Robustness or Efficiency: A Test to Solve the Dilemma

Catherine Dehon, Marjorie Gassner, Vincenzo Verardi*
Université Libre de Bruxelles†

April 8, 2005

Abstract

When dealing with the presence of outliers in a dataset, the problem of choosing between the classical ordinary least squares and robust regression methods is sometimes addressed inadequately. In this article, we propose using a Hausman-type test to determine whether a robust S-estimator is more appropriate than an ordinary least squares one in a multiple linear regression framework, on the basis of the trade-off between robustness and efficiency. An economic example is provided to illustrate the usefulness of the test.

KEYWORDS: Efficiency, Hausman Test, Linear Regression, Robustness, S-estimator
JEL classification codes: C12, C13

*Authors e-mails: cdehon@ulb.ac.be, mgassner@ulb.ac.be, vverardi@ulb.ac.be. The authors would like to thank all their colleagues at ECARES and CEE for helpful comments. All remaining errors are their responsability.

†ECARES (European Center for Advanced Research in Economics and Statistics), CEE (Centre de l'Economie de l'Education) and ISRO (Institut de Statistique et de Recherche Operationnelle). CP 139, Av. F.D. Roosevelt, 50. B-1050 Brussels, Belgium. Tel: +32-2-6503843, Fax: +32-2-6503530.

1 Introduction

In applied economics and econometrics, it has always been highlighted that even if a small amount of data behaves differently from the vast majority of the observations, classical estimations may be affected, leading to results that are not representative of the population. In other words, the presence of outliers might bias the results. Various techniques such as standardized residuals, studentized residuals, Cook distances, etc. have been proposed to identify these non-standard data. Unfortunately, they all suffer from the fact that they are based on residuals that are calculated on a non-robust regression line (or hyperplane). This led several authors to develop methods to estimate regression lines which are not sensible to the presence of outliers. Others suggest the use of graphical tools (based on plotting robust distances against residuals obtained with robust estimation methods) to detect the different types of outliers (Rousseeuw and van Zomeren, 1990). The advantage of robust methods is that they yield estimations resistant to outliers but, unfortunately, the price to pay is a loss of efficiency. An essential question that comes to mind at this point is whether the gain in unbiasedness is more valuable than the corresponding loss in efficiency. The answer to this question is not trivial. To decide if it is more adequate to use a classical regression technique or a robust one, a statistical test is needed, but unfortunately, to the best of our knowledge, no such test exists. The aim of this paper is to create one that will help applied econometricians to decide whether it is more pertinent to use a robust or a standard technique. The

general idea is simple: if the influence of the outliers is limited, the estimated regression parameters obtained by ordinary least squares (LS) and by a robust method should be similar, but LS will be preferred as it is more efficient. In the opposite case, a robust estimator will be preferred.

Durbin (1954) and Wu (1973), introduced the idea that if a model is correctly specified, two consistent methods should produce estimates that are very close. Hausman (1978), following a similar reasoning, developed a test that is based on looking for a statistically significant difference between an estimator that is consistent whether or not the null is true, and an estimator that is efficient (and consistent) under the null hypothesis, but inconsistent otherwise. He proves that asymptotically the test statistic has a chi-square distribution, with a number of degrees of freedom equal to the number of unknown regression parameters when no misspecification is present. This type of test is widely used in econometrics to detect endogeneity or to determine if random effects are appropriate in a panel data framework. In all these cases, the underlying idea is to test for misspecification. What we want to bring forward here is different: imagine we have a well-specified model but a bias appears because of the presence of outliers. As far as we know, no clear test is available to see if, in this context, a robust method is more appropriate than a classical one. In this paper, we show that a Hausman-type test can be used to check for this.

The paper is divided into five sections. After this short introduction, in the second section we introduce the type of test we propose. In the third section we present some simulations, in the fourth we apply the test to some real economic

data, and in the final section, we conclude.

2 A Hausman-type test

Assume we want to estimate a regression model of the type

$$y_i = \theta_0 + x_{i1}\theta_1 + ... + x_{ip-1}\theta_{p-1} + \varepsilon_i \qquad \text{for } i = 1, ..., n \tag{1}$$

where n is the sample size, $x_{i1}, ..., x_{ip-1}$ are the explanatory variables, y_i the dependent variable and ε_i the error term. We suppose that the errors ε_i are independent of the explanatory variables and i.i.d. according to the normal distribution $N(0, \sigma)$, where σ is the residual scale parameter. The vector of regression parameters is $\theta = [\theta_0, ..., \theta_{p-1}]'$. To estimate it, the classical ordinary least squares methodology is the most commonly used; it minimizes the sum of the squared residuals. More precisely:

$$\hat{\theta}_{LS} = \arg\min_{\hat{\theta}} \sum_{i=1}^{n} r_i^2 \quad \text{where } r_i = y_i - \hat{\theta}_0 - x_{i1}\hat{\theta}_1 - ... - x_{ip-1}\hat{\theta}_{p-1} \tag{2}$$

LS estimators are notorious for their sensitivity to outliers. Results can be strongly influenced by the presence of just one "bad" outlier. Several estimation techniques have been developed to reduce the effects of "abnormal" points: Least Median of Squares (LMS), Least Trimmed Squares (LTS), S-estimators (S), MM-estimators (MM), etc. (see Rousseeuw and Leroy, 1987, for a thorough

review of the robust techniques literature). All these estimation techniques have very high breakdown points (roughly speaking, the breakdown point represents the smallest fraction of contaminated data that causes the estimator to take on values arbitrarily far from the "true" unknown parameter) but are less efficient[1]. The class of MM estimators (Yohai, 1987) is very interesting since these estimators combine high breakdown points and high efficiency. However, an estimator with high efficiency will be less robust, more sensitive to outliers than an estimator with lower efficiency, even if its breakdown point is 50%.

This is the reason why we propose to use the very robust S-estimator introduced by Rousseeuw and Yohai (1984). S-estimators form a class of high-breakdown affine equivariant estimators. They are defined as minimizing a scale M-estimator of the residuals. Let $\{r_1, \ldots, r_n\}$ be a sample of residuals. The M-scale estimate $s(r_1, \ldots, r_n)$ is defined as the solution of:

$$\frac{1}{n}\sum_{i=1}^{n}\rho(\frac{r_i}{s}) = b \tag{3}$$

where b is a constant, chosen as $E_\Phi[\rho]$ (Φ is the standard Normal cumulative function) to ensure consistent estimation of σ at normal distribution. Function ρ is assumed to be even and continuously differentiable, with $\rho(0) = 0$ and such that there exists some strictly positive value c for which ρ is strictly increasing on $[0, c]$ and constant on $[c, \infty)$.

[1]For example, LMS has the disadvantage of converging at a rate of $n^{-1/3}$, or the 50% breakdown LTS estimator has a Gaussian efficiency of only 7.1%.

The regression S-estimator is then defined as

$$\hat{\theta}_S = \arg\min_{\hat{\theta}} \quad s(r_1(\hat{\theta}), ..., r_n(\hat{\theta})) \tag{4}$$

and the final scale estimator is

$$\hat{\sigma}_S = s(r_1(\hat{\theta}_S), \dots, r_n(\hat{\theta}_S)). \tag{5}$$

Taking ρ as Tukey's Biweight Function

$$\rho(x) = \begin{cases} \frac{1}{6c^4}x^6 - \frac{1}{2c^2}x^4 + \frac{1}{2}x^2 & if \quad |x| \leq c \\ \frac{c^2}{6} & if \quad |x| > c \end{cases} \tag{6}$$

it can be shown that at a breakdown point of 50% ($c = 1.547$), the Gaussian efficiency of S is 28.7%. Rousseeuw and Yohai (1984) also proved the consistency and the asymptotic normality of the S-estimator, using the fact that it satisfies the first-order necessary conditions of M-estimators defined in Huber (1981).

We have just put forward the key issue underlying the question we want to address in this paper: LS is efficient but not robust while S is robust but inefficient. It is sometimes extremely difficult to determine if the gain in consistency attained using the robust estimator is more valuable than the loss of efficiency due to not using LS. A rule of thumb is that, if the values obtained by the robust and classical estimators are similar, it is better to use the classical one and if they are very different, it is better to use the robust one. This is unfortunately not necessarily pertinent. What we show is that a Hausman-type test may be used to determine if the gain in consistency coming from the use of a robust

estimator overrules the corresponding loss of efficiency (obviously, only if the model is well specified).

The Hausman test (1978) is based on comparing an estimator which is efficient under H_0 of no endogeneity with an estimator that is consistent under the alternative that endogeneity is present. Here, we are interested in comparing the classical efficient LS estimator $\hat{\theta}_{LS}$ under H_0 of no inconsistency due to outliers to the robust S-estimator $\hat{\theta}_S$ that is always consistent. Clearly, if more than 50% percent of the data are contaminated, the robust S-estimator will also break (breakdown point of 50%) but then, can they still really be considered as outliers? Since we are interested in a specific test aimed at discriminating between a robust method and a classical one, we assume that the model is well-specified and that all the Gauss Markov hypotheses are respected (linear functional form, zero mean of disturbance, homoscedasticity, no serial correlation, normality of errors and exogeneity).

From the results of Rousseeuw and Yohai given above[2], it is clear that $\hat{\theta}_{LS}$ and $\hat{\theta}_S$ are both asymptotically normal under H_0. Let $\hat{q}$ denote the difference between the two estimators i.e. $\hat{q} = \hat{\theta}_S - \hat{\theta}_{LS}$. The probability limit of the difference between the two estimators is zero if and only if no outlier is present. Hausman (1978) proved that, when two estimators (one which is always consistent but inefficient, the other efficient but not necessarily consistent) are correlated, the asymptotic variance of their difference is given by the difference of their respective variances.

[2]Asymptotic normality with a convergence rate of $n^{-1/2}$ just as LS.

It is well known that for the classical estimator

$$\hat{\theta}_{LS} \overset{a}{\sim} N(\theta, \sigma^2(X'X)^{-1}) \tag{7}$$

where X is the design matrix i.e. $X = (x_{ij})$ for $i = 1, ..., n$ and $j = 1, ..., p-1$. Similarly, for the robust S-estimator, we have

$$\hat{\theta}_S \overset{a}{\sim} N(\theta, \frac{\sigma^2(X'X)^{-1}}{e}) \tag{8}$$

where e is the efficiency of the S-estimator. Using Tukey's Biweight Function with a 50% breakdown point, the efficiency is $e = 28.7\%$. Denoting the asymptotic variance of $\hat{q}$ by $V(\hat{q})$, we get

$$V(\hat{q}) = V(\hat{\theta}_S) - V(\hat{\theta}_{LS}) = \frac{\sigma^2(X'X)^{-1}}{e} - \sigma^2(X'X)^{-1} \tag{9}$$

where the nuisance parameter σ must be estimated[3]. It is obvious that the estimator of the standard error should be robust itself, otherwise the test might lead to incorrect results under the alternative hypothesis. A first natural choice is the scale estimator obtained by the optimization problem of the S-estimator i.e. $\hat{\sigma}_s$. Its efficiency at Gaussian distributions is equal to 50.59%. We also tried other candidates such as the Median Absolute Deviation estimator (MAD), but it has low efficiency for normal distributions (36.75%), thereby leading to rather unsatisfactory results. Rousseeuw and Croux (1993) introduced an alternative statistic more efficient than the MAD[4] . They propose to use $\hat{\sigma}_{RC} = 1.1926 \underset{i}{med}(\underset{j}{med}\,|x_i - x_j|)$ where the outer median (taken over i) is the median of the n medians of $|x_i - x_j|$, $j = 1, 2, ..., n$. The efficiency of $\hat{\sigma}_{RC}$ at Gaussian

[3]The estimated variance will be denoted by $\hat{V}(\hat{q})$.

[4]Which the authors call S_n.

distributions is 58% which is better than the natural scale estimator obtained when using the S-procedure. We decided to try both to determine which is the best in the variance formula.

The Hausman test statistic is defined as

$$H = \hat{q}' \left[\hat{V}(\hat{q})\right]^{-1} \hat{q} \tag{10}$$

where $\hat{V}(\hat{q})$ is a consistent estimator of $V(\hat{q})$. Hausman (1978) shows that under the null, H is distributed asymptotically as a central χ^2_p where p is the number of unknown parameters. If the latter statistic is higher than the tabulated value of a χ^2_p at a given level of confidence, we reject the hypothesis that the difference between the estimators is not systematic and thus reject the LS estimator. Otherwise, we conclude that the efficiency loss resulting from the use of the S-estimator is more costly than the bias produced by the use of LS.

Note that in (10), $\hat{V}(\hat{q})$ is assumed to be non singular, but, as stated by Chmelarova and Hill (2004) this will almost never hold in practice due to linear restrictions between the elements of $\hat{q}$. To solve this problem, in case of singularity, Hausman and Taylor (1981) and Holly (1982) suggest replacing $\left[\hat{V}(\hat{q})\right]^{-1}$ by some generalized inverse[5] $\left[\hat{V}(\hat{q})\right]^{-}$.

[5]If A is an $m \times n$ matrix, a generalized inverse of A is an $n \times m$ matrix A^- such that $AA^-A = A$ (see Rao and Rao (1998), for example). To ensure uniqueness, Krämer and Sonnberger (1986) propose using the Moore-Penrose pseudo-inverse.

3 Simulations

We will consider two aspects in this section using simulations. First, we look at the finite-sample behavior of the simulated statistics under the null, to check if the approximation of the χ^2 distribution for a small sample is good enough. Secondly, we study the power of the test when different types of outliers are introduced.

The experimental design for the first part of this section is the following: a total of $m = 2000$ samples for each of the sizes $n = 100, 200, 500$ and 700 were generated using the following linear regression

$$y_i = \theta_0 + x_{i1} + \ldots + x_{ip-1} + \varepsilon_i \qquad i \in \{1, \ldots, 2000\} \qquad (11)$$

where each explanatory variable is standard normal, $\varepsilon \sim N(0, 1)$ and $\theta_0 = 1$. For each sample, the test statistic is calculated using the two candidates for the estimation of σ introduced in Section 2. Then, the empirical quantiles of a $\chi^2_{p,0.95}$ are computed. The results of the simulations are given in Table 1.

[INSERT TABLE 1 HERE]

Especially for small sample sizes ($n = 100, 200$), it appears that the approximations using the $\hat{\sigma}_{RC}$ scale estimator are better than those using $\hat{\sigma}_s$ provided by the S-estimator. These results also show that the test is more appropriate if the sample size n is large enough relatively to p the number of parameters. For example, with $p = 5$ and $n = 100$ the difference between the theoretical and the simulated quantiles is quite substantial.

To compare the empirical and theoretical distributions more thoroughly, we use a classical graphical tool: the Quantile Quantile Plot (QQ-Plot). It allows to compare simulated quantiles with the quantiles of the χ^2 distribution with p degrees of freedom. The order of the quantiles chosen are $0.05 \times i$ where $i \in \{1, \ldots, 19\}$. For the graphs of Figure 1, the number of regression parameters is 3 ($p = 3$) and $\hat{\sigma}_{RC}$ was used as the scale estimator (with $\hat{\sigma}_s$ the correspondences are not as good). As can be seen in Figure 1, the empirical quantiles are rather larger than the theoretical ones for $n = 100$. Therefore, with small sample sizes, the use of theoretical quantiles leads to rejecting the null more often than the chosen level α. For $n = 200$, the situation is better and from $n = 500$ on, the match between the two sets of quantiles is rather good.

[INSERT FIGURE 1 HERE]

The second part of the simulations is devoted to the behavior of the test under contamination (H_1). In linear regressions, outliers are classified into three categories: *bad leverage* points, *good leverage* points and *vertical outliers* (see Figure 2 (a)). We will study the power of the test under these three types of contamination. Using $\hat{\sigma}_s$ or $\hat{\sigma}_{RC}$ for the estimation of the nuisance parameter yields very similar results. We report only those obtained with $\hat{\sigma}_{RC}$.

For the simulations, observations were generated according to the model

$$y_i = \theta_0 + x_i + \varepsilon_i \tag{12}$$

where $x \sim N(0, 1)$, $\varepsilon \sim N(0, 1)$ and $\theta_0 = 1$. The sample sizes are again 100, 200, 500 and 700. For all simulations under the alternative, we introduce a very

small percentage of contamination: 1%. Clearly, if the percentage increases, the test will become more powerful.

In a first experiment we replace 1% of the x-values by a constant value C in every data set, hereby creating leverage points. Constant C is assigned each integer value between 0 (corresponding to the null hypothesis) and 9. To calculate the empirical size and power of the test, we generated 400 samples according to the model and computed the percentage of times that the critical value was exceeded. In Table 2, we report the frequency of rejection of the null hypothesis for the simulated data sets and for each value of C. In parentheses, we give the absolute value of the bias of the LS estimator for parameter θ_1.

[INSERT TABLE 2 HERE]

Since the independent variable is computed as $x \sim N(0, 1)$, the C values 0 and 1 are not considered as outliers and, consequently, the percentage of rejection is close to 5% (the confidence level of the test). From values 2 to 9, the percentage of rejection progressively increases (as does the bias of the LS-estimator) to reach 100% rejection. Quite naturally, the power of test also increases with the sample size due to the variance precision. Figure 2 (b) shows how rapidly the percentage of rejection increases as the bad leverage points get further away from the majority of the observations.

[INSERT FIGURE 2 HERE]

The second type of contamination involves replacing 1% of the x-values in the same way as for the first contamination (C values between 0 and 9). But in order to create good leverage points, we simulate the y values using the contaminated x values. Just as in the case of bad leverage points, we compute the empirical size and power of the test with 400 samples according to the model. In Table 3, we report the frequency of rejection of the null hypothesis for the simulated data sets and for each value of C. In parentheses, we give the absolute value of the bias of the LS estimator for parameter θ_1.

[INSERT TABLE 3 HERE]

As predicted, the percentage of rejection for good leverage points is small compared to that of bad leverage points (Figures 2 (b) and 2 (c)). The former type of points generally increase the stability of regression lines implying that the variances of the regression estimators decrease. But as mentioned in Croux et al. (2003) good leverage points can still influence the classical estimator and attract the regression line towards them even if they are not so distant from the "true" regression line. It is therefore not surprising to see that the null hypothesis is sometimes rejected (percentage of rejection close to 15% when $C = 9$).

The last configuration for outliers we look into is the case of vertical outliers (Figure 2 (d)). In fact this kind of contamination is not as "dangerous" as that of bad leverage points. It is well-known for example, that the Least Absolute estimator (L1) is robust with respect to vertical points, but not robust with respect to bad leverage outliers. Nevertheless, if we use LS and a vertical outlier

is far enough, the estimator might be attracted by it and give erroneous results. In this situation, the bias is often more important for the intercept. For the simulations, we contaminated 1% of the y data by replacing them with constant value $D = 3C$ for each integer value of C between 0 and 9. The results are given in Table 4 (in parentheses, the absolute value of the bias of the LS estimator for parameter θ_1, even if the large bias is generally on the intercept in these situations).

[INSERT TABLE 4 HERE]

When comparing Tables 2 and 4, we see that the percentage of rejection of the null is smaller in the case of vertical outliers than in that of bad leverage points. More precisely, the results for $C = 9$ in Table 2 show 100% of rejects (bad leverage points) while those obtained for $D = 9$ or $3*3$ in Table 4 (vertical outliers), correspond to a percentage of rejection between 10% and 37%. This is logical since the LS estimator is less sensitive to this type of contamination. The bias increases as the vertical outliers move further away but becomes strong only for rather big distances.

4 Economic application

To illustrate the usefulness of the test we presented above[6], we use a real dataset on an interesting economic topic: DeLong and Summers (1993) present striking

[6]The Matlab code is available from the authors upon request.

results showing that there is a strong relationship between equipment investment and growth. Similarly to what Zaman, Rousseeuw and Orhan (2001) did, we check if the results of DeLong and Summers are robust to the presence of outliers. To find out if there is a strong relation between growth and equipment investments, the authors propose to run a regression of the type:

$$GDP_i = \theta_0 + \theta_1 LFG_i + \theta_2 GAP_i + \theta_3 EQP_i + \theta_4 NEQ_i + \varepsilon_i \qquad (13)$$

where GDP represents GDP growth per worker over the period 1960-1985, LFG is the labor force growth during the same period, GAP is the relative GDP gap with respect to the United States, EQP and NEQ represent respectively the share of GDP devoted to equipment and non equipment investment over the period 1960-1985. The authors estimated the equation by ordinary least squares. The first thing we want to check is if this technique is appropriate here, or if the eventual presence of outliers might bias the estimation. To do so, we estimate the same equation, but instead of estimating it using ordinary least squares, we estimate it with the S-estimator that we described above. We present the results of the regressions by LS and S, and the differences between the estimated parameters in Table 5.

[INSERT TABLE 5 HERE]

It is hard to tell if the differences between the estimated parameters are significant. To decide if the gain in robustness from the S-estimator is more

valuable than the resulting loss in efficiency, we run our test. We obtain test statistic $H = 2.1163$ associated to a $p - value$ of 0.8328. It is thus clear that the LS estimator is the most appropriate.

Now imagine that for some unexplained reason, a mistake has been made in the GAP variable for Canada, for example a -16 value has been coded instead of the correct value of -0.169. Rerunning the LS and the S estimations, we obtain the results given in Table 6.

[INSERT TABLE 6 HERE]

The differences between the estimated parameters, are of similar magnitude to those in the case presented before. It is thus extremely difficult to decide which technique is the most appropriate, so it may be of interest to run our test. The test statistic here is $H = 41.20$ associated to a $p - value$ of 0.0000. Consequently, our test strongly suggests to reject the use of LS and instead use the S-estimator[7]. Consequently, even if a gross mistake is made, such as the one presented here, we might still find estimated parameters similar to those obtained in the case where no outliers are present.

It may be argued that, instead of using our test, we could have used outlier diagnosis tools. A very interesting one is the display of robust standardized regression residuals versus robust distances. Robust distances on explanatory

[7]Given the relatively small size of the sample, it is more appropriate to compare the calculated statistic H with the empirical quantile provided in Table 1 for $n = 100$ and $p = 5$, i.e. 14.79. The conclusion of the test is the same.

variables allow to identify leverage points, but do not discriminate between good and bad ones (in our case we calculate robust distances using the Minimum Covariance Determinant estimator with a breakdown point of 25%). On the other hand, the robust standardized residuals allow to identify large residuals. All points corresponding to distances higher than the quantile $\sqrt{\chi^2_{p-1,0.975}}$ (as taken by Rousseeuw and van Zomeren, 1990) will be considered as leverage points. For the standardized residuals, we consider, as do the above-mentioned authors, all the robust standardized residuals that lie outside the tolerance band $[-2.5, 2.5]$ as regression outliers. The graphic representation will allow to discriminate between bad leverage points (to the right of the cutoff point and outside the confidence band) from good leverage points (to the right of the cutoff point but within the confidence band), vertical outliers (to the left of the cutoff point but outside the confidence band) and regular observations (to the left of the cutoff point and within the confidence band). In our example, if we use the original data, we obtain the graph of Figure 3 (a). In this case, Cameroon appears to be a mild vertical outlier, while Zambia is a bad leverage point. Looking only at this graph we might conclude that to minimize the influence of the outliers, a robust methodology should be used, but as stated previously, the gain in robustness is clearly overruled by the loss in efficiency. This is a result that could not be deduced by only observing the graph. Now, when we use the artificially contaminated data, the graph (Figure 3 (b)) clearly shows that there is one very bad leverage point i.e. Canada and, in accordance with the results of our test, we find that a robust method is better suited here rather than a

standard linear regression. To conclude, using these graphs to identify outliers is an interesting diagnosis tool, but can by no means solve the problem of the robustness-efficiency trade-off which is tackled by the test we propose.

[INSERT FIGURE 3 HERE]

5 Conclusion

In this article, we propose using a Hausman-type test to determine whether a robust S-estimator is more appropriate than an ordinary least squares one in a multiple linear regression framework, considering the trade-off between robustness and efficiency. Indeed, a very common belief is that, as soon as outliers are detected in the data (at least "bad" outliers), a robust method should be preferred to a classical least squares one. But robustness has a cost: efficiency could be severely reduced. Traditionally, to identify outliers, robust distances are calculated and plotted against robust standardized residuals. When this brings forward bad leverage points or vertical outliers, the typical decision is to reject least squares and turn to a robust method. We show, with a simple economic example, that this is not always the most appropriate choice, since it by no means takes the loss in efficiency into account. The Hausman-type test we propose can be considered as a powerful complementary tool to existing methods. An interesting extension to this paper might be the use of this

robust-to-outliers test to detect other problems that the standard Hausman test already tackles.

References

[1] Chmelarova V. and R.C. Hill, 2004, Finite Sample Properties of the Hausman Test, work in progress, E.J. Ourso College of Business.

[2] Croux, C., S. Van Aelst and C. Dehon, 2003, Bounded Influence Regression using High Breakdown Scatter Matrices, Annals of the Institute of Statistical Mathematics 55, 265-285.

[3] DeLong, J.B. and L.H. Summers, 1993, How Strongly Do Developing Countries Benefit from Equipment Investment. Journal of Monetary Economics 32 (3), 395-415.

[4] Durbin, J., 1954, Errors in Variables, Review of the International Statistical Institute 22, 23-32.

[5] Hausman, J.A., 1978, Specification Tests in Econometrics, Econometrica 46 (6), 1251-1271.

[6] Hausman, J.A. and W. Taylor, 1981, A Generalized Specification Test, Economics Letters 8, 239-245.

[7] Holly, A., 1982, A Remark on Hausman's Specification Test, Econometrica 50, 749-759.

[8] Huber, P.J., 1981, Robust Statistics. Wiley, New York.

[9] Krämer, W. and H. Sonnberger, 1986, The linear regression model under test, Heidelberg : Physica-Verlag.

[10] Rao C.R. and M.B. Rao, 1998, Matrix Algebra and Its Applications to Statistics and Econometrics, World Scientific, Singapore.

[11] Rousseeuw, P.J. and C. Croux, 1993, Alternatives to the Median Absolute Deviation. Journal of the American Statistical Association 88, 1273-1283.

[12] Rousseeuw P.J. and A.M. Leroy, 1987, Robust Regression and Outlier Detection, John Wiley and Sons, New York.

[13] Rousseeuw, P.J. and V.J. Yohai, 1984, Robust regression by means of S-estimators. In: Franke, J., Härdle W., Martin, R.D. (Eds.), Robust and Nonlinear Time Series Analysis, Lecture Notes in Statistics 26, New York: Springer Verlag.

[14] Rousseeuw, P.J. and B.C. van Zomeren, 1990, Unmasking multivariate outliers and leverage points. Journal of the American Statistical Association 85, 633-651.

[15] Yohai, V.J., 1987, High breakdown-point and high efficiency robust estimates. The Annals of Statistics 15, 642-656.

[16] Wu, D.-M., 1973, Alternative Tests of Independence Between Stochastic Regressors and Disturbances", Econometrica 41, 733-750.

[17] Zaman, A., P.J. Rousseeuw and M. Orhan, 2001, Econometric applications of high-breakdown robust regression techniques. Econometrics Letters 71, 1-8.

Table 1: Comparisons between empirical and theoretical quantiles at level $\alpha = 5\%$ using $\hat{\sigma}_S$ and $\hat{\sigma}_{RC}$ as scale estimators.

		n=100	n=200	n=500	n=700	$\chi^2_{p;0.95}$
p=2	$\hat{\sigma}_S$	6.48	6.58	6.07	5.79	
	$\hat{\sigma}_{RC}$	6.25	6.39	5.95	5.73	5.99
p=3	$\hat{\sigma}_S$	9.88	9.10	8.29	7.75	
	$\hat{\sigma}_{RC}$	9.30	8.69	8.05	7.65	7.81
p=4	$\hat{\sigma}_S$	13.57	11.72	10.49	9.85	
	$\hat{\sigma}_{RC}$	12.62	10.60	9.92	10.18	9.49
p=5	$\hat{\sigma}_S$	16.77	13.46	12.04	11.51	
	$\hat{\sigma}_{RC}$	14.79	12.50	11.63	11.25	11.07

Table 2: Power of the test under 1% bad leverage point contamination.

$n \backslash C$	0	1	2	3	4	5	6	7	8	9
100	3.75	4.75	7.25	13.25	23.00	40.00	51.50	69.25	82.50	90.00
	(0.01)	(0.02)	(0.06)	(0.13)	(0.19)	(0.27)	(0.35)	(0.42)	(0.49)	(0.55)
200	7.75	5.25	8.00	16.75	30.75	54.75	73.00	92.50	97.50	99.25
	(0.00)	(0.01)	(0.06)	(0.11)	(0.17)	(0.24)	(0.30)	(0.37)	(0.44)	(0.49)
500	4.50	9.00	16.25	34.25	71.75	92.00	98.00	99.75	100	100
	(0.00)	(0.02)	(0.06)	(0.12)	(0.19)	(0.26)	(0.33)	(0.40)	(0.46)	(0.52)
700	3.25	8.00	8.00	29.75	63.50	90.50	99.50	99.75	100	100
	(0.00)	(0.00)	(0.04)	(0.09)	(0.14)	(0.21)	(0.27)	(0.34)	(0.40)	(0.45)

Table 3: Power of the test under 1% good leverage points contamination

$n \backslash C$	0	1	2	3	4	5	6	7	8	9
100	4.25	3.50	6.75	7.25	7.00	8.50	10.00	10.00	17.25	16.50
	(0.01)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)
200	7.00	4.50	4.00	7.50	5.25	8.00	11.50	8.75	14.25	14.00
	(0.00)	(0.00)	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
500	4.50	6.50	6.25	4.50	6.25	7.00	12.25	12.50	9.00	17.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
700	3.25	6.00	4.50	6.25	4.25	8.25	8.75	13.50	10.25	13.75
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

Table 4: Power of the test under 1% vertical outliers contamination

n\3C	0	3*1	3*2	3*3	3*4	3*5	3*6	3*7	3*8	3*9
100	4.50	3.50	7.25	9.25	12.00	17.25	18.75	23.50	34.00	41.75
	(0.01)	(0.02)	(0.04)	(0.06)	(0.07)	(0.10)	(0.12)	(0.13)	(0.16)	(0.18)
200	6.00	5.50	8.50	15.50	22.25	33.75	44.50	62.25	76.75	83.75
	(0.00)	(0.03)	(0.06)	(0.09)	(0.11)	(0.14)	(0.17)	(0.20)	(0.22)	(0.25)
500	3.25	10.75	19.75	37.00	61.50	76.25	89.25	97.25	99.75	100
	(0.02)	(0.05)	(0.08)	(0.11)	(0.14)	(0.17)	(0.20)	(0.22)	(0.25)	(0.29)
700	4.25	5.50	11.25	24.25	47.25	72.25	89.00	96.00	98.75	99.75
	(0.02)	(0.00)	(0.02)	(0.04)	(0.06)	(0.07)	(0.09)	(0.11)	(0.14)	(0.16)

Table 5: Estimated coefficients on real data

	LS	S	$Difference$
$Constant$	-0.0180	-0.0215	-0.0035
EQP	0.3052	0.2387	-0.0665
NEQ	0.0916	0.1392	0.0476
GAP	-0.0066	-0.0062	0.0004
LFG	0.0849	0.0871	0.0022

Table 6: Estimated coefficients on artificially contaminated data

	LS	S	$Difference$
$Constant$	-0.0013	-0.0216	0.0203
EQP	0.2396	0.2536	-0.0140
NEQ	0.0635	0.1343	-0.0708
GAP	-0.0001	-0.0061	0.0060
LFG	0.1580	0.1197	0.0382

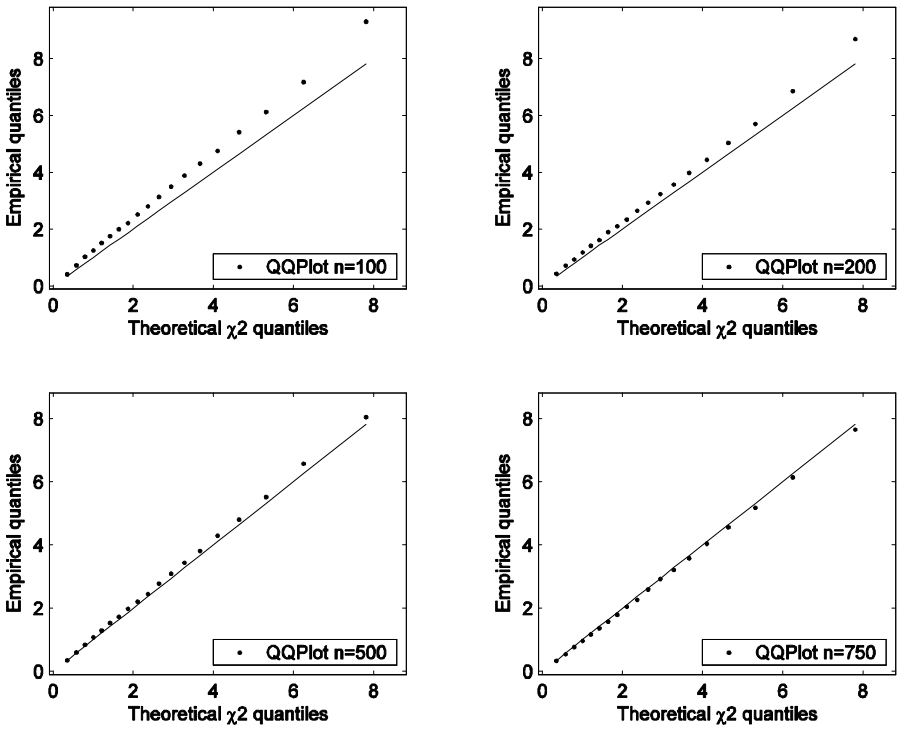

Figure 1: Quantile Quantile Plot when the dimension is $p = 3$ and using $\hat{\sigma}_{RC}$ as the scale estimator

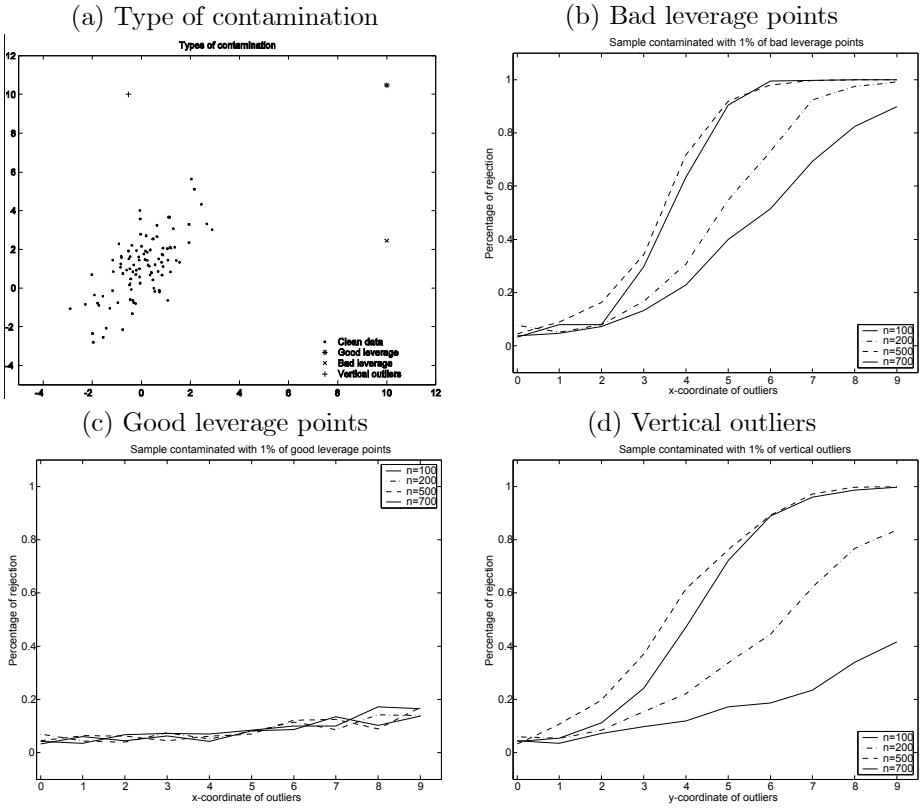

Figure 2: Power of the test under three types of contamination with p=2.

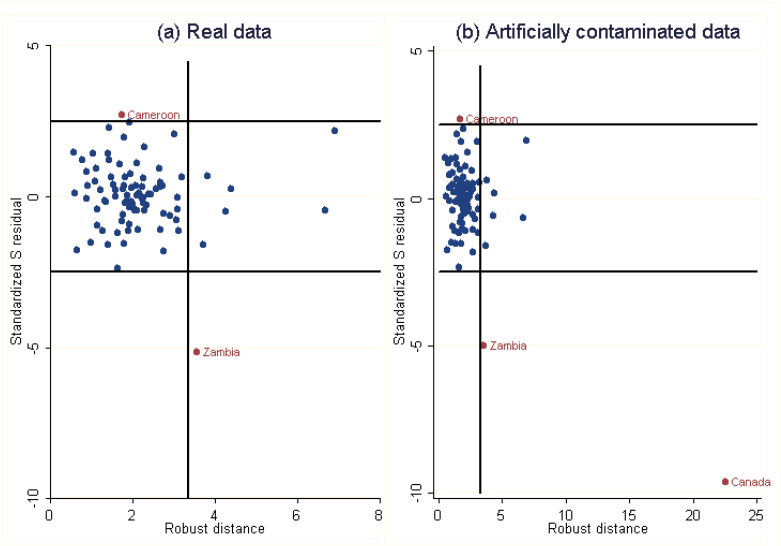

Figure 3: Robust distances versus robust standardized regression residuals