The Canadian Journal of Statistics
Vol. 28, No. 2, 2000, Pages ???-???
La revue canadienne de statistique

Robust estimation of the SUR model*

Martin BILODEAU and Pierre DUCHESNE
Université de Montréal

Key words and phrases: robustness, S-estimators, seemingly unrelated regression.
Mathematics subject classification codes (1991): primary 62J20; secondary 62F35.

ABSTRACT

This paper proposes robust regression to solve the problem of outliers in seemingly unrelated regression (SUR) models. We present an adaptation of S-estimators to SUR models. S-estimators are robust, with high breakdown point, and are much more efficient than other robust regression estimators commonly used in practice. Furthermore, modifications to Ruppert's algorithm allow a fast evaluation of them in this context. The classical example of U.S. corporations is revisited, and it appears that the procedure gives an interesting insight into the problem.

RÉSUMÉ

Nous proposons dans cet article une méthode de régression robuste pour résoudre le problème des valeurs aberrantes dans les modèles SUR. Nous adaptons les S-estimateurs dans les modèles SUR. Les S-estimateurs sont robustes, avec un haut point de rupture, et sont beaucoup plus efficaces que les autres estimateurs robustes de régression utilisés couramment en pratique. De plus, une modification de l'algorithme de Ruppert permet une évaluation rapide de ces estimateurs dans ce contexte. La procédure donne une compréhension intéressante du problème classique sur les compagnies américaines.

1. INTRODUCTION

Since their introduction by Zellner (1962), SUR models have taken an important place in econometrics and in statistics. See for example Judge *et al.* (1985). Srivastava and Giles (1987) give a detailed treatment of estimation and inference in SUR models. However, since the procedure proposed originally by Zellner (1962) is essentially a least squares estimator in a multiple equations model with a particular covariance matrix, it is expected that the estimator is vulnerable to outliers. Robust alternatives using M-estimators of the SUR models are proposed in Koenker and Portnoy (1990). However, the methods studied there are not affine equivariant

*This research was supported in part by a grant from the Natural Sciences and Engineering Research Council of Canada.

and do not take full account of the multivariate nature of the problem. This paper proposes S-estimators which are affine equivariant and can detect multivariate outliers.

Consider a system of q equations

$$\begin{cases} \mathbf{y}_1 = & \mathbf{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\epsilon}_1, \\ \vdots & \\ \mathbf{y}_q = & \mathbf{X}_q\boldsymbol{\beta}_q + \boldsymbol{\epsilon}_q, \end{cases} \tag{1}$$

where $\mathbf{y}_i$ and $\boldsymbol{\epsilon}_i$ are $n \times 1$ vectors, $\mathbf{X}_i$ is a $n \times p_i$ matrix, $\boldsymbol{\beta}_i$ is a $p_i \times 1$ vector and suppose $E(\boldsymbol{\epsilon}_i) = \mathbf{0}$ and $\text{var}(\boldsymbol{\epsilon}_i) = \sigma_{ii}\mathbf{I}_n$, for each equation, $i = 1, \ldots, q$. The particularity of the SUR model is that $\text{cov}(\boldsymbol{\epsilon}_i, \boldsymbol{\epsilon}_j) = \sigma_{ij}\mathbf{I}_n$, $i, j = 1, \ldots, q$.

Using matrix notation, we can write (1) compactly in two ways. First, as $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\mathbf{y} = (\mathbf{y}_1^T, \ldots, \mathbf{y}_q^T)^T$, $\boldsymbol{\epsilon} = (\boldsymbol{\epsilon}_1^T, \ldots, \boldsymbol{\epsilon}_q^T)^T$ and $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \ldots, \boldsymbol{\beta}_q^T)^T$. The design matrix is

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{X}_q \end{pmatrix}.$$

The error term has mean zero and the variance matrix is given by

$$\text{var}(\boldsymbol{\epsilon}) = \mathbf{V} = \boldsymbol{\Sigma} \otimes \mathbf{I}_n,$$

where

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11} & \cdots & \sigma_{1q} \\ \vdots & \ddots & \vdots \\ \sigma_{q1} & \cdots & \sigma_{qq} \end{pmatrix}.$$

The Aitken's estimator of $\boldsymbol{\beta}$, for known $\mathbf{V}$, is given by

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{y} \\ &= [\mathbf{X}^T(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_n)\mathbf{X}]^{-1}\mathbf{X}^T(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_n)\mathbf{y}. \end{aligned}$$

However, since $\mathbf{V}$ is rarely known, a calculable estimator is given by

$$\hat{\boldsymbol{\beta}}_c = [\mathbf{X}^T(\hat{\boldsymbol{\Sigma}}^{-1} \otimes \mathbf{I}_n)\mathbf{X}]^{-1}\mathbf{X}^T(\hat{\boldsymbol{\Sigma}}^{-1} \otimes \mathbf{I}_n)\mathbf{y},$$

where $\hat{\boldsymbol{\Sigma}}$ is a consistent estimator of $\boldsymbol{\Sigma}$. A second equivalent formulation uses multivariate regression $\mathbf{Y} = \tilde{\mathbf{X}}\mathbf{B} + \mathbf{E}$, where $\mathbf{Y} = (\mathbf{y}_1, \ldots, \mathbf{y}_q)$ and $\mathbf{E} = (\mathbf{e}_1, \ldots, \mathbf{e}_n)^T$ with $\mathbf{e}_i$ a $q \times 1$ vector, $\tilde{\mathbf{X}} = (\mathbf{X}_1, \ldots, \mathbf{X}_q)$. The coefficient matrix here has a constrained structure

$$\mathbf{B} = \begin{pmatrix} \boldsymbol{\beta}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\beta}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{\beta}_q \end{pmatrix}.$$

The error term has variance matrix

$$\text{var}(\mathbf{E}) = \text{var}[\text{vec}(\mathbf{E}^T)] = \mathbf{I}_n \otimes \boldsymbol{\Sigma}.$$

For an estimate $\hat{\boldsymbol{\beta}} = (\hat{\boldsymbol{\beta}}_1, \ldots, \hat{\boldsymbol{\beta}}_q)^T$, $\hat{\boldsymbol{\Sigma}}$ uses the inner product matrix of residuals

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{n} \begin{pmatrix} \hat{\boldsymbol{\epsilon}}_1^T \\ \vdots \\ \hat{\boldsymbol{\epsilon}}_q^T \end{pmatrix} (\hat{\boldsymbol{\epsilon}}_1, \ldots, \hat{\boldsymbol{\epsilon}}_q),$$

or equivalently $\hat{\boldsymbol{\Sigma}} = (\mathbf{Y} - \tilde{\mathbf{X}}\hat{\mathbf{B}})^T(\mathbf{Y} - \tilde{\mathbf{X}}\hat{\mathbf{B}})/n = \sum_{i=1}^n \hat{\mathbf{e}}_i \hat{\mathbf{e}}_i^T/n$. Maximum likelihood estimators (MLE) are discussed in Srivastava and Giles (1987); Zellner (1971) used a Bayesian approach. See also Theil (1971) and Judge *et al.* (1985) for econometrics applications.

2. S-ESTIMATORS

We recall S-estimators of regression and S-estimators in the multivariate case. It is well known that the ordinary least squares (OLS) estimator is not robust. A single observation can bring OLS estimators over any arbitrary bound. Because of that, OLS estimators possess zero breakdown point (BP) property. BP is a global measure of the proportion of bad observations that an estimator can handle *before it breaks down* (see Rousseeuw and Leroy (1987, p. 9)). The higher the BP, the higher the global robustness of the estimator. The BP of any estimator is between 0% and 50%. There are several definitions of the BP. In this paper, we will consider the asymptotic version. See Hampel *et al.* (1986, p. 96).

S-estimators of regression have been introduced by Rousseeuw and Yohai (1984). Their name comes from the fact that they are based on estimators of *scale*. S-estimators are equivariant estimators of regression and can reach a BP as high as 50%, meaning that they can handle nearly half of bad observations, giving a good adjustment of the remaining good ones.

Consider the usual regression model $\mathbf{y} = \mathbf{T}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ with constant variance and uncorrelated errors; i.e. model (1) with $q = 1$. The goal is to estimate the regression coefficients with the data $(\mathbf{t}_i, y_i)$, $i = 1, \ldots, n$. Given an arbitrary $\boldsymbol{\beta}$, we can calculate the residuals, $r_i = r_i(\boldsymbol{\beta}) = y_i - \mathbf{t}_i^T\boldsymbol{\beta}$. Then calculate an estimator of the scale s based on the following equation

$$\frac{1}{n}\sum_{i=1}^{n} \rho(r_i/s) = k,$$

where the function ρ is symmetric, twice continuously differentiable and $\rho(0) = 0$. We also suppose that there exists $c > 0$ such that ρ is strictly increasing on $[0, c]$, and constant on $[c, \infty)$. The constant k taken to be $k = E_{F_0}\rho(r)$ assures the consistency of s at the target model F_0. Thus, consistency at the normal distribution is obtained with the $N(0, 1)$ distribution as F_0. An example of ρ is the biweight function (see Example 2.2 of Lopuhaä (1989))

$$\rho(x) = \begin{cases} \frac{x^2}{2} - \frac{x^4}{2c^2} + \frac{x^6}{6c^4}, & |x| \le c, \\ \frac{c^2}{6}, & |x| \ge c. \end{cases}$$

The S-estimator $\hat{\boldsymbol{\beta}}$ is then defined as the solution of the minimization problem

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} s(r_1(\boldsymbol{\beta}), \ldots, r_n(\boldsymbol{\beta})),$$

and the scale estimator is

$$\hat{\sigma} = s(r_1(\hat{\boldsymbol{\beta}}), \ldots, r_n(\hat{\boldsymbol{\beta}})).$$

S-estimators are affine equivariant, possess a BP of $\lambda = k/\rho(c)$, where $0 < \lambda \leq 0.5$ and are asymptotically normal. The choice of λ affects the efficiency of the estimator under a Gaussian model. The higher the BP, the lower the efficiency and vice versa. For example, if we choose a BP of 50%, the efficiency is about 29%, but if we adopt a BP of 10%, the efficiency is then 97% (see Rousseeuw and Yohai (1984)).

S-estimators have been generalized to multivariate estimation of position and dispersion by Davies (1987) and Lopuhaä (1989). The aim is then to estimate the multivariate location and the scatter matrix of a p-dimensional multivariate population. It is usually assumed that $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are i.i.d. from an elliptical distribution, $E_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, with location $\boldsymbol{\mu}$ and scatter matrix $\boldsymbol{\Sigma}$. S-estimators are solutions of

$$\min_{(\boldsymbol{\mu}, \boldsymbol{\Sigma})} |\boldsymbol{\Sigma}|, \text{ subject to } \frac{1}{n} \sum_{i=1}^{n} \rho[\{(\mathbf{x}_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x}_i - \boldsymbol{\mu})\}^{1/2}] = b.$$

The constant b is taken to be $b = E_{F_0}\rho(|\mathbf{r}|)$, where $\mathbf{r} \sim F_0$ follows the $E_p(\mathbf{0}, \mathbf{I})$ distribution.

With S-estimators of regression and S-estimators of position and dispersion applied to the design matrix of a regression model we can produce a diagnostic plot of the adjustment as robust alternative to the classical residual plot. That idea has been introduced by Rousseeuw and van Zomeren (1990) for least median squares (LMS) and minimum volume ellipsoid (MVE) estimators. We will briefly discuss it now. For a given regression model, we can consider the design matrix as a multivariate sample where each line of the design matrix is taken as an observation. If no outlier is present in the design matrix, no observation should be too far from the others, and the robust Mahalanobis distances (MD) should be small (see Rousseeuw and van Zomeren (1990) for the definition of the robust MD). MD permits us to investigate leverage points, and is an alternative to the use of the diagonal of the hat matrix, often called leverages (see Rousseeuw and Leroy (1987) for a discussion of the leverages as a tool to detect leverage points). Furthermore, if the regression model is adequate and there is no outlier, no observation should possess a very large residual. The graphical display plots robust MD versus robust residuals. A cutoff value for MD is often taken as $(\chi^2_{0.975,p-1})^{1/2}$, where p is the number of independant variables including the intercept, and +2.5 and -2.5 for the residuals. The graph gives a way to spot in a single display good observations (small residuals and small MD's), bad leverage points (large residuals and large MD's), good leverage points (small residuals and large MD's) and finally vertical outliers (large residuals and small MD's). We will adapt that graphical display in the case of SUR models in Section 4.

S-estimators are difficult to calculate. Algorithms usually use resampling methods. These algorithms can evaluate a lot of robust estimators, such as LMS or least trimmed squares (LTS) estimators (see Rousseeuw and Leroy (1987)) but S-estimators take longer to calculate. However, Ruppert (1992) gives an improved resampling algorithm. With this algorithm, S-estimators are easier to calculate, with an accuracy comparable to usual resampling algorithms used to calculate LMS or LTS estimators. With Ruppert's algorithm, S-estimators of regression and S-estimators of position and dispersion can be evaluated. We will discuss in the next section how to adapt that algorithm to the context of SUR models.

3. ROBUST ESTIMATION OF THE SUR MODEL

The S-estimator of the SUR model is solution of the optimization problem

$$\min_{(\boldsymbol{\beta},\boldsymbol{\Sigma})} |\boldsymbol{\Sigma}|, \text{ subject to } \frac{1}{n}\sum_{i=1}^{n} \rho[(\mathbf{e}_i^T \boldsymbol{\Sigma}^{-1}\mathbf{e}_i)^{1/2}] = b. \tag{2}$$

Here the constant b is given by $b = E_{F_0}\rho(|\mathbf{r}|)$, where $\mathbf{r} \sim F_0$ follows the $E_q(\mathbf{0},\mathbf{I})$ elliptical distribution. The breakdown point is then $\lambda = b/\rho(c)$. This formulation is between S-estimators of regression and multivariate S-estimators, since we have to minimize a multivariate measure of scale in the presence of q regression models. Following Lopuhaä (1989), the S-estimator must satisfy the estimating equations (see the Appendix)

$$\boldsymbol{\beta} = [\mathbf{X}^T(\boldsymbol{\Sigma}^{-1}\otimes\mathbf{D}_u)\mathbf{X}]^{-1}\mathbf{X}^T(\boldsymbol{\Sigma}^{-1}\otimes\mathbf{D}_u)\mathbf{y} \tag{3}$$

$$\boldsymbol{\Sigma} = q(\mathbf{Y}-\tilde{\mathbf{X}}\mathbf{B})^T\mathbf{D}_u(\mathbf{Y}-\tilde{\mathbf{X}}\mathbf{B})/\sum_{i=1}^{n} v(d_i), \tag{4}$$

where $d_i^2 = \mathbf{e}_i^T\boldsymbol{\Sigma}^{-1}\mathbf{e}_i$, $\mathbf{D}_u = \operatorname{diag}(u(d_i))$, with $u(d) = \rho'(d)/d$, $v(d) = \rho'(d)d - \rho(d) + b$. When $\rho(d) = d^2$ and $b = q$, the S-estimating equations (3), (4) reduce to the normal MLE estimating equations (Srivastava and Giles (1987, p. 155)). Unlike the MLE and Koenker and Portnoy's estimator (1990) the S-estimator gives weight to the n observations in the multivariate regression $\mathbf{Y} = \tilde{\mathbf{X}}\mathbf{B}+\mathbf{E}$ according to multivariate residuals $\mathbf{e}_i$. It is thus suited to detect not only univariate (in each of the q models) outliers but multivariate outliers as well.

It is important to remark that, contrary to LMS and MVE which have a $n^{1/3}$ rate of convergence to a non-normal asymptotic distribution, the S-estimator enjoys a $n^{1/2}$ rate and is asymptotically normal. Simple bootstrap can thus be applied to the S-estimator in order to get estimated standard deviations of the estimates. When $\tilde{\mathbf{X}}$ is considered fixed, standard deviations are estimated by resampling the multivariate centered residuals, $\hat{\mathbf{e}}_i - \bar{\hat{\mathbf{e}}}$, to obtain $\mathbf{E}^*$ and constructing pseudo-observations $\mathbf{Y}^* = \tilde{\mathbf{X}}\hat{\mathbf{B}} + \mathbf{E}^*$. Possible complications of the simple bootstrap for LMS are reported in Simonoff (1994).

The S estimate $(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\Sigma}})$ is evaluated with a modified Ruppert's algorithm. Let $p = \max(p_1,\ldots,p_q)$. For the local improvement step, let

$$\boldsymbol{\Delta}(\tilde{\boldsymbol{\beta}},\tilde{\boldsymbol{\Sigma}}) = (\boldsymbol{\Delta}_1(\tilde{\boldsymbol{\beta}},\tilde{\boldsymbol{\Sigma}}),\boldsymbol{\Delta}_2(\tilde{\boldsymbol{\beta}},\tilde{\boldsymbol{\Sigma}})).$$

The functions $\boldsymbol{\Delta}_1$ and $\boldsymbol{\Delta}_2$ are the values of $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ after one iteration of the estimating equations (3) and (4), respectively. The algorithm goes as follows.

0. Let $\tilde{s} = \infty$. Choose a random subsample of p integers from the first n integers. Denote this set J. Calculate ordinary least squares on the corresponding rows of $\mathbf{Y}$ and $\tilde{\mathbf{X}}$, giving $\tilde{\boldsymbol{\beta}}^{(i)}$, $i = 1,\ldots,q$, and $\tilde{\boldsymbol{\Sigma}} = (\mathbf{Y}-\tilde{\mathbf{X}}\tilde{\mathbf{B}})^T(\mathbf{Y}-\tilde{\mathbf{X}}\tilde{\mathbf{B}})/n$.

   Repeat steps 1) to 4) $Nsamp$ times.

1. Choose a random subsample of p integers from the first n integers. Denote this set J. Calculate ordinary least squares on the corresponding rows of $\mathbf{Y}$ and $\tilde{\mathbf{X}}$, giving $\boldsymbol{\beta}_{J,0}^{(i)}$, $i = 1,\ldots,q$, and $\boldsymbol{\Sigma}_{J,0} = (\mathbf{Y}-\tilde{\mathbf{X}}\mathbf{B}_{J,0})^T(\mathbf{Y}-\tilde{\mathbf{X}}\mathbf{B}_{J,0})/n$.

2. Let $\begin{pmatrix} \boldsymbol{\beta}_{J,j}^{(1)} \\ \vdots \\ \boldsymbol{\beta}_{J,j}^{(q)} \end{pmatrix} = \boldsymbol{\beta}_{J,j}$, $j = 1, \ldots, n_r$ and $\boldsymbol{\Sigma}_{J,j}$ be points on the line segment connecting $(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\Sigma}})$ and $(\boldsymbol{\beta}_{J,0}, \boldsymbol{\Sigma}_{J,0})$.

3. For $j = 0, \ldots, n_r$, $\mathbf{C}_{J,j} \leftarrow |\boldsymbol{\Sigma}_{J,j}|^{-1/q} \boldsymbol{\Sigma}_{J,j}$.

4. For $j = 0, \ldots, n_r$, if $\frac{1}{n}\sum_{i=1}^{n} \rho[(\mathbf{e}_i^T \mathbf{C}_{J,j}^{-1} \mathbf{e}_i)^{1/2}/\tilde{s}] < b$ (Ruppert's condition) then

   (a) $\tilde{\boldsymbol{\beta}} \leftarrow \boldsymbol{\beta}_{J,j}$

   (b) $\tilde{s} \leftarrow s(\tilde{\boldsymbol{\beta}}, \mathbf{C}_{J,j})$, where $s(\tilde{\boldsymbol{\beta}}, \mathbf{C}_{J,j})$ solves

   $$\frac{1}{n}\sum_{i=1}^{n} \rho[(\mathbf{e}_i^T \mathbf{C}_{J,j}^{-1} \mathbf{e}_i)^{1/2}/s(\tilde{\boldsymbol{\beta}}, \mathbf{C}_{J,j})] = b.$$

   (c) $\tilde{\boldsymbol{\Sigma}} \leftarrow \tilde{s}^2 \mathbf{C}_{J,j}$

   (d) (Local improvement) Find the smallest integer $0 \leq m(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\Sigma}}) \leq 10$ such that

      (i) $(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\Sigma}}) \leftarrow (\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\Sigma}})(1 - 2^{-m}) + \boldsymbol{\Delta}(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\Sigma}})2^{-m}$

      (ii) $\mathbf{C} \leftarrow |\tilde{\boldsymbol{\Sigma}}|^{-1/q} \tilde{\boldsymbol{\Sigma}}$

      (iii) $\frac{1}{n}\sum_{i=1}^{n} \rho[(\mathbf{e}_i^T \mathbf{C}^{-1} \mathbf{e}_i)^{1/2}/\tilde{s}] < b$

      If we can find such an m, then $\tilde{s} \leftarrow s(\tilde{\boldsymbol{\beta}}, \mathbf{C})$, $\tilde{\boldsymbol{\Sigma}} \leftarrow \tilde{s}^2 \mathbf{C}$.

The analysis of Grunfeld's data in Section 4 used values of $n_r = 3$ and $Nsamp = 200$ which is more than Ruppert's recommended values (Ruppert (1992), p. 269). The values of b and the truncation point c were chosen with F_0 being a $N(0,1)$ distribution to achieve a 40% breakdown point. This algorithm follows the same steps as Ruppert's algorithm. Some details about this algorithm are in order. If the corresponding rows in steps 0 and 1 do not have full rank, another choice of the set J is made. In step 1, ordinary least squares is used rather than the MLE. At this point a more accurate and computationally intensive estimate is not desirable since the goal is simply to obtain several plausible directions to look at in order to determine the optimal one. The points in step 2 are chosen by dividing the line segment into n_r intervals of equal length and $(\boldsymbol{\beta}_{J,j}, \boldsymbol{\Sigma}_{J,j})$ is uniformly distributed on the jth subinterval. The sequence of $\tilde{\boldsymbol{\Sigma}}$ generated by the algorithm is the sequence of estimates of $\boldsymbol{\Sigma}$. Its determinant is $\tilde{s}^{2q}$ which forms a decreasing sequence since a new estimate $\tilde{\boldsymbol{\Sigma}}$ is calculated *only* when Ruppert's condition is satisfied. The S-PLUS function `s.sur()` for the evaluation of the S-estimator for SUR models is available from the authors.

4. GRUNFELD'S DATA: U.S. CORPORATIONS EXAMPLE

A classical example to illustrate SUR models is the Grunfeld's investment theory applied to U.S. corporations, as described in Grunfeld (1958). The data set consists of annual gross investment of 10 large U.S. corporations from 1935-1954. The dependent variable, annual gross investment, is explained by two independant variables, value of its outstanding shares at the beginning of the year and beginning-of-year real capital stock.

Table 1: Classical analysis using MLE (MLE), Koenker and Portnoy's (1990) analysis using M-estimators (M), and robust analysis using S-estimators (S). Variables are (1): shares at the beginning of the year, and (2): beginning-of-year real capital stock.

	GE			WH		
	Intercept	(1)	(2)	Intercept	(1)	(2)
MLE	-30.749	0.041	0.136	-1.702	0.059	0.056
	(27.346)	(0.013)	(0.024)	(6.928)	(0.013)	(0.049)
M	-11.4	0.026	0.151	5.1	0.039	0.109
	(18.6)	(0.009)	(0.016)	(5.4)	(0.010)	(0.038)
S	-19.323	0.029	0.146	6.008	0.039	0.079
	(33.448)	(0.016)	(0.030)	(8.286)	(0.016)	(0.058)
	[35.248]	[0.017]	[0.029]	[9.754]	[0.019]	[0.065]

A first way to obtain estimates of regression coefficients is to adjust by OLS each equation separately. However, if we treat economy as a whole, one would expect that activities of a corporation in a given year affect the other, if the correlation between the errors of the two regression equations is nonzero. It is a situation where SUR models seem appropriate. That example has been studied many times in the past. See for example Boot and de Wit (1960) and Zellner (1971) using a Bayesian approach. We tried to analyse the 10 corporations simultanously, but the iterative procedure used to calculate the MLE described in Srivastava and Giles (1987, p. 155) converged to a singular variance matrix. The same problem occured with the robust procedure of Section 3. A nonrobust ridge-type adjustment parameter for singularity in SUR models was proposed by Takada *et al.* (1995). We will restrict attention to two U.S. corporations, General Electric and Westinghouse. This particular situation has been analysed by Zellner (1962), Kmenta and Gilbert (1968) and Theil (1971). This example is interesting since these two companies operate in the same branch of the industry and are important competitors. Since a single outlier can badly affect OLS, it seems important to adjust the data set with a more robust procedure. A robust analysis using M-estimation is given in Koenker and Portnoy (1990). We will analyse the data set with the robust procedure of Section 3.

Table 1 gives the regression coefficients for analyses based on the MLE and on the robust procedure described in Section 3. Our MLE analysis agrees with Kmenta and Gilbert (1968). The numbers in parenthesis are the asymptotic standard errors, whereas those between brackets are standard errors based on 2000 bootstrap samples, as described in the preceding section. The estimated variance matrix based on the MLE and on the robust analysis are given, respectively, by

$$\hat{\mathbf{\Sigma}}_{\mathrm{MLE}} = \left(\begin{array}{c|c} \begin{array}{c}702.23\\(222.07)\end{array} & \begin{array}{c}195.35\\(71.43)\end{array} \\ \hline & \begin{array}{c}90.95\\(28.76)\end{array}\end{array}\right), \hat{\mathbf{\Sigma}}_{\mathrm{S}} = \left(\begin{array}{c|c} \begin{array}{c}871.16\\(331.96)\\ [548.78]\end{array} & \begin{array}{c}259.88\\(109.51)\\ [179.22]\end{array} \\ \hline & \begin{array}{c}106.44\\(40.56)\\ [77.86]\end{array}\end{array}\right),$$

giving an estimated correlation of 0.85 with the robust analysis (with an asymptotic 95% confidence interval based on Fisher-z transform given by [0.53,0.90]), which is higher than the estimated correlation of 0.77 obtained with the MLE analysis (with an asymptotic 95% confidence interval based on Fisher-z transform given by [0.60,0.95]). Bootstrapped standard errors seem to agree reasonably well with asymptotic expansions for the estimation of $\boldsymbol{\beta}$. However, bootstrapping functions of a covariance matrix for n as little as $n = 20$ may not be a good idea (Nagao and Srivastava (1992)).

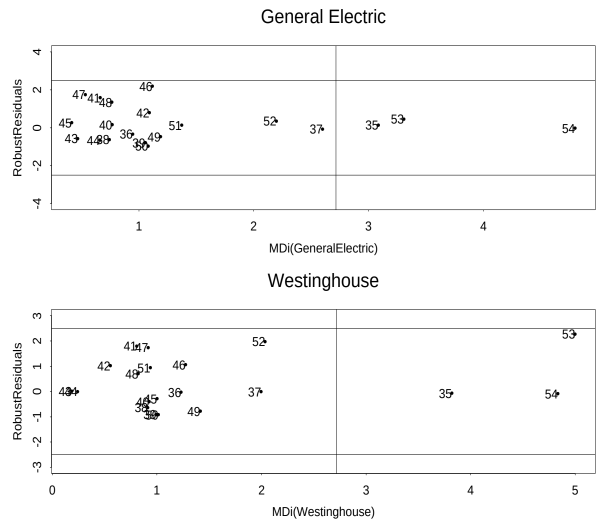

Figure 1: Univariate Diagnostic plots

We also computed S-estimators of the design matrix of each company. We produced robust plots in Figure 1, where standardized robust residuals are plotted against Mahalanobis distances based on observations in the factor space. The plots provide a way to detect univariate outliers for each equation. However since the problem is multivariate in nature, we must also detect possible multivariate outliers in the model. For example, we could be in a situation where ρ_{12} is near one and for observation k $(e_{k,1}, e_{k,2})$ is such that $e_{k,1}$ is positive and $e_{k,2}$ negative. This pair could be outlying without either $e_{k,1}$ or $e_{k,2}$ being univariate outliers. To detect that type of outlier, Figure 2 is a plot of Mahalanobis distance of the bivariate residuals of the robust fit versus years. We did the analyses with a BP ranging from 30%-50%. Choosing the maximal BP is not always desirable in practice, since S-estimators tend to be "attracted" on subsets of the data with unusually small dispersion (Ruppert (1992), p. 269). It is good practice to see the effect of varying the BP on the analysis. Here we produce the results with BP=40%. Relative asymptotic efficiencies of $\hat{\boldsymbol{\beta}}_S$ and $\hat{\boldsymbol{\Sigma}}_S$ with respect to $\hat{\boldsymbol{\beta}}_{MLE}$ and $\hat{\boldsymbol{\Sigma}}_{MLE}$ are, respectively, from Table 2, $\lambda^{-1} \approx 74\%$ and $\sigma_1^{-1} \approx 58\%$. Standard deviations of $\hat{\boldsymbol{\beta}}_S$ and $\hat{\boldsymbol{\beta}}_{MLE}$ in Table 1 are in agreement with a relative asymptotic efficiency of $\lambda^{-1} \approx 74\%$. It is an odd fact, however, that Koenker and Portnoy's standard deviations are smaller than those of $\hat{\boldsymbol{\beta}}_{MLE}$.

Figure 1 seems to indicate that univariate outliers are present in the data set. It shows some leverage points. Note that the outlier years are towards the end of the observation period. Since those outliers represent good leverage points, they seem to agree reasonably well with the model. For example, years 1952-1954, which have good leverages, could possibly be related to the postwar booming economy.

In 1952, General Electric became the largest manufacturer of jet aircraft engines in the U.S. (*Encyclopaedia Britannica*, 1957). Note however that for Westinghouse 1952 is close to being a vertical outlier and 1953 a bad leverage point. This may also be related to the postwar expansion. According to *Business Week* (Oct. 22, 1955, p. 48), Westinghouse was at that time in a second wave of expansion, with an important budget and the construction of several new plants.

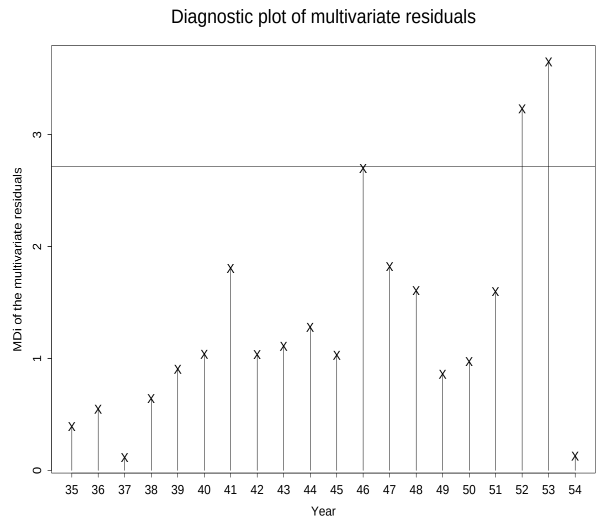

Figure 2: Multivariate diagnostic plot.

Looking at Figure 2, besides 1953, we note that 1946 and 1952 seem to be outliers. These facts were partly hidden in the univariate plots. It is interesting to observe that 1946 is just after World War II, and is a kind of transitional period. An inspection of the data set reveals an important value of gross investment in 1946 for General Electric, whereas the growth was not as high for Westhinghouse. This phenomenon was also perceptible to a lower degree in 1941. Note that year 1946 is almost a vertical outlier in the univariate analysis.

In conclusion, a robust analysis fits a model by giving less weight to some influential observations. If the analysis based on OLS and the robust analysis agreed, it gives use more confidence on the results. For that example, we did not see serious doubt of the model, since good leverages usually give strength to the adjustment. These points had also historical explanations, and it was interesting to see that Grunfeld's theory was not too much affected by that particular historical context. Incidentally, nonparametric additive models were adjusted to each equation by the Fourier smoother (Bilodeau (1992)) which support Grunfeld's linear model.

5. CONCLUDING REMARKS

This paper considered a robust approach to SUR models capable of detecting multivariate outliers. The idea was to adapt the class of high breakdown point S-estimators which has a version for regression as well as for multivariate location and scatter. They are $\sqrt{n}$-consistent and asymptotically normal thus allowing the use of bootstrap methods to evaluate standard errors of S-estimates. However, S-estimators exhibit a trade-off between efficiency and high breakdown point. A referee suggested the use of affine equivariant τ-estimators also defined in both contexts which have high breakdown point and high efficiency. The approach introduced here for SUR models could be adapted to τ-estimators developped by Yohai and Zamar (1988) for regression and later generalized by Lopuhaä (1991a) to the multivariate mean and scatter problem. Our preference was for S-estimators because Ruppert's algorithm allows a fast computational algorithm which does not have a counterpart for τ-estimators. Lopuhaä (1991b) gave a computational algorithm for τ-estimators similar to the one for MVE (Rousseeuw and Leroy (1987), p. 259). However, Ruppert's algorithm evaluates S-estimators in less time than LMS or MVE with the same accuracy.

APPENDIX

We will derive properties of S-estimator in the context of SUR models. Since the S-estimator is defined in terms of the squared Mahalanobis distances, affine equivariance, in the SUR context, immediately follows

$$\begin{aligned}\hat{\boldsymbol{\beta}}_i(a_k\mathbf{y}_k+\mathbf{X}_k\mathbf{g}_k,\mathbf{X}_k;k=1,\ldots,q) &= a_i\hat{\boldsymbol{\beta}}_i(\mathbf{y}_k,\mathbf{X}_k;k=1,\ldots,q)+\mathbf{g}_i,\\ \hat{\sigma}_{ij}(a_k\mathbf{y}_k+\mathbf{X}_k\mathbf{g}_k,\mathbf{X}_k;k=1,\ldots,q) &= a_ia_j\hat{\sigma}_{ij}(\mathbf{y}_k,\mathbf{X}_k;k=1,\ldots,q),\end{aligned}$$

where $a_i>0$ and $\mathbf{g}_i$ is any p_i vector. Kariya (1981) considered this group of transformations, when $\mathbf{X}_i$ is fixed, to derive a locally best invariant test of independence between two SUR models. Let $\mathbf{A}_i=\operatorname{diag}(\mathbf{a}_i,\ldots,\mathbf{a}_i)$ be a $(nq)\times q$ matrix where $\mathbf{a}_i$ is the vector with one in position i and zero elsewhere. Let $\mathbf{e}_i\equiv\mathbf{e}_i(\boldsymbol{\beta})=\mathbf{A}_i^T(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})$, $i=1,\ldots,n$. Let L be the Lagrangian

$$L=\log(|\boldsymbol{\Sigma}|)-\lambda\left\{\frac{1}{n}\sum_{i=1}^{n}\rho[(\mathbf{e}_i^T\boldsymbol{\Sigma}^{-1}\mathbf{e}_i)^{1/2}]-b\right\}. \tag{5}$$

The solution of the problem (2) is obtained by equating partial derivatives $\partial L/\partial\boldsymbol{\beta}$, $\partial L/\partial\boldsymbol{\Sigma}$ and $\partial L/\partial\lambda$ to $\mathbf{0}$. When $\partial L/\partial\boldsymbol{\beta}=\mathbf{0}$, we have that $\boldsymbol{\beta}$ must satisfy

$$\sum_{i=1}^{n}u(d_i)\mathbf{X}^T\mathbf{A}_i\boldsymbol{\Sigma}^{-1}\mathbf{A}_i^T\mathbf{y}=\sum_{i=1}^{n}u(d_i)\mathbf{X}^T\mathbf{A}_i\boldsymbol{\Sigma}^{-1}\mathbf{A}_i^T\mathbf{X}\boldsymbol{\beta}.$$

Noting that $\mathbf{A}_i\boldsymbol{\Sigma}^{-1}\mathbf{A}_i^T=\boldsymbol{\Sigma}^{-1}\otimes\mathbf{a}_i\mathbf{a}_i^T$ and $\sum_{i=1}^n u(d_i)\mathbf{A}_i\boldsymbol{\Sigma}^{-1}\mathbf{A}_i^T=\boldsymbol{\Sigma}^{-1}\otimes\mathbf{D}_u$ we obtain (3). Following exactly Lopuhaä (1989), we obtain also (4). Another form of the estimating equations (3) and (4) is $\frac{1}{n}\sum_{i=1}^n\boldsymbol{\Psi}(\mathbf{A}_i^T\mathbf{X},\mathbf{A}_i^T\mathbf{y},\boldsymbol{\beta},\boldsymbol{\Sigma})=\mathbf{0}$, $\boldsymbol{\Psi}=(\boldsymbol{\Psi}_1^T,\boldsymbol{\Psi}_2^T)^T$ where

$$\begin{aligned}\boldsymbol{\Psi}_1(\mathbf{Z},\mathbf{t},\boldsymbol{\beta},\boldsymbol{\Sigma}) &= u(d)\mathbf{Z}^T\boldsymbol{\Sigma}^{-1}(\mathbf{t}-\mathbf{Z}\boldsymbol{\beta}),\\ \boldsymbol{\Psi}_2(\mathbf{Z},\mathbf{t},\boldsymbol{\beta},\boldsymbol{\Sigma}) &= \operatorname{vec}[q\ u(d)(\mathbf{t}-\mathbf{Z}\boldsymbol{\beta})(\mathbf{t}-\mathbf{Z}\boldsymbol{\beta})^T-v(d)\boldsymbol{\Sigma}],\end{aligned}$$

and $d^2=(\mathbf{t}-\mathbf{Z}\boldsymbol{\beta})^T\boldsymbol{\Sigma}^{-1}(\mathbf{t}-\mathbf{Z}\boldsymbol{\beta})$ as usual. Assuming $(\mathbf{A}_i^T\mathbf{X},\mathbf{A}_i^T\mathbf{y})$, $i=1,\ldots,n$ are i.i.d. we can conclude that the S-estimator of the SUR model satisfies first-order conditions of M-estimators as defined in Huber (1981). They are thus asymptotically normal with convergence rate $n^{1/2}$. The asymptotic variances are now derived. Assuming $\mathbf{e}_i$ has a symmetric distribution independent of $\mathbf{Z}_i=\mathbf{A}_i^T\mathbf{X}$

$$\operatorname{cov}(\boldsymbol{\Psi}_1,\boldsymbol{\Psi}_2)=E\left[q\ u^2(d)\mathbf{Z}^T\boldsymbol{\Sigma}^{-1}\mathbf{e}\ \operatorname{vec}(\mathbf{e}\mathbf{e}^T)\right]-E\left[u(d)v(d)\mathbf{Z}^T\boldsymbol{\Sigma}^{-1}\mathbf{e}\right]\operatorname{vec}(\boldsymbol{\Sigma})=\mathbf{0}.$$

Hence, the central limit theorem gives

$$n^{1/2}\bar{\boldsymbol{\Psi}}(\mathbf{A}_i^T\mathbf{X},\mathbf{A}_i^T\mathbf{y},\boldsymbol{\beta},\boldsymbol{\Sigma})\xrightarrow{d}N\left(\mathbf{0},\operatorname{diag}\left(\operatorname{var}(\boldsymbol{\Psi}_1),\operatorname{var}(\boldsymbol{\Psi}_2)\right)\right),$$

where the asymptotic variance is block diagonal. Now, let

$$\lambda(\boldsymbol{\beta}, \boldsymbol{\Sigma}) = E\ \boldsymbol{\Psi}(\mathbf{Z}, \mathbf{t}, \boldsymbol{\beta}, \boldsymbol{\Sigma})$$

and assume $\lambda(\cdot,\cdot)$ has a nonsingular derivative at the true parameter $\boldsymbol{\beta}_0$, $\boldsymbol{\Sigma}_0$. Then a Taylor expansion gives

$$n^{1/2}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0, \mathrm{vec}(\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}_0)\right) \xrightarrow{d} N\left(\mathbf{0}, (E\ \boldsymbol{\Psi}')^{-1}\mathrm{var}\ \boldsymbol{\Psi}(E\ \boldsymbol{\Psi}')^{-1^T}\right).$$

One easily verifies that $E\partial\boldsymbol{\Psi}_1/\partial\sigma_{kl} = \mathbf{0}$ and $E\partial\boldsymbol{\Psi}_2/\partial\boldsymbol{\beta} = \mathbf{0}$ which implies $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\Sigma}}$ are asymptotically independent. The asymptotic variance of $\hat{\boldsymbol{\beta}}$ is now evaluated assuming further that $\mathbf{e}$ is elliptical. Since $u(d)\boldsymbol{\Sigma}^{-1/2}\mathbf{e}$ is spherical then $\mathrm{var}(u(d)\boldsymbol{\Sigma}^{-1/2}\mathbf{e}) = \alpha\mathbf{I}$ for some α. Taking traces we get

$$\alpha = \frac{1}{q}E\left(\psi^2(|\mathbf{e}_0|)\right),$$

where $\mathbf{e}_0 = \boldsymbol{\Sigma}^{-1/2}\mathbf{e}$ is spherical and $\psi(\cdot) = \rho'(\cdot)$. Hence, $\mathrm{var}(\boldsymbol{\Psi}_1) = \alpha E(\mathbf{Z}^T\boldsymbol{\Sigma}^{-1}\mathbf{Z})$. Similarly, $E\ \partial\boldsymbol{\Psi}_1/\partial\boldsymbol{\beta} = -\beta E(\mathbf{Z}^T\boldsymbol{\Sigma}^{-1}\mathbf{Z})$, where the expression for β is

$$\beta = E\left[\left(1 - \frac{1}{q}\right)u(|\mathbf{e}_0|) + \psi'(|\mathbf{e}_0|)\right].$$

Hence, altogether we get

$$n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{d} N\left(\mathbf{0}, (\alpha/\beta^2)[E(\mathbf{Z}^T\boldsymbol{\Sigma}^{-1}\mathbf{Z})]^{-1}\right).$$

A consistent estimator of $E\mathbf{Z}^T\boldsymbol{\Sigma}^{-1}\mathbf{Z}$ is given by $n^{-1}\mathbf{X}^T(\hat{\boldsymbol{\Sigma}}^{-1}\otimes\mathbf{I})\mathbf{X}$. The equation related to $\boldsymbol{\Psi}_2$ being the same as Lopuhaä (1989), we get immediately

$$n^{1/2}\mathrm{vec}(\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}_0) \xrightarrow{d} N\left(\mathbf{0}, \sigma_1(\mathbf{I} + \mathbf{K}_q)(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma}) + \sigma_2\mathrm{vec}(\boldsymbol{\Sigma})[\mathrm{vec}(\boldsymbol{\Sigma})]^T\right),$$

where

$$\begin{aligned}\sigma_1 &= \frac{q(q+2)E\ \psi^2(|\mathbf{e}_0|)|\mathbf{e}_0|^2}{\left\{E\left[\psi'(|\mathbf{e}_0|)|\mathbf{e}_0|^2 + (q+1)\psi(|\mathbf{e}_0|)|\mathbf{e}_0|\right]\right\}^2},\\ \sigma_2 &= -\frac{2}{q}\sigma_1 + \frac{4E\left(\rho(|\mathbf{e}_0|) - b\right)^2}{\left\{E\ \psi(|\mathbf{e}_0|)|\mathbf{e}_0|\right\}^2}.\end{aligned}$$

Consequently, asymptotic distributions were established assuming elliptical errors without dwelling on regularity conditions for the existence and unicity of S-estimator. Asymptotic values $\lambda = \alpha/\beta^2$, σ_1 and σ_2, at the normal distribution, for q=2,3,...,10, and r=0.1,0.2,0.3,0.4,0.5, which extend the table of Lopuhaä (1989) are reported in Table 2. These values can then be used to provide standard errors of the estimates,

$$\begin{aligned}\mathrm{var}\ n^{1/2}\hat{\sigma}_{ii} &\to (2\sigma_1 + \sigma_2)\sigma_{ii}^2,\\ \mathrm{var}\ n^{1/2}\hat{\sigma}_{ij} &\to \sigma_1(\sigma_{ii}\sigma_{jj} + \sigma_{ij}^2) + \sigma_2\sigma_{ij}^2,\ i \neq j.\end{aligned}$$

Standard error for the correlation coefficient (or its Fisher-z transform) between the two equations will depend only on σ_1 (Tyler (1983))

$$\mathrm{var}\ n^{1/2}\mathrm{arctanh}(\hat{\rho}_{12}) \to \sigma_1.$$

Table 2: Asymptotic values of λ, σ_1 and σ_2 at the normal distribution

		r				
		0.5	0.4	0.3	0.2	0.1
$q=2$	λ	1.725	1.356	1.157	1.055	1.011
	σ_1	2.656	1.735	1.299	1.096	1.018
	σ_2	-1.332	-0.566	-0.222	-0.069	-0.012
$q=3$	λ	1.384	1.188	1.083	1.029	1.006
	σ_1	1.726	1.332	1.137	1.046	1.009
	σ_2	-0.362	-0.160	-0.064	-0.021	-0.004
$q=4$	λ	1.250	1.122	1.054	1.019	1.004
	σ_1	1.424	1.195	1.082	1.028	1.005
	σ_2	-0.151	-0.067	-0.027	-0.009	-0.002
$q=5$	λ	1.182	1.089	1.039	1.014	1.003
	σ_1	1.285	1.132	1.056	1.019	1.004
	σ_2	-0.078	-0.035	-0.015	-0.005	-0.001
$q=6$	λ	1.141	1.069	1.031	1.011	1.002
	σ_1	1.209	1.098	1.042	1.015	1.003
	σ_2	-0.046	-0.021	-0.009	-0.003	-0.001
$q=7$	λ	1.114	1.056	1.025	1.009	1.002
	σ_1	1.162	1.076	1.033	1.012	1.002
	σ_2	-0.030	-0.014	-0.006	-0.002	0.000
$q=8$	λ	1.096	1.047	1.021	1.008	1.002
	σ_1	1.131	1.062	1.027	1.010	1.002
	σ_2	-0.021	-0.010	-0.004	-0.001	0.000
$q=9$	λ	1.082	1.041	1.018	1.007	1.001
	σ_1	1.109	1.052	1.023	1.008	1.002
	σ_2	-0.015	-0.007	-0.003	-0.001	0.000
$q=10$	λ	1.072	1.036	1.016	1.006	1.001
	σ_1	1.093	1.045	1.020	1.007	1.001
	σ_2	-0.011	-0.005	-0.002	-0.001	0.000

The asymptotic variance of the normal MLE is identical to the above expressions for S-estimator but with $\lambda = 1$, $\sigma_1 = 1$ and $\sigma_2 = 0$. Thus λ^{-1} and σ_1^{-1} can be taken as measures of relative efficiencies at the Gaussian model for the estimation of $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$, respectively.

ACKNOWLEDGEMENTS

The authors would like to thank the two referees for helpful comments which helped to improve the presentation.

REFERENCES

Bilodeau, M. (1992). Fourier smoother and additive models. *Canad. J. Statist.*, 20, 257–269.

Boot, J., and de Wit, G. M. (1960). Investment demand: an empirical contribution to the aggregation problem. *International Economic Review*, 1, 3–30.

Davies, P. (1987). Asymptotic behaviour of s-estimates of multivariate location parameters and dispersion matrices. *Ann. Statist.*, 15, 1269–1292.

Grunfeld, Y. (1958). The determinants of corporate investment. Unpublished Ph.D. thesis, University of Chicago, Dept. of Economics.

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

Huber, P. J. (1981), *Robust Statistics.* New-York: Wiley.

Judge, G. G., Hill, R. C., Grittiths, W. E., Lütkepohl, H., and Lee, T.-C. (1985). *The theory and practice of econometrics*, second edn. Wiley: New-York.

Kariya, T. (1981). Tests for the independence between two seemingly unrelated regression equations. *Ann. Statist.*, 9, 381–390.

Kmenta, J., and Gilbert, R. F. (1968). Small sample properties of alternative estimators of seemingly unrelated regressions. *J. Amer. Statist. Assoc.*, 63, 1180–1200.

Koenker, R., and Portnoy, S. (1990). M estimation of multivariate regressions. *J. Amer. Statist. Assoc.*, 85, 1060–1068.

Lopuhaä, H. (1989). On the relation between s-estimators and m-estimators of multivariate location and covariance. *Ann. Statist.*, 17, 1662–1683.

Lopuhaä, H. (1991a). Multivariate τ-estimators for location and scatter. *Canad. J. Statist.*, 19, 307–321.

Lopuhaä, H. (1991b). Breakdown point and asymptotic properties of multivariate S-estimators and τ-estimators: A summary, *in* W. Stahel, and S. Weisberg, eds, *Directions in Robust Statistics and Diagnostics.* Springer-Verlag (Berlin; New-York), 167–182.

Nagao, H., and Srivastava, M. S. (1992). On the distributions of some test criteria for a covariance matrix under local alternatives and bootstrap approximations. *J. Multivariate Anal.*, 43, 331–350.

Rousseeuw, P., and Yohai, V. (1984). Robust regression by means of S-estimators, *in* J. Franke, W. Haerdle, and D. Martin, eds, *Robust and Nonlinear Time Series Analysis.* Springer-Verlag (Berlin; New-York), 256–272.

Rousseeuw, P. J., and Leroy, A. M. (1987). *Robust Regression and Outlier Detection.* New-York: Wiley.

Rousseeuw, P. J., and van Zomeren, B. C. (1990). Unmasking multivariate outliers and leverage points. *J. Amer. Statist. Assoc.*, 85, 633–639.

Ruppert, D. (1992). Computing S estimators for regression and multivariate location/dispersion. *J. of computational and graphical statist.*, 1, 253–270.

Simonoff, J. S. (1994). Review of "An introduction to the bootstrap" by Bradley Efron and Robert J. Tibshirani. *J. Amer. Statist. Assoc.*, 89, 1559–1560.

Srivastava, V., and Giles, D. E. (1987). *Seemingly unrelated regression models: Estimation and inference.* New-York: Dekker.

Takada, H., Ullah, A., and Chen, Y.-M. (1995). Estimation of the seemingly unrelated regression model when the error covariance matrix is singular. *J. Appl. Statist.*, 22, 517–530.

Theil, H. (1971). *Principles of econometrics.* New-York: Wiley.

Tyler, D. E. (1983). Robustness and efficiency properties of scatter matrices. *Biometrika*, 70, 411–420. (Corr: V71 p656).

Yohai, V. J., and Zamar, R. (1988). High breakdown-point of estimates of regression by means of the minimization of an efficient scale. *J. Amer. Statist. Assoc.*, 83, 406–413.

Zellner, A. (1962). An efficient method for estimating seemingly unrelated regressions and tests for aggregation bias. *J. Amer. Statist. Assoc.*, 57, 348–368.

Zellner, A. (1971). *An introduction to Bayesian inference in econometrics.* New-York: Wiley.

Received 16 March 1999
Revised ???
Accepted ???

Département de mathématiques et de statistique
Université de Montréal
C.P. 6128, succursale Centre-ville
Montréal (Québec)
Canada H3C 3J7

e-mail: bilodeau@dms.umontreal.ca

e-mail: duchesne@dms.umontreal.ca