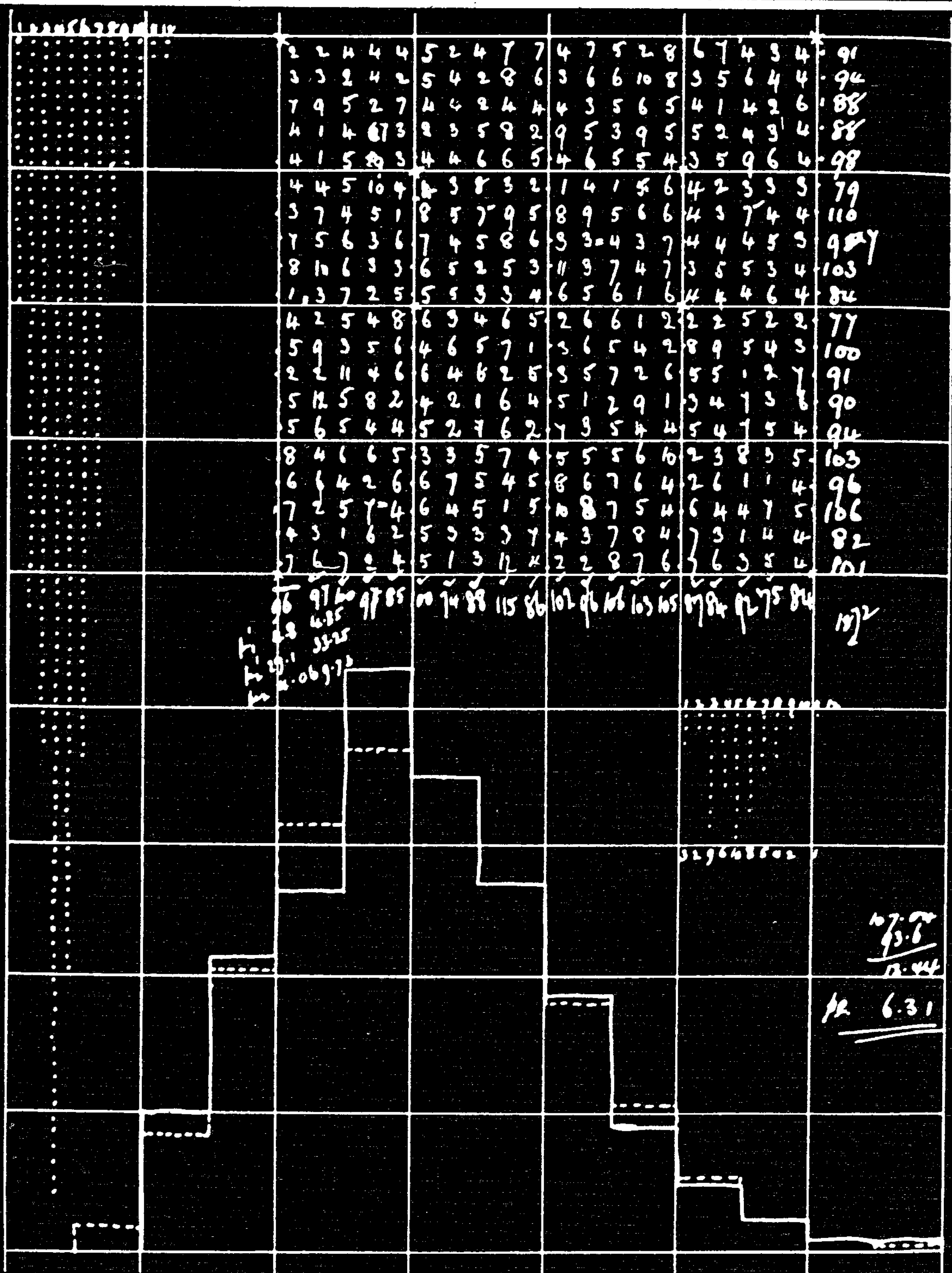
Student

Volume 5 Number 3/4

incorporating
Data & Statistics



September 2006

PAN
Presses Académiques Neuchâtel

Student

Volume 5, Number 3-4 **September 2006**

Student (ISSN 1420-1011) incorporating **Data & Statistics** (ISSN 1420-3308), Volume 5, Number 3-4, September 2006. Published bi-annually by the Institute of Statistics, University of Neuchâtel, Case postale 805, CH-2002 Neuchâtel, Switzerland.

Student (2006), Vol. 5 No. 3-4, 235-246

L_p-norm Estimation: Some Simulation Studies in Presence of Multicollinearity

Massimiliano Giacalone[1] and Rino Richiusa

[1] *University of Naples "Federico II", Naples, Italy*
Email: mgiacalo@unina.it

Abstract : In this paper we propose some simulation studies in order to compare the L_p-norm estimators with the Least Squares method (L_2), introducing a linear regression model in the case of multicollinearity. A comparative analysis is made applying three estimation methods to evaluate the empirical distribution of the regression parameters. Looking at the simulation results we note that improvement using L_p estimators instead of L_2 is more evident in the case of medium collinearity than in the case of strong one.

Key words: Multicollinearity, exponential power function, L_p norm estimators, kurtosis indexes.

1 Introduction

"For non-normal distributions of the ϵ_i's the least-squares estimates have minimum variance among unbiased estimates that are linear combinations of the Y_i's" (Cox and Hinkley, 1968). And in case of multicollinearity?

In this paper we consider the possibility of creating some adaptive robust estimation procedures for the standard linear regression model when the disturbance vector has deviated from normality.

We propose some simulation planes in order to compare the L_p-norm estimators with the least squares method in the case of multicollinearity, a common regression problem introduced in the second section of the paper.

After a brief review on the origins of the Exponential Power Function (E.P.F.), a useful family of symmetrical random error curves, we show their connection with the L_p-norm methods.

In the fourth section we investigate the relationships between the different values of p , the shape parameter of the E.P.F., and some kurtosis indexes that refer to the error distributions, then we analize and propose the $L_{p_{\min}}$ algorithm.

In the subsequent sections we compare three estimation procedures by introducing a linear regression model with collinear regressors. The aim of the simulation study is to choose the best rule for selecting the most appropriate value of p for any given error distribution and to evaluate the influence of multicollinearity in the parameter estimation procedures in terms of bias and variance. Finally, in the last paragraph, some results and comparations are discussed.

2 Multicollinearity in regression

A regression model has two main purposes: 1) it could be used to find out to what extent the outcome (dependent variable) can be predicted by the independent variables and 2) it could be used to determine the strength of a theoretical relationship between the dependent variable and the independent variables.

A regression problem can be affected by multi-collinearity. The absence of multi-collinearity is essential to obtain optimal estimates by a multiple regression model. In regression the multi-collinearity or collinearity problem occours when several regressors are highly correlated. When things are related, we say they are linearly dependent on each other because one can nicely fit a straight regression line to pass through many of the data points of those variables. Collinearity simply means co-dependence. It would have to be eliminated. Doing it is problematic when one's purpose is explanation rather than mere prediction. Indeed when prediction is the goal no problem arizes if, among dependent variables, two regressors have the same 'meaning': simply it is possible eliminating one. On the contrary, when explanation is the goal regressors have been selected according a theoretical rule, so that it is possible eliminate no variable because of the model is theoretical, that is given; it has to be explained exactly, including all selected variables.

Therefore the multi-collinearity problem can and must be eliminated in the case of prediction while is often uneliminable in the case of explanation. In this latter case eliminating a variable leads to a problem of bad-specification of the model. A viable remedy for bad-specification is to prefer a different estimator than the least squares one. This is because the 'classical' Ordinary Least Squares (OLS) method provides estimates which could be statistically not significant in presence of multi-collinearity. It would be better to use alternative methods such as the Ridge Regression (Morris, 1982; Pagel and Lunneberg, 1985) or the Partial Least Squares (Cassel et al., 1999). Here we are going to carry out some simulation studies in order to show some evident differences among three estimation methods (Least Squares estimators and two different L_p-norm estimators) in presence of multi-collinearity.

3 The Exponential Power Function and the L_p-norm estimators

The E.P.F is a family of probability functions proposed by Subbotin in 1923 and studied by Vianelli (1963), Lunetta (1963), Mineo (1989). The density function is:

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

where $M_p = E(z)$ is the *location parameter*, $\sigma_p = \left(E\left[|z-M_p|\right]\right)^{1/p}$ is the *scale parameter*, and $p > 0$ is the *shape parameter*.
Considering the Pearson kurtosis index β_2 we distinguish:

- $0 < p < 1$: double exponential distributions, $\beta_2 > 6$;
- $1 < p < 2$: leptokurtic distributions, $3 < \beta_2 < 6$;
- $p > 2$: platikurtic distributions, $1.8 < \beta_2 < 3$;

For particular values of p we have: the *Laplace distribution* ($p = 1$, $\beta_2 = 6$); the *Gaussian distribution*, ($p = 2$, $\beta_2 = 3$); and the *Uniform distribution* ($p \to \infty$, $\beta_2 \to 1.8$).
Let us consider a sample of n observed data (y_i, x_i), a general linear regression model is:

$$y_i = g\left(x_i, \theta\right) + \epsilon_i, \tag{2}$$

with g a linear function.

The L_p-norm estimators are a mere generalization of the Least Squares replacing exponent 2 by a general exponent p. Therefore they minimize the sum of the p-th power of the absolute deviations of the observed points from the regression function:

$$S_p(\theta) = \sum_{i=1}^{n} |y_i - g(x_i, \theta)|^p \qquad 1 \le p < \infty \tag{3}$$

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

$$l(\theta, \sigma_p, p) = -n \log\left[2p^{1/p}\sigma_p\Gamma(1+1/p)\right] - \left[(p\sigma_p)^{-1}\sum |y_i - g(x_i,\theta)|^p\right] \tag{4}$$

where we consider $z = y_i$ and $\mu_p = g(x_i, \theta)$

$$\delta l/\delta\theta_j = \sum_{i=1}^{n} |y_i - g(x_i,\theta)|^{p-1}\text{sign }\left(y_i - g(x_i,\theta)\right)\frac{\delta g}{\delta\theta_j}$$

$$\sum_{i=1}^{n} |y_i - g(x_i,\theta)|^p = \min \qquad \text{with } p \ge 1 \tag{5}$$

The optimal exponent p for the L_p-norm estimators of the regression parameters is the shape parameter p of the E.P.F.
If it is unknown we have two related problems to consider:
(1) The estimation of the *exponent* p on the sample data.

(2) The choice of the minimization algorithm to obtain the *regression parameters* estimation.

To estimate the exponent p it is possible to find out the following proposals:

- Harter (1977), noting that p depends on $\hat{\beta}_2$ (the sample residual Kurtosis), proposed selecting p with the following rule:
  if $\hat{\beta}_2 > 3.8$ use $p = 1$ (the *least absolute deviations* regression).
  if $2.2 < \hat{\beta}_2 < 3.8$ use $p = 2$ (the *least squares* regression).
  if $\hat{\beta}_2 < 2.2$ use $p = \infty$ (the *minimax* or *Chebychev* regression).

- Money et al. (1982) and Sposito et al. (1983) proposed two different criteria respectively:

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

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

- Mineo A. (1989) proposed the Generalized Kurtosis β_K, as described below.

- Mineo A.M. (1994) considered a new method to estimate p, based on an empirical index called VI.

- Agrò (1995) proposed a maximum likelihood estimation either for the regression parameters or for the p shape parameter.

4 The Exponential Power Function kurtosis indexes

For the density (1), the theoretical moment of order k is a function of the shape parameter p as follows:

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

The *ratios of the moments of order* $2k$ *and the squared moment of order* k only depend on the shape parameter p. This theoretic relation is also called "*Generalized Kurtosis*" (Mineo, 1989) :

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

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

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

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

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

The indexes I and β_2 show a different behaviour according to the variation of p (Giacalone, 1997).
Calculating the sample values of I and β_2 , it is possible to obtain, by inverse interpolation, two different estimations of p.
Gonin and Money (1987), Lunetta, (1966), Kendall-Stuart, (1966) considered the unbiased estimates of the second and fourth order sample moments with correction factors depending on the sample size n :

$$\hat{\mu_2} = \frac{1}{n-1} \sum_i (\epsilon_i - \bar{\epsilon})^2$$

$$\hat{\mu_4} = \frac{(n^2 - 2n + 3)}{(n-1)(n-2)(n-3)} \sum_i (\epsilon_i - \bar{\epsilon})^2 - \frac{3(n-1)(2n-3)}{n(n-2)(n-3)} \hat{\mu}_2^2$$

The ratio of $\hat{\mu_4}$ and $\hat{\mu_2}$ gives the following estimator of β_2:

$$\hat{\beta}_2 = \frac{\hat{\mu_4}}{\hat{\mu_2}^2} \tag{11}$$

For the I empirical index we obtain:

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

5 The $L_{p_{\min}}$ algorithm

It is based on a two-steps alternating procedure : a) minimization procedure to estimate the parameters, b) joint inverse function of I and β_2 to estimate p.
The algorithm is stopped when p does not vary significantly.
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 express the (13) as $[f(p)]^2 + [g(p)]^2 = \min$ to take into account the different variability and average order size related to p , and the standardization factors are the maximum theoretical values.
So using the relation (9) we calculate $\max(\beta_2) = 25.2$ for $p = 0.5$, lower bound in our simulation plan, whilst using the relation (10) we calculate $\max(I) = 0,86054$, for $p = 10$, upper bound in our simulation plan.
The proposed algorithm (Giacalone, 1997) 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 $\epsilon_i = y_i - g(x_i, \theta)$, their average $\bar{\epsilon}$ and insert these quantities in the (13) which is equal to the sum of the two squared

functions to be minimized;

- **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 $\hat{\theta}_i = \theta_{ij}$ 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 (Fletcher and Reeves, 1964). In step 3 a parabolic interpolation method (Everitt, 1987) to find the minimum of the sum of squared functions (13) is adopted.

6 The simulation plan

We consider 500 samples of sizes n = 50, 100, generated from E.P.F., and 6 values of p, ranging from 1.0 to 3.5 with step 0.5. The algorithm for generating the ϵ_i (for $p \geq 1$) from an E.P.F. is suggested by Chiodi (1986).
The values of y_i are given by the multiple regression model:

$$y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \epsilon_i \tag{14}$$

with X_1, X_2 identically distributed and independent variables from a Gaussian standardized distribution and X_3 a linear combination of X_1 and X_2.

$$X_3 = X_1 + X_2 + Z \text{ with } Z \sim N(0, \sigma_z) \tag{15}$$

Therefore we can write the related variance and covariance matrix: It is easy to see

$$\left\| \begin{array}{cccc} & X_1 & X_2 & X_3 \\ X_1 & 1 & 0 & 1 \\ X_2 & 0 & 1 & 1 \\ X_3 & 1 & 1 & 2+\sigma_z^2 \end{array} \right\|$$

that: $E(X_3^2) = E(X_1^2) + E(X_2^2) + \sigma_z^2 = 2 + \sigma_z^2$
and the correspondent correlation matrix is equal to:

$$r = \left\| \begin{array}{ccc} 1 & 0 & 1/\sqrt{2+(\sigma_z^2)} \\ 0 & 1 & 1/\sqrt{2+(\sigma_z^2)} \\ 1/\sqrt{2+(\sigma_z^2)} & 1/\sqrt{2+(\sigma_z^2)} & 1 \end{array} \right\|$$

where $r_{13} = \text{cov }(X_1, X_3)/\sqrt{\text{var }(X_1)\text{var }(X_3)} = 1/\sqrt{1(2+\sigma_z^2)} = r_{23}$
and $R^2_{3.12} = 1 - \det A/\det A_{33} = 2/(2+\sigma_z^2)$
In the simulation model one can see that the rate of multicollinearity is (inversely)

proportional to σ_z^2.
Let be the parameter values $\beta_0 = 1$, $\beta_1 = 2$, $\beta_2 = 3$, $\beta_3 = 4$, a comparative analysis is made applying the following three estimation methods on the same samples, with the results reported in the Tab 2-7 :

(1) Least squares estimators (L_2).

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

(3) L_p-norm estimators with p as in our proposal (Section 5) ($L_{p_{\min}}$).

One can see that for any p and for any method, the parameter estimates of β_0, β_1, β_2 , β_3, are biased when $n = 50$ but their variances decrease for increasing values of n. That depends on the multicollinearity of the model. It is possible noting the unbiasedness of the estimates only for middle-large samples sizes.
The L_p-norm estimators give us better parameter estimates for the parameters β_1 and β_2 compared to the least squares method especially for p far from 2 (see tab 7 and tab 8).
The gain in efficiency using the L_p-norm estimators is higher for the L_p method in all the cases considered except for the case $p = 2$ where the error is generated by a Gaussian distribution.
The $L_{p_{\min}}$ method could be considered as half-way between the L_2 and L_p methods because we estimate the exponent p on the sample data.

CASE 1 $\sigma_z = 1$ $R^2_{3.12} = 0.66$

p	$M(\beta_0)$	$V(\beta_0)$	$M(\beta_1)$	$V(\beta_1)$	$M(\beta_2)$	$V(\beta_2)$	$M(\beta_3)$	$V(\beta_3)$
				$n = 50$				
1.0	1.162	.928	2.117	.964	3.124	.849	3.748	1.334
1.5	0.973	.816	1.877	.834	2.889	.826	4.042	1.083
2.0	1.029	.799	1.956	.786	2.835	.778	3.987	1.037
2.5	1.037	.678	1.934	.713	2.903	.645	4.059	0.968
3.0	0.956	.665	1.873	.685	2.918	.616	3.764	0.864
3.5	1.235	.558	2.097	.609	3.221	.593	3.952	0.798
				$n = 100$				
1.0	1.043	.128	1.957	.164	2.934	.024	3.838	.033
1.5	1.034	.116	1.942	.153	2.919	.022	4.039	.028
2.0	1.018	.099	2.001	.126	2.965	.017	4.022	.023
2.5	1.025	.078	2.034	.101	3.015	.014	3.959	.019
3.0	1.028	.065	1.977	.078	2.953	.011	3.853	.016
3.5	1.039	.058	1.963	.069	2.941	.009	3.978	.014

Table 1: Mean and Variance of β_0,β_1,β_2,β_3 for a multiple regression model ($\beta_0 = 1$,$\beta_1 = 2$,$\beta_2 = 3$,$\beta_3 = 4$) estimated with L_p-method on 500 samples of size $n = 50$, $n = 100$.

CASE 2 $\sigma_z = 0.01$ $R^2_{3.12} = 0.99$

p	$M(\beta_0)$	$V(\beta_0)$	$M(\beta_1)$	$V(\beta_1)$	$M(\beta_2)$	$V(\beta_2)$	$M(\beta_3)$	$V(\beta_3)$
$n = 50$								
1.0	1.276	1.281	2.202	1.323	3.135	1.481	3.801	1.634
1.5	1.053	1.016	1.856	1.053	2.859	1.213	4.052	1.466
2.0	1.049	.949	1.942	1.002	2.845	1.187	3.963	1.255
2.5	1.047	.878	1.921	.938	2.883	1.045	4.072	1.109
3.0	0.934	.723	2.143	.801	3.108	.912	3.792	1.021
3.5	1.262	.622	2.104	.712	3.221	.796	3.938	0.934
$n = 100$								
1.0	1.067	.625	1.944	.744	2.944	.834	3.855	.941
1.5	1.045	.514	1.931	.633	2.931	.711	4.043	.844
2.0	1.022	.397	2.005	.503	2.922	.566	4.051	.654
2.5	1.037	.255	2.051	.401	3.029	.533	4.065	.592
3.0	1.045	.155	1.983	.277	2.934	.345	3.839	.416
3.5	1.061	.103	1.942	.188	2.928	.233	4.044	.307

Table 2: Mean and Variance of β_0,β_1,β_2,β_3 for a multiple regression model ($\beta_0 = 1$,$\beta_1 = 2$,$\beta_2 = 3$,$\beta_3 = 4$) estimated with L_p-method on 500 samples of size $n = 50$, $n = 100$.

CASE 3 $\sigma_z = 1$ $R^2_{3.12} = 0.66$

p	$M(\beta_0)$	$V(\beta_0)$	$M(\beta_1)$	$V(\beta_1)$	$M(\beta_2)$	$V(\beta_2)$	$M(\beta_3)$	$V(\beta_3)$
$n = 50$								
1.0	1.143	.865	2.111	.883	3.133	.821	3.802	1.277
1.5	0.981	.802	1.854	.799	2.898	.766	4.032	1.055
2.0	1.033	.745	1.949	.746	2.866	.722	3.985	0.965
2.5	0.995	.655	1.912	.695	2.915	.601	4.044	0.929
3.0	0.967	.624	1.892	.633	2.932	.587	3.811	0.833
3.5	1.195	.507	2.075	.576	3.198	.544	3.898	0.776
$n = 100$								
1.0	1.028	.101	1.968	.134	2.955	.021	3.854	.031
1.5	1.025	.099	1.955	.124	2.901	.019	3.998	.026
2.0	1.013	.093	1.998	.119	2.977	.015	4.018	.021
2.5	1.019	.075	2.029	.098	3.011	.012	3.966	.017
3.0	1.024	.062	1.975	.074	2.951	.009	3.866	.013
3.5	1.033	.056	1.961	.067	2.939	.007	3.975	.011

Table 3: Mean and Variance of β_0,β_1,β_2,β_3 for a multiple regression model ($\beta_0 = 1$,$\beta_1 = 2$,$\beta_2 = 3$,$\beta_3 = 4$) estimated with L_p-method on 500 samples of size $n = 50$, $n = 100$.

CASE 4 $\sigma_z = 0.01$ $R^2_{3.12} = 0.99$

p	$M(\beta_0)$	$V(\beta_0)$	$M(\beta_1)$	$V(\beta_1)$	$M(\beta_2)$	$V(\beta_2)$	$M(\beta_3)$	$V(\beta_3)$
				$n = 50$				
1.0	1.164	.921	2.122	.945	3.119	.841	3.744	1.329
1.5	0.969	.820	1.875	.831	2.878	.822	4.039	1.077
2.0	1.025	.794	1.958	.783	2.832	.769	3.978	1.032
2.5	1.035	.677	1.936	.709	2.915	.643	4.044	0.958
3.0	0.944	.655	1.876	.679	2.921	.620	3.788	0.871
3.5	1.244	.566	2.098	.611	3.219	.588	3.961	0.787
				$n = 100$				
1.0	1.045	.122	1.944	.161	2.941	.023	3.855	.032
1.5	1.031	.115	1.941	.149	2.933	.023	4.041	.027
2.0	1.009	.095	1.999	.121	2.971	.016	4.019	.021
2.5	1.024	.076	2.031	.099	3.018	.013	3.962	.018
3.0	1.025	.062	1.981	.074	2.951	.011	3.844	.015
3.5	1.034	.056	1.973	.067	2.945	.008	3.981	.013

Table 4: Mean and Variance of β_0,β_1,β_2,β_3 for a multiple regression model ($\beta_0 = 1$,$\beta_1 = 2$,$\beta_2 = 3$,$\beta_3 = 4$) estimated with L_p-method on 500 samples of size $n = 50$, $n = 100$.

CASE 5 $\sigma_z = 1$ $R^2_{3.12} = 0.66$

p	$M(\beta_0)$	$V(\beta_0)$	$M(\beta_1)$	$V(\beta_1)$	$M(\beta_2)$	$V(\beta_2)$	$M(\beta_3)$	$V(\beta_3)$
				$n = 50$				
1.0	1.155	1.098	2.122	1.088	3.136	1.023	3.698	1.529
1.5	0.955	.905	1.855	.923	2.901	.958	4.051	1.208
2.0	1.033	.745	1.949	.746	2.866	.722	3.985	0.965
2.5	1.031	.701	1.921	.751	2.922	.688	4.123	0.955
3.0	0.942	.688	1.855	.713	2.944	.651	3.728	0.888
3.5	1.303	.599	2.102	.654	3.234	.611	3.911	0.833
				$n = 100$				
1.0	1.041	.256	1.932	.231	2.569	.058	3.855	.053
1.5	1.031	.178	1.929	.189	2.933	.039	4.055	.037
2.0	1.013	.093	1.998	.119	2.977	.015	4.018	.021
2.5	1.033	.069	2.043	.109	3.022	.013	3.961	.020
3.0	1.033	.066	1.966	.081	2.966	.012	3.899	.017
3.5	1.044	.061	1.944	.072	2.955	.011	3.985	.015

Table 5: Mean and Variance of β_0,β_1,β_2,β_3 for a multiple regression model ($\beta_0 = 1$,$\beta_1 = 2$,$\beta_2 = 3$,$\beta_3 = 4$) estimated with L_p-method on 500 samples of size $n = 50$, $n = 100$.

CASE 6 $\sigma_z = 0.01$ $R^2_{3.12} = 0.99$

p	$M(\beta_0)$	$V(\beta_0)$	$M(\beta_1)$	$V(\beta_1)$	$M(\beta_2)$	$V(\beta_2)$	$M(\beta_3)$	$V(\beta_3)$
$n = 50$								
1.0	1.164	.921	2.121	.967	3.134	.855	3.337	1.345
1.5	0.988	.820	1.888	.836	2.885	.829	4.041	1.091
2.0	1.025	.794	1.958	.783	2.832	.769	3.978	1.032
2.5	1.039	.733	1.942	.737	2.921	.691	4.077	0.991
3.0	0.966	.703	1.887	.701	2.933	.612	3.755	0.903
3.5	1.246	.588	2.115	.651	3.233	.609	3.945	0.847
$n = 100$								
1.0	1.044	.145	1.966	.188	2.945	.029	3.844	.038
1.5	1.036	.121	1.951	.169	2.901	.024	4.045	.032
2.0	1.009	.095	1.999	.121	2.971	.016	4.019	.021
2.5	1.022	.082	2.039	.105	3.021	.015	3.933	.020
3.0	1.027	.073	1.975	.083	2.977	.013	3.799	.017
3.5	1.033	.062	1.961	.071	2.933	.011	3.965	.015

Table 6: Mean and Variance of β_0,β_1,β_2,β_3 for a multiple regression model ($\beta_0 = 1$,$\beta_1 = 2$,$\beta_2 = 3$,$\beta_3 = 4$) estimated with L_p-method on 500 samples of size $n = 50$, $n = 100$.

CASE 1,3,5 $\sigma_z = 1$ $R^2_{3.12} = 0.66$

p	1.0		1.5		2.0		2.5		3.0		3.5	
	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2
$n = 50$												
L_p	2.2	1.8	1.3	1.2	1.0	1.0	1.1	1.2	1.4	1.2	1.8	1.3
$L_{p_{\min}}$	2.1	1.7	1.3	1.4	0.8	0.9	0.8	0.9	1.3	1.2	1.7	1.3
$n = 100$												
L_p	1.8	1.6	1.4	1.1	1.0	1.0	1.1	1.1	1.2	1.1	1.4	1.5
$L_{p_{\min}}$	1.8	1.5	1.3	1.0	0.9	0.9	1.1	1.0	1.2	1.1	1.3	1.3

Table 7: Relative efficiency of L_p-norm estimators compared to the least squares (parameter β_1, β_2).

CASE 2,4,6 $\sigma_z = 0.01$ $R^2_{3.12} = 0.99$

p	1.0		1.5		2.0		2.5		3.0		3.5	
	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2
$n = 50$												
L_p	1.2	1.1	1.2	1.2	1.0	1.0	1.1	1.2	1.3	1.2	1.5	1.2
$L_{p_{\min}}$	1.3	1.2	1.3	1.3	0.9	0.8	0.9	1.1	1.4	1.2	1.4	1.2
$n = 100$												
L_p	1.2	1.1	1.1	1.1	1.0	1.0	1.2	1.1	1.2	1.1	1.3	1.1
$L_{p_{\min}}$	1.1	1.1	1.2	1.2	1.1	0.9	1.0	1.0	1.2	1.1	1.3	1.2

Table 8: Relative efficiency of L_p-norm estimators with respect to the least squares (parameter β_1, β_2).

7 Conclusions

In this ending section we underline the objectives reached with our simulation plan. In this study the L_p-norm methods are considered not in order to reduce the multicollinearity in the model, but with the aim of seeing the improvements in the parameter estimation.
Looking at the simulation results we note that the improvement using L_p estimators instead of least squares is more evident in the case $\sigma_z = 1$ $R^2_{3.12} = 0,66$ (medium multicollinearity) than in the case $\sigma_z = 0,01$ $R^2_{3.12} = 0,99$ (strong multicollinearity).
That is due to the characteristics of L_p methods that are adaptive procedures related to the erratic component of the model and not related to the deterministic one.
Finally, the simulation studies show how our algorithm ($L_{p_{\min}}$) achieves more efficient estimates for the regression parameters when compared with the least squares procedure.
In particular, using the L_p and the $L_{p_{\min}}$ method , a better performance, in terms of the variances of the parameter estimates, is always obtained in the case of nonnormal symmetric distributions compared the least squares situation also considering a model with collinear regressors.

Acknowledgments

This paper was partially financed by the P.R.I.N 2000 Project *Data Mining and Simbolical Data Analysis* on M.U.R.S.T. funds; National Coordinator: Prof. N. Carlo Lauro, University of Naples "Federico II".

References

[1] Agrò G. (1995): "Maximum likelihood estimation for the exponential power function parameters". *Comm. Statist - Simula. Computa.,24, pp. 523,536.*

[2] Cassel, C.; Westlund, A. H. and Hackl, P. (1999): "Robustness of partial least-

squares method for estimating latent variable quality structures". *Journal of Applied Statistics*, 26, 435-448.

[3] Chiodi M. (1986): "Procedures for Generating Pseudo-Random Numbers from a Normal Distribution of Order p" *Riv. Stat. Applicata*, 19, pp. 7-26.

[4] Cox D.R., Hinkley D.V. (1968): "A note on the Efficiency of Least Squares Estimates". *Journal of Royal Statistical Society*, B; 30, 841.

[5] Everitt B. and Hand D. (1987): "The Statistical Consultant in Action", New York: Cambridge University Press.

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

[7] Giacalone M. (1997): "L_p-norm estimators for nonlinear regression models" *Student, vol 2, pp 119-130.*

[8] Gonin R. and Money A.H. (1987): "A Review of Computational Methods for Solving the Nonlinear L1 Norm Estimation Problem". *S.D.A.B. on the L1 Norm and Related Methods.* Dodge ed al., North Holland.

[9] Harter H.L. (1977): "The nonuniqueness of absolute values regression". *Commun. Statist. Simula. Computa.* A6, pp. 829-838.

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

[11] Lunetta G. (1963): "Di una Generalizzazione dello Schema della Curva Normale" *Ann. Facoltà Econ. Comm.* Palermo, 17, pp. 237-244.

[12] Lunetta G. (1966): "Di Alcune Distribuzioni Deducibili da una Generalizzazione dello Schema della Curva Normale". *Ann. Facoltà Econ. Comm.* Palermo, 20, pp. 117-143.

[13] Mineo A. (1989): "The Norm-p Estimation of Location, Scale and Simple Linear Regression Parameters". *Lectures Notes in Statistics-Statistical Modelling* Trento Pringer-Verlag, pp. 222-223.

[14] Mineo A.M. (1994): "Un nuovo metodo di stima di p per una corretta valutazione dei parametri di intensità e di scala di una curva normale di ordine p". *Atti della XXXVII Riunione Scientifica della SIS*, vol.2, San Remo pp. 147-154.

[15] Money A.H. et al. (1982): "The Linear Regression Model: L_p-Norm Estimation and the Choice of p". *Commun. Statist. - Simula. Computa*, 11, pp. 89-109.

[16] Morris, J.D. (1982): "Ridge regression and some alternative weighting techniques: A comment on Darlington." *Psychological Bulletin*, 91, 203-210.

[17] Pagel, M.D. and Lunneberg, C.E. (1985): "Empirical evaluation of ridge regression". *Psychological Bulletin*, 97,342-355.

[18] Sposito V.A. (1983): "On Unbiased L_p Regression". *J.A.S.A.*, 77.

[19] Subbotin M. (1923): "On the Law of Frequency of Error". *Mathematicheskii Sbornik.*

[20] Vianelli S. (1963): "La Misura della Variabilità Condizionata in uno Schema Generale delle Curve Normali di Frequenza". *Statistica*, XXIII n. 4.