, The linear regression model: Lp norm estimation and the choice of p, Commun. Stat. Simul. Comput. 11 (1) (1982) 89–109.
[34] S. Nadarajah, A generalized normal distribution, J. Appl. Stat. 32 (7) (2005) 685–694.
[35] H. Nyquist, The optimal Lp norm estimator in linear regression models, Commun. Stat.-Theory Methods 12 (21) (1983) 2511–2524.
[36] D. Rosadi, P. Filzmoser, Robust second-order least-squares estimation for regression models with autoregressive errors, Stat. Pap. 60 (1) (2019) 105–122.
[37] C.S. Savva, P. Theodossiou, The risk and return conundrum explained: international evidence, J. Financ. Econometr. 16 (3) (2018) 486–521.
[38] V.A. Sposito, M.L. Hand, B. Skarpness, On the efficiency of using the sample kurtosis in selecting optimal Lp estimators, Commun. Stat. - Simul. Comput. 12 (3) (1983) 265–272.
[39] M.T. Subbotin, On the law of frequency of error, Математический сборник 31 (2) (1923) 296–301.
[40] P. Theodossiou, Skewed generalized error distribution of financial assets and option pricing, Multinatl. Financ. J. 19 (4) (2015) 223–266.
[41] Vasudevay, R., & Kumari, J. V. (2013). On general error distributions. In Prob-Stat Forum.
[42] A. Wilhelmsson, GARCH forecasting performance under different distribution assumptions, J. Forecast. 25 (8) (2006) 561–578.
[43] R. Zeckhauser, M. Thompson, Linear regression with non-normal error terms, Rev. Econ. Stat. (1970) 280–286.
[44] D. Zhu, V. Zinde-Walsh, Properties and estimation of asymmetric exponential power distribution, J. Econometr. 148 (1) (2009) 86–99.