EUI WORKING PAPERS IN ECONOMICS

EUI Working Paper ECO No. 95/2

Robust Estimation: An Example

JEROEN HINLOOPEN
and
RIEN WAGENVOORT

European University Institute, Florence

EUROPEAN UNIVERSITY INSTITUTE, FLORENCE

ECONOMICS DEPARTMENT

EUI Working Paper **ECO** No. 95/2

Robust Estimation: An Example

JEROEN HINLOOPEN
and
RIEN WAGENVOORT

BADIA FIESOLANA, SAN DOMENICO (FI)

Printed in Italy in January 1995
European University Institute
Badia Fiesolana
I – 50016 San Domenico (FI)
Italy

Robust Estimation: An Example

Jeroen Hinloopen*

Rien Wagenvoort

European University Institute
Department of Economics
Badia Fiesolana
San Domenico di Fiesole (FI)
Italy

November 1994

Abstract

The use of Ordinary Least Squares and its generalizations is widespread among economists, although it should be common knowledge that a single outlying observation can cause this technique to produce arbitrary estimates and hence incorrect t-values. In this paper a robust alternative, an example of a General M estimator, is discussed. Not only has this estimator a breakdown point of 50%, it also yields consistent estimates and, as we show by means of a simulation experiment, is more efficient than Rousseeuw's Least Median of Squares estimator. We also propose a specific correction factor which improves both the *resampling* algorithm and the *projection* algorithm for computing the Minimum Volume Ellipsoid estimator.

* We are much indebted to Teun Kloek (Erasmus University Rotterdam) for guiding us into the field of robust estimation. André Lucas (Erasmus University Rotterdam) provided valuable comments. Robert Waldmann (European University Institute) increased our understanding. The usual disclaimer applies.

1. INTRODUCTION

The aim of a regression method is to reveal statistical patterns presumably contained in some data set. For this purpose, the use of Ordinary Least Squares (OLS) and its generalizations is widespread among economists, because of computational ease and because of the appealing statistical properties of OLS when data are (unrealistically) 'smooth'[1]. However, it should be common knowledge that adding a single outlying observation can cause this estimation technique to produce dramatically different estimates. And these outlying observations can be present in any data set due to a number of reasons: they may emerge from typos or measurement errors, the error term may come from a fat and/or long tailed distribution, or the fitted model may be only appropriate for a sub set of the data.

Ever since the discovery of the *least squares* criterium (which was around 1800), researchers were aware of its shortcomings if data are erratic. In 1805 Legendre noted that "If among these errors are some which appear too large to be admissible, then those observations which produced these errors will be rejected, as coming from too faulty experiments, and the unknowns will be determined by means of the other observations, which will then give much smaller errors"[2]. Legendre points here at the two core issues of robust estimation: which observation needs to be adjusted and in what way. An early attempt to generate robust estimates was made in 1887 by Edgeworth who introduced the *least absolute values* estimator. This estimator is defined as minimizing the sum of *absolute* errors (instead of minimizing the sum of *squared* errors) over the unknown parameters. Although this estimator is less sensitive than OLS to certain types of outliers (see Judge (1988, p.899)), it can still be tricked by one outlying observation.

By now, a considerable literature has developed on estimation techniques which are little affected by outlying observations. These so called robust estimators produce more or less the same estimates as classical estimators when smooth data are involved but keep producing adequate estimates when data are contaminated.

---

[1] Smooth data do not contain leverage points and/or vertical outliers (see Section 2).

[2] Cited by Rousseeuw and Leroy (1987).

In principle there are two ways of constructing a robust estimate, and although the aim of both approaches is the same (revealing the statistical pattern of the majority of the data), the way of achieving it is quite opposite. The first approach is to develop an estimator which is not so easily affected by obscure observations. According to this robust estimate outliers are detected. The second approach is to construct regression diagnostics to reveal outlying observations. The data are then adjusted accordingly and a classical regression is performed on the altered data set. Unfortunately it is very difficult to identify outlying observations when there are many of them, or when the dimension of the problem exceeds three (in which case we can no longer rely on visual perception). In this paper we focus on the first approach.

Rousseeuw's (1984) Least Median of Squares (LMS) estimator is an example of the first category. Instead of minimizing the *sum* of squares over the unknown parameters, it is the *median* of squares that is minimized[3]. This estimator has the appealing property that its breakdown point (the maximum fraction of data contamination which leaves the estimator undisturbed[4]) is 50%. By definition this is the highest percentage achievable[5], since beyond this limit the distinction between 'good' and 'bad' data becomes arbitrary. The breakdown point of OLS is 0%, i.e. one outlying observation can cause this technique to produce completely arbitrary estimates.

In this paper another example of a robust estimation technique, a General M estimator, is presented. In effect we work out a specific example of the general (and complex) case presented by Simpson, Ruppert and Carroll (1992). This might be considered superfluous, but it is not at all a straightforward exercise to translate their article (and associated ones) into workable (GAUSS) computer programs. In doing so we try to bridge the gap between (advanced) statistics and applied economics, and provide a relatively simple outline for performing robust regression analyses.

---

[3] In Appendix 4 this estimator is described in more detail.

[4] In Appendix 2 we give a formal definition of this concept.

[5] Strictly speaking this is only true for affine equivariant estimators.

The idea of the GM estimator is to start with a high breakdown point (HBP) estimator, followed by some Newton-Raphson (NR) iterations to solve the first order condition associated with minimizing the (weighted) sum of squares. A natural HBP estimator is Rousseeuw's LMS technique. Simpson, Ruppert and Carroll (1992) show that under mild conditions[6] the breakdown point of the preliminary estimator(s) carries over to the final GM estimator. They also demonstrate that the GM estimator yields consistent estimates[7]. In addition, we show in this paper, by means of a simulation experiment, that the GM estimator we use is more efficient than it's preliminary (LMS) estimator when different kinds of data contamination are involved.

The GM estimator we employ, involves computation of Minimum Volume Ellipsoid (MVE) location and scale estimates (see Rousseeuw and Leroy (1987))[8]. Since there is no analytical expression for this estimator we must rely on numerical approximations. In the literature two algorithms are developed for this purpose (see Rousseeuw and Van Zomeren (1990)): the *resampling* and the *projection* algorithm. We derive a correction factor which improves both algorithms for computing the Minimum Volume Ellipsoid estimator.

The set up of this paper is as follows. In Section 2 we give a heuristic explanation of the GM estimator, followed by Section 3, in which we indicate which models and what kind of data are suitable to apply this estimator on. The subsequent section illustrates the specific problems associated with OLS. Simulation results regarding the efficiency of the GM estimator are presented in Section 5. Section 6 presents a real data example while conclusions are stated in Section 7. Five appendices conclude this paper. The first contains an analytical derivation of the GM estimator, while some of its statistical properties (breakdown point and consistency) together with its covariance matrix are discussed in Appendix 2. Appendix 3 contains the description of a robust estimator for the variance of a regression which is followed by an analytical description of Rousseeuw's LMS estimator in Appendix 4. Finally, in Appendix 5 we give a formal definition of the MVE estimator, and also derive and test the above mentioned correction factor.

---

[6] In Appendix 2 these conditions will be discussed in more detail.

[7] See Appendix 2.

[8] In Appendix 5 this estimator is explained in detail.

2. A SIMPLE OUTLINE OF A GM ESTIMATOR

Robust regression methods originate from the mere existence of outlying observations. These outliers may corrupt classic statistical analyses and therefore need to be examined in more detail. Following Rousseeuw and Van Zomeren (1990) we distinguish two types of outliers: leverage points and vertical outliers (see Figure 1). Leverage points are data for which the explanatory variable lies far from the bulk of explanatory observations; vertical outliers are observations which are positioned far from the majority of the data, but whose explanatory component is not necessarily a leverage point. Of course, an observation can be both a vertical outlier and a leverage point.

The principle of the estimator is to detect both kinds of outlying observations and to diminish their impact[9]. In order to do so each observation is given a weight less than or equal to 1, according to its relative location in the data set. For each point it is examined how far its independent component lies from the majority of the explanatory variables (i.e. to what extent it is a leverage point) and a weight is given accordingly. By definition, the identification of leverage points is independent of the estimate of the unknown parameters. Also, given some preliminary robust estimate, the residual of each observation is examined in terms of size (i.e. to what extent it is a vertical outlier). Given initial weights based on the identification of leverage points and the preliminary robust estimate[10], the first order condition associated with minimizing the weighted sum of squares is iteratively solved using NR. Since the estimate depends on the weights based on vertical outliers, which in turn are identified according to the estimate, we update these weights in each NR iteration.

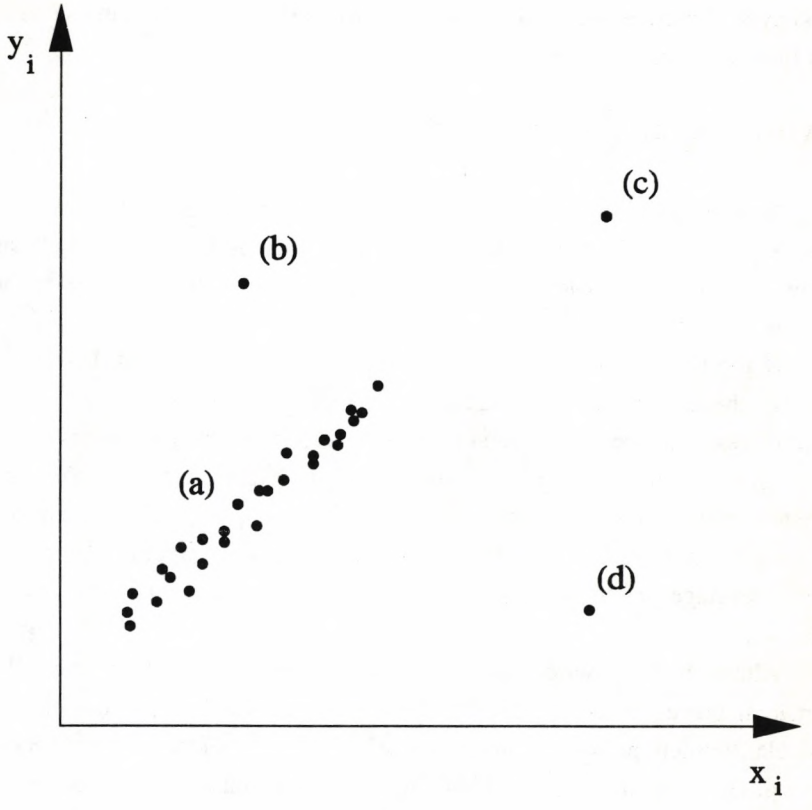
To detect leverage points (and determine weights correspondingly), a measure of distance is needed. Classical measures, such as the entries on the diagonal of the hat matrix or the mahalanobis distance, are inappropriate since these indicators themselves can be corrupted by leverage points. Consequently,

---

[9] As opposed to L_p and M estimators which only correct for vertical outliers (see e.g. Judge et al. (1988, Chapter 22)).

[10] In Appendix 1 precise definitions of these weights are given.

Figure 1 Simple Regression Example with (a) Regular Observation
(b) Vertical Outlier, (c) Leverage Point and
(d) Vertical Outlier and Leverage Point



mild leverage points or a cluster of huge outliers in the x-dimension can escape easily from being discovered (for a prominent exposition of this phenomenon see Rousseeuw and Van Zomeren (1990)). To overcome this *masking effect* we use Minimum Volume Ellipsoid (MVE) estimates to compute a robust measure of distance. Indeed, the MVE estimator has a breakdown point of 50% whereas the classic measures can be tricked by a single observation.

Vertical outliers are defined relative to the statistical relationship contained in the data (for instance, point (c) in Figure 1 is not a vertical outlier since it fits perfectly to the relation between **x** and **y**). However, in general this relation is not known. Therefore, we use a robust preliminary estimate (Rousseeuw's LMS) as a first approximation.

3. WHEN TO USE A GM ESTIMATOR

In economics, data can be qualitative or quantitative and be measured either at micro (e.g. firm or household) or macro (e.g. aggregate) level. In what follows we will discuss each of these cases and indicate when the GM estimator is to be used.

If the response variable is qualitative (giving rise to probit, logit, etc. models), the GM estimation technique is inappropriate. In this case weights based on residuals are of no meaning while correcting for leverage points is not at all straightforward[11]. On the other hand, if explanatory variables are dichotomous (i.e. dummies) and the dependent variable is not, the GM estimator can be applied. Note however that qualitative explanatory variables can never contain leverage points and are therefore left out when computing the MVE-distances.

Although time series often contain outlying observations, the GM estimation procedure is not developed to estimate ARIMA type models. In particular, outliers present in time series are of a fundamental different nature (see e.g. Gómez and Maravall (1994)) than the ones considered in this paper.

[11] In fact, to the best of the authors' knowledge, there is in the literature no research to be found on robust estimation techniques when the dependent variable is of qualitative nature.

Using the GM approach is thus only appropriate when a structural model is to be estimated. For a discussion on joint estimation of model parameters and outlier effects in time series see Chen and Liu (1993).

Micro data are notorious for their erratic behaviour. Especially in this case a robust estimation technique is called for. Aggregation of micro data can remove some of the irregularities, but also in this case robust estimation is still desired since classic and robust estimates are almost identical when data are smooth. On the other hand, if the aggregate data remain erratic, the GM estimator yields more reliable estimates compared to classic techniques. Moreover, cross-country (macro economic) surveys are likely to contain (severe) outliers despite their aggregate nature.

To summarize, the GM estimator described in this paper is appropriate for estimating a structural model in which the response variable is not dichotomous.

4. OF LITTLE SIGNIFICANCE

As an illustration of the effect of leverage points and/or vertical outliers on OLS-estimates, we present in this section some examples of simple regressions when different types of data contamination are involved. In each of the cases considered we also compare OLS with the GM estimator and conclude that the latter is little affected by data pollution while the former in many cases generates unreliable estimates.

4.1 PARAMETER ESTIMATES

First data are generated according to

$$y_t = \beta x_t + e_t, \quad t = 1,..,100, \tag{1}$$

where e_t is *iid* N(0,1), $\beta = 1$ and x_t is *iid* N(0,10). Re-estimating β, using both OLS and GM results[12]

[12] Standard errors are within parentheses.

$$
\begin{aligned}
OLS &: \hat{y}_t = \underset{(0.036)}{1.0094 x_t}, \; t=1,..,100, \; \bar{R}^2=0.87, \\
GM &: \hat{y}_t = \underset{(0.041)}{1.0095 x_t}, \; t=1,..,100, \; \bar{R}^2=0.89.
\end{aligned} \tag{2}
$$

The estimates of β are almost identical and are both significant. Also, the respective adjusted coefficients of determination ($\bar{R}^2$) are approximately the same[13]. Indeed, when data are smooth, both techniques lead to the same results. Note however that in this case OLS is more efficient.

To consider the effect of outlying observations on the estimates of β, one observation is adjusted such that it becomes both a leverage point and a vertical outlier[14]. In particular we set x_{50} at 26 and y_{50} at -26. Re-estimating (1) with this adjusted data set gives us

$$
\begin{aligned}
OLS &: \hat{y}_t = \underset{(0.033)}{0.0952 x_t}, \; t=1,..,100, \; \bar{R}^2=0.01, \\
GM &: \hat{y}_t = \underset{(0.042)}{1.0107 x_t}, \; t=1,..,100, \; \bar{R}^2=0.89.
\end{aligned}
$$

We see that the GM estimator produces almost exactly the same estimate as before, while the OLS estimate of β is completely wrong for the majority of the data.

In applied econometric studies it is not unusual to include the square of an explanatory variable into the regression to model non-linearities. However, when outliers are present this can lead to dramatic changes in the estimation results. To expose this phenomenon we first add the square of **x** as an explanatory variable to (1) and re-estimate the equation. This gives us

---

[13] In case of GM estimation the adjusted coefficient of determination is, apart from the obvious correction for the number of explanatory variables, defined as the variance of the weighted estimated responses divided by the variance of the weighted response variable.

[14] In Section 5 we consider case (a) through (d) of Figure 1 more rigorously.

$$OLS : \hat{y}_t = \underset{(0.042)}{1.0066 x_t} + \underset{(0.007)}{0.0010 x_t^2},\ t=1,..,100,\ \bar{R}^2=0.87,$$

$$GM : \hat{y}_t = \underset{(0.053)}{0.9949 x_t} - \underset{(0.012)}{0.0037 x_t^2},\ t=1,..,100,\ \bar{R}^2=0.79.$$

For both estimation procedures the inclusion of the square of **x** does not significantly change the results reported in (2). Both techniques reject the significance of the added explanatory variable and the estimated coefficient for **x** changes little. We now proceed with adding the square of **x** as an explanatory variable with again however x_{50} set equal to 26 and y_{50} to -26. Re-estimating the model we get the following results

$$OLS : \hat{y}_t = \underset{(0.019)}{1.1626 x_t} - \underset{(0.001)}{0.0793 x_t^2},\ t=1,..,100,\ \bar{R}^2=0.85,$$

$$GM : \hat{y}_t = \underset{(0.053)}{0.9950 x_t} - \underset{(0.012)}{0.0035 x_t^2},\ t=1,..,100,\ \bar{R}^2=0.79.$$

OLS is obviously tricked by the outlying observation. Taking the square amplifies the effect this observation has on this estimation technique. On the other hand the GM estimator still reveals the true statistical pattern between **x** and **y**. In passing we can observe that OLS not only gives an unreliable parameter estimate for **x** and the square of **x**, but there is also a dramatic increase in the associated t-value for both explanatory variables. Indeed, this indicates yet another problem of OLS: misleading significance.

4.2 SPURIOUS REGRESSION

To reveal this problem more prominently, consider the following illustration. Two random samples (*iid* N(0,1)) of length 100, ***x*** and ***y***, are regressed on each other, giving the following results

$$OLS : \hat{y}_t = \underset{(0.031)}{-0.0248 x_t}, \; t=1,..,100, \; \bar{R}^2=0.00,$$

$$GM : \hat{y}_t = \underset{(0.100)}{-0.0530 x_t}, \; t=1,..,100, \; \bar{R}^2=0.00.$$

As is to be expected, both techniques present insignificant estimates and find no statistical relationship between x and y. Again we introduce one outlying observation by setting x_{50} and y_{50} equal to 10. Re-estimation with this contaminated data set leads to

$$OLS : \hat{y}_t = \underset{(0.027)}{0.4801 x_t}, \; t=1,..,100, \; \bar{R}^2=0.23,$$

$$GM : \hat{y}_t = \underset{(0.103)}{-0.0420 x_t}, \; t=1,..,100, \; \bar{R}^2=0.00.$$

Although for 99% of the observations there is no statistical relation between x and y, OLS returns a very significant estimate. Again however the GM approach is not significantly affected by the outlier.

4.3 NON-SPURIOUS REGRESSION

Finally, we consider the case in which OLS fails to find the statistical relation contained in some data set. In order to do so we change two observations of the data set used to generate (2), such that the estimate of the coefficient (by OLS) little changes. In particular we set y_{50} at -535 and y_{51} at 100. Re-estimating the relationship gives us

$$OLS : \hat{y}_t = \underset{(0.605)}{1.0095 x_t}, \; t=1,..,100, \; \bar{R}^2=0.00,$$

$$GM : \hat{y}_t = \underset{(0.043)}{1.0137 x_t}, \; t=1,..,100, \; \bar{R}^2=0.89.$$

Even though the estimate of OLS is almost the same, the significance of the statistical relation is not revealed. On the other hand, the GM estimator is relatively insensitive to the pollution of the data.

Routine data are thought to contain 1 to 10 percent contamination (Hampel et al. (1986, p.28)). The examples of this section indicate that OLS should be used with care in any applied research. On the other hand, the GM estimator results reliable parameter estimates and standard errors under various kinds of data pollution.

5. EFFICIENCY

In Appendix 2 the breakdown point of the GM estimator is presented together with its covariance matrix. Also it is shown that the estimator yields consistent estimates if the errors follow an approximately normal distribution. In this section we present our simulation results regarding the efficiency of the GM estimator.

To examine the efficiency of the GM estimator we have performed a simulation, the results of which are summarized in Table 1 and Table 2. A single experiment began with generating a matrix of explanatory variables, consisting of a constant and observations drawn from a standard normal distribution. Then a response variable was created according to (1) with β set equal to unity. Given these data we re-estimated β using the OLS, LMS and GM estimators. Next we successively corrupted the explanatory variables, the response variable and both the independent and dependent variables by replacing (randomly drawn) 10% of the observations by random values drawn from a normal distribution with zero mean and variance 100. In terms of figure 1 we first added points like (d) and (b) separately and then simultaneously. The last corruption may have involved adding observations like (c). Given these polluted data sets we again re-estimated β with all three estimators. For a number of explanatory variables this process was repeated a thousand times[15].

[15] Table 1 and Table 2 display the outcomes of the experiment for p is 2 and 3 respectively. We have performed also simulations for p is 4 and 5. In general, this led to the same results as those depicted in Table 1 and Table 2, and are therefore left out.

Table 1. Simulation Results, n=200, 10% Pollution and p=2[a]

Parameter	Normal			Leverage Points and Vertical Outliers			Vertical Outliers			Leverage Points or Vertical Outliers or Leverage Points and Vertical Outliers		
	OLS	LMS	GM	OLS	LMS	GM	OLS	LMS	GM	OLS	LMS	GM
C	0.9993393	0.9914205	0.9986054	0.0908890	0.9792566	0.9932006	0.8917015	1.0010682	0.9982748	0.0853386	0.9563297	0.9912552
	0.0047838	0.0406848	0.0051607	0.8289588	0.0460491	0.0066695	0.0635336	0.0394229	0.0060993	0.8444700	0.0497642	0.0081347
	0.0047833	0.0406112	0.0051587	0.0024760	0.0456188	0.0066232	0.0518051	0.0394218	0.0060963	0.0078645	0.0478571	0.0080582
β_1	1.0025388	1.0020433	1.0024995	0.0925122	0.9693673	0.9270890	0.9038399	0.9983132	1.0007254	0.0875231	0.9703895	0.9088464
	0.0047410	0.0386258	0.0053079	0.8261456	0.0477279	0.0158638	0.0622094	0.0360389	0.0058890	0.8399073	0.0518598	0.0216360
	0.0047345	0.0386216	0.0053016	0.0026114	0.0467895	0.0105478	0.0529626	0.0360361	0.0058885	0.0072932	0.0509830	0.0133270

[a] The entries in each cell are the mean estimated value, the mean squared error, and the variance over the 1000 runs.

Table 2. Simulation Results, n=200, 10% Pollution and p=3[a]

Parameter	Normal			Leverage points and Vertical Outliers			Vertical outliers			Leverage Points or Vertical Outliers or Leverage Points and Vertical Outliers		
	OLS	LMS	GM	OLS	LMS	GM	OLS	LMS	GM	OLS	LMS	GM
C	0.9993873	1.0011929	0.9998889	0.0904873	0.9785818	0.9866330	0.9137580	1.0034837	1.0003961	0.0840982	0.9664711	0.9841070
	0.0053047	0.0424127	0.0056318	0.8302599	0.0478710	0.0094409	0.0651984	0.0407042	0.0062480	0.8469589	0.0524271	0.0116083
	0.0053043	0.0424113	0.0056362	0.0030467	0.0474122	0.0092623	0.0577607	0.0406920	0.0062478	0.0080827	0.0513029	0.0113557
β_1	0.9990165	0.9880156	0.9983373	0.0930456	0.9551168	0.8620082	0.9044555	0.9950322	0.9979770	0.0815791	0.9356814	0.8182970
	0.0051675	0.0395368	0.0055620	0.8258630	0.0578371	0.0417308	0.0650630	0.0396203	0.0065344	0.8515437	0.0673916	0.0687978
	0.0051666	0.0393932	0.0055593	0.0032968	0.0558226	0.0226891	0.0559342	0.0395957	0.0065303	0.0080468	0.0632547	0.0357818
β_2	1.0010880	1.0023044	1.0011883	0.0922960	0.9535330	0.8633396	0.8940574	1.0067276	1.0023724	0.0818572	0.9370641	0.8186708
	0.0052888	0.0374614	0.0057190	0.8271868	0.0635356	0.0433264	0.0686845	0.0389701	0.0068082	0.8505429	0.0714211	0.0674646
	0.0052877	0.0374561	0.0057176	0.0032602	0.0613764	0.0246503	0.0574607	0.0389249	0.0068026	0.0075566	0.0674602	0.0345843

[a] The entries in each cell are the mean estimated value, the mean squared error, and the variance over the 1000 runs.

From Table 1 and Table 2 we see that when there is no pollution OLS is most efficient, i.e. the estimates of OLS have the lowest variance. However, in this case the variance of the GM estimates just exceeds that of OLS, while that of the LMS estimates is considerably greater (Indeed, the poor efficiency of the LMS estimator is considered a serious set back of this estimator (see e.g. Simpson, Ruppert and Carroll (1992)).

Moving on to the more realistic situations we see the dramatic decay of OLS. Especially when leverage points are involved this method breaks down completely. It is little consolidation that in these cases OLS is most efficient. Notice that vertical outliers have much less influence on OLS than leverage points (this is also observed by Rousseeuw and Leroy (1987)). Nevertheless, also when there are only vertical outliers OLS produces unreliable estimates. All in all, when data contain outlying observations OLS is not to be recommended as can be concluded from the high mean squared error in all of these cases[16].

In terms of efficiency the GM estimator is always superior to Rousseeuw's LMS. In some cases however the LMS estimates are closer to the real value of β, but in all cases the associated mean squared error exceeds that of the GM estimator. Based on statistical performance the GM estimator is therefore to be preferred to the LMS technique.

6. STARS, LIGHT AND HEAT

To illustrate the importance of robust estimation when data contain outlying observations, we use an example from the astronomy. In this field of science it is well known that the star cluster CYG OB1, which contains 47 stars in the direction of Cygnus, comprises four conspicuous stars, so called giants (see Rousseeuw and Leroy (1987)). The scatterplot of the logarithm of the effective temperature at the surface of a star (T_e) and the logarithm of its light intensity (L/L_0), the Hertzsprung-Russell diagram, for the star cluster CYG OB1, reveals that the celestial bodies can be divided into two groups: the majority of the stars, which are lying along a positively sloped band, and the four giant stars in the upper left corner (see Figure 2). Table 3 summarizes the data of star cluster CYG OB1 (observations 11, 20, 30 and 34 are giants).

[16] Note that this statistic takes into account both the variance of the estimates and the extent to which the estimates differ from the true value of the unknown parameter.

Figure 2 Hertzsprung-Russel Diagram of the Star Cluster CYG OB1

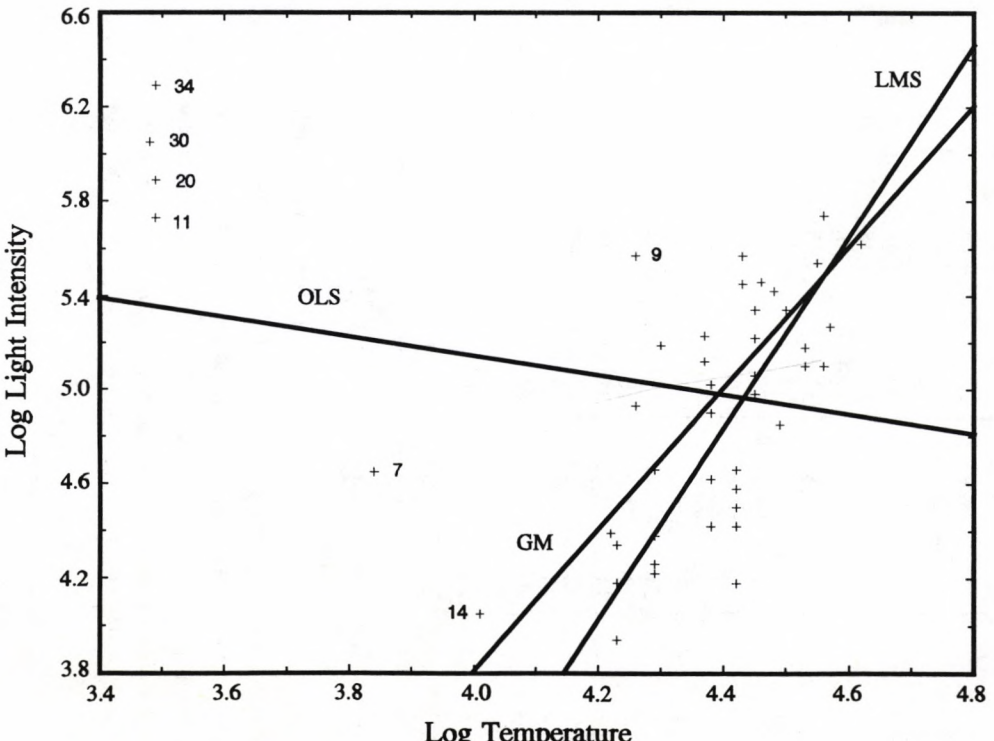


Table 3 Data for the Hertzsprung-Russell Diagram

Star Index	log T_e (x_i)	log(L/L_0) (y_i)	Star Index	log T_e (x_i)	log(L/L_0) (y_i)
1	4.37	5.23	25	4.38	5.02
2	4.56	5.74	26	4.42	4.66
3	4.26	4.93	27	4.29	4.66
4	4.56	5.74	28	4.38	4.90
5	4.30	5.19	29	4.22	4.39
6	4.46	5.46	30	3.48	6.05
7	3.84	4.65	31	4.38	4.42
8	4.57	5.27	32	4.56	5.10
9	4.26	5.57	33	4.45	5.22
10	4.37	5.12	34	3.49	6.29
11	3.49	5.73	35	4.23	4.34
12	4.43	5.45	36	4.62	5.62
13	4.48	5.42	37	4.53	5.10
14	4.01	4.05	38	4.45	5.22
15	4.29	4.26	39	4.53	5.18
16	4.42	4.58	40	4.43	5.57
17	4.23	3.94	41	4.38	4.62
18	4.42	4.18	42	4.45	5.06
19	4.23	4.18	43	4.50	5.34
20	3.49	5.89	44	4.45	5.34
21	4.29	4.38	45	4.55	5.54
22	4.29	4.22	46	4.45	4.98
23	4.42	4.42	47	4.42	4.50
24	4.49	4.85			

Source: Rousseeuw and Leroy (1987, p.27).

OLS reveals a negative relation between the light intensity and the temperature of a star ($\hat{y} = 6.793 - 0.413x$), although for 43 stars (which in the astronomy are said to lie on the main sequence) this relation is positive. Clearly, the four giants trick OLS. The LMS estimate is insensitive to these huge objects and yields a line which fits the majority of observations properly ($\hat{y} = -12.964 + 4.046x$)[17]. Also the GM estimator ignores these observations. However,

[17] Rousseeuw and Leroy (1987) find a somewhat different LMS line ($\hat{y} = -12.298 + 3.898x$). This is due to the enormous increase in computer technology over the years which enables us to perform much more drawings to compute the LMS estimator. The difference in LMS estimates does not affect the main conclusions of this section.

Table 4 Estimation Results, Hertzsprung-Russell Data[a]

Dependent Variable: Log Light Intensity			
Explanatory Variables	OLS	LMS	GM
Constant	6.793 (1.237)	-12.964	-7.132 (3.023)
Log Temperature	-0.413 (-0.286)	4.046	2.741 (0.680)
$\bar{R}^2$	0.023		0.98
$\hat{\sigma}$	0.552	0.368[b]	0.440[c]

[a] Standard errors are within parentheses.
[b] See appendix 4 for a description of this estimator.
[c] See appendix 3 for a description of this estimator.

since the GM estimator is more efficient than Rousseeuw's LMS, it uses more information contained in the data. In particular, observations 7 and 14 attract the GM-line ($\hat{y} = -7.132 + 2.741x$). On the other hand, they cannot reverse the relation between light intensity and temperature, as exhibited by the majority of the stars.

Table 4 summarizes the estimation results. It is striking to see that the OLS estimate is not only completely wrong (both coefficients have the wrong sign), but is also insignificant for the main explanatory variable. A well known phenomenon in the astronomy is not revealed by this estimation technique. The LMS and GM estimators find the true, positive, relation, whereas the latter technique leaves no doubt as to the significance of the relation between the light intensity and the temperature of a star.

The scatterplot of standardized residuals versus x-distances is an aid in visualizing leverage points and vertical outliers (see Rousseeuw and Van Zomeren (1990)). For our regression results with OLS, LMS and GM these plots are depicted in Figure 3, 4 and 5 respectively. Leverage points are observations with a (robust) x-distance exceeding the square root of the 97.5th

Figure 3 Scatterplot of Standardised OLS-Residuals versus Mahalanobis Distances

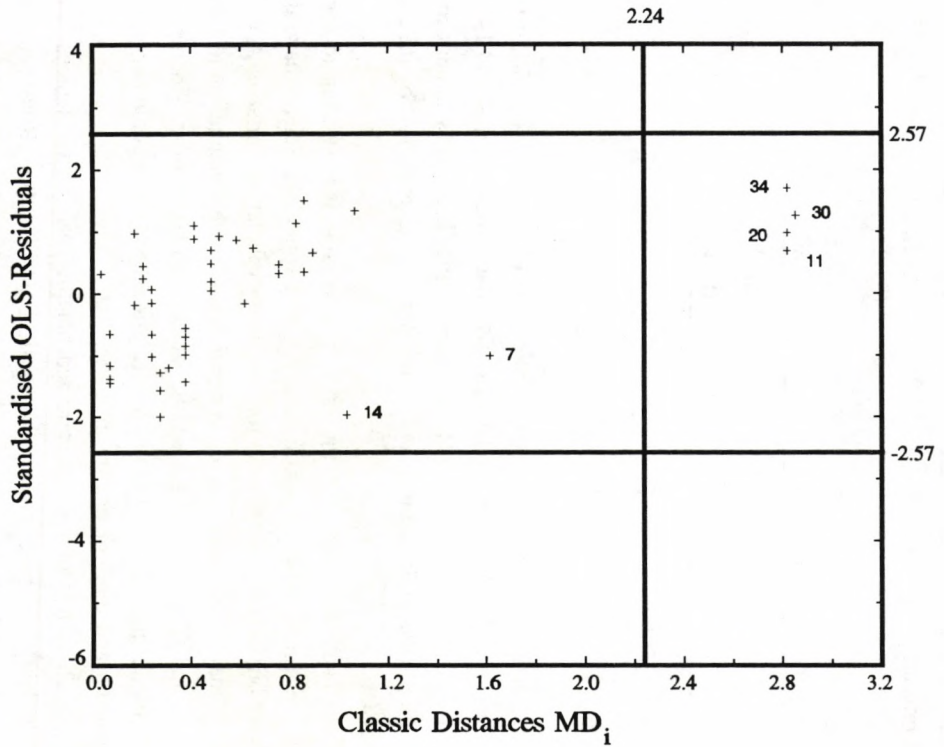


Figure 4 Scatterplot of Standardised LMS-Residuals versus MVE-Distances

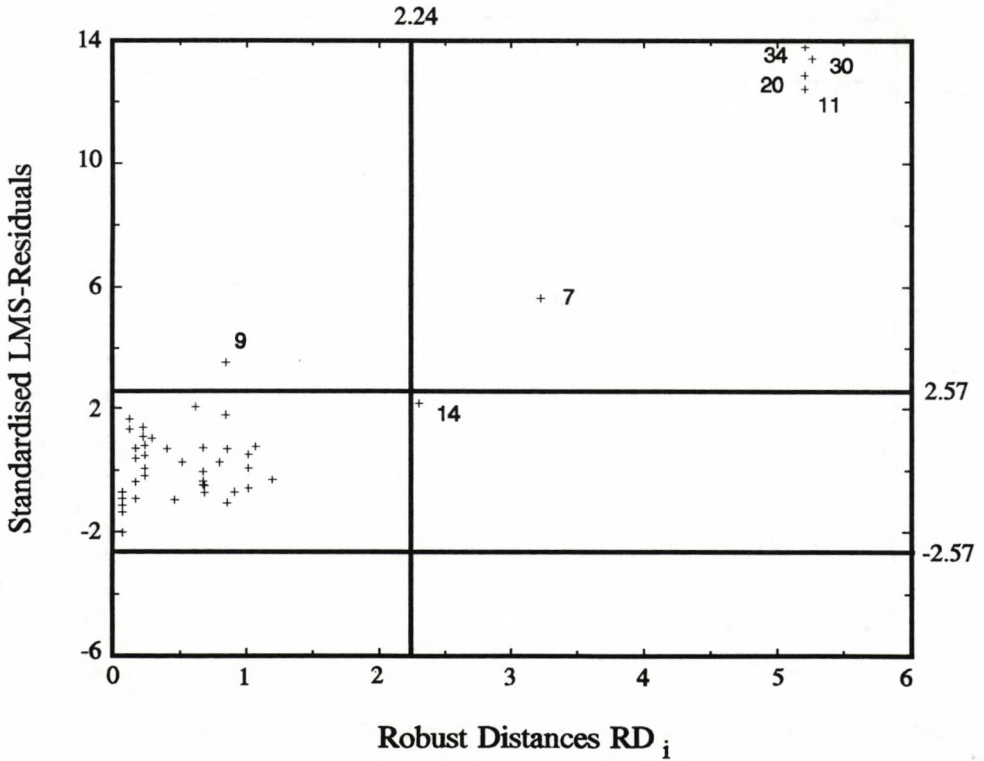


Figure 5 Scatterplot of Standardised GM-Residuals versus MVE-Distances

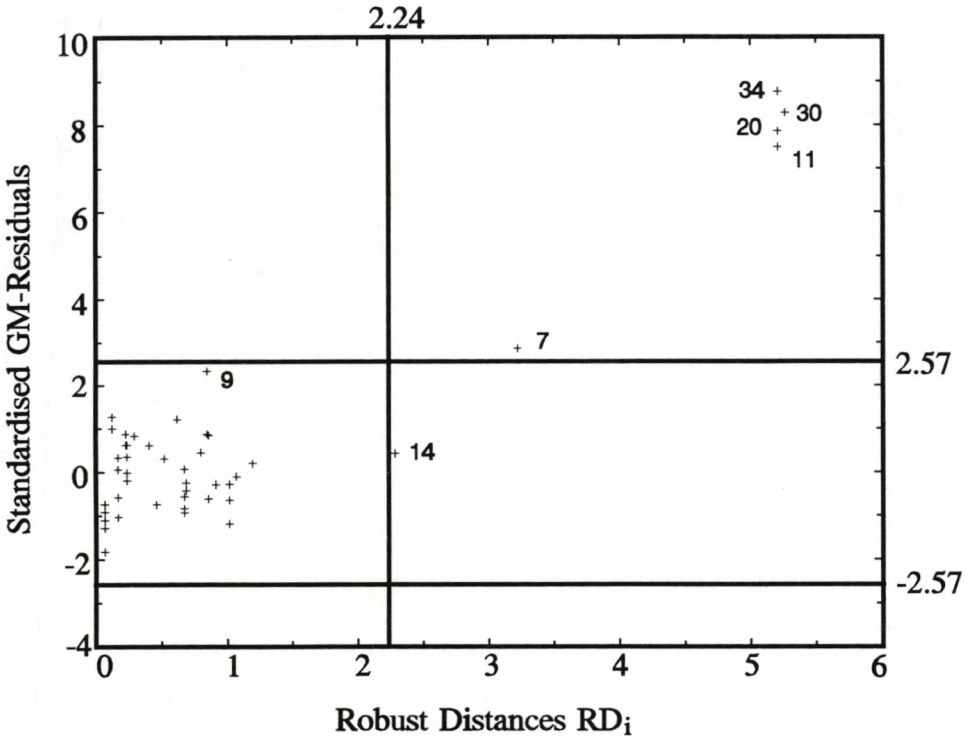


Table 5 Outliers in the CYG OB1 Data Identified by LMS and GM

Estimation Technique	Leverage Points	Vertical Outlier	Leverage Point and Vertical Outlier	Weights based on x-distances $w_x(x_i)$	Weights based on residuals $w_r(r_i)$
			7	0.695	0.220
			11	0.430	0
GM	14			0.991	0.976
			20	0.430	0
			30	0.430	0
			34	0.430	0
			7		
		9			
LMS			11		
	14				
			20		
			30		
			34		

percentile of the $\chi^2(p)$ distribution[18]. If p is 1, this critical value is 2.24. In terms of Figures 3, 4 and 5, points to the right of the vertical line with x-distance 2.24 are leverage points. The standardized residuals are expected to follow a standard normal distribution. For this probability density function the absolute value of 99% of its elements is less than 2.57. Points lying outside the strip indicated by this critical value are thus vertically outlying. The link between Figures 3, 4 and 5 and Figure 1 is that points in the mid west areas of the former figures coincide with points (a) in the latter, points in the mid east are like (c), points in the north east area are conform (d) and points in the south west area correspond to (b).

According to the GM and LMS estimates, five observations, including the four giants, are both a leverage point and a vertical outlier (see also Table 5). Star 14 is only a leverage point. LMS also identifies observation 9 as vertically outlying. On the other hand, OLS only recognizes the giant stars as leverage points but not as vertical outliers. Note that this implies that, according to the

[18] See Appendix 1.

OLS result, the four giants are not characterized by a relation between light intensity and temperature which differs much from that associated with the majority of the data (compare Figure 1, in which leverage points (c) are in line with the regular observations (a)). Moreover, star 7 and 14 are not identified as outlying. The inability of the classical (mahalanobis) x-distance to reveal these two observation as leverage points is due to the masking effect. The four giants corrupt the mahalanobis measure such that mild leverage points remain undiscovered (see also Appendix 5).

6. CONCLUSIONS

The success of OLS and its generalizations in applied economic research is not justified by its performance on 'contaminated' (real) data. If leverage points and/or vertical outliers are present, OLS not only produces unreliable parameter estimates but also incorrect t-values. Especially in economics, where data are in many cases far from smooth, the use of OLS can lead to seriously wrong conclusions.

In this paper we have described a reliable alternative, a General M estimator, which has a high breakdown point, is consistent and, as our simulation results indicate, is more efficient than Rousseeuw's LMS estimator. An example from the astronomy, involving real data, illustrates again that OLS should be used with care in applied research.

Also we have derived analytically a correction factor which, according to our simulation results, improves both the *resampling* and *projection* algorithm in approximating the MVE estimates.

On the other hand, much research still needs to be done to present a full alternative to OLS. Especially the lack of specification tests related to robust estimators is an omission.

REFERENCES

Beaton, A.E. and Tuckey, J.W., 1974, "The fitting of Power Series, Meaning Polynomials, Illustrated on Band-Spectro-Scopic Data", *Technometrics,* Vol.16, 147-85.

Chen, C. and Liu, L.M., 1993, "Joint Estimation of Model Parameters and Outlier Effects in Time Series", *Journal of the American Statistical Association,* Vol.88, p.284-97.

Gasko, M. and Donoho, D., 1982, "Influential Observation in Data Analysis", *Proceedings of the Business and Economic Statistics Section, American Statistical Association,* 104-9.

Judge, G.C., Hill, R.C., Griffiths, W.E., Lütkepohl, H. and Lee, T.-C., 1988, "Introduction to the Theory and Practice of Econometrics", *Wiley*, New York.

Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. and Stahel, W.A., 1986, "Robust Statistics. The Approach Based on Influence Functions", *Wiley,* New York.

Gómez, V. and Maravall, A., 1994, "Program Tramo: Time Series Regression with Arima Noise, Missing Observations, and Outliers; Instructions for the User", Working Paper ECO 94/31, *European University Institute.*

Rousseeuw, P.J., 1984, "Least Median of Squares Regression", *Journal of the American Statistical Association,* Vol. 79, No.388, December, 871-80.

_____, 1985, "Multivariate Estimation with High Breakdown Point", in *Mathematical Statistics and Applications,* Grossman, W., Pflug, G., Vincze, I. and Wertz, W. (eds.), *Riedel Publishing,* Dordrecht, Vol.B, 283-97.

_____ and Leroy, A.M., 1987, "Robust Regression and Outlier Detection", *Wiley,* New York.

_____ and Zomeren, B.C. van, 1990a, "Unmasking Multivariate Outliers and Leverage Points", *Journal of American Statistical Association,* Vol.85, No.411, September, 633-39.

_____ and _____, 1991, "Robust Distances: Simulations and Cutoff Values", in Stahel, W. and Weisberg, S. (eds.), "Directions in Robust Statistics and Diagnostics, Part II", *Springer Verlag*, New York (NY), p.195-203.

Simpson, D.G., Ruppert, D. and Carroll, R.J., 1992, "On One-Step GM Estimates and Stability of Inferences in Linear Regression", *Journal of the American Statistical Association,* Vol.87, No.418, June, 439-50.

APPENDIX 1. A GENERAL M ESTIMATOR DERIVED

In this appendix we derive a General M estimator (such as presented by Simpson, Ruppert and Carroll (1992)) which, by definition, downweights both vertical outliers and leverage points. Consider the classical linear model

$$q_i = z_i\gamma + v_i \quad i = 1,..,n,$$

where the v_i's are approximately *iid* $N(0,\sigma^2)$. q_i is an observable dependent variable and z_i a row vector of length p of observable explanatory variables. We first scale the model with a robust estimate[19] of σ to get

$$y_i = \boldsymbol{x}_i\beta + \varepsilon_i \quad i = 1,..,n, \tag{A1}$$

causing the ε_i's to be approximately *iid* $N(0,1)$. To estimate the finite unknown parameter vector β we have as objective

$$\min_{\beta} \sum_{i=1}^{n} \rho(r_i), \tag{A2}$$

where $r_i = y_i - \boldsymbol{x}_i\beta$ and $\rho \in C^2[-c,c]$, $\rho(\cdot):\mathbf{R}\rightarrow\mathbf{R}_+$, $c \in \mathbf{R}$. Mallows[20] proposes to state the first order condition as

$$\sum_{i=1}^{n} \boldsymbol{x}_i^T w_x(\boldsymbol{x}_i) r_i w_r(r_i) = 0, \tag{A3}$$

where $w_x(\cdot):\mathbf{R}^p\rightarrow\mathbf{R}_+$ and $w_r(\cdot):\mathbf{R}\rightarrow\mathbf{R}_+$ are weight functions based on the identification of leverage points and vertical outliers respectively (note that $w_r = (\partial\rho(r_i)/\partial r_i)/r_i$). $\boldsymbol{x}_i^T$ is the transpose of $\boldsymbol{x}_i$. Since both leverage points and vertical outliers are downweighted, the estimator belongs to the class of General M estimators (see e.g. Rousseeuw and Leroy (1987, p.13)).

We use the Newton-Raphson (NR) algorithm to solve (A3). In order to do so we have to differentiate the LHS of (A3) with respect to β, which results in

---

[19] See appendix 4 for a description of this estimator.

[20] See Hampel et al. (1986, p.322).

$$-\sum_{i=1}^{n} \boldsymbol{x}_i^T w_x(\boldsymbol{x}_i)[\boldsymbol{x}_i w_r(r_i) - r_i w_r'(r_i)],$$

where

$$w_r'(r_i) = \frac{\partial w_r(r_i)}{\partial r_i}\frac{\partial r_i}{\partial \beta}. \tag{A4}$$

An initial robust approximation of β is obtained by using Rousseeuw's LMS estimator[21].

A1.1 WEIGHTS BASED ON LEVERAGE POINTS

The identification of leverage points is independent of the estimate of β. What we need in order to detect leverage points and to determine weights correspondingly is a measure of distance defined over the explanatory variables. A classical measure is the diagonal of the hat matrix (see e.g. Judge et al. (1988, p.892)).

$$h_i = \boldsymbol{x}_i (X^T X)^{-1} \boldsymbol{x}_i^T \quad i = 1,..,n, \tag{A5}$$

where $X = (\boldsymbol{x}_1,..,\boldsymbol{x}_n)^T$. This measure can be interpreted as the extent to which an observation deviates from the sample average. Related to (A5) is the Mahalanobis distance[22]

$$MD_i = \sqrt{[\boldsymbol{x}_i - T(X)]C(X)^{-1}[\boldsymbol{x}_i - T(X)]^T} \quad i = 1,..,n, \tag{A6}$$

where $T(\cdot)$ is the arithmetic mean and $C(\cdot)$ the sample covariance matrix. Both (A5) and (A6) can however be corrupted by leverage points. Consequently, mild leverage points or a cluster of huge outliers can escape easily from being discovered when these classical non-robust indicators are used (For a prominent exposition of this phenomenon see Rousseeuw and Van Zomeren (1990). See also the example of Section 6). To overcome this *masking effect* we insert robust estimates of $T(X)$ and $C(X)^{-1}$ in (A6); the so called Minimum Volume Ellipsoid

[21] See Appendix 4 for a description of this estimator.

[22] Rousseeuw and Van Zomeren (1990) mention that $h_i = MD_i^2/(n-1) + 1/n$.

(MVE) estimates[23]. This results in robust distances RD_i since the MVE estimator has a breakdown point of 50% (see Rousseeuw and Leroy (1987)).

Weights based on leverage points are now given by

$$w_x(\boldsymbol{x}_i) = \min.\{1, \frac{\sqrt{\chi^2_{0.975}(p)}}{RD_i}\}, \quad i=1,..,n. \tag{A7}$$

Notice that all observations with a robust distance exceeding the 97.5th percentile of the $\chi^2(p)$ distribution are identified as leverage points and receive weights less than unity.

A1.2 WEIGHTS BASED ON VERTICAL OUTLIERS

The derivative of the objective function $\rho(\cdot)$ with respect to r_i, $\psi(r_i)$, measures the rate of change of the objective (A2) due to a infinitely small change in r_i. In case of OLS the objective function can be written such that $\psi(r_i)$ equals r_i. As a consequence, large residuals have a strong influence on (A2), i.e. OLS is highly sensitive to vertical outliers. To overcome this problem we propose $\psi(r_i)$ to be as depicted in Figure 6; the bi-square function. Notice that the influence of an observation first increases as its associated residual increases, but then diminishes to become zero for very large residuals.

To derive the bi-square function suppose $\rho(\cdot)$ can be described by a polynomial of degree 6 on some interval and as a constant elsewhere[24]

$$\begin{aligned} \rho(r_i) &= \alpha_1 r_i^2 + \alpha_2 r_i^4 + \alpha_3 r_i^6 \quad |r_i| < c \\ &= \gamma \qquad\qquad\qquad\quad |r_i| \geq c. \end{aligned} \tag{A8}$$

Differentiating (A8) results

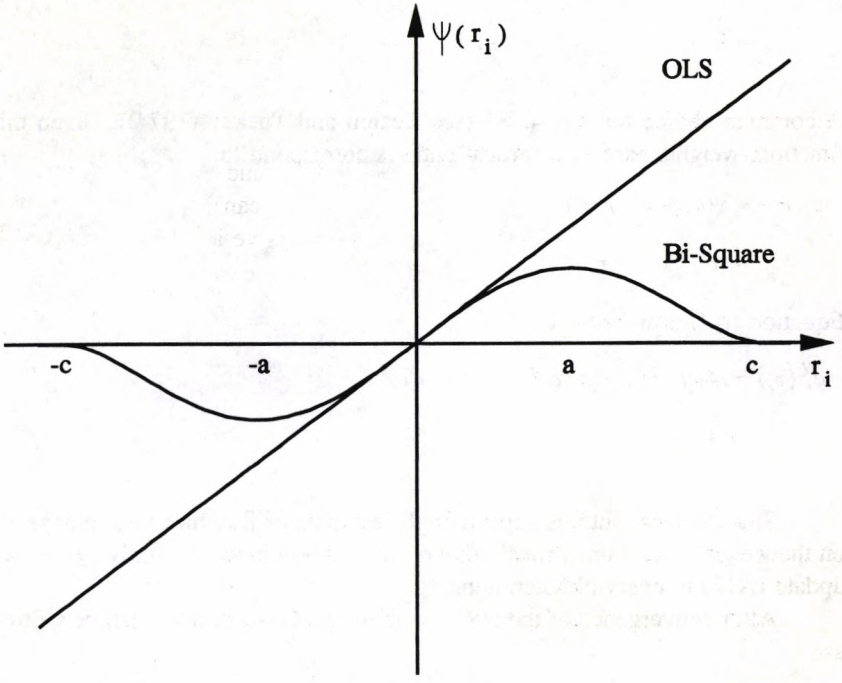
$$\begin{aligned} \psi(r_i) &= 2\alpha_1 r_i + 4\alpha_2 r_i^3 + 6\alpha_3 r_i^5 \quad |r_i| < c \\ &= 0 \qquad\qquad\qquad\quad |r_i| \geq c. \end{aligned} \tag{A9}$$

---

[23] See appendix 5 for a description of this estimator.

[24] From Figure 6 it is clear that the primitive of the bi-square function is indeed a polynomial of at least degree 6.

Figure 6 The Bi-Square Function and OLS' Psi-Function



The second derivative equals

$$\frac{\partial \psi(r_i)}{\partial r_i} = 2\alpha_1 + 12\alpha_2 r_i^2 + 30\alpha_3 r_i^4 \quad |r_i| < c$$
$$= 0 \qquad\qquad |r_i| \geq c. \tag{A10}$$

Forcing (A9) and (A10) to be continuous functions and setting $\frac{\partial \psi(r_i)}{\partial r_i}\Big|_{r_i=0}$ equal to 1 leads to the bi-square function

$$\psi(r_i) = r_i[1-(r_i/c)^2]^2 \quad |r_i| < c$$
$$= 0 \qquad\qquad |r_i| \geq c. \tag{A11}$$

A common choice for c is 4.685 (see Beaton and Tuckey (1974)). Given this function, weights based on vertical outliers correspond to

$$w_r(r_i) = \psi(r_i)/r_i \quad r_i \neq 0$$
$$= 1 \qquad r_i = 0. \tag{A12}$$

Equation (A4) now becomes

$$w_r'(r_i) = 4r_i/c^2[1-(r_i/c)^2]\boldsymbol{x}_i \quad |r_i| < c$$
$$= 0 \qquad\qquad |r_i| \geq c.$$

Since vertical outliers depend on the estimate of β, which in turn depends on the weights based on vertical outliers, $w_r(r_i)$ is computed iteratively, i.e. we update (A12) in every NR iteration.

After convergence of the NR algorithm, the GM estimator can be written as

$$\hat{\beta}_{GM} = \frac{\sum_{i=1}^{n} \boldsymbol{x}_i^T w_x(\boldsymbol{x}_i) w_r(r_i) y_i}{\sum_{i=1}^{n} \boldsymbol{x}_i^T w_x(\boldsymbol{x}_i) w_r(r_i) \boldsymbol{x}_i}, \tag{A13}$$

where $w_x(\boldsymbol{x}_i)$ is given by (A7) and $w_r(r_i)$ by (A12). In obvious matrix notation (A13) reads

$$\hat{\beta}_{GM} = [X^T W_x(X) W_r(r) X]^{-1} [X^T W_x(X) W_r(r) y], \tag{A14}$$

where $W_x(X)$ is a diagonal matrix comprising $w_x(x_i)$, $i=1,..,n$, and $W_r(r)$ is a diagonal matrix with entries $w_r(r_i)$, i=1,..,n.

APPENDIX 2. STATISTICAL PROPERTIES OF THE GM ESTIMATOR

A2.1 t-VALUES

To derive the covariance matrix of the GM estimator, $cov(\hat{\beta})$, we use a Taylor expansion around $\hat{\beta} = \beta$ for the first order condition (A3)

$$\begin{aligned} 0 &= \sum_{i=1}^{n} \psi(r_i) x_i^T w_x(x_i) \\ &\approx \sum_{i=1}^{n} \psi(e_i) x_i^T w_x(x_i) + \sum_{i=1}^{n} \frac{\partial \psi(r_i)}{\partial r_i} \frac{\partial r_i}{\partial \beta} x_i^T w_x(x_i)[\hat{\beta} - \beta]. \end{aligned} \tag{A15}$$

Evaluating (A15) in $\hat{\beta} = \beta$ and performing obvious manipulations results

$$\sqrt{n}[\hat{\beta} - \beta] \approx \sqrt{n} \left\{ \frac{1}{n} \sum_{i=1}^{n} \frac{\partial \psi(e_i)}{\partial e_i} x_i^T x_i w_x(x_i) \right\}^{-1} \left\{ \frac{1}{n} \sum_{i=1}^{n} \psi(e_i) x_i^T w_x(x_i) \right\}.$$

In large samples, the covariance matrix of $\hat{\beta}$ approximately equals

$$COV(\hat{\beta}) = [X^T V(X) W_x(X) X]^{-1} X^T W_x(X) \Psi W_x(X) X [X^T V(X) W_x(X) X]^{-1}, \tag{A16}$$

where Ψ is a diagonal matrix consisting of $\psi(r_i)^2$, $i=1,..,n$, $V(X)$ is a diagonal matrix containing $\partial\psi(r_i)/\partial r_i$, $i=1,..,n$ and $W_x(X)$ is as defined as in Appendix 1. With (A16) the usual t-values can be computed.

A2.2 BREAKDOWN POINT

To what extent a regression estimator is affected by corrupted data is measured by its breakdown point: the smallest amount of contamination that can cause the estimator to take on arbitrary values. Suppose $T(\cdot)$ is an estimator. Define $\eta(\delta; T(\cdot), X)$ to be the supremum of $\|T(X') - T(X)\|$ for all X', where X' corresponds to the original data set with a fraction δ (=m/n) of the observations

replaced by arbitrary values. The breakdown point of $T(\cdot)$ on X is then defined as (Rousseeuw (1984))

$$b(T(\cdot),X) = \min\{\frac{m}{n};\eta(m;T(\cdot),X)=\infty\}, \tag{A17}$$

where n is the number of cases in X and m the number of original data points replaced by arbitrary values. (From (A17) it is clear that the OLS estimator has a breakdown point of zero; a single outlying observation can cause this technique to produce any estimate).

Simpson, Ruppert and Carroll show that "under reasonably general conditions the regression parameter estimates (GM estimates) inherit the breakdown properties of the preliminary estimates of the regression parameters and the multivariate location and scale estimates of the design x's" (1992, p.446). In particular it must be the case that

$$\frac{\psi(r_i)}{r_i} > 0 \quad |r_i|<a, \tag{A18}$$

$$\frac{\partial\psi(r_i)}{\partial r_i} > 0 \quad |r_i| \quad \textit{for at least } n-m-\frac{n}{2} \textit{ good points}, \tag{A19}$$

$$\frac{\partial\psi(r_i)}{\partial r_i} \geq 0 \quad |r_i|<a \quad \textit{for all points}, \tag{A20}$$

the set of 'good' points (i.e. $|r_i|<a$*) must contain a linearly independent subset of size p,* (A21)

and a must strictly exceed some tuning constant κ (A22)

From (A11) it is straightforward to recognize that (A18) holds (see also Figure 6). For 'bad' points (i.e. $a<|r_i|<c$) it is the case that $\partial\psi(r_i)/\partial r_i<0$, and (A20) is violated (see also (A28)). However, if (A19) holds it is always possible to

manipulate these 'bad' points such that (A20) holds[25]. Since the majority of the r_i's are assumed to come from a normal distribution, condition (A19) is always met. Finally, if c equals 4.685 than a is equal to 2.095 and exceeds κ (κ being 0.6745, see Appendix 3)).

To summarize, for the bi-square function conditions (A18)-(A22) are fulfilled. Therefore, the described GM estimator can handle up to 50% pollution of the data since both the LMS and MVE estimators have breakdown points of 50%.

A2.3 CONSISTENCY

Simpson, Ruppert and Carroll (1992) show that under certain conditions the GM estimator is asymptotic normal and produces root-n consistent estimates of the unknown parameters, if the identification of leverage points is based on MVE estimates. In particular it must be that

The score function $\psi(r_i)$ *is bounded and continuous,* (A23)

$E[\psi(\varepsilon v)] = 0$ *and* $E[\varepsilon v \frac{\partial \psi(\varepsilon v)}{\partial r_i}] = 0$ *for any nonnegative scalar* v, (A24)

$\psi(r_i)$ *has derivative* $\frac{\partial \psi(r_i)}{\partial r_i}$ *such that*

$$\| \frac{\partial \psi(r_i)}{\partial r_i} \|_{sup} < \infty \text{ and } \| r_i \frac{\partial \psi(r_i)}{\partial r_i} \|_{sup} < \infty, \quad \text{(A25)}$$

where $\|.\|_{sup}$ *is the supremum norm,*

$$\frac{\partial \psi(r_i)}{\partial r_i} \text{ has derivative } \frac{\partial^2 \psi(r_i)}{\partial r_i^2} \text{ such that}$$
$$\| \frac{\partial^2 \psi(r_i)}{\partial r_i^2} \|_{sup} < \infty, \| r_i \frac{\partial^2 \psi(r_i)}{\partial r_i^2} \|_{sup} < \infty \text{ and } \| r_i^2 \frac{\partial^2 \psi(r_i)}{\partial r_i^2} \|_{sup} < \infty \quad \text{(A26)}$$

and

[25] In particular these 'bad' data points can be set equal to a forcing their contribution to the NR-version of the hessian to be zero (see also Ruppert, Simpson and Carroll (1992, p.441, Remark 2.1)).

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} var[\psi(\varepsilon_i)]w_x(x_i)^2 x_i^T x_i = A \text{ and}$$

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} E[\frac{\partial\psi(\varepsilon_i)}{\partial\varepsilon_i}]w_x(x_i)x_i^T x_i = B \qquad \text{(A27)}$$

for some symmetric positive definite matrices A and B.

Condition (A23) and (A24) are checked easily using (A11), noting that ϵ_i's are approximately *iid* N(0,1) and remembering that all odd moments of the normal distribution are zero. Further note that the derivative of $\psi(\cdot)$ equals

$$\frac{\partial\psi(r_i)}{\partial r_i} = [1-(r_i/c)^2][1-5(r_i/c)^2] \quad |r_i|<c$$
$$= 0 \qquad\qquad |r_i|\geq c. \qquad \text{(A28)}$$

Clearly (A28) is bounded and becomes 0 for large residuals. As a result (A25) and (A26) follow immediately. Finally, since the errors are assumed to approximately follow a (standard) normal distribution, and because both $\psi(\varepsilon_i)$ and $\partial\psi(\varepsilon_i)/\partial\varepsilon_i$ are bounded, condition (A27) is satisfied.

APPENDIX 3. A ROBUST ESTIMATOR FOR THE VARIANCE OF A REGRESSION

Consider a normally distributed random variable, u, with mean μ and variance σ^2. For a sample of n observations of u (denoted by $\boldsymbol{u}$) the sample median is a robust estimator of the location parameter μ. Consequently u - med(u) is symmetrically distributed around zero. If we take the median of the absolute value of $\boldsymbol{u}$ - med($\boldsymbol{u}$), we have a robust estimate of the 75th percentile of the initial distribution of u. Therefore it is straightforward to propose

$$s = med(|u-med(u)|)/0.6745, \qquad \text{(A29)}$$

as a robust estimator for σ where 0.6745 is the 75th percentile of the standard normal distribution.

Table A1 shows that (A29) is indeed relatively insensitive to outlying observations. In order to generate this table standard normal vectors of length 500 where generated for which both the usual standard deviation and (A29) where calculated. We then contaminated the series by replacing 50 randomly

Table A1. Simulation Results, n=500, 10% Pollution[a]

Normal		Contaminated	
Classic	Robust	Classic	Robust
0.99997468	0.99791024	3.29625390	1.11684750
0.00095094	0.00277199	5.36803770	0.01693631
0.00095094	0.00276763	0.09525565	0.00328297

[a] The entries in each cell are the mean estimated value, the mean squared error and the variance over the 1000 runs.

drawn observations by elements of a normal distribution with mean zero and variance 100. For these contaminated series we also calculated both (A29) and the classical standard variation.

In case of uncontaminated data the classical estimator is of course most efficient since it is the uniformly minimum variance unbiased estimator of σ. However, when the data are less smooth it is striking to see that this estimator more than three times over-estimates the standard deviation of 90 percent of the data. In this case the robust estimator keeps producing reliable estimates.

We use as final estimation of the variance of the regression (see e.g. Table 4) the square of (A29), where u_i is defined as

$$u_i = y_i - \boldsymbol{x}_i \beta_{GM} \quad i = 1,..,n. \tag{A30}$$

APPENDIX 4. LEAST MEDIAN OF SQUARES REGRESSION

Rousseeuw (1984) introduced a simple yet elegant and robust estimator for the unknown parameter vector β in the classical linear model

$$y_i = \boldsymbol{x}_i \beta + \varepsilon_i \quad i = 1,..,n.$$

The conventional least squares technique has as objective

$$\min_{\beta} \frac{1}{n} \sum_{i=1}^{n} (y_i - \boldsymbol{x}_i \beta)^2. \tag{A31}$$

Replacing the summation in (A31) by the median over all squares yields the Least Median of Squares estimator

$$\min_{\beta} med(y_i - \boldsymbol{x}_i\beta)^2. \tag{A32}$$

Rousseeuw (1984) shows that for n observations and p explanatory variables the breakdown point of the LMS estimator is $([n/2]-p+2)/n$, which is as high as 50% when n goes to infinity (the notation [r] stands for the largest integer less than or equal to r).

To approximate (A32) we compute the OLS estimator of a randomly drawn sample of size $p+1$ (this results a perfect fit). Given this estimate, β_{LMS}, we calculate the objective value $med(y_i - \boldsymbol{x}_i\beta_{LMS})^2$ and proceed with the next drawing. The estimate with the lowest objective value is the first approximation of the LMS estimate. We then proceed by calculating a refinement for the intercept estimate. In particular, this final estimate is given by

$$\beta_{LMS_1} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \boldsymbol{x}_i^*\beta_{LMS}^*), \tag{A33}$$

where $\boldsymbol{x}_i^* = x_{i,2}..x_{i,p}$ and $\beta_{LMS}^* = (\beta_{LMS_2},..,\beta_{LMS_p})^T$ (see Rousseeuw and Leroy (1987)).

If all distinct samples were to be enumerated, there were far too many estimates to be made. However Rousseeuw and Leroy (1987) show that the probability of getting a sub sample consisting of p non-outlying observations when the fraction of contamination equals δ is

$$1-(1-(1-\delta)^p)^m, \tag{A34}$$

where m is the required number of independent sub samples. Rewriting (A34) gives the number of drawings needed to get a 'good' sub sample with probability Λ

$$m = \ln(1-\Lambda)/\ln(1-(1-\delta)^p).$$

In our GAUSS program Λ is set equal to 0.99 and δ is set equal to 0.5 (However, when approximating the LMS estimator we used 10 times as much trials as prescribed by (A34)).

The standard deviation estimate associated with the LMS technique is the result of a two stage procedure (see Rousseeuw and Leroy (1987)). An initial estimate is made according to

$$s^0 = (1+5/(n-p))\sqrt{med[r_i(\hat{\beta}_{LMS})^2]}\ /0.6745.$$

Given s^0 weights are determined

$$w_i = \begin{cases} 1 & if\ |r_i/s^0| \le 2.5 \\ 0 & otherwise \end{cases}.$$

The second step yields the estimate of the standard deviation

$$s_{LMS} = \sqrt{(\sum_{i=1}^{n} w_i r_i^2)/(\sum_{i=1}^{n} w_i - p)}\ . \tag{A35}$$

We use (A35) as scale factor to get (A1).

APPENDIX 5. THE MINIMUM VOLUME ELLIPSOID ESTIMATOR

A5.1 DEFINITION OF THE ESTIMATOR

The MVE estimator (see Rousseeuw and Leroy (1987) and Rousseeuw and Van Zomeren (1990)) is based on the (hyper) ellipsoid of minimum (hyper) area containing at least half of the observations. The location estimate corresponds to the centre of this (hyper) ellipsoid while the corresponding covariance estimate is the (hyper) ellipsoid multiplied by some factor to obtain consistency.

Since the (absolute value of the) determinant of a matrix can be interpreted as the hyper volume of the space comprised by the columns of this matrix, the MVE estimator, the pair (T,C), is formally defined as

$$\min_{(T,C)} |\det(C)| \tag{A36}$$

$$\ni:\ \#\{i;[\boldsymbol{x}_i - T]C^{-1}[\boldsymbol{x}_i - T]^T \le \delta^2\} \ge h$$

where $h = [(n+p+1)/2]$. If it is assumed that the majority of the data comes from a normal distribution, δ^2 is set equal to the 50th percentile of the $\chi^2(p)$ distribution.

Rousseeuw and Leroy (1987) prove that the breakdown point of the MVE estimator is $([n/2]-p+1)/n$, which is 50% as the number of observations, n, goes to infinity.

A5.2 DERIVATION OF A CORRECTION FACTOR

To approximate (A36) we have tested two algorithms; the *resampling* algorithm and the *projection* algorithm (see Rousseeuw and Van Zomeren (1990, 1991) and Gasko and Donoho (1982)). In the one-dimensional version of the *projection* algorithm we look at

$$RD_i = \frac{|x_i - (x_j + x_{j-h+1})/2|}{x_j - x_{j-h+1}} \quad i = 1,..,n, \tag{A37}$$

where $h = [(n+p+1)/2]$, $x_j - x_{j-h+1}$ is the smallest of the differences of

$$x_h - x_1, x_{h+1} - x_2, ..., x_n - x_{n-h+1}$$

and $x_1 \leq x_2 \leq ... \leq x_n$. If x *iid* $N(\beta, \sigma^2)$, then the median of $(x-\beta)/\sigma$ must equal the square root of $\chi^2_{0.50}(1)$ or 0.675. On the other hand, by definition of h, 50% of the x_i's are in between x_{j-h+1} and x_j. Therefore, the median difference (over i) between x_i and $(x_{j-h+1} + x_j)/2$ equals

$$x_j - (x_{j-h+1} + x_j)/2 = (x_{j-h+1} + x_j)/2 - x_{j-h+1} = (x_j - x_{j-h+1})/2. \tag{A38}$$

Inserting (A38) into (A37) reveals that the median of RD_i equals 0.5 in stead of 0.675. To restore this deviation we propose as correction factor[26]

$$\frac{\sqrt{\chi^2_{0.50}(p)}}{\underset{i}{med.}(RD_i)}. \tag{A39}$$

We have chosen the 50th percentile of the χ^2-distribution since the MVE estimator has a breakdown point of 50%.

A5.3 SIMULATION RESULTS

We have tested different versions of both the *resampling* and *projection* algorithm to approximate (A36). Table A2 reports our simulation results. Each experiment consisted of generating a standard normal random variable for which

[26] This factor is originally conceived by Teun Kloek.

the MVE-distances were approximated, using various versions of the resampling algorithm (lines 1-4) and the projection algorithm (lines 5-6). It was then examined how many of these distances exceeded the square of the 97.5th percentile of the $\chi^2(p)$-distribution ($\hat{\alpha}$). This was repeated a thousand times both for a small sample (n = 50) and a large sample (n = 200), as well as for a number of explanatory variables. In Table A2 the mean outcomes of $1-\hat{\alpha}$ are reported. Further the variances of the different $1-\hat{\alpha}$ were calculated with which the hypothesis H_0: $\hat{\alpha}=0.025$ could be tested. The absolute values of the test statistic

$$\frac{0.975-(1-\frac{1}{1000}\sum_{i=1}^{1000}\hat{\alpha}_i)}{\sqrt{var(\hat{\alpha})/1000}}$$

are also reported in Table A2.

The rows labelled (1) of Table A2 contain the results obtained when using the resampling algorithm as described in Rousseeuw and van Zomeren (1990, Appendix). Apart from the one variable case this algorithm does not adequately approximate the MVE-distances. The second entries (2) involved the resampling algorithm with the one-step improvement as suggested by Rousseeuw and Leroy (1987, p.260). This however does not significantly improve the performance of the resampling algorithm. In fact, only for the small sample with p=1 it yields reliable distances. Algorithm (3) consists of applying correction factor (A39) on (1). This however does not give any satisfactory result. Both the one-step improvement of Rousseeuw and Leroy (1987) and (A39) are used in algorithm (4). Surprisingly, this version of the resampling algorithm has the best performance of all versions considered[27]. For the large sample the hypothesis that $1-\hat{\alpha}$ = 0.975 cannot be rejected. And although it is rejected for p = 3,..,5 when n = 50, it is only by a small margin. The projection algorithm (5) almost never yields appropriate distance approximations (of course, for p = 1 this follows immediately from the last section). Applying correction factor (A39) to this algorithm does improve its performance as can be seen from the rows marked (6). If the sample is large and there are a substantial number of explanatory variables (i.e. p exceeds 2), H_0 cannot be rejected when using this adjusted version of the projection algorithm.

[27] The authors cannot give an analytical explanation for this result.

Table A2 Simulation Results of the Resampling and Projection Algorithms.

Algo-rithm	n=50									
	p=1		p=2		p=3		p=4		p=5	
	1-α	t	1-α	t	1-α	t	1-α	t	1-α	t
(1)[a]	0.9819	**1.39**	0.9576	2.64	0.9505	3.46	0.9488	3.80	0.9501	3.61
(2)	0.9675	**1.39**	0.9562	2.97	0.9519	3.42	0.9525	3.43	0.9563	2.97
(3)	0.9355	5.93	0.8783	12.11	0.8569	14.39	0.8426	16.20	0.8419	16.34
(4)	0.9677	**1.25**	0.9650	**1.70**	0.9611	2.25	0.9599	2.50	0.9628	2.08
(5)	0.9849	2.12	0.9742	**0.15**	0.9875	2.92	0.9927	4.75	0.9979	8.49
(6)	0.9352	5.95	0.9768	**0.33**	0.9680	**1.26**	0.9544	3.15	0.9445	4.62
	n=200									
(1)	0.9730	**0.51**	0.9637	2.50	0.9584	3.48	0.9523	4.66	0.9418	6.33
(2)	0.9609	3.47	0.9605	3.49	0.9607	3.53	0.9602	3.67	0.9586	3.91
(3)	0.9605	3.39	0.9462	5.82	0.9372	7.22	0.9274	8.90	0.9129	11.01
(4)	0.9727	**0.59**	0.9712	**1.00**	0.9716	**0.92**	0.9706	**1.18**	0.9703	**1.24**
(5)	0.9940	7.52	0.9933	7.22	0.9965	10.20	0.9984	13.39	0.9994	18.04
(6)	0.9601	3.48	0.9855	3.16	0.9811	**1.71**	0.9737	**0.31**	0.9658	**1.98**

[a] (1) Resampling
(2) Resampling with one-step improvement
(3) Resampling with correction factor (A39)
(4) Resampling with one-step improvement *and* correction factor (A39)
(5) Projection
(6) Projection with correction factor (A39)

To conclude, in all cases considered the resampling algorithm with Rousseeuw and Leroy's one step improvement combined with correction factor (A39) yields reliable approximations of the MVE distances. On the other hand, if the sample is large and the number of explanatory variables not too small, applying correction factor (A39) to the projection algorithm also results in appropriate distance approximations. In this case we recommend using the latter since the projection algorithm is much faster than the resampling algorithm (see also Rousseeuw and Van Zomeren (1991)).

Working Papers of the Department of Economics
Published since 1993

ECO No. 93/1
Carlo GRILLENZONI
Forecasting Unstable and Non-Stationary Time Series

ECO No. 93/2
Carlo GRILLENZONI
Multilinear Models for Nonlinear Time Series

ECO No. 93/3
Ronald M. HARSTAD/Louis PHLIPS
Futures Market Contracting When You Don't Know Who the Optimists Are

ECO No. 93/4
Alan KIRMAN/Louis PHLIPS
Empirical Studies of Product Markets

ECO No. 93/5
Grayham E. MIZON
Empirical Analysis of Time Series: Illustrations with Simulated Data

ECO No. 93/6
Tilman EHRBECK
Optimally Combining Individual Forecasts From Panel Data

ECO NO. 93/7
Víctor GÓMEZ/Agustín MARAVALL
Initializing the Kalman Filter with Incompletely Specified Initial Conditions

ECO No. 93/8
Frederic PALOMINO
Informed Speculation: Small Markets Against Large Markets

ECO NO. 93/9
Stephen MARTIN
Beyond Prices Versus Quantities

ECO No. 93/10
José María LABEAGA/Angel LÓPEZ
A Flexible Demand System and VAT Simulations from Spanish Microdata

ECO No. 93/11
Maozu LU/Grayham E. MIZON
The Encompassing Principle and Specification Tests

ECO No. 93/12
Louis PHLIPS/Peter MØLLGAARD
Oil Stocks as a Squeeze Preventing Mechanism: Is Self-Regulation Possible?

ECO No. 93/13
Pieter HASEKAMP
Disinflation Policy and Credibility: The Role of Conventions

ECO No. 93/14
Louis PHLIPS
Price Leadership and Conscious Parallelism: A Survey

ECO No. 93/15
Agustín MARAVALL
Short-Term Analysis of Macroeconomic Time Series

ECO No. 93/16
Philip Hans FRANSES/Niels HALDRUP
The Effects of Additive Outliers on Tests for Unit Roots and Cointegration

ECO No. 93/17
Fabio CANOVA/Jane MARRINAN
Predicting Excess Returns in Financial Markets

ECO No. 93/18
Iñigo HERGUERA
Exchange Rate Fluctuations, Market Structure and the Pass-through Relationship

ECO No. 93/19
Agustín MARAVALL
Use and Misuse of Unobserved Components in Economic Forecasting

ECO No. 93/20
Torben HOLVAD/Jens Leth HOUGAARD
Measuring Technical Input Efficiency for Similar Production Units: A Survey of the Non-Parametric Approach

ECO No. 93/21
Stephen MARTIN/Louis PHLIPS
Product Differentiation, Market Structure and Exchange Rate Passthrough

ECO No 93/22
F. CANOVA/M. FINN/A. R. PAGAN
Evaluating a Real Business Cycle Model

ECO No 93/23
Fabio CANOVA
Statistical Inference in Calibrated Models

ECO No 93/24
Gilles TEYSSIÈRE
Matching Processes in the Labour Market in Marseilles. An Econometric Study

ECO No 93/25
Fabio CANOVA
Sources and Propagation of International Business Cycles: Common Shocks or Transmission?

ECO No. 93/26
Marco BECHT/Carlos RAMÍREZ
Financial Capitalism in Pre-World War I Germany: The Role of the Universal Banks in the Financing of German Mining Companies 1906-1912

ECO No. 93/27
Isabelle MARET
Two Parametric Models of Demand, Structure of Market Demand from Heterogeneity

ECO No. 93/28
Stephen MARTIN
Vertical Product Differentiation, Intra-industry Trade, and Infant Industry Protection

ECO No. 93/29
J. Humberto LOPEZ
Testing for Unit Roots with the k-th Autocorrelation Coefficient

ECO No. 93/30
Paola VALBONESI
Modelling Interactions Between State and Private Sector in a "Previously" Centrally Planned Economy

ECO No. 93/31
Enrique ALBEROLA ILA/J. Humberto LOPEZ/Vicente ORTS RIOS
An Application of the Kalman Filter to the Spanish Experience in a Target Zone (1989-92)

ECO No. 93/32
Fabio CANOVA/Morten O. RAVN
International Consumption Risk Sharing

ECO No. 93/33
Morten Overgaard RAVN
International Business Cycles: How much can Standard Theory Account for?

ECO No. 93/34
Agustín MARAVALL
Unobserved Components in Economic Time Series *

ECO No. 93/35
Sheila MARNIE/John MICKLEWRIGHT
"Poverty in Pre-Reform Uzbekistan: What do Official Data Really Reveal?"

ECO No. 93/36
Torben HOLVAD/Jens Leth HOUGAARD
Measuring Technical Input Efficiency for Similar Production Units: 80 Danish Hospitals

ECO No. 93/37
Grayham E. MIZON
A Simple Message for Autocorrelation Correctors: DON'T

ECO No. 93/38
Barbara BOEHNLEIN
The Impact of Product Differentiation on Collusive Equilibria and Multimarket Contact

ECO No. 93/39
H. Peter MØLLGAARD
Bargaining and Efficiency in a Speculative Forward Market

***

*out of print

ECO No. 94/1
Robert WALDMANN
Cooperatives With Privately Optimal Price Indexed Debt Increase Membership When Demand Increases

ECO No. 94/2
Tilman EHRBECK/Robert WALDMANN
Can Forecasters' Motives Explain Rejection of the Rational Expectations Hypothesis?

ECO No. 94/3
Alessandra PELLONI
Public Policy in a Two Sector Model of Endogenous Growth *

ECO No. 94/4
David F. HENDRY
On the Interactions of Unit Roots and Exogeneity

ECO No. 94/5
Bernadette GOVAERTS/David F. HENDRY/Jean-François RICHARD
Encompassing in Stationary Linear Dynamic Models

ECO No. 94/6
Luigi ERMINI/Dongkoo CHANG
Testing the Joint Hypothesis of Rationality and Neutrality under Seasonal Cointegration: The Case of Korea

ECO No. 94/7
Gabriele FIORENTINI/Agustín MARAVALL
Unobserved Components in ARCH Models: An Application to Seasonal Adjustment *

ECO No. 94/8
Niels HALDRUP/Mark SALMON
Polynomially Cointegrated Systems and their Representations: A Synthesis

ECO No. 94/9
Mariusz TAMBORSKI
Currency Option Pricing with Stochastic Interest Rates and Transaction Costs: A Theoretical Model

ECO No. 94/10
Mariusz TAMBORSKI
Are Standard Deviations Implied in Currency Option Prices Good Predictors of Future Exchange Rate Volatility?

ECO No. 94/11
John MICKLEWRIGHT/Gyula NAGY
How Does the Hungarian Unemployment Insurance System Really Work?

ECO No. 94/12
Frank CRITCHLEY/Paul MARRIOTT/Mark SALMON
An Elementary Account of Amari's Expected Geometry

ECO No. 94/13
Domenico Junior MARCHETTI
Procyclical Productivity, Externalities and Labor Hoarding: A Reexamination of Evidence from U.S. Manufacturing

ECO No. 94/14
Giovanni NERO
A Structural Model of Intra-European Airline Competition

ECO No. 94/15
Stephen MARTIN
Oligopoly Limit Pricing: Strategic Substitutes, Strategic Complements

ECO No. 94/16
Ed HOPKINS
Learning and Evolution in a Heterogeneous Population

ECO No. 94/17
Berthold HERRENDORF
Seigniorage, Optimal Taxation, and Time Consistency: A Review

ECO No. 94/18
Frederic PALOMINO
Noise Trading in Small Markets

ECO No. 94/19
Alexander SCHRADER
Vertical Foreclosure, Tax Spinning and Oil Taxation in Oligopoly

ECO No. 94/20
Andrzej BANIAK/Louis PHLIPS
La Pléiade and Exchange Rate Pass-Through

ECO No. 94/21
Mark SALMON
Bounded Rationality and Learning; Procedural Learning

*out of print

ECO No. 94/22
Isabelle MARET
Heterogeneity and Dynamics of Temporary Equilibria: Short-Run Versus Long-Run Stability

ECO No. 94/23
Nikolaos GEORGANTZIS
Short-Run and Long-Run Cournot Equilibria in Multiproduct Industries

ECO No. 94/24
Alexander SCHRADER
Vertical Mergers and Market Foreclosure: Comment

ECO No. 94/25
Jeroen HINLOOPEN
Subsidising Cooperative and Non-Cooperative R&D in Duopoly with Spillovers

ECO No. 94/26
Debora DI GIOACCHINO
The Evolution of Cooperation: Robustness to Mistakes and Mutation

ECO No. 94/27
Kristina KOSTIAL
The Role of the Signal-Noise Ratio in Cointegrated Systems

ECO No. 94/28
Agustín MARAVALL/Víctor GÓMEZ
Program SEATS "Signal Extraction in ARIMA Time Series" - Instructions for the User

ECO No. 94/29
Luigi ERMINI
A Discrete-Time Consumption-CAP Model under Durability of Goods, Habit Formation and Temporal Aggregation

ECO No. 94/30
Debora DI GIOACCHINO
Learning to Drink Beer by Mistake

ECO No. 94/31
Víctor GÓMEZ/Agustín MARAVALL
Program TRAMO "Time Series Regression with ARIMA Noise, Missing Observations, and Outliers" - Instructions for the User

ECO No. 94/32
Ákos VALENTINYI
How Financial Development and Inflation may Affect Growth

ECO No. 94/33
Stephen MARTIN
European Community Food Processing Industries

ECO No. 94/34
Agustín MARAVALL/Christophe PLANAS
Estimation Error and the Specification of Unobserved Component Models

ECO No. 94/35
Robbin HERRING
The "Divergent Beliefs" Hypothesis and the "Contract Zone" in Final Offer Arbitration

ECO No. 94/36
Robbin HERRING
Hiring Quality Labour

ECO No. 94/37
Angel J. UBIDE
Is there Consumption Risk Sharing in the EEC?

ECO No. 94/38
Berthold HERRENDORF
Credible Purchases of Credibility Through Exchange Rate Pegging: An Optimal Taxation Framework

ECO No. 94/39
Enrique ALBEROLA ILA
How Long Can a Honeymoon Last? Institutional and Fundamental Beliefs in the Collapse of a Target Zone

ECO No. 94/40
Robert WALDMANN
Inequality, Economic Growth and the Debt Crisis

ECO No. 94/41
John MICKLEWRIGHT/ Gyula NAGY
Flows to and from Insured Unemployment in Hungary

ECO No. 94/42
Barbara BOEHNLEIN
The Soda-ash Market in Europe:
Collusive and Competitive Equilibria
With and Without Foreign Entry

ECO No. 94/43
Hans-Theo NORMANN
Stackelberg Warfare as an Equilibrium
Choice in a Game with Reputation Effects

ECO No. 94/44
Giorgio CALZOLARI/Gabriele
FIORENTINI
Conditional Heteroskedasticity in
Nonlinear Simultaneous Equations

ECO No. 94/45
Frank CRITCHLEY/Paul MARRIOTT/
Mark SALMON
On the Differential Geometry of the Wald
Test with Nonlinear Restrictions

ECO No. 94/46
Renzo G. AVESANI/Giampiero M.
GALLO/Mark SALMON
On the Evolution of Credibility and
Flexible Exchange Rate Target Zones

***

ECO No. 95/1
Paul PEZANIS-CHRISTOU
Experimental Results in Asymmetric
Auctions - The 'Low-Ball' Effect

ECO No. 95/2
Jeroen HINLOOPEN/Rien
WAGENVOORT
Robust Estimation: An Example