Working Paper 94-37
Statistics and Econometrics Series 14
November 1994

Departamento de Estadística y Econometría
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (341) 624-9849

ROBUST ESTIMATION IN SIMULTANEOUS EQUATIONS MODELS

Ricardo A. Maronna and Víctor J. Yohai*

Abstract

In this paper we review existing work on robust estimation for simultaneous equations models. Then we discuss three strategies for obtaining estimators with a high breakdown point, a controllable efficiency, and a reasonable computational cost: (a) robustifying Three-Stages Least Squares, (b) robustifying the Full Information Maximum Likelihood method by minimizing the determinant of a robust covariance matrix of residuals, and (c) generalizing multivariate tau-estimators (Lopuhaä 1991) to these models. The latter seems the most promising approach.

*Maronna, Universidad Nacional de la Plata, Argentina; Yohai, Universidad de San Andrés, Argentina and Universidad de Buenos Aires. The authors were partially supported by C.I.P.B.A. and C.O.N.I.C.E.T., respectively. Most of this work was done while the first author was a visiting professor at Universidad Carlos III de Madrid, España.

Robust Estimation in Simultaneous Equations Models

Ricardo A. Maronna [1] and Víctor J. Yohai [2]
[1]Universidad Nacional de La Plata, C.C. 172, 1900 La Plata, Argentina
[2]Universidad de San Andrés, V. Dumas 284, 1644 Victoria, Argentina
and Universidad de Buenos Aires *

Abstract

In this paper we review existing work on robust estimation for simultaneous equations models. Then we discuss three strategies for obtaining estimators with a high breakdown point, a controllable efficiency, and a reasonable computational cost: (a) robustifying Three-Stages Least Squares, (b) robustifying the Full Information Maximum Likelihood method by minimizing the determinant of a robust covariance matrix of residuals, and (c) generalizing multivariate tau-estimators (Lopuhaä 1991) to these models. The latter seems the most promising approach.

1 Introduction

Simultaneous equations models are an important tool in Econometrics. They are an extension of the multivariate linear model (MLM). While their correctness in specific situations may be open to criticisms, there is no doubt that they constitute an interesting field of research for statisticians. In particular, research on the effects of outliers on estimation procedures, and on methods robust with respect to outliers, seems to be scanty. In this paper we review the main classical ideas on estimation in this model, and the most relevant approaches to robust

*The authors were partially supported by C.I.P.B.A. and C.O.N.I.C.E.T., respectively. Most of this work was done while the first author was a visiting professor at Universidad Carlos III, Madrid.

estimation therein. We then propose some robust estimators with a high breakdown point, a controllable efficiency, and a feasible computational cost.

Simultaneous equations models have the form

$$\mathbf{\Gamma}\mathbf{y}_i = \mathbf{B}\mathbf{x}_i + \mathbf{e}_i, \quad i = 1, \ldots, n, \tag{1}$$

where $\mathbf{x}_i \in \mathrm{R}^p$ are called the vectors of *exogenous variables*, $\mathbf{y}_i \in \mathrm{R}^q$ are the *endogenous variables*, and the disturbances $\mathbf{e}_i \in \mathrm{R}^q$ are i.i.d random vectors with mean $\mathbf{0}$ and covariance matrix $\mathbf{\Sigma}$. The matrices $\mathbf{B} \in \mathrm{R}^{q \times p}$, $\mathbf{\Gamma} \in \mathrm{R}^{q \times q}$ and $\mathbf{\Sigma} \in \mathrm{R}^{q \times q}$ are the unknown parameters of the system. We can write (1) –the so-called *structural form* of the model– more compactly as

$$\mathbf{Y}\mathbf{\Gamma}' = \mathbf{X}\mathbf{B}' + \mathbf{E}, \tag{2}$$

where $\mathbf{Y}$, $\mathbf{X}$ and $\mathbf{E}$ are the matrices with rows $\mathbf{y}_i'$, $\mathbf{x}_i'$ and $\mathbf{e}_i'$, respectively. The vectors $\mathbf{x}_i$'s are either random and uncorrelated with the $\mathbf{e}_i$'s, or nonrandom and such that $\lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} \mathbf{x}_i \mathbf{e}_i' = 0$ in probability.

Model specification includes restrictions on $\mathbf{B}$ and $\mathbf{\Gamma}$ – which usually specify that some coefficients must be 0 or 1– without which they would not be identifiable. Put for brevity $\mathbf{\Delta} = (\mathbf{B}, \mathbf{\Gamma})$. Then the restrictions will be expressed by stating that $\mathbf{\Delta}$ must belong to a set $\mathcal{D} \subseteq \mathrm{R}^{q \times (p+q)}$. In particular, $\mathbf{B}$ is nonsingular. If we specify $\mathbf{\Gamma} = \mathbf{I}$, we have the MLM.

It may always be assumed that

$$\mathbf{\Gamma}_{ii} = 1. \tag{3}$$

This condition usually arises naturally in econometric models. Let us denote by $x_{.j}$ and $e_{.j}$ the j-th columns of $\mathbf{X}$ and $\mathbf{E}$, by $\mathbf{Y}_j$ the matrix $\mathbf{Y}$ without column j, by β_j the j-th row of $\mathbf{B}$, and by γ_j the j-th row of $\mathbf{\Gamma}$ without element $\mathbf{\Gamma}_{ii}$. Then under (3), (1) can be expressed as a set of q systems, the j-th system being

$$\mathbf{y}_{.j} = -\mathbf{Y}_j \gamma_j + \mathbf{X}_j \beta_j + \mathbf{e}_{.j}. \tag{4}$$

The *reduced form* of the model is

$$\mathbf{y}_i = \mathbf{\Pi}\mathbf{x}_i + \mathbf{u}_i, \tag{5}$$

where $\mathbf{\Pi} = \mathbf{\Gamma}^{-1}\mathbf{B}$ and $\mathbf{u}_i = \mathbf{\Gamma}^{-1}\mathbf{e}_i$.

Now we review the main estimation methods. For further references, see (Amemiya, 1985) or (Judge et al., 1985).

The matrix $\mathbf{\Pi}$ can be consistently estimated from (5) by applying the least squares estimator (LSE) to each coordinate. but one cannot in general get $\mathbf{\Gamma}$ and $\mathbf{B}$ from it, except in certain circumstances ("exact identifiability").

Ordinary least squares (OLS) consists of estimating for each j the parameters in the right-hand side of (4) by least squares. It is not consistent, for the regressors are not uncorrelated with the disturbances.

The method of *two-stage least squares (2SLS)* avoids this pitfall by first regressing the $\mathbf{y}$'s on the $\mathbf{x}$'s (first stage) and then estimating the parameters by applying OLS (with the restrictions) to (4), but with the $\mathbf{y}$'s on the right-hand side replaced by the fitted values $\hat{\mathbf{Y}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ (second stage). This method is consistent. However, it is in general not asymptotically efficient, for the estimation for equation j does not take into account the information contained in the other equations.

The inefficiency of 2SLS comes from treating the q systems as if they were independent. To avoid this drawback, the *three-stage least squares* (3SLS) method proceeds by first estimating $\mathbf{\Sigma}$ from the residuals from 2SLS, and then using it to estimate the parameters by generalized least squares. It is asymptotically efficient.

The maximunm likelihood estimator corresponding to multivariate normal $\mathbf{e}_i$'s is called *full-imformation maximum likelihood* method (FIML). Let $\mathbf{r}_i$ be the reduced form residuals:

$$\mathbf{r}_i(\mathbf{\Delta}) = \mathbf{y}_i - \mathbf{\Gamma}^{-1}\mathbf{B}\mathbf{x}_i; \tag{6}$$

and let

$$\mathbf{C}(\mathbf{\Delta}) = \sum_i \mathbf{r}_i(\mathbf{\Delta})\mathbf{r}_i(\mathbf{\Delta})'. \tag{7}$$

Then FIML is equivalent to

$$\mathbf{\Delta}^* = \arg \min_{\mathbf{\Delta}\in\mathcal{D}} \det(\mathbf{C}(\mathbf{\Delta})), \tag{8}$$

i.e., one tries to make residuals "as small as possible". The minimization is performed iteratively; one can take 2SLS or 3SLS as starting point. It is asymptotically efficient.

Let the vector δ be a parametrisation of $\mathbf{\Delta} \in \mathcal{D}$ without restrictions (in the usual case that some parameters are set to 0 or 1, δ will simply consist of the remaining free parameters); and let δ^* the corresponding estimator. Then, for both 3SLS and FIML, it is proved that $\sqrt{n}(\delta^* - \delta)$ tends in distribution to the multivariate normal with covariance matrix

$$(\mathbf{R}'(\mathbf{\Sigma}^{-1} \otimes \mathbf{Q})\mathbf{R})^{-1}, \tag{9}$$

where "$\otimes$" is the Kronecker product, $\mathbf{Q} = \lim_{n\to\infty} \mathbf{X}'\mathbf{X}/n$, and $\mathbf{R}$ depends only on Π and the restrictions.

It is an interesting fact that, if the $\mathbf{e}_i$'s are normal, the FIML estimator has no finite moments of any order; for 2SLS and 3SLS, existence of moments depends on the restrictions (Mariano, 1982). This shows that the mean square error may be meaningless, and hence that more robust criteria are needed to evaluate the estimators.

2 Available robust methods

All the above estimators are clearly sensitive to atypical observations. Kuh and Welsch (1980) apply regression diagnostics to the detection of outliers. However, work on specifical robust methods for simultaneous equations is scarce. We now review the main published approaches.

It is natural to try to robustify 2SLS by replacing the LS regressions involved by robust regression estimators. Amemiya (1982) treats two types of estimators based essentially on applying L_1 regression to the second stage. He shows that under certain forms of non-normality of the distrubances, both may be much more efficient that 2SLS. It is clear however that these procedures have a low efficiency for normal disturbances, and do not consider outliers in the $\mathbf{x}$'s.

Another approach is to robustify FIML by replacing it by Maximum Likelihood estimation corresponding to a distribution with heavier tails than the normal. Prucha and Kelejian (1984) do this by considering the family of Maximum Likelihood estimators for $\mathbf{e}_i$'s having a multivariate Student distribution with ν degees of freedom, the case $\nu = \infty$ corresponding to FIML; and linearized versions thereof.

These type of estimators are robust against outliers in the $\mathbf{y}$'s, and can be tuned to attain a high efficiency for normal $\mathbf{e}$'s. However, they are not robust against leverage points in the $\mathbf{x}$'s, so that they have null breakdown point if the $\mathbf{x}$'s are unbounded.

In order to take leverage points into account, Krasker and Welsch (1985) propose a robust version of 2SLS, which essentially consists of using a bounded-influence GM estimator for the second stage. Krasker (1986) completes this idea: he proposes another method, in which all LS steps in both stages are replaced by GM estimation. While Krasker's procedure is much more robust than its predecessors, it inherits the drawbacks of regression GM estimators (see Maronna, Bustos and Yohai, 1979): (a) their efficiency for normal disturbances can be chosen only for a *given* distribution of the regressors, and they can have an arbitrarily low efficiency for heavy-tailed regressors, (b) their breakdown point tends to 0 when the number of regressors increases, (c) this becomes still worse when the residual scale is estimated (Maronna and Yohai, 1991), and much worse still due to the necessary estimation of a scaling matrix for the regressors.

Koenker and Portnoy (1990) consider M estimators for the MLM, which consist of applying a regression M estimator to each coordinate of the responses; they apply it to the model of "Seemingly Unrelated Regressions". Again, this procedure has two disadvantages: low efficiency when the responses are correlated, and a null breakdown point.

3 Some robust proposals

We now consider some strategies to attain estimators with high efficiency with respect to FIML under normal disturbances, and high breakdown point for large p and q. It is also a desirable property that the estimators be equivariant in the particular case of the MLM.

3.1 Robustifying 3SLS

A natural strategy is to robustify all steps in 3SLS. For this, first perform a robust version of 2SLS by replacing in both stages all LS regressions by a robust and highly efficient estimator like Yohai and Zamar's (1988) τ-estimators. Then apply a high breakdown point estimator of multivariate scale to the residuals, such as the projection estimators of (Maronna, Stahel and Yohai, 1992) or the Stahel-Donoho estimator (Maronna and Yohai, 1995a), to obtain a robust estimate of $\mathbf{\Sigma}$. Then use this estimate to scale the equations, and finally obtain the coefficients by applying a robust regression to them.

Although this estimator is conceptually simple, its asymptotic distribution seems complicated; and it may not be computationally cheaper than the estimators described below.

3.2 FIML with robust covariance

The second strategy is to robustify FIML by minimizing the determinant of a robust version of (7), i.e.,

$$\mathbf{\Delta}^* = \arg \min_{\mathbf{\Delta} \in \mathcal{D}} \det(\mathbf{C}(\mathbf{\Delta})),, \tag{10}$$

where $\mathbf{C}$ is a robust scale matrix.

The simplest choice would be to use a multivariate M estimator. Recall that M estimators have breakdown point $\delta^* \leq 1/(q+1)$ (Maronna, 1976); and thus one would expect this to be an upper bound for the breakdown point of $\mathbf{\Delta}^*$. Unfortunately, it turns out that, even for moderate q, a reasonable efficiency entails δ^* much lower than the upper bound (Maronna and Yohai, 1991). The reasons can be understood by noting that for $q = 1$ the estimator reduces to a regression S estimator, for which it is known (Hosjer, 1972) that one cannot have at the same time a high efficiency and a high breakdown point.

Another choice is the Stahel-Donoho estimator:

$$\mathbf{C}_{SD} = \sum_i w_i \mathbf{r}_i \mathbf{r}_i', \tag{11}$$

with

$$w_i = W(z_i) \tag{12}$$

where z_i measures the "outlyingness" of $\mathbf{r}_i$:

$$z_i = \sup_{\|\mathbf{a}\|=1} \frac{|\mathbf{a}'\mathbf{r}_i|}{s(\mathbf{a})}, \tag{13}$$

with $s(\mathbf{a}) = \sigma(|\mathbf{a}'\mathbf{r}_1|, \ldots, |\mathbf{a}'\mathbf{r}_n|)$, where σ is a robust scale; and W is a weight function such that $W(z)z^2$ is bounded. (Since the residuals are centered around $\mathbf{0}$, no centering of $\mathbf{a}'\mathbf{r}$ is needed here).

Let σ be an M estimator of scale, defined as solution of

$$\text{ave } \rho_1(z_i/\sigma) = b,$$

where ρ is a nondecreasing bounded function. Then for $q = 1$ –i.e., linear regression– this reduces to a τ-estimator. One can thus expect these estimators to keep for the general case the same properties of high efficiency and high breakdown point of τ-estimators.

Unfortunately, the theoretical treatment of (10) becomes extremely awckward, the main difficulty being differentiating z with respect to the parameters. Computation of the estimators would also be very expensive, since $\mathbf{C}_{SD}$ would have to be computed at each iteration of the minimization process.

A feasible estimator can be obtained by one-step reweighting, as follows.

1. Estimate each row of $\mathbf{\Pi}$ from the reduced form, using a regression estimator with high efficiency and high breakdown point (e.g., τ-estimators). Call $\mathbf{\Pi}^*$ the resulting matrix.

2. Compute residuals: $\mathbf{r}_{0t} = \mathbf{y}_i - \mathbf{\Pi}^*\mathbf{x}_i$.

3. Compute the outlyingness z_i of each $\mathbf{r}_{0t}$ as in (13), and weights w_i as in (12).

4. Compute 3SLS with weights w_i.

5. Let now $C_W(\boldsymbol{\Delta}) = \sum_i w_i \mathbf{r}_i \mathbf{r}_i'$. Minimize $\det(C_{W})$, with 3SLS from step 4 as starting values.

Note that the weights are computed *only once*. This approach requires q p-variate regessions, plus computing the Stahel-Donoho weights; the cost of the final minimisation is only marginal. If one has a program that computes 3SLS and FIML, it suffices to apply it to $w_i^{1/2}\mathbf{x}_i, w_i^{1/2}\mathbf{y}_i$.

It is not difficult to prove that this estimator is Fisher-consistent if the **e**'s have an ellipsoidal distribution. It is complicated, but feasible, to obtain a heuristic derivation of its asymptotic distribution for ellipsoidal **e**'s; as it happens with one-step reweighting, it depends on the initial estimators. The estimator is not equivariant under the MLM.

3.3 Multivariate τ-estimators

A more promising approach is to extend Lopuhaä's (1991) multivariate τ-estimators to this case, defining the estimators as solutions of

$$\det(\mathbf{C})\ (\sum_{i=1}^{n} \rho_2(d_i))^q = \min, \tag{14}$$

subject to

$$\text{ave}\ \rho_1(d_i) = 1, \tag{15}$$

where

$$d_i = \mathbf{r}_i(\boldsymbol{\Delta})'\mathbf{C}^{-1}\mathbf{r}_i(\boldsymbol{\Delta}). \tag{16}$$

Here ρ_1 and ρ_2 are nondecreasing functions with $\rho_i(0) = 0$. For univariate regression, Yohai and Zamar show that under reasonable conditions -which include that both functions be bounded- one can have an arbitrarily high normal efficiency with an arbitrarily high $\delta^* \leq 0.5$. Lopuhaä (1991) applies the same ideas to multivariate location and scatter. In the MLM, if both functions are equal to the identity, we have LS. It is not difficult to prove that for simultaneous equations, $\rho_1(r) = \rho_2(r) = r$ yields FIML.

It can be proved (see Maronna and Yohai, 1995b) that under the model, these estimators have an asymptotic covariance matrix of the form (9), with $\boldsymbol{\Sigma}$ replaced by a matrix $\mathbf{S}$ which has the form of the asymptotic covariance matrix of a multivariate location M-estimator. While the general expression is complicated, for symmetric **e**'s it reduces to

$$\mathbf{S} = \mathbf{D}^{-1}\mathbf{A}(\mathbf{D}^{-1})', \tag{17}$$

with

$$\mathbf{D} = 2\,\mathrm{E}\,u_1'(d)\,\mathbf{ee}'\mathbf{V}^{-1} + \mathrm{E}\,u_1(d)\,\mathbf{I} \tag{18}$$

and

$$\mathbf{A} = \mathrm{E}\,u_1(d)^2\,\mathbf{ee}', \tag{19}$$

where $d = \mathbf{e}'\mathbf{V}^{-1}\mathbf{e}$ and $\mathbf{V}$ is defined by

$$\mathbf{V} = \mathrm{E}\,u_2(d)\,\mathbf{ee}', \tag{20}$$

the functions u_1 and u_2 depending on ρ_1, ρ_2 and the distribution of $\mathbf{e}$.

Approximate confidence intervals for the parameters can be obtained by estimating $\mathbf{S}$ from (17)-(20), replacing expectations by sample averages, and the $\mathbf{e}_i$'s by the residuals $\mathbf{r}_i$'s.

Computation of these estimators can be performed by using essentially the same approach as for univariate τ-estimators. A rough approximation is first obtained by subsampling, and then an iterative search is used to converge to a local -hopefully global- minimum. If one had to deal only with the MLM, it would suffice to take subsamples of size $\max(q,p)$ to perform LSE. For simultaneous equations, let p_j be the number of parameters in the j-th equation of (4). Then we would take subsambples of size $\max(q,p,p_1,\ldots,p_q)$ and for each one fit the parameters of the j-th equation from a sub-subsample of size q_j.

While much work remains to be done, concerning the choice of ρ_1 and ρ_2 and the evaluation of the bias and variability of these estimators, we think that this approach is the most promising one.

4 References

Amemiya, T. (1982), "Two Stage Least Absolute Deviations Estimators". *Econometrica* 50, 689–711.

Amemiya, T. (1985), *"Advanced Econometric Methods"*, Blackwell.

Hosjer, W.S. (1992), "On the Optimality of S-estimators", *Statistics and Probability Letters*, 12, 413-419.

Judge, G., Griffiths, W., Hill, R., Lütkepohl, H. and Lee, T., (1985), *"The Theory and Practice of Econometrics"*, John Wiley and Sons, New York.

Koenker, R, and Portnoy, S. (1990), "M Estimation of Multivariate Regressions", *Journal of the American Statistical Association*, 85, 1060–1068.

Krasker, W.S. (1986), "Two-Stage Bonded-Influence Estimators for Simultaneous-Equations Models", *Journal of Business & Economic Statistics*, 4, 437–444.

Krasker, W.S. and Welsch, R.E. (1985), "Resistant Estimation for Simultaneous-Equations Models Using Weighted Instrumental Variables", *Econometrica*, 53, 1475–1488.

Kuh, E. and Welsch, R.E. (1980), "Econometric Models and their Assessment for Policy: Some New Diagnostics Applied to the Translog Energy Demand in Manufacturing", in *Proceedings of the Workshop on Validation and Assessment Issues of Energy Models*, ed. by S. Gass, Washington D.C.: National Bureau of Standards, 445-475.

Lopuhaä, H. (1991), "Estimation of Location and Covariance with High Breakdown Point", Ph.D. Thesis, Technische Universiteit Delft.

Mariano, R.S. (1982), "Analytical Small-Sample Distribution Theory in Econometrics: the Simultaneous-Equations Case", *International Economic Review*, 23, 503–534.

Maronna, R.A. (1976), "Robust M–estimators of Multivariate Location and Scatter", *The Annals of Statistics*, 4, 51–67.

Maronna, R.A., Bustos, O., and Yohai, V.J., (1979), "Bias- and Efficiency-Robustness of Generalized M-Estimators for Regression with Random Carriers", in *Smoothing Techniques for Curve Estimation*, (T. Gasser, an M. Rosenblatt, eds.), 91-116. Lecture Notes in Mathematics 757. Springer Verlag, New York.

Maronna, R.A. and Yohai, V.J. (1991) "The Breakdown Point of Simultaneous GM-Estimates of Regression and Scale", *Journal of the American Statistical Association*, 586, 699-703.

Maronna, R.A. and Yohai, V.J. (1991), "Robust Estimation in Simultanous-Equations Models" (in Spanish), communication presented at the First Latin American Meeting os Statistical Societies, Valparaíso, Chile.

Maronna, R.A., Stahel, W.A. and Yohai, V.J. (1992), "Bias–Robust Estimators of Multivariate Scatter Based on Projections", *Journal of Multivariate Analysis*, 42, 141–161.

Maronna, R.A. and Yohai, V.J. (1995a), "The Behavior of the Stahel–Donoho Robust Multivariate Estimator". To appear in *Journal of the American Statisticsl Association.*

Maronna, R.A. and Yohai, V.J. (1995b), "Tau-estimators for Simultanoeus Equations Models", work in preparation.

Prucha, I.R. and Kelejian, H.H. (1984), "The Structure of Simultaneous Equation Estimators: a Generalization Towards Nonnormal Disturbances", *Econometrica*, 52, 721–736.

Yohai, V.J. and Zamar, R.H. (1988), "High Breakdown-Point Estimates of Regression by Means of the Minimization of an Efficient Scale", *Journal of the American Statistical Association*, 83, 406–413.