

# Multicollinearity in regression: an efficiency comparison between $L_p$ -norm and least squares estimators

Massimiliano Giacalone<sup>1</sup> · Demetrio Panarello<sup>2</sup> · Raffaele Mattera<sup>1</sup>

Published online: 12 September 2017  
© Springer Science+Business Media B.V. 2017

**Abstract** Multicollinearity is one of the most important issues in regression analysis, as it produces unstable coefficients' estimates and makes the standard errors severely inflated. The regression theory is based on specific assumptions concerning the set of error random variables. In particular, when errors are uncorrelated and have a constant variance, the ordinary least squares estimator produces the best estimates among all linear estimators. If, as often happens in reality, these assumptions are not met, other methods might give more efficient estimates and their use is therefore recommendable. In this paper, after reviewing and briefly describing the salient features of the methods, proposed in the literature, to determine and address the multicollinearity problem, we introduce the  $L_{pmin}$  method, based on  $L_p$ -norm estimation, an adaptive robust procedure that is used when the residual distribution has deviated from normality. The major advantage of this approach is that it produces more efficient estimates of the model parameters, for different degrees of multicollinearity, than those generated by the ordinary least squares method. A simulation study and a real-data application are also presented, in order to show the better results provided by the  $L_{pmin}$  method in the presence of multicollinearity.

**Keywords** Multicollinearity · Exponential power function ·  $L_p$ -norm estimators ·  $L_{pmin}$  method · Kurtosis indexes · Relative efficiency

## 1 Introduction

Generally, a regression model has two main purposes (Saville and Wood 1991):

---

✉ Massimiliano Giacalone  
massimiliano.giacalone@unina.it

<sup>1</sup> Department of Economics and Statistics, University of Naples 'Federico II', Complesso Universitario di Monte Sant'Angelo, Via Cintia 26, 80126 Naples, Italy

<sup>2</sup> Department of Economic and Legal Studies, Parthenope University of Naples, Palazzo Pacanowski, Via Generale Parisi 13, 80132 Naples, Italy

- a. it could be used to determine to what extent the dependent variable can be predicted by the independent variables;
- b. it could be used to capture the strength of a theoretical relationship between the dependent variable and the independent variables.

The first use of the term “multicollinearity” in the literature was by Frisch (1934). We refer to multicollinearity when there is a strong linear dependence between two or more regressors (Leahy 2000).

Multicollinearity is not just present or absent, but it is necessary to check for its degree, or severity, ranging from no collinearity to perfect collinearity. Normally, some degree of multicollinearity is always present in real data. Perfect multicollinearity is an extreme and uncommon case, occurring when two or more independent variables in a regression model exhibit a perfect linear relationship. In the latter case, the regression coefficients are indeterminate and their standard errors are infinite.

Multicollinearity may lead to several issues, summarized in the following points (Besley et al. 1980; Shieh 2011; Montgomery et al. 2012):

- Estimates issues: considering the ordinary least squares (OLS) method, the regression coefficients’ estimates are given by:

$$\hat{\beta}_{OLS} = (X'X)^{-1}X'Y$$

In the presence of a high degree of multicollinearity, the  $X'X$  matrix is quasi-singular (ill-conditioned); hence, there is a misalignment of the equation system that should provide the parameters’ estimates, which are therefore inaccurate. In the case of perfect multicollinearity, the matrix is non-invertible; this causes computational problems as well, as any statistical software would give warnings such as “The matrix is singular” and will be unable to proceed with the estimation.

- Forecasting issues: from a practical point of view, it is useless to make any kind of prediction for a value of the response variable  $y$ , due to the uncertainty of the values of  $\hat{\beta}$ .
- Interpreting issues: in the presence of multicollinearity, the coefficients’ estimates are less efficient and their interpretation becomes tricky. Regressors don’t explain the variance in the dependent variable as they should normally do.

When multicollinearity is detected, it becomes important to understand which explanatory variables are causing the issue, and the easiest solution would be to exclude them from the analysis. However, doing so is not always a good solution, as it depends on whether the purpose of the analysis is explanation or prediction. When prediction is the goal, no problem arises if, among dependent variables, two regressors have the same ‘meaning’: it is possible to just drop one of them. When, instead, explanation is the goal, regressors are selected according to a theoretical rule, so that it is not possible to eliminate any variable: the model has to be explained exactly, including every variable that was selected. Literature has recently introduced other methods to deal with multicollinearity without the need of dropping variables from the model, that are described in the following section.

A widely used estimator is the ordinary least squares (OLS) one, which is a valid alternative to the Maximum Likelihood estimator, especially when no information is given about the error distribution. According to the Gauss-Markov theorem, the OLS estimator is BLUE (Best Linear Unbiased Estimator) when  $E(\varepsilon) = 0$  and  $\text{var}(\varepsilon) = \sigma^2 I_n$ . This means

that, among all the linear estimators, it will be the one with the least variance, guaranteeing furthermore that the estimates are unbiased.

When the set of error random variables doesn't respect the Gauss-Markov assumptions, the OLS estimator is not BLUE and might therefore provide less efficient estimates than the ones provided by other methods, even though it continues to be unbiased. The variance of the residuals will be biased, invalidating significance tests, confidence intervals and the  $R^2$  coefficient.

In this paper, we are going to carry out some simulations in order to show the differences, in terms of efficiency, between the OLS estimator and two different  $L_p$ -norm estimators—whose properties will be described in the following sections—in the presence of multicollinearity.

This paper is organized as follows. In the next section, after a brief review on the origins and some new tools to deal with the multicollinearity problem, we illustrate some new methods, proposed in the literature, to solve this issue in the case of stochastic regression, typical in the analysis of economic and social data. In the third section, we introduce the exponential power function (E.P.F.), a useful family of symmetrical random error curves, and its connection with the  $L_p$ -norm method. In the fourth section, we investigate the relationship 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 analyze and propose the  $L_{pmin}$  algorithm in the fifth section. In the following one (6), we propose a simulation study, with the aim of choosing the best rule to select 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 variance. Finally, in Sect. 7, we propose a real data application, showing that our method gives better estimates than OLS. In the last Sect. (8), some results from the empirical study and comparisons are discussed.

## 2 Solving the multicollinearity problem: a review

Literature has devoted a lot of attention to the multicollinearity problem and the different ways to solve it.

When multicollinearity is detected in a regression model, the easiest solution is to eliminate the problematic variables from the model, so that it doesn't include variables that are correlated (Bowerman and O'Connell 1993; Bring 1996).

But, sometimes, it is hard to choose which variables to exclude. Another classical remedy to the multicollinearity problem is the Principal Component Analysis (Maddala 1977; Gower and Blasius 2005), that converts the original variables into a new set of linearly uncorrelated variables.

Since eliminating a variable leads to a problem of bad-specification of the model, a viable remedy is the use of a different estimator than the least squares one. This is because the 'classical' OLS method provides estimates that could be statistically not significant in the presence of multicollinearity. Some common alternatives are the partial least squares (Cassel et al. 1999) and the Ridge Regression (Pagel and Lunneberg 1985; Dijkstra 2014; Masson et al. 2014). The latter, in particular, was introduced by Hoerl and Kennard (1970), and has the advantage of reducing the variance term of the slope parameters (Muniz and Kibria 2009; Akdeniz et al. 2015). In the most recent literature, though, some robust estimators are preferred. There are many types of robust regression models and, traditionally, the least squares criterion has always been the most used one. The Least Absolute

Value (LAV) is a regression coefficient that minimizes the sum of the absolute values of the residuals:

$$\hat{\beta}_{LAV} = \arg \min_{\beta} \sum |y_i - x'_i \beta|$$

However, when multicollinearity is present, other methods should be preferred, such as the Ridge Least Absolute Value (RLAV), handling the problem of multicollinearity and outliers simultaneously. A new approach to solve the multicollinearity problem considers the use of a ridge regression based on Bisquare Ridge LTS estimators (BRLTS). Through this method, it is possible to obtain good estimates, removing multicollinearity when it is “moderate or high” (Pati et al. 2016).

In a regression model, we expect a high variance explained ( $R^2$ ). The higher the variance explained, the better the model is. Though, in the presence of multicollinearity, the variance, standard errors and parameter estimates may be all inflated.

Other authors suggest the so called “Nested estimate procedure” (Lin 2008), that is a relatively new method, based on OLS, consisting in an iterative estimation procedure of independent variable parameters. Garcia et al. (2011) presented the “raise method” as an alternative to the Nested estimation, that keeps all information, which could be highly recommended in some cases, and compares the results with other procedures.

There are no specific tests to detect multicollinearity, but some characteristics of the estimated model can reveal it:

- a high  $R^2$  with non-significant regression coefficients (t-scores);
- a high correlation between the single variables;
- a high VIF (Variance Inflation Factor).

The Variance Inflation Factor (VIF) is calculated for each variable in the model, based on the expression:

$$VIF_j = \frac{1}{(1 - R_j^2)}$$

where  $R_j^2$  is the multiple  $R^2$  for the regression of the j-th element of the vector of estimates on the other covariates.

A high VIF reveals linear dependence between the j-th column and the remaining columns of the X matrix and, thus, the presence of multicollinearity (Lazaridis 2007). However, as Curto and Pinto (2011) point out, the real impact on the variance can be overestimated by the VIF and, for this reason, CVIF can be used in place of VIF. In formulas:

$$CVIF_j = \frac{((1 - R^2)TSS/(n - k)(1 - R_j^2)TSS_j)}{((1 - R_0^2)TSS/((n - k)TSS_j))}$$

or:

$$CVIF_j = VIF_j \times \frac{1 - R^2}{1 - R_0^2}$$

where  $R_0^2 = R_{yx_2}^2 + R_{yx_3}^2 + \dots + R_{yx_k}^2$ .

In particular,  $R_{yx_k}^2$  is the square of the correlation coefficient between the dependent variable and the independent variables, keeping TSS and TSS<sub>j</sub> constant. In general,  $R_0^2$  can

be lower or higher than  $R^2$ , but when the independent variables are orthogonal we have  $R_0^2 = R^2$ .

When  $R_0^2 > R^2$ , CVIF ranges from 1 to  $+\infty$  and  $CVIF_j > VIF_j$ ; when  $R_0^2 < R^2$ , CVIF ranges from 0 to 1 and  $CVIF_j < VIF_j$ . About interpreting issues, as suggested by Curto and Pinto (2011), we can use the rule-of-thumb  $CVIF_j > 10$  to indicate the presence of severe multicollinearity.

Another common issue in economics is the spurious relationship between variables (Armor et al. 2017). As Chatelain and Ralf (2014) show, by increasing the number of observations it is possible to bring a spurious inference of highly correlated classical suppressors. In particular, they propose a tool to solve these issues: the Parameter Inflation Factor (PIF), obtained by analyzing a trivariate regression model as follows:

$$x_1 = \beta_{12}x_2 + \beta_{13}x_3 + \varepsilon_{1,23}$$

PIF is defined as:

$$PIF_{12} = \left(1 - \frac{r_{13}}{r_{12}}r_{23}\right)VIF_{12}$$

where  $r_{12}$ ,  $r_{13}$  and  $r_{23}$  are the correlation coefficients of the regressors and VIF is the Variance Inflation Factor.

While the Variance Inflation Factor only depends on the correlation between the independent variables, the Parameter Inflation Factor also considers the correlation between the regressors and the dependent variable. In the presence of a high PIF value, there may be highly correlated classical suppressors.

When regressors and error are correlated, the OLS estimator becomes less efficient because its variance increases (Vargha et al. 2013; Griffiths and Hajargasht 2016), even though it remains consistent (Krone et al. 2017).

Another estimator that can be used in the presence of multicollinearity is the Classical Linear Regression Model (CLRM) (Ayinde 2007):

$$y_t = \beta_0 + \beta_1x_{1t} + \beta_2x_{2t} + \beta_3x_{3t} + \varepsilon_t$$

where  $t = 1, 2, \dots, n$  and  $\varepsilon_t \sim (0, \sigma^2)$  and assuming that  $x_1$  is correlated with error  $\varepsilon$  and  $x_2$  is correlated with  $x_1$ .

In addition, we can consider a generalized least squares (GLS) model:

$$y_t = \beta_0 + \beta_1x_{1t} + \beta_2x_{2t} + \beta_3x_{3t} + u_t$$

where  $t = 1, 2, \dots, n$ ;  $u_t = \rho u_{t-1} + \varepsilon_t$  and  $\rho$  is the correlation.

Starting from this point, Ayinde (2007) shows that CLRM and GLS models are equivalent in the case of zero correlation. In addition, comparing CLRM, GLS and Maximum Likelihood (ML) estimators, using the MSE (Mean Square Error) criteria in a Monte Carlo simulation plan, the author shows that ML and GLS are better than an OLS estimator with low replication, but when there is a high number of replications the OLS method is to be favored.

Another study by Tran and Tsionas (2013) proposes the Generalized Method of Moments (GMM) estimator, showing that if there is no correlation between the variables, it performs as the standard Maximum Likelihood estimator. However, in the case of strong correlation, the MLE becomes biased while the GMM estimator remains unbiased.

Finally, as Giacalone and Richiusa (2006) point out,  $L_p$ -norm estimators allow to manage the residuals more efficiently, particularly when there is some degree of

multicollinearity and regressors are correlated with residuals. This is because the  $L_p$ -norm methods are adaptive procedures with respect to the error component of the model and not to the deterministic one.

As a result, using  $L_p$ -norm estimators in the presence of stochastic regressors, it is possible to obtain a better efficiency in the estimates than using OLS.

### 3 The exponential power function and the $L_p$ -norm estimators

The exponential power function (E.P.F.) is a family of probability functions proposed by Subbotin (1923) and studied, among others, by Vianelli (1963), Lunetta (1966), Mineo (1989), Gonin and Money (1989), Chiodi (1995) and Bottazzi and Secchi (2011).

The E.P.F. constitutes a valid generalization of the Gaussianity hypothesis, that is usually assumed even though, depending on the data we have at disposal, it is not always fully supportable. In the literature, it is also known as “Normal distribution of order  $p$ ” or “Generalized Error Distribution”, and it constitutes a parametric alternative to the robust methods (Mineo 2003).

The density function of the E.P.F. is:

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

where  $M_p = E(z)$  is the location parameter,  $\sigma_p = (E[|z - M_p|])^{1/p}$  is the scale parameter and  $p > 0$  is the shape parameter.

Considering the Pearson kurtosis index  $\beta_2$ , we distinguish:

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

For particular values of  $p$ , we have:

- the Laplace distribution ( $p = 1$ ,  $\beta_2 = 6$ );
- the Gaussian distribution ( $p = 2$ ,  $\beta_2 = 3$ );
- the Uniform distribution ( $p \rightarrow \infty$ ,  $\beta_2 \rightarrow 1.8$ ).

Considering a sample of  $n$  observed data  $(x_i, y_i)$ , a general linear regression model is:

$$y_i = g(x_i, \theta) + \varepsilon_i \quad (2)$$

with  $g(\cdot)$  linear function.

$L_p$ -norm estimators are useful generalizations of ordinary least squares estimators, obtained by replacing the exponent 2 with a general exponent  $p$  (Ekblom and Henriksson 1969; Forsythe 1972). Therefore, they minimize the sum of the  $p$ -th power of the absolute deviations of the observed points from the regression function:

$$S_p(\theta) = \sum_{i=1}^n |y_i - g(x_i, \theta)|^p \quad \text{with } 1 \leq p < \infty \quad (3)$$

Under the regular assumptions, the log-likelihood associated with the sample, where  $z = y_i$ ;  $M_p = g(x_i, \theta)$  is given by:

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

$$\begin{aligned} \frac{\partial l}{\partial \theta_j} &= \sum_{i=1}^n |y_i - g(x_i, \theta)|^{p-1} \text{sign}(y_i - g(x_i, \theta)) \frac{\partial g}{\partial \theta_j} = 0 \\ \sum_{i=1}^n |y_i - g(x_i, \theta)|^p &= \min \text{ with } p \geq 1 \end{aligned} \tag{5}$$

When the order  $p$  is specified, all the terms in the (4), except for the last part containing the vector  $\theta$ , are constant. Therefore, Maximum Likelihood estimators are equivalent to  $L_p$ -norm estimators (5).

If  $p$  is unknown, there are 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' estimates.

About the procedures to estimate  $p$ , it is possible to find out the following proposals in the literature:

- Harter (1977), noting that  $p$  depends on  $\hat{\beta}_2$  (the sample residual kurtosis), proposed to select  $p$  using the following rules:  
 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 (1982) proposed two different criteria respectively:

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

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

- Mineo (1989) proposed the Generalized Kurtosis  $\beta_K$ , as described in Sect. 4.
- Mineo (1994) considered a new method to estimate  $p$ , based on an empirical index called VI. This estimation method is based on a two-steps algorithm in which, iteratively, the estimates of  $p$  and the other parameters are evaluated: the structure parameter is estimated by numerically solving an equation in which the theoretical value of the  $I$  index is equaled to the empirical one; the other parameters are estimated from the corresponding maximum likelihood estimators as a function of the current value of  $\hat{p}$ .
- Agrò (1995) proposed a maximum likelihood estimation either for the E.P.F. parameters or for the  $p$  shape parameter. It is a two-steps process which, however, is essentially suitable for medium to large samples ( $n > 50$ ).
- Giacalone (1997) proposed an algorithm based on a two-steps alternating procedure that firstly estimates the  $\theta$  parameter vector by means of the classical conjugated gradient algorithm (Fletcher and Reeves 1964) and secondly estimates  $p$  using a joint inverse function of  $I$  and  $\beta_2$ , obtained comparing empirical and theoretical moments,

matching (10) with (12) and (9) with (11). The minimization algorithm stops when the variation of  $p$  is not significant (Everitt and Hand 1987).

- Agrò (1999) proposed an adjustment of the likelihood estimation, based on Cox and Reid (1987), which consists in a reparametrization that leads to asymptotically uncorrelated estimators and, consequently, to better results (at least, for  $n > 30$ ).
- Finally, Mineo and Ruggieri (2005) developed an R package for dealing with the exponential power function, with functions to compute the density function, the distribution function and the quantiles from an E.P.F., available on the Comprehensive R Archive Network (CRAN).

### 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:

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

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 known as ‘‘Generalized Kurtosis’’ (Mineo 1989):

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

If  $k = 2$ , we get the Pearson’s kurtosis index:

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

If  $k = 1$ , considering the square root of the reciprocal, we get the Geary’s length of tails index (Geary 1936):

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

The indexes  $I$  and  $\beta_2$  show a different behavior according to the variation of  $p$  (Giacalone 1996, 1997). Calculating the sample values of  $I$  and  $\beta_2$ , it is possible to obtain, by inverse interpolation, two different estimations of  $p$ .

Gonin and Money (1987), Kendall and Stuart (1966) and Lunetta (1966) considered the unbiased estimates of the second and fourth order sample moments, with correction factors depending on the sample size  $n$ :

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

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

where  $\varepsilon_i$  and  $\bar{\varepsilon}$  are, respectively, the estimated residuals and their average.

The ratio of  $\hat{\mu}_4$  and  $\hat{\mu}_2$  gives the following estimator of  $\beta_2$ :



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

For the I empirical index we obtain:

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

### 5 The L<sub>pmin</sub> algorithm

The proposed algorithm is based on a two-steps alternating procedure:

1. minimization procedure to estimate the parameters;
2. joint inverse function of I and  $\beta_2$  to estimate p.

The algorithm stops when p does not vary significantly. To obtain this 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 (Counihan 1985; Giacalone 2002).

The function used to estimate p is therefore the following:

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

where  $I, \hat{I}, \beta_2, \hat{\beta}_2$  are given, respectively, by (10), (12), (9) and (11). For simplicity, we can express (13) as follows:

$$[f(p)]^2 + [g(p)]^2 = \min.$$

As the indexes  $I$  and  $\beta_2$  have different orders of magnitude and different variance, it is necessary to eliminate this difference in order to obtain a joint estimation of p, 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.

Considering p ranging from 0.5 to 10, 25.2 is the value of  $\beta_2$  when  $p = 0.5$ , and 0.86054 is the value of  $I$  when  $p = 10$ . This way it is possible to obtain a joint estimator made up of two squared functions. The two kurtosis indexes are used as it was observed that norm 1 kurtosis is a valid choice for the estimation of p in the presence of outliers, while norm 2 kurtosis performs better when the sample values gather around the center of the distribution.

So, using the relation (9) we calculate  $\max(\beta_2) = 25.2$ , for  $p = 0.5$ , whilst using the relation (10) we calculate  $\max(I) = 0.86054$ , for  $p = 10$ .

The proposed algorithm is then specified in the following steps (Giacalone 1997):

1. set  $i = 0$  and  $p_0 = 2$ ;
2. fit the model (2) to the data using the previous step value  $p_i$ ;
3. compute the estimated residuals  $\varepsilon_i = y_i - g(x_i, \theta)$ , their average  $\bar{\varepsilon}$  and insert these quantities in the (13) that is equal to the sum of the two squared functions to be minimized;
4. minimize the function (13) to obtain  $p_{i+1}$ , new estimate of p;

5. 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 the steps 2–5;
6. otherwise: stop the algorithm assuming  $\hat{\theta}_i = \theta_{ij}$  as  $L_p$ -norm estimator for the parameter  $\theta_i$  and the value  $p = p_i$  as joint estimation of  $p$ .

In step 2, a nonlinear  $L_p$ -norm estimation is examined. The problem could be solved using the optimality conditions encountered in unconstrained optimization (McCormick 1983). The minimization algorithm (Fletcher and Reeves 1964) is used to take the special structure of the problem directly into account.

In step 3, we calculate both empirical and theoretical I and  $\beta_2$  kurtosis indexes to obtain the values to estimate  $p$  from (13).

In step 4, a parabolic interpolation method (Everitt and Hand 1987) to find the minimum sum of squared functions (13) is used. The convergence of the proposed algorithm was empirically verified by simulating 5000 samples of different sizes ( $n = 30, 50, 100, 200, 500, 1000$ ) for six fixed theoretical values of  $p$ .

### 6 A simulation study for efficiency comparison

After looking at the more recent literature (e.g. Alabi et al. 2008, 2014), we chose to consider 5000 samples of size  $n = 30, 50, 100, 200, 500$  and 1000, generated from an E.P.F., and 6 values of  $p$  (1.1, 1.5, 2.0, 2.5, 3.0, 3.5) for three different degrees of multicollinearity (low  $R^2 = 0.33$ , medium  $R^2 = 0.66$ , high  $R^2 = 0.99$ ).

The algorithm to generate the  $\varepsilon_i$  (for  $p \geq 1$ ) from an E.P.F. is suggested by Chiodi (1986). The values of  $y_i$  are given by the following multiple regression model:

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

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

$$X_3 = X_1 + X_2 + Z \text{ with } Z \sim N(0, \sigma_z^2) \tag{15}$$

Therefore, we can write the associated variance and covariance matrix:

$$S = \begin{vmatrix} & X_1 & X_2 & X_3 \\ X_1 & 1 & 0 & 1 \\ X_2 & 0 & 1 & 1 \\ X_3 & 1 & 1 & 2 + \sigma_z^2 \end{vmatrix}$$

It is easy to notice 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:

$$A = \begin{vmatrix} & X_1 & X_2 & X_3 \\ X_1 & 1 & 0 & 1/\sqrt{2 + (\sigma_z^2)} \\ X_2 & 0 & 1 & 1/\sqrt{2 + (\sigma_z^2)} \\ X_3 & 1/\sqrt{2 + (\sigma_z^2)} & 1/\sqrt{2 + (\sigma_z^2)} & 1 \end{vmatrix}$$

where  $A_{13} = \frac{\text{cov}(X_1, X_3)}{\sqrt{\text{var}(X_1)\text{var}(X_3)}} = \frac{1}{\sqrt{2 + (\sigma_z^2)}} = A_{23}$ ,  $R_{3,12}^2 = 1 - \frac{\det A}{\det A_{33}} = \frac{2}{2 + (\sigma_z^2)}$ , and  $\det A_{33}$  is the cofactor of the element occupying the same position in the correlation matrix (Leti 1983).

In this particular regression model, the degree of multicollinearity is inversely proportional to  $\sigma_z^2$ .

Specifically, in our simulation we have put as parameter values:  $\beta_0 = 1$ ,  $\beta_1 = 2$ ,  $\beta_2 = 3$ ,  $\beta_3 = 4$ , and a comparative analysis was performed applying the following three estimation methods on the same samples:

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 ( $L_{pmin}$ ).

The results of the simulation plan are reported in Tables 10, 11, 12, 13, 14 and 15 in “Appendix 1”, where the OLS simulation results are marked in bold; furthermore, in Tables 13, 14 and 15, the number of samples with a difficult convergence (“Conv.”) is also presented. In the tables, we report mean (M) and variance (V) of the regression model parameters. We can see that, for any value of  $p$  and for any method, the estimates of  $\beta_0$ ,  $\beta_1$ ,  $\beta_2$  and  $\beta_3$  are less efficient for small samples (e.g. for  $n = 30$  and  $n = 50$ ) but their variances decrease as  $n$  increases.

We calculated the relative efficiency for every value of  $p$  and  $n$ , and the results are shown in Tables 1A, 1B, 2A, 2B, 3A and 3B. From the tables, we notice that  $L_p$ -norm estimators ( $L_p$  and  $L_{pmin}$ ) give better estimates for every parameter, compared to the least squares method, especially when the value of  $p$  is not close to 2: there is a gain in efficiency in all the cases considered, except for the case  $p = 2$ , in which the error follows a Gaussian distribution.

However, the  $p$  parameter of any real dataset is rarely equal to 2. When the value of  $p$  is different than 2, the OLS method always brings less efficient estimates than both the  $L_p$  and  $L_{pmin}$  methods.

Indeed, in most cases, the  $L_{pmin}$  method produces more efficient estimates than OLS, but less efficient than the  $L_p$  method, that uses the true value of  $p$ . Therefore,  $L_{pmin}$  could be considered halfway between the  $L_2$  and  $L_p$  methods, because the  $p$  parameter is estimated on the data sample.

The relative efficiency indexes (Tables 1A, 1B, 2A, 2B, 3A, 3B) show that, for medium and high degrees of multicollinearity ( $R^2 = 0.66$ ;  $R^2 = 0.99$ ),  $L_p$ -norm estimators are more efficient than OLS. However, in the presence of a low degree of multicollinearity ( $R^2 = 0.33$ ), the OLS estimator’s efficiency is sometimes higher than  $L_p$ -norm.

Since we know the distribution of residuals, both OLS and  $L_p$ -norm estimators are also Maximum Likelihood estimators in our case. As the sample size increases to infinity, a Maximum Likelihood estimator is always efficient, since it achieves the Cramér-Rao lower bound. That said, given that all of these estimators benefit from asymptotic efficiency property, the efficiency gain is larger for small samples.

## 7 A real data application

To better show the difference between OLS and  $L_p$ -norm estimates, we have built a dataset made up of four variables, as in Mouza and Targoudtzidis (2012): GDP per capita, unemployment rate and working hours are examined together in order to show a connection with labor accidents in the UK. Working hours are considered along with unemployment rate, as the latter measures any occupation of the individuals throughout a period of time, without giving any information about the nature of this occupation (e.g. part-time and occasional workers are considered employed even if they spend a small amount of time

**Table 1** Relative efficiency of  $L_p$ -norm estimators compared to OLS (parameters  $\beta_0, \beta_1, \beta_2, \beta_3$ ) on 5000 samples of size  $n = 30, n = 50, n = 100, n = 200, n = 500, n = 1000$  ( $\sigma_z = 2, R^2 = 0.33$ )

p	1.1				1.5				2.0			
	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$
A For p = 1.1, 1.5 and 2												
n = 30												
$L_p$	1.22	1.13	1.28	1.27	0.93	0.77	0.82	0.84	1.00	1.00	1.00	1.00
$L_{pmin}$	1.12	1.04	1.21	1.29	0.85	0.75	0.74	0.81	0.74	0.81	0.73	0.88
n = 50												
$L_p$	1.35	1.17	1.22	1.34	0.84	0.82	0.72	0.84	1.00	1.00	1.00	1.00
$L_{pmin}$	1.24	1.13	1.14	1.26	0.71	0.65	0.63	0.73	0.82	0.87	0.81	0.89
n = 100												
$L_p$	1.23	1.15	1.12	1.22	1.03	0.93	0.88	0.97	1.00	1.00	1.00	1.00
$L_{pmin}$	1.16	1.08	1.05	1.11	0.94	0.85	0.84	0.93	0.87	0.95	0.82	0.78
n = 200												
$L_p$	1.26	1.23	1.09	1.14	1.15	1.12	1.18	1.16	1.00	1.00	1.00	1.00
$L_{pmin}$	1.19	1.17	1.05	1.02	0.97	0.96	0.94	0.98	0.96	0.87	0.83	0.95
n = 500												
$L_p$	1.23	1.18	1.14	1.07	1.16	1.23	1.16	0.97	1.00	1.00	1.00	1.00
$L_{pmin}$	1.17	1.12	1.05	1.02	1.11	1.15	1.07	0.88	0.95	0.92	0.97	0.94
n = 1000												
$L_p$	1.14	1.06	1.13	1.04	1.03	1.12	1.13	1.08	1.00	1.00	1.00	1.00
$L_{pmin}$	1.07	0.97	1.05	0.95	0.96	1.08	1.07	1.03	0.96	0.93	0.95	0.93
p	2.5				3.0				3.5			
	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$
B For p = 2.5, 3 and 3.5												
n = 30												
$L_p$	0.94	0.88	0.98	0.88	1.06	1.12	1.33	1.24	1.23	1.16	1.33	1.26
$L_{pmin}$	0.87	0.84	0.93	0.84	1.03	1.02	1.25	1.17	1.15	1.13	1.28	1.19
n = 50												
$L_p$	0.88	0.95	0.88	0.95	0.98	1.08	1.13	0.93	1.06	0.96	0.98	1.14
$L_{pmin}$	0.86	0.91	0.84	0.92	0.87	0.99	0.95	0.89	1.04	0.88	0.93	1.07
n = 100												
$L_p$	1.06	0.96	0.94	0.96	1.08	1.16	1.34	1.25	1.16	1.14	1.05	1.13
$L_{pmin}$	1.03	0.94	0.91	0.90	1.05	0.98	1.22	1.09	1.08	0.89	0.97	1.05
n = 200												
$L_p$	1.14	1.06	0.98	0.92	1.08	1.14	1.37	1.26	1.14	1.18	1.07	1.16
$L_{pmin}$	1.07	0.98	0.93	0.84	1.03	1.05	1.28	1.19	1.06	1.09	1.02	1.13
n = 500												
$L_p$	1.14	1.02	1.14	1.08	1.06	1.15	1.24	1.18	1.13	1.04	0.96	1.04
$L_{pmin}$	1.11	0.96	1.06	1.03	1.01	1.07	1.17	1.09	1.07	0.95	0.93	1.01
n = 1000												
$L_p$	1.08	1.09	1.06	1.14	1.05	1.16	1.14	1.07	1.03	1.12	1.07	1.03
$L_{pmin}$	1.04	1.05	1.02	1.10	1.02	1.08	1.07	1.02	1.01	1.04	1.05	1.01

**Table 2** Relative efficiency of  $L_p$ -norm estimators compared to OLS (parameters  $\beta_0, \beta_1, \beta_2, \beta_3$ ) on 5000 samples of size  $n = 30, n = 50, n = 100, n = 200, n = 500, n = 1000$  ( $\sigma_z = 1, R^2 = 0.66$ )

p	1.1				1.5				2.0			
	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$
A For $p = 1.1, 1.5$ and $2$												
n = 30												
$L_p$	2.07	1.78	2.14	1.86	1.46	1.58	1.38	1.57	1.00	1.00	1.00	1.00
$L_{pmin}$	1.84	1.65	1.83	1.77	1.34	1.54	1.27	1.54	0.87	0.94	0.73	0.76
n = 50												
$L_p$	1.84	1.66	2.07	1.67	1.38	1.56	1.34	1.51	1.00	1.00	1.00	1.00
$L_{pmin}$	1.67	1.59	1.78	1.73	1.33	1.51	1.23	1.38	0.85	0.92	0.86	0.83
n = 100												
$L_p$	1.43	1.36	1.64	1.56	1.24	1.53	1.38	1.42	1.00	1.00	1.00	1.00
$L_{pmin}$	1.24	1.31	1.52	1.52	1.21	1.47	1.17	1.28	0.94	0.97	0.85	0.88
n = 200												
$L_p$	1.23	1.26	1.44	1.45	1.16	1.41	1.34	1.39	1.00	1.00	1.00	1.00
$L_{pmin}$	1.18	1.22	1.42	1.37	1.13	1.34	1.17	1.36	0.91	0.93	0.88	0.94
n = 500												
$L_p$	1.22	1.17	1.28	1.37	1.16	1.26	1.24	1.21	1.00	1.00	1.00	1.00
$L_{pmin}$	1.14	1.15	1.24	1.25	1.07	1.25	1.17	1.08	0.96	0.97	0.93	0.90
n = 1000												
$L_p$	1.19	1.17	1.05	1.14	1.04	1.05	1.02	1.11	1.00	1.00	1.00	1.00
$L_{pmin}$	1.15	1.14	1.02	1.07	1.02	1.03	1.01	1.08	0.99	0.98	0.96	0.94
p	2.5				3.0				3.5			
	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$
B For $p = 2.5, 3$ and $3.5$												
n = 30												
$L_p$	0.95	1.12	1.24	1.14	1.14	1.37	1.51	1.64	1.52	1.64	1.43	1.86
$L_{pmin}$	0.77	0.88	0.95	0.89	1.10	1.25	1.37	1.52	1.34	1.48	1.31	1.64
n = 50												
$L_p$	1.12	1.13	1.34	1.24	1.14	1.23	1.47	1.43	1.42	1.53	1.38	1.61
$L_{pmin}$	0.95	0.96	1.13	1.07	1.11	1.15	1.36	1.38	1.37	1.34	1.34	1.45
n = 100												
$L_p$	1.23	1.34	1.43	1.36	1.18	1.14	1.32	1.34	1.36	1.53	1.25	1.54
$L_{pmin}$	1.12	1.28	1.24	1.17	1.12	1.09	1.23	1.25	1.27	1.37	1.35	1.23
n = 200												
$L_p$	1.14	1.24	1.33	1.25	1.08	1.13	1.25	1.23	1.24	1.41	1.16	1.33
$L_{pmin}$	1.11	1.18	1.24	1.16	1.05	1.06	1.14	1.15	1.18	1.27	1.14	1.27
n = 500												
$L_p$	1.14	1.12	1.24	1.08	1.04	1.13	1.12	1.15	1.23	1.34	1.18	1.24
$L_{pmin}$	1.07	1.05	1.18	1.04	1.01	1.06	1.08	1.06	1.14	1.17	1.16	1.18
n = 1000												
$L_p$	1.06	1.08	1.14	1.05	1.07	1.14	1.04	1.12	1.13	1.21	1.16	1.12
$L_{pmin}$	1.03	1.06	1.08	1.02	1.05	1.08	0.95	1.06	1.05	1.08	1.14	1.04

**Table 3** Relative efficiency of  $L_p$ -norm estimators compared to OLS (parameters  $\beta_0, \beta_1, \beta_2, \beta_3$ ) on 5000 samples of size  $n = 30, n = 50, n = 100, n = 200, n = 500, n = 1000$  ( $\sigma_z = 0.01, R^2 = 0.99$ )

p	1.1				1.5				2.0			
	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$
A For $p = 1.1, 1.5$ and $2$												
n = 30												
$L_p$	1.43	1.34	1.38	1.47	1.24	1.29	1.52	1.64	1.00	1.00	1.00	1.00
$L_{pmin}$	1.37	1.27	1.24	1.44	1.22	1.27	1.31	1.48	0.94	0.85	0.88	0.75
n = 50												
$L_p$	1.26	1.14	1.25	1.34	1.24	1.18	1.42	1.44	1.00	1.00	1.00	1.00
$L_{pmin}$	1.24	1.12	1.17	1.28	1.16	1.16	1.28	1.37	0.85	0.97	0.78	0.82
n = 100												
$L_p$	1.16	1.18	1.24	1.23	1.15	1.16	1.34	1.24	1.00	1.00	1.00	1.00
$L_{pmin}$	1.14	1.16	1.17	1.16	1.14	1.07	1.26	1.15	0.95	0.98	0.78	0.89
n = 200												
$L_p$	1.14	1.07	1.13	1.16	1.06	1.01	1.24	1.14	1.00	1.00	1.00	1.00
$L_{pmin}$	1.12	1.05	1.11	1.05	1.02	0.95	1.18	1.13	0.88	0.95	0.84	0.76
n = 500												
$L_p$	1.07	1.04	1.15	1.06	1.09	1.07	1.23	1.07	1.00	1.00	1.00	1.00
$L_{pmin}$	1.04	1.03	1.12	1.02	1.03	1.05	1.17	1.05	0.95	0.99	0.29	0.97
n = 1000												
$L_p$	1.06	0.98	1.07	1.04	1.05	1.04	1.05	0.96	1.00	1.00	1.00	1.00
$L_{pmin}$	1.03	0.96	1.03	1.02	1.02	1.02	1.02	0.95	0.97	0.95	0.94	0.98
p	2.5				3.0				3.5			
	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$
B For $p = 2.5, 3$ and $3.5$												
n = 30												
$L_p$	1.13	0.97	0.84	1.07	1.33	1.14	1.43	1.18	1.42	1.64	1.82	1.29
$L_{pmin}$	0.95	0.88	0.76	1.04	1.17	0.95	1.27	1.15	1.37	1.55	1.46	1.27
n = 50												
$L_p$	0.95	1.04	0.98	0.95	1.24	1.34	1.37	1.43	1.74	1.93	1.64	1.24
$L_{pmin}$	0.87	0.95	0.94	0.87	1.16	1.25	1.18	1.18	1.53	1.44	1.45	1.12
n = 100												
$L_p$	1.12	1.05	1.24	1.16	1.14	1.24	1.25	1.34	1.24	1.84	1.44	1.24
$L_{pmin}$	0.99	1.03	1.07	1.15	1.05	1.16	1.23	1.26	1.12	1.67	1.36	1.22
n = 200												
$L_p$	1.23	1.32	1.42	1.22	1.24	1.16	1.08	1.24	1.14	1.43	1.34	1.18
$L_{pmin}$	0.95	1.16	1.14	1.15	1.16	1.13	1.05	1.14	1.06	1.36	1.25	1.15
n = 500												
$L_p$	1.14	1.24	1.34	1.14	1.36	1.34	1.14	1.14	1.19	1.34	1.28	1.17
$L_{pmin}$	1.03	1.21	1.25	1.05	1.07	1.26	1.07	1.08	1.13	1.22	1.16	1.05
n = 1000												
$L_p$	1.09	1.14	1.23	1.05	1.14	1.15	1.04	1.14	1.14	1.13	1.18	1.15
$L_{pmin}$	1.05	1.03	1.04	1.04	1.04	1.03	1.03	1.05	1.06	1.05	1.14	1.08

at the working place). The data, presented in “Appendix 2” (Table 16), refers to the United Kingdom, in the period 1971–2007 (yearly, 37 observations).

More in detail, the variables used in our application are the following:

1. GDP—Gross Domestic Product per capita (thousand US dollars, current prices and purchasing power parity), source: OECD databank;
2. UR—unemployment rate (percentage), source: OECD databank;
3. AH—average hours worked (hours per year per worker), source: OECD databank;
4. FI—fatal injuries of employees (absolute numbers), source: UK national statistics database, health and safety executive.

To study the relationship between the economic cycle and workplace accidents, we estimated the following regression model:

$$FI = \beta_0 + \beta_1 GDP + \beta_2 UR + \beta_3 AH + \varepsilon$$

The  $R^2$  for this regression is 0.886, and the Adjusted  $R^2$  is 0.875. In this model, the parameters to be estimated are 4 (including the intercept), just like in our simulation study.

In the following table (Table 4), we show the VIF values, for each explanatory variable taken as dependent and the remaining as independent, in order to get the corresponding  $R^2$ :

About VIF interpretation, considering the limitations highlighted by O’Brien (2007), we opted for the “rule of 10” proposed by Menard (1995). Since none of the VIFs exceed 10, we can conclude that there is no severe collinearity in this regression, but only a medium degree of collinearity. In this case, looking at the results of the simulation study, our method brings better estimates than OLS.

The results of the OLS estimation ( $p = 2$ ) are (Table 5):

Thus, the OLS method returns the following model:

$$\hat{Y} = -255.40 - 13.97X_1 - 15.14X_2 + 0.54X_3 \quad (16)$$

Using the  $L_p$ -norm estimation method proposed in this paper ( $L_{p_{\min}}$ ), we estimated  $p = 1.696$  (leptokurtic distribution of residuals), and the results are (Table 6):

Thus, the  $L_p$ -norm method returns the following model:

$$\hat{Y} = -25.01 - 14.68X_1 - 16.06X_2 + 0.42X_3 \quad (17)$$

As data points out, the intercept ( $\beta_0$ ) value gets closer to zero when estimated through the  $L_p$ -norm method. This is already proof of a better accuracy of the estimates: when the value of the other variables (e.g. GDP and Average Working Hours) is zero, it is natural to expect, for Fatal Injuries, a value that is as close as possible to zero.

In Table 7, by minimizing the objective function expressed by (3), we clearly show that the  $L_p$ -norm method brings a large improvement in the estimates’ efficiency.

**Table 4**  $R^2$ , adjusted  $R^2$  and VIF for each explanatory variable in the model

Variable as dependent	$R^2$	Adj. $R^2$	VIF
GDP	0.855	0.847	6.916
UR	0.583	0.558	2.398
AH	0.875	0.868	7.987

**Table 5** OLS regression results

Residuals			
Range	1Q	Median	3Q
-95.55 to 142.42	-47.15	-10.35	32.99
Coefficients			
$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$
-255.3976	-13.9667	-15.1445	0.5437

**Table 6**  $L_p$ -norm regression results

Residuals			
Range	1Q	Median	3Q
-91.911 to 149.528	-42.493	-4.577	36.283
Coefficients			
$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$
-25.0052	-14.6777	-16.0582	0.4217

**Table 7** Minimization of the objective function

OLS	$L_p$ -norm
138,073.50	36,133.23

The results from Table 7 indicate that our estimates are better than the ones obtained by OLS, as they are characterized by a significantly lower variability in the residuals. The objective function represents the stochastic component of the model; the deterministic component is therefore better explained by the  $L_p$ -norm method, given the lower value of the objective function. We can justify the relevant difference resulting from the minimization with the important difference in the estimated values for the intercept.

Below, for each value of  $p$ , we present the scatter plot of residuals against fitted values (Fig. 1). As the plots suggest, the variances of the error terms are not equal.

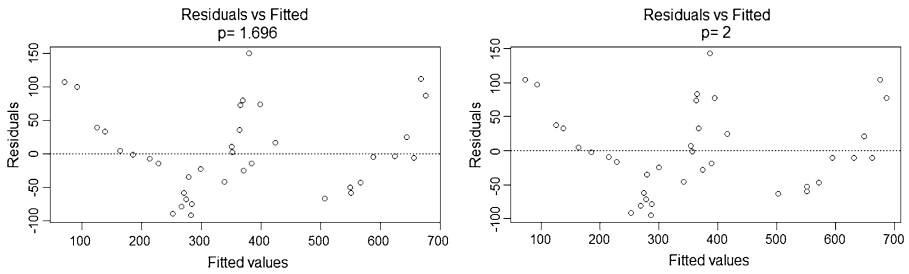
The following figure (Fig. 2) shows, for each value of  $p$ , the Normal distribution Quantile–Quantile plot. As the succession of points doesn't form a straight line, the plots suggest that the data doesn't come from a Normal distribution and, therefore, a different distribution should be considered.

Dynamic graphic techniques could also be used, for a graphical comparison of the two different estimation methods (Destefanis and Porzio 1999).

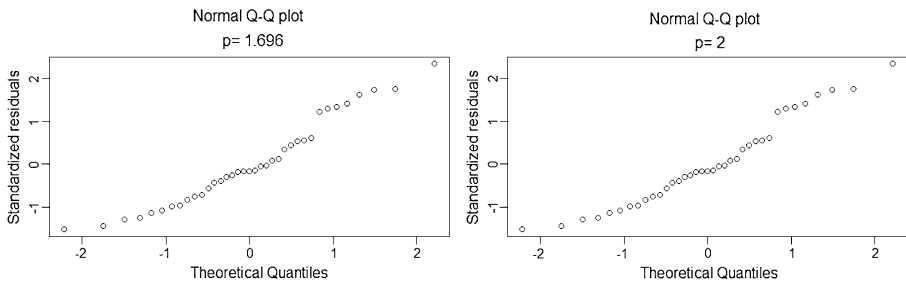
Moreover, we performed a regression analysis in terms of elasticity. The results, shown in Tables 8 and 9, are kind of impressive, since they suggest, as also evidenced by Mouza and Targoutzidis (2012), that fatal injuries are somewhat related to, and explained by, labor market indicators.

It is very interesting to note that the signs of the estimated coefficients are the same with both our estimation method and OLS: this evidences the robustness of our analysis.





**Fig. 1** Scatter plot of residuals against fitted values, for  $p = 1.696$  and  $p = 2$



**Fig. 2** Normal quantile–quantile plot, for  $p = 1.696$  and  $p = 2$

Fatal Injuries appear inelastic in relation to GDP per capita and Unemployment Rate but, at the same time, Fatal Injuries are elastic in relation to Average Working Hours. In particular, although FI is positively influenced by an increase in unemployment, this appears inelastic. Overall, we might expect that rising unemployment rates would increase employee marginal working hours, leading to an increase in workplace accidents. Indeed, the AH variable is strictly positive and elastic: when AH increases, FI increases as well. The reason behind this result is, by the way, statistically intuitive: as time spent at the working place increases, chance of injury increases consequently.

In this application, although there are some improvements in the estimates by using  $L_p$ -norm methods, principally linked to a better explanation of the deterministic part of the regression model, the results are in line with what happens when using OLS. This is because data is affected by only a medium degree of collinearity and the estimated  $p$  is close to two.

**Table 8** OLS regression results (in terms of elasticity)

Residuals			
Range	1Q	Median	3Q
-0.344 to 0.483	-0.0558	-0.003317	0.0809
Coefficients			
$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$
-0.335	-0.694	0.112	1.037

**Table 9**  $L_p$ -norm regression results (in terms of elasticity)

Residuals			
Range	1Q	Median	3Q
-0.349 to 0.483	-0.0547	-0.00036	0.0727
Coefficients			
$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$
-0.386	-0.678	0.163	1.034

## 8 Conclusions

Multicollinearity indicates a situation in which the independent variables in a regression model are highly correlated, leading to instability and large variance of the OLS estimator. For this reason, the absence of multicollinearity is essential to obtain optimal estimates from a multiple regression model.

Ordinary least squares, one of the simplest and most used estimation methods, depends on different binding assumptions regarding the random variables and the nature of independent variables.

In this paper, the  $L_p$ -norm methods are considered not only in order to reduce multicollinearity in the model, but also in order to see some improvements in the parameter estimation.

By using  $L_p$ -norm, a better performance in terms of variance of parameter estimates is always gained in the case of non-Normal symmetric distributions, compared to the least squares situation, even when considering a model with collinear regressors.

When interpreting the simulation results, we notice that the improvements using  $L_p$ -norm estimators in place of least squares are more evident in the case  $\sigma_z = 0.01$  and  $R_{3,12}^2 = 0.99$  (high collinearity) than in the case  $\sigma_z = 1$  and  $R_{3,12}^2 = 0.66$  (medium collinearity). Only in the case of low collinearity, the OLS method gives more efficient estimates.

So, the use of  $L_p$ -norm methods in the presence of stochastic regressors is strongly recommended. This is linked to the characteristics of these methods, that are adaptive procedures with respect to the error component of the model and not to the deterministic one.

Finally, more interesting is the real data application, as it confirms that the coefficients estimated by  $L_p$ -norm can explain reality better than the ones obtained through OLS. It is evident that  $L_p$ -norm estimators, characterized by a lower value of the objective function compared to OLS, allow for a better explanation of the deterministic component of the model.

All the simulations were made using R 3.3.2 and Stata 12.0.

## Appendix 1

See Tables [10](#), [11](#), [12](#), [13](#), [14](#) and [15](#).

**Table 10** Mean and variance of  $\beta_0, \beta_1, \beta_2, \beta_3$  for a multiple regression model ( $\beta_0 = 1, \beta_1 = 2, \beta_2 = 3, \beta_3 = 4$ ) estimated by  $L_p$ -norm method on 5000 samples of size  $n = 30, n = 50, n = 100, n = 200, n = 500, n = 1000$  ( $\sigma_z = 2, R^2 = 0.33$ )

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 = 30								
1.1	0.998	2.418	2.001	2.287	2.994	2.983	4.003	2.923
1.5	0.979	1.872	2.001	2.261	3.004	2.300	4.005	2.221
<b>2.0</b>	<b>1.046</b>	<b>1.579</b>	<b>1.993</b>	<b>1.832</b>	<b>2.974</b>	<b>1.856</b>	<b>3.996</b>	<b>1.832</b>
2.5	0.983	1.331	2.002	1.583	3.015	1.590	4.004	1.514
3.0	1.000	1.107	1.994	1.373	2.998	1.313	3.978	1.336
3.5	0.994	0.980	2.012	1.197	2.997	1.172	4.006	1.171
n = 50								
1.1	0.997	1.251	1.989	1.496	3.008	1.464	4.010	1.517
1.5	0.986	1.062	2.005	1.237	2.994	1.036	4.021	1.257
<b>2.0</b>	<b>0.993</b>	<b>0.855</b>	<b>2.009</b>	<b>1.006</b>	<b>2.994</b>	<b>1.036</b>	<b>4.004</b>	<b>1.026</b>
2.5	0.998	0.736	1.977	0.879	3.016	0.872	3.991	0.881
3.0	1.020	0.626	1.992	0.759	2.980	0.749	3.983	0.727
3.5	0.993	0.548	1.990	0.643	3.004	0.675	4.002	0.648
n = 100								
1.1	1.005	0.556	2.009	0.665	2.975	0.689	4.011	0.689
1.5	0.980	0.479	2.008	0.578	3.015	0.567	4.007	0.593
<b>2.0</b>	<b>1.011</b>	<b>0.407</b>	<b>2.004</b>	<b>0.498</b>	<b>2.999</b>	<b>0.496</b>	<b>3.975</b>	<b>0.479</b>
2.5	0.994	0.345	2.003	0.421	3.004	0.414	4.001	0.409
3.0	0.999	0.295	1.995	0.354	2.993	0.343	4.004	0.353
3.5	1.005	0.243	1.997	0.289	2.998	0.301	4.001	0.294
n = 200								
1.1	1.011	0.259	1.993	0.305	2.982	0.313	3.999	0.305
1.5	1.004	0.246	2.003	0.290	2.999	0.286	3.995	0.294
<b>2.0</b>	<b>0.996</b>	<b>0.207</b>	<b>1.984</b>	<b>0.251</b>	<b>3.013</b>	<b>0.243</b>	<b>4.004</b>	<b>0.240</b>
2.5	0.996	0.167	1.995	0.205	3.007	0.200	3.998	0.200
3.0	0.994	0.139	2.002	0.167	3.006	0.162	4.003	0.168
3.5	1.004	0.118	1.998	0.144	3.000	0.144	3.989	0.144
n = 500								
1.1	0.995	0.096	2.005	0.116	2.998	0.115	4.004	0.113
1.5	0.993	0.093	1.998	0.112	3.009	0.111	4.002	0.113
<b>2.0</b>	<b>0.999</b>	<b>0.081</b>	<b>2.006</b>	<b>0.093</b>	<b>2.994</b>	<b>0.096</b>	<b>4.001</b>	<b>0.099</b>
2.5	1.002	0.066	2.003	0.081	2.993	0.078	4.000	0.081
3.0	1.004	0.054	1.996	0.067	2.996	0.065	4.002	0.067
3.5	0.999	0.047	1.998	0.056	3.000	0.056	3.998	0.056
n = 1000								
1.1	1.000	0.094	1.995	0.119	3.001	0.134	3.999	0.118
1.5	1.004	0.072	2.004	0.103	2.999	0.097	4.001	0.092
<b>2.0</b>	<b>1.006</b>	<b>0.063</b>	<b>1.996</b>	<b>0.098</b>	<b>2.995</b>	<b>0.095</b>	<b>3.991</b>	<b>0.087</b>
2.5	0.999	0.046	2.005	0.075	3.001	0.083	4.000	0.065
3.0	0.992	0.028	2.000	0.030	3.009	0.035	4.009	0.035

**Table 10** continued

p	M( $\beta_0$ )	V( $\beta_0$ )	M( $\beta_1$ )	V( $\beta_1$ )	M( $\beta_2$ )	V( $\beta_2$ )	M( $\beta_3$ )	V( $\beta_3$ )
3.5	1.000	0.013	1.999	0.032	2.999	0.024	3.999	0.031

OLS simulation results are in bold

**Table 11** Mean and variance of  $\beta_0, \beta_1, \beta_2, \beta_3$  for a multiple regression model ( $\beta_0 = 1, \beta_1 = 2, \beta_2 = 3, \beta_3 = 4$ ) estimated by  $L_p$ -norm method on 5000 samples of size  $n = 30, n = 50, n = 100, n = 200, n = 500, n = 1000$  ( $\sigma_z = 1, R^2 = 0.66$ )

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 = 30								
1.1	0.992	0.567	2.014	0.711	2.987	0.685	4.010	0.702
1.5	1.008	0.483	1.996	0.565	3.003	0.590	3.995	0.550
<b>2.0</b>	<b>0.987</b>	<b>0.382</b>	<b>2.012</b>	<b>0.456</b>	<b>3.005</b>	<b>0.450</b>	<b>4.005</b>	<b>0.468</b>
2.5	0.997	0.332	2.000	0.391	2.997	0.382	4.007	0.409
3.0	1.006	0.275	1.996	0.337	2.995	0.332	4.001	0.342
3.5	1.006	0.249	1.996	0.292	3.003	0.299	3.990	0.296
n = 50								
1.1	0.991	0.304	2.003	0.379	3.007	0.369	4.005	0.368
1.5	1.003	0.264	1.997	0.315	3.010	0.322	3.983	0.321
<b>2.0</b>	<b>1.014</b>	<b>0.226</b>	<b>1.994</b>	<b>0.269</b>	<b>2.985</b>	<b>0.261</b>	<b>3.991</b>	<b>0.262</b>
2.5	1.009	0.184	1.995	0.221	2.990	0.217	3.998	0.214
3.0	0.998	0.153	2.005	0.185	2.995	0.186	4.001	0.179
3.5	1.006	0.134	1.999	0.162	2.993	0.163	3.999	0.161
n = 100								
1.1	1.008	0.141	2.003	0.176	2.986	0.167	3.998	0.164
1.5	1.005	0.122	1.987	0.145	3.001	0.147	4.000	0.148
<b>2.0</b>	<b>0.997</b>	<b>0.107</b>	<b>2.004</b>	<b>0.126</b>	<b>2.999</b>	<b>0.128</b>	<b>3.998</b>	<b>0.124</b>
2.5	0.995	0.087	1.999	0.102	3.000	0.105	4.008	0.106
3.0	0.999	0.073	2.003	0.087	2.996	0.087	3.999	0.087
3.5	1.003	0.063	1.996	0.076	3.001	0.074	3.998	0.073
n = 200								
1.1	1.009	0.062	1.993	0.074	2.994	0.076	3.995	0.079
1.5	1.001	0.058	2.003	0.068	3.002	0.071	3.994	0.072
<b>2.0</b>	<b>0.997</b>	<b>0.051</b>	<b>1.999</b>	<b>0.059</b>	<b>3.006</b>	<b>0.060</b>	<b>4.001</b>	<b>0.063</b>
2.5	1.001	0.041	2.001	0.049	2.997	0.051	3.998	0.050
3.0	0.992	0.034	2.008	0.041	3.004	0.042	4.000	0.041
3.5	0.999	0.030	2.002	0.035	2.994	0.037	4.002	0.036
n = 500								
1.1	0.998	0.023	1.998	0.028	3.000	0.028	4.001	0.029
1.5	1.001	0.022	1.996	0.027	2.997	0.027	4.002	0.028
<b>2.0</b>	<b>1.000</b>	<b>0.019</b>	<b>2.000</b>	<b>0.023</b>	<b>3.001</b>	<b>0.022</b>	<b>3.999</b>	<b>0.023</b>
2.5	1.001	0.016	1.998	0.019	3.000	0.019	4.001	0.020
3.0	0.999	0.013	2.000	0.016	2.999	0.017	4.001	0.016

**Table 11** continued

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 = 1000								
3.5	0.997	0.011	2.001	0.014	3.002	0.013	4.000	0.013
1.1	0.998	0.012	2.002	0.014	2.999	0.014	4.000	0.013
1.5	0.998	0.011	2.001	0.013	3.003	0.014	4.001	0.014
<b>2.0</b>	<b>1.002</b>	<b>0.010</b>	<b>2.000</b>	<b>0.012</b>	<b>2.997</b>	<b>0.012</b>	<b>4.000</b>	<b>0.012</b>
2.5	0.997	0.008	2.001	0.009	3.003	0.009	3.999	0.010
3.0	0.998	0.006	2.000	0.008	3.001	0.008	3.999	0.008
3.5	1.001	0.006	2.001	0.007	2.999	0.007	3.999	0.007

OLS simulation results are in bold

**Table 12** Mean and variance of  $\beta_0, \beta_1, \beta_2, \beta_3$  for a multiple regression model ( $\beta_0 = 1, \beta_1 = 2, \beta_2 = 3, \beta_3 = 4$ ) estimated by  $L_p$ -norm method on 5000 samples of size n = 30, n = 50, n = 100, n = 200, n = 500, n = 1000 ( $\sigma_z = 0.01, R^2 = 0.99$ )

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 = 30								
1.1	0.981	0.599	1.999	0.491	3.000	0.585	3.999	0.597
1.5	1.009	0.590	2.000	0.473	2.999	0.573	4.000	0.561
<b>2.0</b>	<b>1.004</b>	<b>0.487</b>	<b>2.000</b>	<b>0.452</b>	<b>3.000</b>	<b>0.470</b>	<b>4.000</b>	<b>0.561</b>
2.5	1.007	0.278	1.999	0.433	3.999	0.380	4.000	0.490
3.0	1.006	0.273	1.999	0.421	2.999	0.342	4.000	0.445
3.5	1.003	0.245	1.999	0.403	2.999	0.304	4.000	0.391
n = 50								
1.1	1.007	0.309	1.999	0.384	2.999	0.360	3.999	0.315
1.5	0.989	0.261	1.999	0.315	3.000	0.316	3.999	0.310
<b>2.0</b>	<b>1.018</b>	<b>0.220</b>	<b>1.000</b>	<b>0.260</b>	<b>3.000</b>	<b>0.264</b>	<b>4.000</b>	<b>0.258</b>
2.5	1.003	0.206	2.000	0.186	3.000	0.185	3.999	0.222
3.0	1.007	0.152	1.999	0.172	3.000	0.185	3.999	0.185
3.5	0.999	0.135	1.999	0.164	2.999	0.151	4.000	0.166
n = 100								
1.1	0.982	0.136	1.999	0.166	3.000	0.163	4.000	0.164
1.5	0.994	0.128	1.999	0.151	2.999	0.147	3.999	0.148
<b>2.0</b>	<b>0.979</b>	<b>0.088</b>	<b>2.000</b>	<b>0.125</b>	<b>2.999</b>	<b>0.133</b>	<b>3.999</b>	<b>0.123</b>
2.5	0.999	0.064	2.000	0.104	2.999	0.130	4.000	0.104
3.0	1.000	0.055	2.000	0.054	3.000	0.052	3.000	0.056
3.5	0.999	0.050	2.000	0.054	2.999	0.038	3.999	0.049
n = 200								
1.1	1.000	0.055	1.999	0.056	2.999	0.061	4.000	0.056
1.5	0.999	0.053	1.999	0.054	2.999	0.053	3.999	0.053
<b>2.0</b>	<b>1.001</b>	<b>0.049</b>	<b>2.000</b>	<b>0.054</b>	<b>3.000</b>	<b>0.050</b>	<b>3.999</b>	<b>0.048</b>
2.5	0.999	0.040	1.999	0.050	3.000	0.047	4.000	0.044
3.0	1.000	0.036	2.000	0.047	3.000	0.045	4.000	0.042

**Table 12** continued

p	M( $\beta_0$ )	V( $\beta_0$ )	M( $\beta_1$ )	V( $\beta_1$ )	M( $\beta_2$ )	V( $\beta_2$ )	M( $\beta_3$ )	V( $\beta_3$ )
3.5	1.000	0.035	2.000	0.040	2.999	0.041	4.000	0.039
n = 500								
1.1	1.000	0.038	1.999	0.036	2.999	0.036	2.999	0.034
1.5	1.000	0.033	2.000	0.033	3.000	0.032	3.999	0.030
<b>2.0</b>	<b>0.999</b>	<b>0.030</b>	<b>2.000</b>	<b>0.029</b>	<b>3.000</b>	<b>0.030</b>	<b>4.000</b>	<b>0.029</b>
2.5	0.999	0.028	2.000	0.025	3.000	0.026	3.999	0.026
3.0	1.000	0.026	1.999	0.024	2.999	0.026	4.000	0.026
3.5	1.000	0.026	2.000	0.023	3.000	0.024	4.000	0.024
n = 1000								
1.1	1.000	0.034	2.000	0.036	2.999	0.035	4.000	0.031
1.5	0.999	0.031	2.000	0.032	2.999	0.030	3.999	0.026
<b>2.0</b>	<b>0.999</b>	<b>0.025</b>	<b>1.999</b>	<b>0.027</b>	<b>2.999</b>	<b>0.027</b>	<b>3.999</b>	<b>0.021</b>
2.5	1.000	0.017	2.000	0.026	3.000	0.022	4.000	0.018
3.0	1.000	0.012	2.000	0.021	3.000	0.019	4.000	0.013
3.5	1.000	0.011	1.999	0.015	2.999	0.013	4.000	0.011

OLS simulation results are in bold

**Table 13** Mean and variance of  $\beta_0, \beta_1, \beta_2, \beta_3$  for a multiple regression model ( $\beta_0 = 1, \beta_1 = 2, \beta_2 = 3, \beta_3 = 4$ ) estimated by  $L_p$ -norm method ( $L_{pmin}$ ) on 5000 samples of size  $n = 30, n = 50, n = 100, n = 200, n = 500, n = 1000$  ( $\sigma_z = 2, R^2 = 0.33$ )

p	M( $\beta_0$ )	V( $\beta_0$ )	M( $\beta_1$ )	V( $\beta_1$ )	M( $\beta_2$ )	V( $\beta_2$ )	M( $\beta_3$ )	V( $\beta_3$ )	M( $\hat{p}$ )	V( $\hat{p}$ )	Conv.
n = 30											
1.1	0.965	2.604	2.058	3.009	2.998	3.137	4.008	3.149	1.428	1.922	21
1.5	1.022	2.200	1.957	2.684	2.992	2.740	3.992	2.647	1.994	4.328	44
2.0	0.994	1.837	1.999	2.270	2.994	2.129	4.044	2.271	2.530	6.931	55
2.5	1.012	1.562	1.984	1.978	2.973	1.908	4.007	1.919	3.126	9.316	94
3.0	0.982	1.396	2.034	1.699	3.027	1.751	3.694	1.763	3.625	11.078	107
3.5	1.021	1.260	2.012	1.494	2.971	1.515	3.995	1.549	4.096	12.903	129
n = 50											
1.1	1.025	1.285	2.008	1.579	2.991	1.569	3.959	1.550	1.203	0.241	3
1.5	0.998	1.140	2.003	1.436	2.978	1.395	3.998	1.392	1.604	1.126	8
2.0	0.989	1.014	2.007	1.234	2.992	1.207	4.006	1.224	2.198	2.747	19
2.5	1.022	0.849	1.997	0.997	2.997	1.062	3.968	1.025	2.932	5.406	28
3.0	0.994	0.722	2.014	0.885	3.009	0.911	3.992	0.882	3.506	7.138	55
3.5	1.013	0.656	1.990	0.788	2.980	0.780	4.002	0.774	4.032	8.671	79
n = 100											
1.1	1.005	0.544	2.012	0.669	2.981	0.648	3.997	0.673	1.137	0.036	2
1.5	1.004	0.535	2.012	0.596	2.996	0.611	3.980	0.635	1.489	0.056	3
2.0	0.998	0.454	1.985	0.544	3.011	0.528	4.006	0.530	2.021	0.577	9
2.5	0.998	0.377	1.989	0.450	3.012	0.456	4.002	0.455	2.632	1.520	12
3.0	0.999	0.322	1.993	0.375	3.012	0.405	4.000	0.381	3.232	2.800	15
3.5	0.999	0.279	1.999	0.337	3.001	0.332	4.000	0.331	3.942	5.043	27

**Table 13** continued

p	M( $\beta_0$ )	V( $\beta_0$ )	M( $\beta_1$ )	V( $\beta_1$ )	M( $\beta_2$ )	V( $\beta_2$ )	M( $\beta_3$ )	V( $\beta_3$ )	M( $\hat{\rho}$ )	V( $\hat{\rho}$ )	Conv.
n = 200											
1.1	1.001	0.258	1.990	0.307	2.998	0.311	4.005	0.316	1.118	0.017	0
1.5	1.010	0.248	1.986	0.298	2.994	0.301	4.001	0.308	1.489	0.056	3
2.0	1.002	0.218	1.989	0.251	3.002	0.255	3.999	0.250	1.992	0.146	2
2.5	0.998	0.175	1.996	0.210	2.996	0.204	4.011	0.216	2.523	0.387	3
3.0	1.004	0.148	1.995	0.168	2.999	0.180	3.997	0.183	3.101	0.950	6
3.5	1.000	0.128	2.004	0.158	2.996	0.150	3.998	0.150	3.677	1.850	10
n = 500											
1.1	1.000	0.092	1.996	0.113	3.004	0.112	3.997	0.113	1.010	0.007	0
1.5	0.999	0.097	1.997	0.110	3.000	0.110	3.998	0.119	1.498	0.121	0
2.0	1.000	0.081	2.003	0.097	3.003	0.100	3.994	0.098	1.998	0.051	1
2.5	1.000	0.068	2.006	0.079	2.998	0.082	3.997	0.082	2.497	0.105	2
3.0	0.998	0.055	2.008	0.069	2.994	0.067	3.999	0.065	3.027	0.233	4
3.5	1.001	0.046	2.004	0.055	2.998	0.058	3.997	0.055	3.555	0.355	5
n = 1000											
1.1	1.002	0.083	1.996	0.117	3.000	0.119	3.997	0.126	1.099	0.004	0
1.5	0.996	0.083	2.010	0.126	2.998	0.120	3.999	0.126	1.498	0.010	0
2.0	0.999	0.066	1.995	0.096	3.011	0.089	3.995	0.097	2.000	0.023	0
2.5	1.000	0.037	1.999	0.062	3.002	0.067	4.000	0.065	2.502	0.051	1
3.0	0.999	0.014	2.006	0.036	2.993	0.035	4.002	0.036	3.012	0.105	0
3.5	0.995	0.009	2.008	0.015	2.998	0.012	4.001	0.012	3.528	0.201	1

**Table 14** Mean and variance of  $\beta_0, \beta_1, \beta_2, \beta_3$  for a multiple regression model ( $\beta_0 = 1, \beta_1 = 2, \beta_2 = 3, \beta_3 = 4$ ) estimated by  $L_p$ -norm method ( $L_{p_{min}}$ ) on 5000 samples of size  $n = 30, n = 50, n = 100, n = 200, n = 500, n = 1000$  ( $\sigma_z = 1, R^2 = 0.66$ )

p	M( $\beta_0$ )	V( $\beta_0$ )	M( $\beta_1$ )	V( $\beta_1$ )	M( $\beta_2$ )	V( $\beta_2$ )	M( $\beta_3$ )	V( $\beta_3$ )	M( $\hat{\rho}$ )	V( $\hat{\rho}$ )	Conv.
n = 30											
1.1	1.021	0.637	1.985	0.794	2.986	0.719	3.994	0.774	1.428	1.978	10
1.5	0.994	0.576	1.994	0.726	3.031	0.707	3.989	0.709	1.660	3.106	21
2.0	0.999	0.451	2.009	0.547	2.985	0.575	4.016	0.556	2.516	6.837	63
2.5	0.995	0.410	1.996	0.499	3.008	0.486	4.006	0.498	3.160	9.523	70
3.0	0.985	0.360	1.997	0.427	3.009	0.422	4.023	0.448	3.751	11.871	119
3.5	0.980	0.326	2.015	0.388	3.009	0.398	4.019	0.393	4.085	12.835	132
n = 50											
1.1	1.002	0.317	1.983	0.388	3.009	0.327	4.005	0.387	1.204	0.332	1
1.5	0.984	0.299	2.000	0.361	3.012	0.350	4.014	0.351	1.591	1.080	7
2.0	1.002	0.255	1.989	0.310	3.015	0.313	3.987	0.306	2.242	3.058	25
2.5	1.009	0.211	1.993	0.254	2.987	0.259	3.997	0.258	2.895	5.142	34
3.0	1.007	0.180	1.990	0.217	2.999	0.225	3.994	0.217	3.520	7.231	51
3.5	0.997	0.158	2.002	0.193	2.996	0.196	4.009	0.193	4.093	9.154	66
n = 100											
1.1	1.000	0.141	1.996	0.167	3.000	0.170	4.001	0.171	1.145	0.040	0

**Table 14** continued

p	M( $\beta_0$ )	V( $\beta_0$ )	M( $\beta_1$ )	V( $\beta_1$ )	M( $\beta_2$ )	V( $\beta_2$ )	M( $\beta_3$ )	V( $\beta_3$ )	M( $\hat{\rho}$ )	V( $\hat{\rho}$ )	Conv.
1.5	0.998	0.132	2.002	0.156	2.996	0.160	4.003	0.160	1.495	0.164	1
2.0	1.002	0.113	1.995	0.141	2.999	0.132	3.998	0.131	2.038	0.569	3
2.5	0.997	0.092	1.994	0.113	3.013	0.114	3.997	0.118	2.653	1.603	3
3.0	0.998	0.083	2.004	0.100	2.996	0.099	4.004	0.098	3.248	2.978	11
3.5	1.000	0.066	2.002	0.082	3.000	0.080	3.999	0.081	3.970	4.760	21
n = 200											
1.1	0.997	0.065	1.998	0.077	3.002	0.078	4.002	0.077	1.121	0.017	0
1.5	0.991	0.061	2.004	0.075	3.008	0.074	4.001	0.075	1.489	0.056	0
2.0	1.002	0.053	1.999	0.063	3.000	0.064	3.996	0.063	2.005	0.149	3
2.5	0.997	0.043	1.999	0.052	3.007	0.051	4.000	0.052	2.547	0.417	1
3.0	0.997	0.036	2.001	0.044	3.001	0.045	4.000	0.043	3.105	0.934	4
3.5	0.996	0.032	2.001	0.038	3.003	0.038	4.002	0.036	3.682	1.958	3
n = 500											
1.1	1.000	0.023	2.001	0.029	3.001	0.029	3.995	0.028	1.104	0.007	0
1.5	0.999	0.022	1.999	0.028	2.999	0.028	3.995	0.027	1.494	0.020	2
2.0	1.004	0.020	1.996	0.024	2.993	0.025	4.000	0.024	1.994	0.050	0
2.5	0.999	0.017	2.000	0.020	3.001	0.021	3.998	0.020	2.509	0.010	0
3.0	0.998	0.013	1.999	0.017	3.001	0.017	4.001	0.016	3.023	0.231	3
3.5	0.998	0.012	1.999	0.014	2.999	0.014	4.003	0.015	3.457	0.500	1
n = 1000											
1.1	0.998	0.011	1.999	0.014	3.002	0.013	4.000	0.014	1.100	0.004	0
1.5	1.000	0.011	2.000	0.014	3.000	0.014	3.998	0.014	1.498	0.009	0
2.0	1.000	0.010	2.000	0.011	3.001	0.012	3.996	0.012	1.997	0.023	0
2.5	1.001	0.008	1.998	0.010	3.000	0.009	3.999	0.010	2.505	0.050	1
3.0	0.999	0.006	2.001	0.008	3.001	0.008	3.998	0.008	3.006	0.100	0
3.5	0.998	0.005	1.999	0.007	3.000	0.007	4.002	0.006	3.522	0.203	1

**Table 15** Mean and variance of  $\beta_0, \beta_1, \beta_2, \beta_3$  for a multiple regression model ( $\beta_0 = 1, \beta_1 = 2, \beta_2 = 3, \beta_3 = 4$ ) estimated by  $L_p$ -norm method ( $L_{pmin}$ ) on 5000 samples of size  $n = 30, n = 50, n = 100, n = 200, n = 500, n = 1000$  ( $\sigma_z = 0.01, R^2 = 0.99$ )

p	M( $\beta_0$ )	V( $\beta_0$ )	M( $\beta_1$ )	V( $\beta_1$ )	M( $\beta_2$ )	V( $\beta_2$ )	M( $\beta_3$ )	V( $\beta_3$ )	M( $\hat{\rho}$ )	V( $\hat{\rho}$ )	Conv.
n = 30											
1.1	0.998	0.629	1.999	0.739	3.000	0.741	4.000	0.755	1.400	1.509	0
1.5	1.000	0.527	2.000	0.642	2.999	0.643	3.999	0.652	1.889	3.555	1
2.0	0.998	0.429	2.000	0.526	2.999	0.541	4.001	0.536	2.463	5.443	1
2.5	0.999	0.369	1.999	0.404	2.999	0.451	3.999	0.473	2.969	7.292	2
3.0	1.000	0.326	2.000	0.404	3.000	0.400	4.000	0.404	3.492	8.910	1
3.5	1.000	0.301	1.999	0.349	2.999	0.372	4.000	0.359	3.820	9.966	4
n = 50											
1.1	1.000	0.310	1.999	0.380	3.000	0.374	4.000	0.384	1.215	0.326	0
1.5	1.000	0.289	1.999	0.349	3.000	0.350	3.999	0.346	1.594	0.935	0
2.0	1.000	0.239	1.999	0.285	3.000	0.295	3.999	0.295	2.206	2.467	0



**Table 15** continued

p	M( $\beta_0$ )	V( $\beta_0$ )	M( $\beta_1$ )	V( $\beta_1$ )	M( $\beta_2$ )	V( $\beta_2$ )	M( $\beta_3$ )	V( $\beta_3$ )	M( $\hat{\rho}$ )	V( $\hat{\rho}$ )	Conv.
2.5	1.000	0.203	1.999	0.243	2.999	0.253	4.000	0.249	2.824	4.165	1
3.0	0.999	0.180	2.000	0.213	3.000	0.214	3.999	0.214	3.386	5.913	1
3.5	0.999	0.156	2.000	0.188	3.000	0.191	3.999	0.182	3.935	7.458	1
n = 100											
1.1	1.000	0.131	2.000	0.162	3.000	0.164	3.999	0.163	1.139	0.038	0
1.5	1.000	0.127	2.000	0.156	2.999	0.152	3.999	0.157	1.487	0.151	0
2.0	1.000	0.096	1.999	0.133	3.000	0.130	4.000	0.131	2.041	0.506	0
2.5	0.999	0.059	2.000	0.132	2.999	0.129	4.000	0.114	2.610	1.237	0
3.0	0.999	0.035	2.000	0.122	3.000	0.046	3.999	0.069	3.276	2.723	0
3.5	1.000	0.027	1.999	0.064	3.000	0.037	4.000	0.061	3.824	3.942	0
n = 200											
1.1	0.999	0.064	2.000	0.069	3.000	0.077	4.000	0.97	1.119	0.018	0
1.5	0.999	0.056	2.000	0.066	3.000	0.062	4.000	0.089	1.502	0.058	0
2.0	1.000	0.041	1.999	0.044	3.000	0.048	3.999	0.059	1.998	0.147	0
2.5	1.000	0.040	1.999	0.039	2.999	0.042	3.999	0.048	2.558	0.414	0
3.0	0.999	0.039	2.000	0.037	2.999	0.039	4.000	0.044	3.086	0.903	0
3.5	1.000	0.027	1.999	0.029	3.000	0.033	4.000	0.037	3.683	1.741	0
n = 500											
1.1	0.999	0.032	2.000	0.033	2.999	0.033	4.000	0.031	1.103	0.009	0
1.5	1.000	0.029	2.000	0.030	2.999	0.031	4.000	0.029	1.502	0.033	0
2.0	0.999	0.026	1.999	0.027	3.000	0.027	3.999	0.025	1.978	0.161	0
2.5	1.000	0.023	2.000	0.025	2.999	0.026	3.999	0.024	2.502	0.311	0
3.0	0.999	0.020	2.000	0.023	2.999	0.025	4.000	0.021	3.091	0.937	0
3.5	0.999	0.016	1.999	0.020	3.000	0.023	3.999	0.019	3.588	1.457	0
n = 1000											
1.1	0.999	0.015	2.000	0.017	3.000	0.016	4.000	0.013	1.108	0.055	0
1.5	1.000	0.012	1.999	0.013	2.999	0.013	4.000	0.011	1.501	0.049	0
2.0	1.000	0.007	2.000	0.010	2.999	0.008	3.999	0.010	2.002	0.063	0
2.5	0.999	0.007	2.000	0.009	3.000	0.005	4.000	0.006	2.505	0.107	0
3.0	1.000	0.005	1.999	0.007	3.000	0.005	3.999	0.005	2.998	0.146	0
3.5	1.000	0.004	2.000	0.006	3.000	0.004	3.999	0.003	3.532	0.208	0

## Appendix 2

See Table 16.

**Table 16** Fatal occupational accidents (FI), gross domestic product (GDP), unemployment rate (UNR) and working hours (WH), for UK (1971–2007). *Source* Mouza and Targoudtzidis (2012)

Year	FI	GDP	UNR	WH
1971	780	3.823	3.4	1905
1972	669	4.117	3.7	1871
1973	763	4.651	2.7	1927
1974	651	4.989	2.6	1888
1975	620	5.431	4.1	1883
1976	584	5.904	5.4	1866
1977	524	6.425	5.8	1847
1978	499	7.108	5.7	1826
1979	492	7.891	4.7	1818
1980	440	8.413	5.8	1773
1981	441	9.077	8.8	1715
1982	472	9.843	10.1	1730
1983	448	10.599	10.9	1717
1984	438	11.272	10.9	1733
1985	400	12.005	11.2	1766
1986	355	12.736	11.3	1768
1987	361	13.656	10.3	1758
1988	529	14.805	8.5	1798
1989	370	15.674	7.1	1786
1990	346	16.361	6.9	1771
1991	297	16.638	8.6	1767
1992	276	17.003	9.8	1732
1993	245	17.740	10.2	1726
1994	191	18.842	9.3	1740
1995	209	19.755	8.5	1743
1996	207	20.977	7.9	1742
1997	212	22.435	6.8	1740
1998	188	23.311	6.1	1734
1999	162	24.249	5.9	1723
2000	213	26.041	5.4	1711
2001	206	27.585	5	1714
2002	183	28.888	5.1	1696
2003	168	29.863	5	1677
2004	172	31.747	4.7	1672
2005	164	32.695	4.8	1676
2006	191	34.137	5.4	1669
2007	178	35.669	5.3	1670

## References

- Agrò, G.: Maximum likelihood estimation for the exponential power function parameters. *Commun. Stat. Simul. Comput.* **24**, 523–536 (1995)
- Agrò, G.: Parameter orthogonality and conditional profile likelihood: the exponential power function case. *Commun. Stat. Theory Methods* **28**(8), 1759–1768 (1999)
- Akdeniz, F., Akdeniz Duran, E., Roozbeh, M., Arashi, M.: Efficiency of the generalized difference-based Liu estimators in semiparametric regression models with correlated errors. *J. Stat. Comput. Simul.* **85**(1), 147–165 (2015)
- Alabi, O.O., Ayinde, K., Olatayo, T.O.: On effects of multicollinearity on the power rates of the ordinary least squares estimators. *J. Math. Stat.* **4**, 75–80 (2008).
- Alabi, O.O., Olatayo, T.O., Afolabi, F.R.: Empirical determination of the tolerable sample size for Ols estimator in the presence of multicollinearity ( $\rho$ ). *Appl. Math.* **5**(13), 1870 (2014)
- Armor, D.J., Cotla, C.R., Stratmann, T.: Spurious relationships arising from aggregate variables in linear regression. *Qual. Quant.* **51**(3), 1359–1379 (2017)
- Ayinde, K.: A comparative study of the performances of the OLS and some GLS estimators when stochastic regressors are both collinear and correlated with error terms. *J. Math. Stat.* **3**(4), 196–200 (2007)
- Besley, D.A., Kuh, E., Welsch, R.E.: *Regression diagnostics: identifying influential data and sources of collinearity*. Wiley, New York (1980)
- Bottazzi, G., Secchi, A.: A new class of asymmetric exponential power densities with applications to economics and finance. *Ind. Corp. Change* **20**(4), 991–1030 (2011)
- Bowerman, B.L., O'Connell, R.T.: *Forecasting and Time Series: An Applied Approach*, 3rd edn (1993)
- Bring, J.: A Geometric Approach to Compare Variables in a Regression Model. *The American Statistician*, vol. 50, pp. 57–62. Wiley, New York (1996)
- Cassel, C., Westlund, A.H., Hackl, P.: Robustness of partial least squares method for estimating latent variable quality structures. *J. Appl. Stat.* **26**, 435–448 (1999)
- Chatelain, J., Ralf, K.: Spurious regressions and near-multicollinearity, with an application to aid, policies and growth. *J. Macroecon.* **39**, 85–96 (2014)
- Chiodi, M.: Procedures for generating pseudo-random numbers from a normal distribution of order  $p$ . *Riv. Stat. Appl.* **19**, 7–26 (1986)
- Chiodi, M.: Generation of pseudo random variates from a normal distribution of order  $p$ . *Stat. Appl. (Italian J. Appl. Stat.)* **7**(4), 401–416 (1995)
- Counihan, M.: A structured approach to fortran 77 programming, TMR ellis, Addison-Wesley International Computer Science Series. *Softw. Pract. Exp.* **15**(3), 319 (1985)
- Cox, D.R., Reid, N.: A note of efficiency of least squares estimates. *J. R. Stat. Soc. B* **30**, 841 (1987)
- Curto, J., Pinto, J.: The corrected VIF (CVIF). *J. Appl. Stat.* **38**(7), 1499–1507 (2011)
- Destefanis, S., Porzio, G.C.: Dynamic graphics and model validation: an application to best-practice production functions. *Appl. Stoch. Models Bus. Ind.* **15**(4), 259–267 (1999)
- Dijkstra, T.K.: Ridge regression and its degrees of freedom. *Qual. Quant.* **48**(6), 3185–3193 (2014)
- Eklblom, H., Henriksson, S.:  $L_p$ -criteria for the estimation of location parameters. *SIAM J. Appl. Math.* **17**(6), 1130–1141 (1969)
- Everitt, B., Hand, D.: *The Statistical Consultant in Action*. Cambridge University Press, New York (1987)
- Fletcher, R., Reeves, C.M.: Function minimization by conjugate gradients. *Comput. J.* **7**(2), 149–154 (1964)
- Forsythe, A.B.: Robust estimation of straight line regression coefficients by minimizing  $p$ th power deviations. *Technometrics* **14**, 159–166 (1972)
- Frisch, R.: *Statistical Confluence by Means of Complete Regression Systems*, vol. 5. Universitetes økonomiske Institut, Copenhagen (1934)
- Garcia, C., Pérez, J.G., Liria, J.S.: The raise method. An alternative procedure to estimate the parameters in presence of collinearity. *Qual. Quan.* **45**, 403–423 (2011)
- Geary, R.C.: Moments of the ratio of the mean deviation to the standard deviation for normal samples. *Biometrika* **28**, 295–307 (1936)
- Giacalone, M.: Parameter evaluation of an exponential power function by simulation study. In: *Proceedings in Computational Statistics, Compstat 1996, Barcelona, Spain* (1996)
- Giacalone, M.:  $L_p$ -norm estimators for nonlinear regression models. *Student* **2**, 119–130 (1997)
- Giacalone, M.:  $L_p$ -norm estimation for linear regression models: a comparative study in presence of multicollinearity. In: *Proceedings in Computational Statistics, Compstat 2002, Berlin, Germany* (2002)
- Giacalone, M., Richiusa, R.:  $L_p$ -norm estimation: some simulation studies in presence of multicollinearity. *Student* **5**, 235–247 (2006)

- Gonin, R., Money, A.H.: A review of computational methods for solving the nonlinear L1 norm estimation problem. In: Dodge Y. (ed.) *Statistical Data Analysis Based on the L1 Norm and Related Methods*, Dodge, North Holland (1987)
- Gonin, R., Money, A.H.: *Nonlinear  $L_p$ -Norm Estimation*. Marcel Dekker Inc., New York (1989)
- Gower, J.C., Blasius, J.: Multivariate prediction with nonlinear principal components analysis: theory. *Qual. Quant.* **39**(4), 359–372 (2005)
- Griffiths, W.E., Hajargasht, G.: Some models for stochastic frontiers with endogeneity. *J. Econom.* **190**(2), 341–348 (2016)
- Harter, H.L.: The nonuniqueness of absolute values regression. *Commun. Stat. Simul. Comput.* **A6**, 829–838 (1977)
- Hoerl, A.E., Kennard, R.W.: Ridge regression: applications to non orthogonal problems. *Technometrics* **12**(1), 69–82 (1970)
- Kendall, M.C., Stuart, A.: *The Advanced Theory of Statistics*. Griffin, London (1966)
- Krone, T., Albers, C.J., Timmerman, M.E.: A comparative simulation study of AR (1) estimators in short time series. *Qual. Quant.* **51**(1), 1–21 (2017)
- Lazaridis, A.: A note regarding the condition number: the case of spurious and latent multicollinearity. *Qual. Quant.* **41**(1), 123–135 (2007)
- Leahy, K.: Multicollinearity: when the solution is the problem. In: Rud, O.P. (ed.) *Data Mining Cookbook*, pp. 106–108. Wiley, London (2000)
- Leti, G.: *Statistica Descrittiva. Il Mulino*, Bologna (1983)
- Lin, F.J.: Solving multicollinearity in the process of fitting regression model using the nested estimate procedure. *Qual. Quant.* **42**(3), 417–426 (2008)
- Lunetta, G.: Di Alcune Distribuzioni Deducibili da una Generalizzazione dello Schema della Curva Normale. *Ann. Fac. Econ. Commun. Palermo* **20**, 117–143 (1966)
- Maddala, G.S.: *Econometrics*. McGraw Hills Publishing Company, New York (1977)
- Menard, S.: *Applied Logistic Regression Analysis: Sage University Series on Quantitative Applications in the Social Sciences*. Sage, Thousand Oaks (1995)
- Masson, K., Shukur, G., Sjolander, P.: A new ridge regression causality test in the presence of multicollinearity. *Commun. Stat. Theory Methods* **43**, 235–248 (2014)
- McCormick, G.P.: *Nonlinear Programming: Theory, Algorithms, and Applications*. Wiley, New York (1983)
- Mineo, A.: The Norm-p Estimation of Location, Scale and Simple Linear Regression Parameters. *Lectures Notes in Statistics-Statistical Modelling*, pp. 222–223. Springer, Trento (1989)
- Mineo, A.M.: 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* **2**, 147–154 (1994)
- Mineo, A.M.: On the estimation of the structure parameter of a normal distribution of order p. *Statistica* **1**, 109–122 (2003)
- Mineo, A.M., Ruggieri, M.: A software tool for the exponential power distribution: the normalp package. *J. Stat. Softw.* **12**(4), 1–24 (2005)
- Montgomery, D.C., Peck, E.A., Vining, G.G.: *Introduction to Linear Regression Analysis*, vol. 821. Wiley, London (2012)
- Money, A.H., Affleck-Graves, J.F., Hart, M.L., Barr, G.D.I.: The linear regression model:  $L_p$  norm estimation and the choice of p. *Commun. Stat. Simul. Comput.* **11**(1), 89–109 (1982)
- Mouza, A.M., Targoutzidis, A.: The impact of the economic cycle on fatal injuries. The case of UK 1971–2007. *Qual. Quant.* **46**, 1917–1929 (2012)
- Muniz, G., Kibria, B.G.: On some ridge regression estimators: an empirical comparison. *Commun. Stat. Simul. Comput.* **38**(3), 621–630 (2009)
- O'Brien, R.M.: A caution regarding rules of thumb for variance inflation factors. *Qual. Quant.* **41**(5), 673–690 (2007)
- Pagel, M.D., Lunneberg, C.E.: Empirical evaluation of ridge regression. *Psychol. Bull.* **97**, 342–355 (1985)
- Pati, K.D., Adnan, R., Rasheed, B.A., MD, J. M. A.: Estimation parameters using Bisquare weighted robust ridge regression BRLTS estimator in the presence of multicollinearity and outliers. In: Salleh, S., Aris, N.A., Bahar, A., Zainuddin, Z.M., Maan, N., Lee, M.H., Ahmad, T. (eds.) *AIP Conference Proceedings*. AIP Publishing (2016)
- Saville, D., Wood, G.R.: *Statistical Methods: The Geometric Approach*. Springer, New York (1991)
- Shieh, G.: Clarifying the role of mean centring in multicollinearity of interaction effects. *Br. J. Math. Stat. Psychol.* **64**(3), 462–477 (2011)
- Sposito, V.A.: On unbiased  $L_p$  regression estimators. *J. Am. Stat. Assoc.* **77**(379), 652–653 (1982)
- Subbotin, M.: On the law of frequency of error. *Математический сборник.* **31**(2), 296–301 (1923)

- Tran, K., Tsionas, E.: GMM estimation of stochastic frontier model with endogenous regressors. *Econ. Lett.* **118**, 233–236 (2013)
- Vargha, A., Bergman, L.R., Delaney, H.D.: Interpretation problems of the partial correlation with non-normally distributed variables. *Qual. Quant.* **47**(6), 3391–3402 (2013)
- Vianelli, S.: La Misura della Variabilità Condizionata in uno Schema Generale delle Curve Normali di Frequenza. *Statistica* **23**(4), 447–474 (1963)