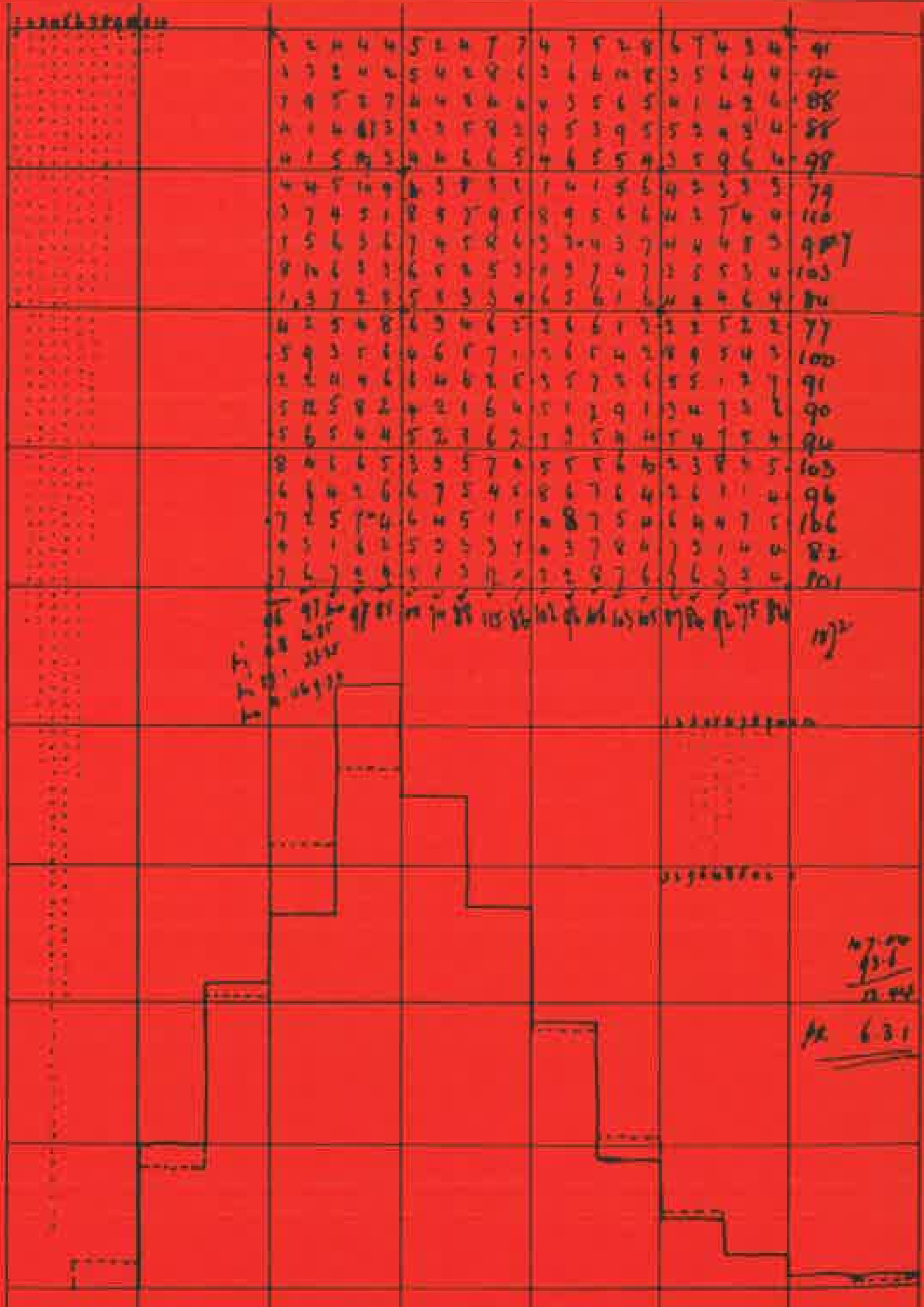
Student

Volume 2 Number 2

incorporating

Data & Statistics



November 1997

PAN

Presses Académiques Neuchâtel

The aim of ***Student*** is to present the statistical thoughts of graduate students to the community of researchers, teachers and practitioners of statistics. The journal welcomes the submission of articles in all areas of the theory, methodology and applications of statistics and probability, as well as papers on the history of statistics. Articles of an expository or tutorial nature are very much welcomed.

Submissions should be sent in two copies to the address below. It is assumed that the manuscripts have not previously been published and are not simultaneously being considered for publication elsewhere.

The language used must be either English or French. Manuscripts should be typed on one side of the paper in double spacing with wide margins. They should generally not exceed 20 pages including figures and tables. Papers must be accompanied by two summaries, in English and French respectively, each one at most 200 words, followed by about five key words.

In addition, if possible, a LaTeX input file should be sent by electronic mail to **student@seco.unine.ch**. If electronic mail is not feasible, a standard floppy disk may be submitted. Manuscripts submitted in the LaTeX format should use the "article" style and should not use any special macros. All papers are refereed.

Student is published twice a year. The annual subscription price of ***Student*** is SF 25.00 (postage included).

For advertising, subscription and general enquiries, please contact:

Groupe de Statistique
Pierre-à-Mazel 7
CH-2000 Neuchâtel
Switzerland
Phone: +41-32-718 13 80
Fax: +41-32-718 13 81

Edited by Statistics Group and published
by the Presses Académiques Neuchâtel

Cover: Page from Gosset's (Student) notebook containing the analysis of Haemacytometer counts (page 370 of "Studies in the History of Statistics and Probability" (1970) edited by E.S. Pearson and M.G. Kendall). Reproduced by permission of Hodder and Stoughton publishers Ltd, London, England.

fiu. 13

Student

Volume 2, Number 2 **November 1997**

Special Issue on Statistical Analysis Based on the L1-Norm and Related Methods

Data & Statistics

Student (ISSN 1420-1011) incorporating **Data & Statistics** (ISSN 1420-3308), Volume 2, Number 2, November 1997. Published bi-annually by Statistics Group, University of Neuchâtel, Pierre-à-Mazel 7, CH-2000 Neuchâtel, Switzerland.

Editorial

The Third International Conference on L_1-Norm and Related Methods was held in Neuchâtel, Switzerland, from August 11-15, 1997, as a Satellite Meeting to the 51st International Statistical Institute Session in Istanbul. Thirty seven of the ninety two papers presented at the Conference appeared in Volume 31 of the *Lectures Notes - Monograph Series* published by the Institute of Mathematical Statistics, Hayward, California.

The present issue of *Student* contains a selection of another thirteen refereed papers presented during the Conference. As *Student* is used to offer a variety of statistical materials, we thought it would be instructive to bring together here a collection of papers which would demonstrate the necessity of future research in different fields of statistics using as base other measures of distance than the traditional L_2-norm.

The Editor

Nearly all the participants of the Third L_1-Norm Conference, August 11, 1997.

Student (1997), Vol.2, No.2, 119-130

L_p-Norm Estimation for Nonlinear Regression Models

Massimiliano Giacalone

University of Naples "Federico II", Italy

Abstract: We examine the use of L_p-norm estimators in the framework of nonlinear regression models, assuming an Exponential Power Function as error distribution. In this work, we suggest a new criterion to jointly estimate the L_p-norm exponent p and the regression parameters. This approach is motivated by theoretical error distribution considerations. These distributions are elements of a class of density functions (E.P.F.), which are related to L_p-norm estimators. Finally, we present a simulation study that leads us to conclude that L_p-norm estimators are a suitable tool for studying nonlinear regression problems in the case of nonnormal symmetric error distributions.

Key words: Exponential Power Function, adaptive procedures, kurtosis indexes.

1 Introduction

This study deals with the construction of an adaptive estimation procedure for nonlinear regression models. In this context, the L_p-norm estimation methods are investigated for different values of p over a range of error distributions with varying kurtosis. A previous work showed the same procedures applied to the location model (Giacalone, 1996).

As it is well known, in order to obtain the L_p-norm estimator of the unknown regression parameter vector $\underline{\theta}$, we minimize the sum of the p-th power of the absolute deviations of the observed points from the regression function:

$$S_p(\underline{\theta}) = \sum_{i=1}^{n} |y_i - g(\underline{x}_i, \underline{\theta})|^p \qquad 1 < p < \infty \tag{1}$$

A rule for selecting the most appropriate value of p for any given error distribution is proposed, based on the Geary length of tails index and on the Pearson kurtosis of the error distribution. The relative efficiencies of the different examined estimators are discussed. In general, the true error distribution, and hence the kurtosis, is not known. In order to overcome this problem, the performances of the relationships between p and the two Kurtosis sample indexes as well as the related algorithm are examined. The new adaptive scheme suggested is then compared with that of several commonly used proposed alternatives. The E.P.F., here considered as underlying error distribution, is a family of density functions proposed by Subbotin (1923) and studied by Vianelli (1963), Lunetta (1963) and Mineo (1989). A brief introduction is given at this point. The density function is:

$$f_p(z) = \frac{1}{2p^{1/p}\sigma_p\Gamma\left(1+\frac{1}{p}\right)} \exp\left[-\frac{1}{p}\left|\frac{z-M_p}{\sigma_p}\right|^p\right] \quad (2)$$

with $\sigma_p > 0, \quad p > 0, \quad -\infty < z < +\infty$,
where $M_p = E(z)$ is the location parameter, $\sigma_p = (E[|z - M_p|^p])^{1/p}$ is the scale parameter, and p is the shape parameter. In fact, as p varies from 0 to ∞ the (2) assumes several shapes with different length of tails and kurtosis. Considering the Pearson kurtosis index β_2 we distinguish:

1. $0 < p < 1$: double exponential distributions, cuspidate, very long tailed and $\beta_2 > 6$;
2. $p = 1$: the Laplace distribution, cuspidate, long tailed with $\beta_2 = 6$;
3. $1 < p < 2$: leptokurtic distributions with long tails and $3 < \beta_2 < 6$;
4. $p = 2$: the Gaussian normal distribution with $\beta_2 = 3$;
5. $p > 2$: platikurtic distributions with short tails and $1.8 < \beta_2 < 3$;
6. $p \to \infty$: the uniform distribution with $\beta_2 \to 1.8$.

As previously stated, we assume that the error distribution of the regression model is a member of this class of symmetric functions. In this case, the optimal exponent p for the L_p-norm estimators of the regression parameters is the shape parameter p of the E.P.F. Our estimate of p is based on the two indexes, length of tails and kurtosis, strictly related to the shaped parameter and to the sample residuals.

2 L_p-norm nonlinear regression

Let us consider a sample of n observed data $(\underline{x}_i, y_i)$, where y_i is the dependent variable and $\underline{x}_i$ the independent nonrandom predictors. The general nonlinear regression model is:

$$y_i = g\left(\underline{x}_i, \underline{\theta}\right) + e_i \tag{3}$$

where g is a derivable function, $\underline{\theta} = (\theta_0, \theta_1, ..., \theta_k)$ is the unknown real parameter vector to be estimated, and the random errors e_i are independent and identically distributed variables according to E.P.F. scheme with mean zero and σ_p constant.

Under the supposed assumptions the loglikelihood related to the sample is given by:

$$\begin{aligned} L\left(\underline{\theta}, \sigma_p, p\right) &= -n \log\left[2p^{1/p}\sigma_p \Gamma\left(1 + \frac{1}{p}\right)\right] + \\ &\quad - \left[(p\sigma_p)^{-1} \sum |y_i - g\left(\underline{x}_i, \underline{\theta}\right)|^p\right] \end{aligned} \tag{4}$$

where we consider $z = y_i$ and $M_p = g(\underline{x}_i, \underline{\theta})$

When p is known it is easy to calculate the first partial derivatives with respect to θ to get the system of n nonlinear equations with n equations and $k+1$ variables:

$$\frac{\delta L}{\delta \theta_j} = \sum_{i=1}^{n} |y_i - g\left(\underline{x}_i, \underline{\theta}\right)|^{p-1} \text{sign } \left(y_i - g\left(\underline{x}_i, \underline{\theta}\right)\right) \frac{\delta g}{\delta \theta_j} = 0$$

The solution of this system gives us the maximum likelihood estimators of the regression parameters. The same equations 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:

$$\sum |y_i - g\left(\underline{x}_i, \underline{\theta}\right)|^p = \min \qquad \text{with } \; p \geq 1 \tag{5}$$

This result shows that the optimal exponent p is equal to the shape parameter of the E.P.F. assumed as underlying error distribution and is very useful in connecting the L_p-norm estimators to the E.P.F. theoretical scheme. When the value of p is unknown, we consider two related problems: the estimate of a suitable exponent p based on the sample data and the choice of the minimization algorithm used to obtain the regression parameters estimation. Even if the least squares, the least absolute deviations and the minimax estimators are particular cases of L_p-norm estimators (respectively with $p = 1$, $p = 2$, $p = \infty$), there is no theoretical reason for values of p other than 1, 2, ∞, not to be considered. We do not consider the case $p < 1$, that is the quasi-norm problem (Ekblom et al., 1969).

It should be noted that $p = 1$, $p = 2$, $p = \infty$ provide exact solutions, while the other values of p give rise to a nonlinear programming problem, whose solution can only be found in a given level of convergence. Different values of p have been proposed in the literature.

Forsythe (1972) suggested that $p = 1.5$ might be a good compromise value, as it provides estimates which are substantially better than the least squares when the error distribution has long tails, and is not bad when the errors have a Normal distribution. Harter (1977) proposed p with the following rule: if $\hat{\beta}_2 > 3.8$ use $p = 1$ (the regression of the least absolute values), if $2.2 < \hat{\beta}_2 < 3.8$ use $p = 2$ (the regression of the least squares) if $\hat{\beta}_2 < 2.2$ use $p = \infty$ (minimax or Chebychev regression), where $\hat{\beta}_2$ is the sample kurtosis.

Money et al. (1982) and Sposito et al. (1983) respectively obtained the following criteria for the choice of p by means of an extended simulation study:

$$\hat{p} = 9/\hat{\beta}_2^2 + 1 \qquad \text{for} \qquad 1 \le p < \infty \tag{6}$$

$$\hat{p} = 6/\hat{\beta}_2 \qquad \text{for} \qquad 1 \le p < 2 \tag{7}$$

Recently, the choice of a suitable value of p, for the linear regression case, was based on the relationships between some particular indexes of the E.P.F. and the sample residuals. This approach (Mineo, 1989) introduced the "generalized kurtosis" index to estimate p assuming the E.P.F as residual distribution. The likelihood estimation of p was considered by Agrò (1995).

3 The shape parameter, the Pearson kurtosis, and the length of tails.

For the density (2), we observe that the theoretical moment of order k is a function of the shape parameter p:

$$E\left|z - M_p\right|^k = \left(p\sigma_p^p\right)^{-k/p} \frac{\Gamma\left(\dfrac{k+1}{p}\right)}{\Gamma\left(\dfrac{1}{p}\right)} = \mu_k \tag{8}$$

This important relation shows that the ratios of the moments of order $2k$ and k^2 depend only on the shape parameter p.

From (8) we obtain the theoretic relation called "Generalized Kurtosis":

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

if $k = 2$ we can write the Pearson Kurtosis index:

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{\Gamma\left(\frac{1}{p}\right)\Gamma\left(\frac{5}{p}\right)}{\left[\Gamma\left(\frac{3}{p}\right)\right]^2} \tag{9}$$

If $k = 1$, considering the square root of the reciprocal, we have the Geary length of tails index:

$$I = \frac{\mu_1}{\sqrt{\mu_2}} = \frac{\Gamma\left(\frac{2}{p}\right)}{\sqrt{\Gamma\left(\frac{1}{p}\right)\Gamma\left(\frac{3}{p}\right)}} \tag{10}$$

The indexes I and β_2 show a different behavior according to the variation of p. For example if p increases from 1 to 4 the index β_2 decreases from 6 to 2.1884, while I increase from 0.7071 to 0.8409. (see Table 1).

Let us now consider the following sample kurtosis of the resulting residuals obtained by the ratios of empirical moments. An estimation of β_2 is given by:

$$\beta_2^* = \frac{n\sum_i(\varepsilon_i - \bar{\varepsilon})^4}{\left[\sum_i(\varepsilon_i - \bar{\varepsilon})^2\right]^2}$$

This estimation is strongly influenced in samples with many outliers because of the fourth moment, while the one based on I,

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

is particularly disturbed in samples with many values centered around the location parameter value. In Table 1 are shown the values of I and β_2 for p varying from 0.5 to 10 derived from equations (9) and (10). In the opposite way, calculating the sample values of I and β_2, it is possible to obtain, by inverse interpolation, two different estimations of p.

p	β_2	I	p	β_2	I	p	β_2	I
0.5	25.20000	.54772	3.7	2.24068	.83771	6.9	1.95463	.85587
0.6	15.57876	.59685	3.8	2.22208	.83879	7.0	1.95099	.85611
0.7	11.06208	.63441	3.9	2.20470	.83990	7.1	1.94748	.85634
0.8	8.56514	.66392	4.0	2.18844	.84090	7.2	1.94409	.85657
0.9	7.02556	.68766	4.1	2.17320	.84184	7.3	1.94082	.85678
1.0	6.00000	.70711	4.2	2.15889	.84273	7.4	1.93767	.85699
1.1	5.27660	.72330	4.3	2.14543	.84357	7.5	1.93462	.85719
1.2	4.74348	.73696	4.4	2.13276	.84436	7.6	1.93167	.85739
1.3	4.33681	.74861	4.5	2.12081	.84511	7.7	1.92883	.85758
1.4	4.01786	.75891	4.6	2.10952	.84583	7.8	1.92607	.85776
1.5	3.76195	.76738	4.7	2.09885	.84651	7.9	1.92341	.85794
1.6	3.55270	.77503	4.8	2.08875	.84715	8.0	1.92083	.85811
1.7	3.37882	.78178	4.9	2.07918	.84776	8.1	1.91833	.85827
1.8	3.23236	.78776	5.0	2.07010	.84834	8.2	1.91592	.85843
1.9	3.10751	.79310	5.1	2.05327	.84942	8.3	1.91357	.85859
2.0	3.00000	.79788	5.2	2.04547	.84993	8.4	1.91130	.85874
2.1	2.90460	.80206	5.3	2.03803	.85041	8.5	1.90910	.85889
2.2	2.82473	.80609	5.4	2.03095	.85087	8.6	1.90696	.85903
2.3	2.75252	.80963	5.5	2.02418	.85131	8.7	1.90489	.85917
2.4	2.68841	.81285	5.6	2.01773	.85173	8.8	1.90288	.85930
2.5	2.63116	.81580	5.7	2.01155	.85213	8.9	1.90093	.85943
2.6	2.57977	.81849	5.8	2.00565	.85252	9.0	1.89903	.85956
2.7	2.53342	.82097	5.9	2.00000	.85289	9.1	1.89719	.85968
2.8	2.49143	.82326	6.0	1.99459	.85324	9.2	1.89539	.85970
2.9	2.45325	.82537	6.1	1.98930	.85358	9.3	1.89452	.85981
3.0	2.41840	.82732	6.2	1.98442	.85391	9.4	1.89365	.85991
3.1	2.38648	.82914	6.3	1.97965	.85422	9.5	1.89196	.86002
3.2	2.35716	.83082	6.4	1.97506	.85452	9.6	1.89031	.86013
3.3	2.33015	.83239	6.5	1.97065	.85481	9.7	1.88871	.86024
3.4	2.30827	.83993	6.6	1.96641	.85500	9.8	1.88715	.86034
3.5	2.28286	.83522	6.7	1.96234	.85536	9.9	1.88564	.86044
3.6	2.26064	.83651	6.8	1.95841	.85562	10.0	1.88416	.86054

Table 1: Theoretical values of β_2 and I, evaluated for p varying from 0.5 to 10.

Gonin and Money (1987) considered the unbiased estimates of the second and fourth order sample moments with correction factors depending on the sample size n:

$$\hat{\mu}_2 = \frac{1}{n-1} \sum_i (\varepsilon_i - \bar{\varepsilon})^2$$

$$\hat{\mu}_4 = \frac{(n^2 - 2n + 3)}{(n-1)(n-2)(n-3)} \sum_i (\varepsilon_i - \bar{\varepsilon})^4 - \frac{3(n-1)(2n-3)}{n(n-2)(n-3)} \hat{\mu}_2^2$$

The same result is available using the r-th K-statistics and their relations with the sample moments (Lunetta, 1966, Kendall-Stuart, 1966).

The ratio of $\hat{\mu}_4$ and $\hat{\mu}_2^2$ gives an unbiased estimator of β_2:

$$\hat{\beta}_2 = \frac{\hat{\mu}_4}{\hat{\mu}_2^2} \tag{11}$$

For the I empirical index we can only consider the correction factor of the second sample moment, consequently we use:

$$\hat{I} = \frac{\sum_i |\varepsilon_i - \bar{\varepsilon}|}{\sqrt{\sum_i |\varepsilon_i - \bar{\varepsilon}|^2}} \frac{\sqrt{n-1}}{n} \tag{12}$$

4 The proposed algorithm

The proposed algorithm is based on a two steps alternating procedure that firstly estimates the $\underline{\theta}$ parameter vector by means of the classical conjugated gradient algorithm (Fletcher-Reeves, 1964) and secondly estimates p using a joint inverse function of I and β_2 obtained comparing empirical and theoretical moments. The algorithm stops when the variation of p is not significant. In order to obtain our estimate, we minimize the difference between empirical and theoretical indexes (13) to avoid some convergence problems encountered when we investigate a different algorithm considering this difference equal to zero (Giacalone, 1994).

The function used to estimate p is therefore the following:

$$\left[\left(I - \hat{I}\right) : 0.86054\right]^2 + \left[\left(\beta_2 - \hat{\beta}_2\right) : 25.2\right]^2 = \min \tag{13}$$

where I, $\hat{I}$, β_2, $\hat{\beta}_2$ are respectively given by (10), (12), (9), (11).

For simplicity we can express this condition as $\left[f\left(p\right)\right]^2 + \left[g\left(p\right)\right]^2 = \min$.

We observe that the indexes I and β_2 show different variability and different average order size related to the variation of p. For a joint evaluation we have to eliminate the difference in average order size, therefore setting $0 < f(p) < 1$ and $0 < g(p) < 1$.

The maximum theoretical values of $f(p)$ and $g(p)$ are the chosen standardization factors. Unfortunately the β_2 index diverges to ∞, for $p \rightarrow 0$. In order to standardize $g(p)$, we use the max β_2 value equal to 25.2 and corresponding to $p = 0.5$, which is the lower bound in our simulation study. In an analogous way, to standardize $f(p)$, we use the max I value equal to 0.86054 and related to $p = 10$, which is the upper bound in our simulation study.

The proposed algorithm is then specified in the following steps:

STEP 0: Set $i = 0$ and $p_0 = 2$;

STEP 1: Fit the model to the data using the previous step value p_i;

STEP 2: Compute the estimated residuals $\varepsilon_i = y_i - g(\underline{x}_i, \underline{\theta})$, their average $\overline{\epsilon}$, and insert these quantities in the (13) which is equal to the sum of two squared functions to minimize;

STEP 3: Minimize the function (13) to obtain p_{i+1}, new estimate of p;

STEP 4: Compare the estimated p_{i+1} with the previous p_i, and if $|p_{i+1} - p_i| > 0.01$ then set $i = i + 1$ and repeat Steps 1-4, otherwise:

STEP 5: Stop the algorithm assuming the values $\hat{\theta}_i = \theta_{ji}$, as L_p-norm estimators for the parameters θ_i' and the value $p = p_i$ as joint estimation of p.

In Step 1 a nonlinear L_p-norm estimation is considered. The problem could be resolved using the optimality conditions encountered in unconstrained optimization (McCormick, 1983). The minimization algorithm (Fletcher-Reeves, 1964) is used because it takes the special structure of the problem (1) directly into account. In Step 3 we use the parabolic interpolation method (Everitt, 1987) to find the minimum sum of squared functions (13). The convergence of the proposed algorithm was empirically verified by simulating samples of different sizes for fixed theoretical values of p.

5 Design of the simulation study and conclusive notes

The performance of the above method was tested using a simulation study. The unbiasedness and the asymptotic behavior of the new estimation procedure for nonlinear regression model parameters and for the shape parameter p was verified by a Montecarlo experiment (see Tables 2 and 4). We considered 500 samples of sizes $n = 50, 100, 200$, generated from E.P.F., for 6 different values of p, ranging from 1.0 to 3.5 with step 0.5. The algorithm used to generate the pseudo-random standardized deviates ε_i (for $p \geq 1$) was proposed by Chiodi (1986). The samples x_i $(i = 1, 2, \ldots n)$ were generated from a uniform distribution (0.5, 1.5). The values of $z_i = y_i$ are given by the simple exponential model:

$$y_i = \theta_2 e^{\theta_1 x_i} + \varepsilon_i \qquad i = 1, 2 \ldots n \tag{14}$$

We used model (14) with 500 samples estimates and $\theta_1' = 0.5$, $\theta_2' = 1.0$, $\sigma_p = 1.0$ as parameter values, to obtain the corresponding empirical frequency distributions and the relative analysed constants (mean, variance and mean squared error).

p	$M(\theta_1)$	$V(\theta_1)$	$M(\theta_2)$	$V(\theta_2)$	$M(\theta_1)$	$V(\theta_1)$	$M(\theta_2)$	$V(\theta_2)$
	$n = 50$				$n = 200$			
1.0	.4516	.1283	1.0671	.1645	.4975	.0249	1.0186	.0338
1.5	.4791	.1167	1.0426	.1534	.4974	.0226	1.0037	.0283
2.0	.4738	.0994	1.0548	.1266	.4915	.0178	1.0158	.0232
2.5	.4747	.0782	1.0418	.1013	.4918	.0145	1.0089	.0196
3.0	.4744	.0657	1.0579	.0785	.4927	.0116	1.0106	.0168
3.5	.5204	.0587	1.0409	.0692	.4998	.0093	1.0049	.0143

Table 2: Mean and variance of θ_1, θ_2, for a simple exponential model (14) ($\theta_1' = 0.5$, $\theta_2' = 1.0$, $\sigma_p = 1$) estimated with $L_{p_{\min}}$ method, on 500 samples of size $n = 50$, $n = 200$.

Here we present a comparative analysis using five different L_p-norm estimations:

Least squares estimators (L_2);

L_p-norm estimators with theoretical p of the E.P.F. (L_p);

L_p-norm estimators with p estimated as in (6) (Gonin and Money, 1989) ($L_{p_{gm}}$);

L_p-norm estimators using the maximum likelihood estimate of p (Agrò, 1995) (L_{p+}).

L_p-norm estimators with p estimated as in our proposal (13) ($L_{p_{\min}}$).

From the experimental results, partially reported in Tables 2 and 3, we can observe that for any p, the parameter estimates of θ_1 and θ_2 are biased for $n = 50$. Their variances decrease for increasing values of n. This is true for all the methods used 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 samples sizes (see when $n = 200$). Considering the relative efficiency (see Table 3) we note that all the L_p-norm estimators give us better parameter estimates with respect to the least squares method especially for theoretical values of p away from 2.

We can also observe the gain in efficiency using the L_p-norm estimators that is higher for the L_p method in all cases considered except when $p = 2$ where L_2 and L_p methods are equal. The $L_{p_{gm}}$, L_{p+}, $L_{p_{\min}}$ methods give us different efficiency parameter values depending on the criterion used to estimate p and could be considered half way between L_2 and L_p methods. The most important difference between the L_p-norm estimators examined are presented in Table 4 where we consider the empirical sample distributions of p estimates.

For all methods considered, the estimate of p is generally biased. This bias is higher for $p > 2$, and for $n = 50$. In an analogous way, the variance of the p estimates increases as p rises and decreases in proportion to the increase in sample size n. For the cases $p = 1.0$ and $p = 1.5$ the proposed method $L_{p_{\min}}$ seems to be the most efficient considering the variance and the mean squared error of the p estimate. The worst behavior of the $L_{p_{gm}}$ method depends on the bias of the p estimate that is reflected in the values of the mean squared error.

p	1.0		1.5		2.0		2.5		3.0		3.5	
	θ_1	θ_2	θ_1	θ_2	θ_1	θ_2	θ_1	θ_2	θ_1	θ_2	θ_1	θ_2
	$n = 50$											
L_p	2.25	1.88	1.33	1.25	1.00	1.00	1.11	1.24	1.45	1.29	1.82	1.39
$L_{p_{gm}}$	1.88	1.69	1.61	1.36	0.69	0.93	1.07	1.32	1.22	1.18	1.61	1.28
L_{p+}	1.76	1.66	1.41	1.27	0.68	0.95	0.81	1.11	1.08	1.03	1.65	1.29
$L_{p_{\min}}$	2.16	1.77	1.34	1.48	0.85	0.98	0.89	0.99	1.33	1.24	1.73	1.32
	$n = 200$											
L_p	1.88	1.65	1.45	1.19	1.00	1.00	1.19	1.16	1.25	1.17	1.48	1.50
$L_{p_{gm}}$	1.27	1.22	1.23	1.08	0.96	0.95	1.03	1.01	1.07	1.08	1.08	1.01
L_{p+}	1.85	1.54	1.09	1.07	0.98	0.95	1.02	1.05	1.13	1.11	1.28	1.26
$L_{p_{\min}}$	1.81	1.52	1.31	1.09	0.97	0.94	1.13	1.07	1.20	1.14	1.32	1.31

Table 3: Relative efficiency of L_p-norm estimators with respect to the least squares.

Case $p = 2$ proves to be the most efficient with $L_{p_{gm}}$ related to $L_{p_{\min}}$ and L_{p+} for the given sample sizes. Gonin and Money's method achieves, for all the examined theoretical values of p, estimates around the value $p = 2$. For the cases $p = 2.5$ $p = 3.0$ and $p = 3.5$ Gonin and Money's procedure gives a lower variance and mean squared error compared to the L_{p+} and $L_{p_{\min}}$ methods, but presents a very high bias for $n = 200$ sample size. The most important result is related to the asymptotic behavior of the p estimates. The methods L_{p+} and $L_{p_{\min}}$ seem to show a possible asymptotic convergence to a Normal distribution while the $L_{p_{gm}}$ method shows that the bias increases even when the sample size increases (see $p = 3.5$).

Looking at the simulation results, it seems reasonable to distinguish two cases. For $p < 2$, the $L_{p_{\min}}$ method here proposed gives us the best performance. The case $p > 2$ is well dealt using either the $L_{p_{\min}}$ or the L_{p+} methods. Agrò's method achieves the best results when the sample size is high as the maximum likelihood methods generally do.

Finally, our simulation study shows how our algorithm achieves good efficient estimates for regression and shape parameters when compared to the results obtained by the Gonin and Money's procedure. In particular, the asymptotic unbiasedness is a fundamental aspect for the proper function of every estimation procedure. In non-normal symmetric distribution, better performance of variances of p and parameter estimates are obtained when using the proposed $L_{p_{\min}}$ method.

p	Methods: $L_{p_{gm}}$, L_{p_+}, $L_{p_{\min}}$ 1.0	1.5	2.0	2.5	3.0	3.5
	$n = 50$					
M(p_{gm})	1.3349	1.8927	2.2161	2.5012	2.6938	2.8734
V(p_{gm})	0.1097	0.1546	0.2236	0.2479	0.2566	0.2637
MSE(p_{gm})	0.2219	0.3088	0.2703	0.2479	0.3504	0.6563
M(p_+)	1.4957	1.9985	2.5561	3.0392	3.5245	4.0822
V(p_+)	0.3515	0.4846	1.8152	2.1054	2.4296	3.1064
MSE(p_+)	0.5972	0.7331	2.1244	2.3961	2.7047	3.4453
M($p_{\min}$)	1.2374	1.7485	2.2656	2.7222	3.2114	3.4099
V($p_{\min}$)	0.1221	0.3577	0.6451	1.0566	1.4437	1.6687
MSE($p_{\min}$)	0.1784	0.4135	0.7156	1.1060	1.4884	1.6768
	$n = 200$					
M(p_{gm})	1.3307	1.7143	2.0596	2.3458	2.5733	2.7622
V(p_{gm})	0.0207	0.0446	0.0559	0.0589	0.0514	0.0585
MSE(p_{gm})	0.1301	0.0905	0.0594	0.0827	0.2335	0.6028
M(p_+)	1.2157	1.7386	2.0861	2.5182	2.9188	3.3094
V(p_+)	0.0495	0.0221	0.0624	0.1608	0.2747	0.3564
MSE(p_+)	0.0960	0.0790	0.0698	0.1611	0.2813	0.3927
M($p_{\min}$)	1.0784	1.5553	2.0641	2.5311	3.0423	3.5289
V($p_{\min}$)	0.0348	0.0476	0.0975	0.2741	0.4996	0.6135
MSE($p_{\min}$)	0.0411	0.0507	0.1016	0.2751	0.5014	0.6143

Table 4: Mean, variance and mean squares error of p estimated on samples of sizes $n = 50$, $n = 200$ generated on E.P.F., considering different methods to estimate p.

References

[1] Agrò, G. (1995). Maximum likelihood estimation for the exponential power function parameters. *Comm. Statist. Simul. Comput.* **24**, 523-536.

[2] Chiodi, M. (1986). Procedures for generating pseudo-random numbers from a normal distribution of order *p*. *Riv. Stat. Applic.* **19**, 7-26.

[3] Ekblom and Henriksson, S. (1969). L_p criteria for the estimation of location parameters. *SIAM J. Appl. Math.* **17**, 1130-1141.

[4] Everitt, B.S. (1987). *Introduction to Optimization Methods and their Application in Statistics.* Chapman and Hall, London.

[5] Fletcher, R. and Reeves, C.M. (1964). Function minimization by coniugate gradients. *Comput. J.* **7**, 149-154.

[6] Forsythe, A.B. (1972). Robust estimation of straight line regression coefficients by minimizing p-th power deviations. *Technometrics* **14**, 159-166.

[7] Giacalone, M. (1994). Stimatori di norma L_p. Applicazioni alla domanda di beni alimentari in Italia. In *Convegno: La Statistica per la Gestione Aziendale e l'Analisi di Mercato*, 141-157. Bressanone.

[8] Giacalone, M. (1996). Parameter evaluation of an exponential power function by simulation study. In *Shorts Communications and Posters, Compstat 96.* Barcelona.

[9] Gonin, R. and Money, A.H. (1987). A review of computational methods for solving the nonlinear L_1 norm estimation problem. In *Statistical Data Analysis Based on the L_1 Norm and Related Methods*, Ed. Y. Dodge. North Holland.

[10] Gonin, R. and Money A.H. (1989). *Nonlinear L_p-Norm Estimation.* Marcel Dekker, New York.

[11] Harter, H.L. (1977). The nonuniqueness of absolute values regression. *Commun. Statist. Simul. Comput.* **6**, 829-838.

[12] Kendall, M.C. and Stuart, A. (1966). *The Advanced Theory of Statistics.* Griffin, London.

[13] Lunetta, G. (1963). Di una Generalizzazione dello schema della curva normale. *Ann. Facoltà Econ. Comm.* **17**, 237-244. Palermo.

[14] Lunetta, G. (1966). Di alcune distribuzioni deducibili da una generalizzazione dello schema della curva normale. *Ann. Facoltà Econ. Comm.* **20**, 117-143. Palermo

[15] Mineo, A. (1989). The norm-p estimation of location, scale and simple linear regression parameters. In *Lectures Notes in Statistics-Statistical Modelling,* 222-223. Trento Springer-Verlag.

[16] Money, A.H. et al. (1982). The linear regression model: L_p-norm estimation and the choice of p. *Commun. Statist. Simul. Comput.* **11**, 89-109.

[17] Sposito, V.A. (1982). On unbiased L_p regression. *J. Amer. Statist. Assoc.* **77**, 652-653.

[18] Subbotin, M. (1923). On the law of frequency of error. *Mathematicheskii Sbornik.*

[19] Vianelli, S. (1963). La misura della variabilità condizionata in uno schema generale delle curve normali di frequenza. *Statistica* **23**.

Student

Volume 2, Number 2 **November 1997**

Special Issue on Statistical Analysis Based on the L1-Norm and Related Methods

Data & Statistics

A Statistical Journal for Graduate Students

ISSN 1420-1011
ISSN 1420-3308