BOUNDED-INFLUENCE ESTIMATORS FOR THE SURE MODEL

by

Franco Peracchi

Department of Economics
University of California
Los Angeles

September 1988
Working Paper No. 521

JEL Classification No. 211
Keywords: Robust estimation, Influence function, SURE model

Abstract

This paper considers robust estimation of the seemingly unrelated regression equations (SURE) model. Following the infinitesimal approach to robustness, we characterize the class of optimal bounded-influence estimators and we propose two computationally simple estimators. We also study the behavior of these estimators.

<u>1. Introduction</u>

In recent years much research has been devoted to developing regression estimators which are robust, that is, not too sensitive to violations of some of the underlying statistical assumptions [see e.g. Krasker (1980), Huber (1981, 1983), Krasker and Welsch (1982), Hampel et al. 1986)]. There have been also some interesting econometric applications of these estimators . For example, Krasker, Kuh and Welsch (1983) and Small (1986) estimate hedonic price models for housing, Swartz and Welsch (1986) estimate and forecast energy demand, Thomas (1987) uses a very large data set to estimate Engel curves for food. All these studies show that robust methods can lead to significant differences with respect to ordinary least squares (LS) in terms of point estimates, inference and forecasts. This is mainly due to the fact that robust methods are much less sensitive than LS to local violations of the model assumptions. All the available studies, however, assume that the disturbances have a scalar covariance matrix at the 'central' model. This rules out important practical cases, such as systems of seemingly unrelated regression equations (SURE).

In this paper we consider robust estimation of the SURE model. Following the infinitesimal approach to robustness [Hampel et al. (1986)], we present a class of regular (that is, consistent and asymptotically normal) M-estimators that have a bounded and continuous influence function (IF). All these estimators are therefore qualitatively robust [Hampel (1971)], that is, small perturbations of the assumed statistical model can only have small effects on the distribution of the estimates. This desirable local stability property is not shared by the conventional maximum likelihood (ML) estimator based on assumption of normal (Gaussian) disturbances. In the class of estimators with a bounded and continuous IF, the ones proposed in this paper have minimum asymptotic mean square error (MSE) at the Gaussian model, and therefore attain the best trade-off between efficiency and robustness.

This paper is organized as follows. Section 2 examines the robustness properties of the Gaussian ML estimator. Section 3 characterizes the class of optimal bounded-influence estimators. Section 4 presents some Monte Carlo results. Section 5 contains the conclusions.

2. Robustness properties of the Gaussian ML estimator

Consider a system of q regression equations of the form

$$y_{in} = x_{in}' \beta_i + u_{in}, \qquad i = 1,\dots,q; \qquad n = 1,\dots,N,$$

where β_i is a $k_i \times 1$ vector of possibly unknown parameters and x_{in} is a $k_i \times 1$ vector of exogenous regressors. The disturbances $\{u_{in}\}$ are serially independent with zero mean and finite variance, but are contemporaneously correlated across regression equations. The system can be rewritten as

$$y_n = x_n' \beta_0 + u_n, \qquad n = 1,\dots,N, \quad (1)$$

where y_n is a $q \times 1$ vector, $\beta_0 = (\beta_1',\dots,\beta_q')'$ is a $k \times 1$ vector of parameters with $k = \Sigma_{i=1}^{q} k_i$,

$$x_n = \begin{bmatrix} x_{1n} & & \\ & \ddots & \\ & & x_{qn} \end{bmatrix}$$

is a k×q matrix of exogenous regressors, and $\{u_n\}$ are q×1 independently and identically distributed vectors with zero mean and finite variance Σ_0, a symmetric, positive definite (p.d.) q×q matrix. Let $\sigma_0 = \text{vec}\ \Sigma_0$ [1], and let $\Theta \subset R^{k+q^2}$ and $\theta_0 = (\beta_0', \sigma_0')' \in \Theta$ denote respectively the parameter space and the vector of parameters to be estimated. The parameter of interest is β_0, whereas σ_0 is a nuisance parameter.

The parameter θ_0 can be estimated using various methods, including single-equation LS or, if efficiency gains are sought, Zellner's (1962) feasible GLS procedure and the method of ML. In the ML case it is commonly assumed that the disturbances in (1) have a q-variate normal (Gaussian) distribution. Let $\{F_\theta\}$ denote the Gaussian parametric model for the observations, and let $\hat{\theta}_{ML} = (\hat{\beta}'_{ML}, \hat{\sigma}'_{ML})'$ denote the Gaussian ML estimator of θ_0. It is well known that the Zellner's estimator is asymptotically equivalent to $\hat{\theta}_{ML}$ and, if the Gaussian model is correctly specified and possesses a finite, positive definite (p.d.) Fisher information matrix, $\hat{\theta}_{ML}$ is consistent and asymptotically efficient.

To investigate the robustness properties of the Gaussian ML estimator, let $\hat{\theta}_{N-1}$ denote the estimator corresponding to a sample of size N - 1, and consider adding to the sample an additional observation z = (y, (vec x)')' $\in Z$, where Z denotes the sample space. Let $\hat{\theta}_N$ denote the resulting estimator. The rescaled difference $N(\hat{\theta}_N - \hat{\theta}_{N-1})$ can be shown to converge in probability to a finite limit. This limit, viewed as a function of z, is called the influence function (IF) of $\hat{\theta}_{ML}$ [Hampel (1974)], and is a measure of the asymptotic bias of $\hat{\theta}_{ML}$ when the distribution of the observations is subject to an infinitesimal amount of contamination at a given point in the sample space. The IF therefore provides a description of the local robustness properties of an estimator. It follows from standard results [see e.g. Serfling (1980)] that the IF of $\hat{\theta}_{ML}$ (and of any other estimator asymptotically equivalent to $\hat{\theta}_{ML}$), evaluated at the Gaussian model $\{F_\theta\}$, is given by

$$IF(z,\hat{\beta}_{ML},F_\theta) = J_\beta(\theta)^{-1} \times V r$$

$$IF(z,\hat{\sigma}_{ML},F_\theta) = (1/2)\ J_\sigma(\theta)^{-1} [V \otimes V'] \text{ vec } (rr' - I_q)$$

where $r = V'(y - x'\beta)$ is the vector of standardized disturbances, V is a finite, p.d. matrix such that $VV' = \Sigma^{-1}$, $J_\beta(\theta)$ and $J_\sigma(\theta)$ are the diagonal blocks of the Fisher information matrix, and $\otimes$ denotes the Kronecker product. Notice that $IF(z,\hat{\beta}_{ML},F_\theta)$ and $IF(z,\hat{\sigma}_{ML},F_\theta)$ are both unbounded functions of z. This reflects the fact that one large disturbance or one gross-error in x are sufficient to completely spoil the estimates. Also notice that the influence of a single disturbance on $\hat{\beta}_{ML}$ is linear, whereas its effect on $\hat{\sigma}_{ML}$ is quadratic.

3. Bounded influence estimation

We now consider the class of M-estimators of θ_0. An estimator in this class is a root of an implicit equation of the form

$$\textstyle\sum_{n=1}^{N} \eta_N(z_n,\theta) = 0, \tag{2}$$

where the vector function $\eta_N(\cdot,\theta)$, mapping $Z \times \Theta$ into R^{k+q^2}, is called the score function associated with the estimator. Clearly, we obtain the ML estimator when $\eta_N(\cdot,\theta)$ is equal to the likelihood score.

As it is well known, the efficiency and robustness properties of an M-estimator are closely related to the properties of its IF [see e.g. Serfling (1980) and Huber (1981)]. Given a parametric model, an M-estimator is efficient if and only if its IF is a non-singular linear transformation of the likelihood score. On the other hand, an M-estimator is qualitatively robust in the sense of Hampel (1971) if and only if its IF is bounded and continuous. Qualitative robustness is a desirable property, because it ensures that small departures from the assumed statistical model can only have small effects on the distribution of an estimator. The Gaussian ML estimator is clearly not qualitatively robust. A natural quantitative measure of the robustness of an M-estimator is given by the sup-norm of its IF, called the estimator's sensitivity. This measure provides an upper bound on the asymptotic bias that may arise under small departures from the model assumptions. An estimator with a bounded IF or, equivalently, a finite sensitivity, is called a bounded-influence estimator. In this paper we allow for separate sensitivity bounds for the estimators of β_0 and σ_0, and we consider the class of bounded-influence estimators $\tilde{\theta} = (\tilde{\beta}', \tilde{\sigma}')'$ such that

$$\sup_{z\in Z} \|IF(z,\tilde{\beta},F_\theta)\|_{B_1} \leq \gamma_1 \tag{4}$$

$$\sup_{z\in Z} \|IF(z,\tilde{\sigma},F_\theta)\|_{B_2} \leq \gamma_2 \tag{5}$$

where (B_1, B_2) are p.d. matrices, (γ_1, γ_2) are finite constants, and $\|x\|_B = (x'B\ x)^{1/2}$ denotes the norm of the vector x in the metric of the p.d. matrix B. An estimator in this class is called optimal if it has minimum asymptotic mean square error (MSE) at the assumed Gaussian model $\{F_\theta\}$.

Optimal bounded-influence estimators can be characterized by their score function. Peracchi (1987) showed that the optimal score has a relatively simple form when the MSE criterion is defined in the metric of a block-diagonal matrix with sub-blocks equal to (B_1, B_2). In particular, the symmetry about zero of the Gaussian error distribution implies that the first k components of the optimal score $\eta(z,\theta)$ are given by

$$\eta_1(z,\theta) = w_1(z,\theta)\ x\ V\ r, \tag{6}$$

where $w_1(z,\theta)$ is a scalar weight function defined by

$$w_1(z,\theta) = \min\{1,\ \gamma_1/\|A_1\ x\ V\ r\|_{B_1}\},$$

and A_1 is a p.d. matrix root of the equation

$$E\ \min\left\{1,\ \frac{\gamma_1}{\|A_1\ x\ V\ r\|_{B_1}}\right\} x\ V\ r\ r'V'x' - A_1^{-1} = 0,$$

with E denoting expectations taken with respect to the Gaussian model (1). It can be shown that this equation has a solution only if

$$\gamma_1 \geq (\text{trace } B_1)\ /\ E\ \|x\ V\ r\|_{B_1}.$$

The remaining q^2 components of the optimal score are given by

$$\eta_2(z,\theta) = w_2(z,\theta)\ \text{vec}\ (rr' - I_q), \tag{7}$$

where $w_2(z,\theta)$ is a scalar weight function defined by

$$w_2(z,\theta) = \min\{1,\ \gamma_2/\|A_2 \text{ vec } (rr' - I_q)\|_{B_2}\},$$

and A_2 is a p.d. matrix root of the equation

$$E_\Phi \min\left\{1,\ \frac{\gamma_2}{\|A_2 \text{ vec}(rr'-I)\|_{B_2}}\right\} \text{vec}(rr'-I)\ \text{vec}(rr'-I)' - A_2^{-1} = 0,$$

with expectations taken with respect to the multivariate standard normal distribution. It can be shown that this equation has a solution only if

$$\gamma_2 \geq (\text{trace } B_2) \,/\, E_\Phi \|\text{vec } (rr' - I_q)\|_{B_2}.$$

Since the optimal score function $\eta(\cdot,\theta)$ is bounded and continuous, $\tilde{\theta}$ is qualitatively robust. If $(\gamma_1, \gamma_2) \longrightarrow \infty$, then $\tilde{\theta}$ reduces to the Gaussian ML estimator of θ_0. Notice that if γ_1 or γ_2 are finite, system estimation is required even when each equation contains exactly the same regressors and there are neither cross-equations nor covariance restrictions.

The optimal estimator $\tilde{\theta}$ can be interpreted as a weighted ML estimator. Geometrically, the likelihood score for one observation is shrunk so as to satisfy the robustness constraints (3) and (4). The exact form of the weights applied to the likelihood score depends on the choice of the matrices (B_1, B_2). For example, when B_1 and B_2 are both equal to the identity matrix we obtain the SURE analogue of the Hampel-Krasker estimator of regression [Hampel (1978), Krasker (1980)]. When $B_1 = AV(\tilde{\beta},F_\theta)^{-1}$ and $B_2 = AV(\tilde{\sigma},F_\theta)^{-1}$ we obtain the analogue of the regression estimator of Krasker and Welsch (1982).

Computation of $\tilde{\theta}$ for a given sample can be expensive because the equation (2) must be solved numerically, and each iteration requires solving two implicit matrix equations for A_1 and A_2. Following the suggestion of Peracchi (1987), the arbitrariness of the choice of B_1 and B_2 can be exploited to simplify the computation. In particular, B_1 and B_2 can be chosen such that

$$w_1(z,\theta) = \min\{1,\ \gamma_1/\|x\ V\ r\|\} \tag{8}$$

$$w_2(z,\theta) = \min\{1,\ \gamma_2/\|\mathrm{vec}\ (rr' - I_q)\|\}, \tag{9}$$

where $\|\cdot\|$ denotes the Euclidean norm. The resulting estimator of β_0, denoted by BI1, is not invariant under a reparameterization of the model. An invariant estimator, denoted by BI2, can be obtained by choosing B_1 such that

$$w_1(z,\theta) = \min\{1,\ \gamma_1/\|x\ V\ r\|_{J_\beta(\theta)^{-1}}\} \tag{10}$$

where $J_\beta(\theta) = E\ x\ \Sigma^{-1}\ x'$.

In this paper we shall also consider a simple generalization of Huber's M-estimator of regression. The score function of this estimator has the same form as (6) and (7), with

$$w_1(z,\theta) = \min\{1,\ \gamma_1/\|r\|\} \tag{11}$$

and $w_2(z,\theta)$ given by (9). This estimator has a bounded IF only if the regressors take values in a bounded set, but should have good robustness properties when disturbances have a thick-tail distribution.

Let $\bar{\theta} = (\bar{\beta}, \bar{\sigma})$ denote any of the BI1, BI2 and Huber-type estimators. Given a sample of size N, estimates of β_0 and Σ_0 can be obtained by a simple iteratively reweighted LS algorithm, with the (i + 1)-th iteration given by

$$\bar{\beta}^{(i+1)} = [\sum_{n=1}^{N} w_{1n}^{(i)} x_n \bar{\Sigma}_{(i)}^{-1} x_n']^{-1} \sum_{n=1}^{N} w_{1n}^{(i)} x_n \bar{\Sigma}_{(i)}^{-1} y_n$$

$$\bar{\Sigma}_{(i+1)} = [\sum_{n=1}^{N} w_{2n}^{(i)}]^{-1} \sum_{n=1}^{N} w_{2n}^{(i)} (y_n - x_n' \bar{\beta}^{(i)}) (y_n - x_n' \bar{\beta}^{(i)})'$$

where $w_{1n}^{(i)}$ is given by (8), (10) or (11), $w_{2n}^{(i)}$ is given by (9), and β, Σ and V are replaced by the corresponding values obtained from the i-th iteration. The algorithm can be started at the single-equation LS estimate of β_0. Starting at some robust estimates is however preferable, and is essential if the algorithm is iterated only a few times.

A simple modification of the argument in Maronna and Yohai (1981) can be used to establish consistency and asymptotic normality of $\bar{\theta}$ under general conditions. The asymptotic variance matrix of $\bar{\theta}$ can be estimated consistently by $P_N^{-1} Q_N P_N^{-1}$, where $P_N = N^{-1} \sum_{n=1}^{N} (\partial/\partial\theta) \eta(z_n, \bar{\theta}_N)$ and $Q_N = N^{-1} \sum_{n=1}^{N} \eta(z_n, \bar{\theta}_N) \eta(z_n, \bar{\theta}_N)'$. If the distribution of the disturbances in (1) is symmetric about zero, the asymptotic variance matrix of $\bar{\theta}$ is block diagonal with respect to $\bar{\beta}$ and $\bar{\sigma}$, which implies that $\bar{\beta}$ and $\bar{\sigma}$ are asymptotically independent.

The asymptotic normality of $\bar{\theta}$, and the asymptotic independence of $\bar{\beta}$ and $\bar{\sigma}$, lead to simple Wald-type tests of hypothesis concerning the regression parameters. Score tests can also be constructed, based on the average score evaluated at the restricted estimates. The asymptotic normality of $\overline{\theta}$ implies that the Wald- and score-test statistics have an asymptotic χ^2 distribution under the null hypothesis. Since $\overline{\theta}$ is a bounded-influence estimator all these tests are robust, that is, their level and power are relatively insensitive to small deviations from the assumed statistical model [Peracchi (1987)]. This property is not shared by tests based on the Gaussian ML or the Huber-type estimators.

The difference between $\overline{\theta}$ and the Gaussian ML estimator of θ_0 can be used as the basis for specification tests of the type proposed, among others, by Hausman (1978). In the case of bounded influemce estimators, such specification tests are likely to be quite powerful because, while $\overline{\theta}$ is only slightly less efficient than $\hat{\theta}_{ML}$ at the assumed model, the difference between $\overline{\theta}$ and $\hat{\theta}_{ML}$ can be large when the model is misspecified.

Finally, in the case of bounded-influence estimators, the weights $\{w_{1n}\}$ and $\{w_{2n}\}$ summarize all the information on the influence of a particular observation, and can therefore be used as effective diagnostics for outliers and influential observations. Since the weights are jointly computed with the estimates, no further calculation is required.

4. Monte Carlo results

In this Section we report the results of a set of Monte Carlo experiments, carried out in order to study the behavior of the various estimators under small departures from the assumed Gaussian model (1). The departures considered include non-normal disturbances and gross-errors in the data.

The 'central' model consists of two simple regression equations, with intercept and slope all equal to 1. The regressors in each equation have been randomly drawn from the $U(-\xi,\xi)$ distribution [2], and experiments have been carried out with different values of ξ. The 2×1 disturbance vector has been generated as $u_n = C\ r_n$, where C is the Cholesky decomposition of the 2×2 p.d. matrix Σ_0 and the elements of r_n are independently drawn from a $N(0,1)$ distribution. Different choices of Σ_0 are considered so as to allow for different amount of correlation across regression equations.

The estimators considered include the single-equation LS estimator, Zellner's feasible GLS estimator, the Gaussian ML estimator, Huber-type estimator and the BI1 and BI2 estimators. The starting values for bounded-influence estimation are given by the Huber-type estimates. The sensitivity bounds for bounded-influence and Huber-type estimators have been chosen so as to attain an average weight of approximately 95% at the central model. This choice results is an asymptotic relative efficiency of about 95% at the Gaussian model.

Each Monte Carlo experiment consists of 1000 replications, and each set of experiments has been carried out for different samples sizes [3].

To avoid bothering the reader with too many tables, we choose to present our results in graphical form. Detailed numerical tables are available from the Author upon request.

Non-normal disturbances

First we examine the behavior of the various estimators under small departures of the distribution of the disturbances from normality. The distributions that we consider are ordered in Table 1 by the index of tail length suggested by Rosenberger and Gasko (1983), namely [4]

$$\tau(F) = \frac{F^{-1}(.99) - F^{-1}(.5)}{\Phi^{-1}(.99) - \Phi^{-1}(.5)} \Big/ \frac{F^{-1}(.75) - F^{-1}(.5)}{\Phi^{-1}(.75) - \Phi^{-1}(.5)} .$$

Because all distributions are symmetric, no bias arises and the relevant issue is the precision of the various estimators.

The top part of Figure 1 shows the root mean square error (RMSE) of alternative estimators of the slope parameter in the first equation, for samples of size 25 e 50 respectively. The spread of the distribution of the regressors corresponds to a population R^2 of .50 for each equation. Single-equation LS, Zellner's feasible GLS and the Gaussian ML estimator all behave very similarly and so we only report results for Zellner's estimator. The BI1 and the Huber-type estimator behave almost identically and so we only report results for the former. The bottom part of the Figure shows the ratio of the RMSE of the various estimators to the one of Zellner's estimator. It is easily seen that Zellner's estimator looses its efficiency very quickly as the tail length of the error distribution increases from 25 to 50. Also notice that the BI2 and Huber-type estimators are more efficient than the BI1 estimator for heavy tailed distributions, but this difference reduces considerably as the sample size increases. In the case of distributions with long tails, the median absolute error (MAE) of the estimates provides a better measure of an estimator's variability around the true parameter value. The results for the MAE case, shown in Figure 2, agree with the ones from Figure 1, except that the decline in the efficiency of Zellner's estimator relative to more robust estimators is less pronounced.

Very similar results are obtained for the slope parameter in the second equation and the intercept parameters, and for different choices of the variance of the regressors and the contemporaneous correlation across regression equations. These results are not presented here.

Gross-errors

Next we report some Monte Carlo results that illustrate the robustness properties of the various estimators under contamination by gross-errors. Observations are generated from a simple errors-in-variables model, where measurement errors only occur with probability π. More precisely, observations are generated as follows

$$y^*_{in} = \beta_{i1} + \beta_{i2}x^*_{in} + u_{in},$$

$$y_{in} = y^*_{in} + v_{in},$$

$$x_{in} = x^*_{in} + w_{in}, \qquad i = 1,2,$$

where $\beta_{i1} = \beta_{i2} = 1$, the regression disturbances $\{u_{in}\}$ are serially uncorrelated and normally distributed with mean zero and unit variance, and Cov $(u_{1n}, u_{2n}) = 0.5$. The measurement errrors v_{in} and w_{in} are equal to zero with probability $1 - \pi$, and with probability π are randomly drawn from a normal distribution with mean zero and standard deviation equal to δ and ω respectively. To study the effects of changes in the amount and type of contamination, we considered different choices of π, δ and ω. When $\delta = \omega = 0$ we obtain the Gaussian model (1). When the regressors are measured with errors (that is, $\omega > 0$), a bias arises. Thus, we focus attention on the behavior of the various estimators in terms of their bias and imprecision. Bias is measured by the difference between the

(Monte Carlo) mean or median and the true parameter value, and imprecision is measured by the RMSE and the MAE.

Figures 3 and 4 refer to the case when $\pi = .05$ and only the regressors are measured with error, that is, $\delta = 0$. For simplicity we only present results for the sample size of 25. Thebias and the imprecision increase with ω for all estimators, but this increase is very moderate for the two bounded-influence estimators (and especially for the BI1 estimator). Notice that the Huber-type and the Gaussian ML estimators now behave very similarly in terms of both bias and imprecision, particularly for large values of ω. Also notice that the superiority of the bounded-influence estimators is even clearer if median bias and MAE are considered (Figure 4).

Our justification for bounded-influence estimators is that they offer protection against small departures from the assumed statistical model. We now investigate the effects of increasing the 'degree of misspecification' of the model, by increasing the measurement error probability and/or the variance of the contaminating distributions. First the measurement error probability π is increased from .05 to .10 (Figures 5 and 6). Our general conclusions do not change, but the superiority of the bounded-influence estimators is now somewhat reduced. This can also be observed in Figures 7 and 8, where the measurement error probability is kept at .05, but both the dependent variables and the regressors are now measured with error (the results are shown for the case when $\delta = 3$).

5. Conclusions

The derivation of optimal bounded-influence estimators relies on asymptotic arguments. The Monte Carlo results presented in this paper indicate that these estimators maintain good efficiency and robustness properties even in small samples. This provides further evidence in favor of using optimal bounded-influence estimators when an investigator seeks protection against small departures from the model assumptions, while retaining high efficiency at the central model. The price that must be paid, namely a small efficiency loss if the assumed model is exactly correct, seems small compared with the potentially large gains in terms of reduced bias and imprecision of the estimates of the parameters of interest. Further, the investigator is free to choose the efficiency loss that he/she is willing to tolerate.

References

Hampel, F.R. (1971), "A general qualitative definition of robustness", *Annals of Mathematical Statistics*, 42, 1087-1096.

Hampel, F.R. (1974), "The influence function and its role in robust estimation", *Journal of the American Statistical Association*, 69, 383-393.

Hampel, F.R. (1978), "Optimally bounding the gross-error-sensitivity and the influence of position in factor space", *Proceedings of the ASA Statistical Computing Section*, ASA, Washington, D.C.

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

Hausman, J.A. (1978), "Specification tests in econometrics", *Econometrica*, 46, 1251-1272.

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

Krasker, W.S. (1980), "Estimation in linear regression models with disparate data points", *Econometrica*, 48, 1333-1346.

Krasker, W.S. and Welsch, R.E. (1982), "Efficient bounded-influence regression estimation", *Journal of the American Statistical Association*, 77, 595-604.

Krasker, W.S., Kuh, E., and Welsch, R.E. (1983), "Estimation for dirty data and flawed models", in Z. Griliches and M.D. Intriligator (eds.), *Handbook of Econometrics*, Vol. 1, North-Holland, Amsterdam.

Maronna, R. and Yohai, V. (1981), "Asymptotic behavior of general M-estimates for regression and scale with random carriers", *Zeitschrift fur Wahrscheinlichkeitstheorie und verwandte Gebiete*, 58, 7-20.

Peracchi, F. (1987), "Bounded influence methods in econometrics with an application to the censored regression model", Ph.D. Thesis, Princeton University.

Rosenberger, J.L. and Gasko, M. (1983), "Comparing location estimators: Trimmed means, medians, and trimeans", in D.C. Hoaglin, F. Mosteller and J.W. Tukey (eds.), *Understanding Robust and Exploratory Data*

Analysis, Wiley, New York.

Serfling, R.J. (1980), *Approximation Theorems of Mathematical Statistics*, Wiley, New York.

Small, K. (1986), "Effects of the 1979 gasoline shortages on Philadephia housig prices", *Journal of Urban Economics*, 19, 371-381.

Swartz, S. and Welsch, R.E. (1986), "Applications of bounded-influence and diagnostic methods in energy modeling", in D.A. Belsley and E. Kuh (eds.), *Model Reliability*, MIT Press, Cambridge, Mass.

Thomas, D. (1987), "Robust regression of the food share Engel curve", Economic Growth Center, Yale University, mimeo.

Zellner, A. (1962), "An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias", *Journal of the American Statistical Association*, 57, 348-368.

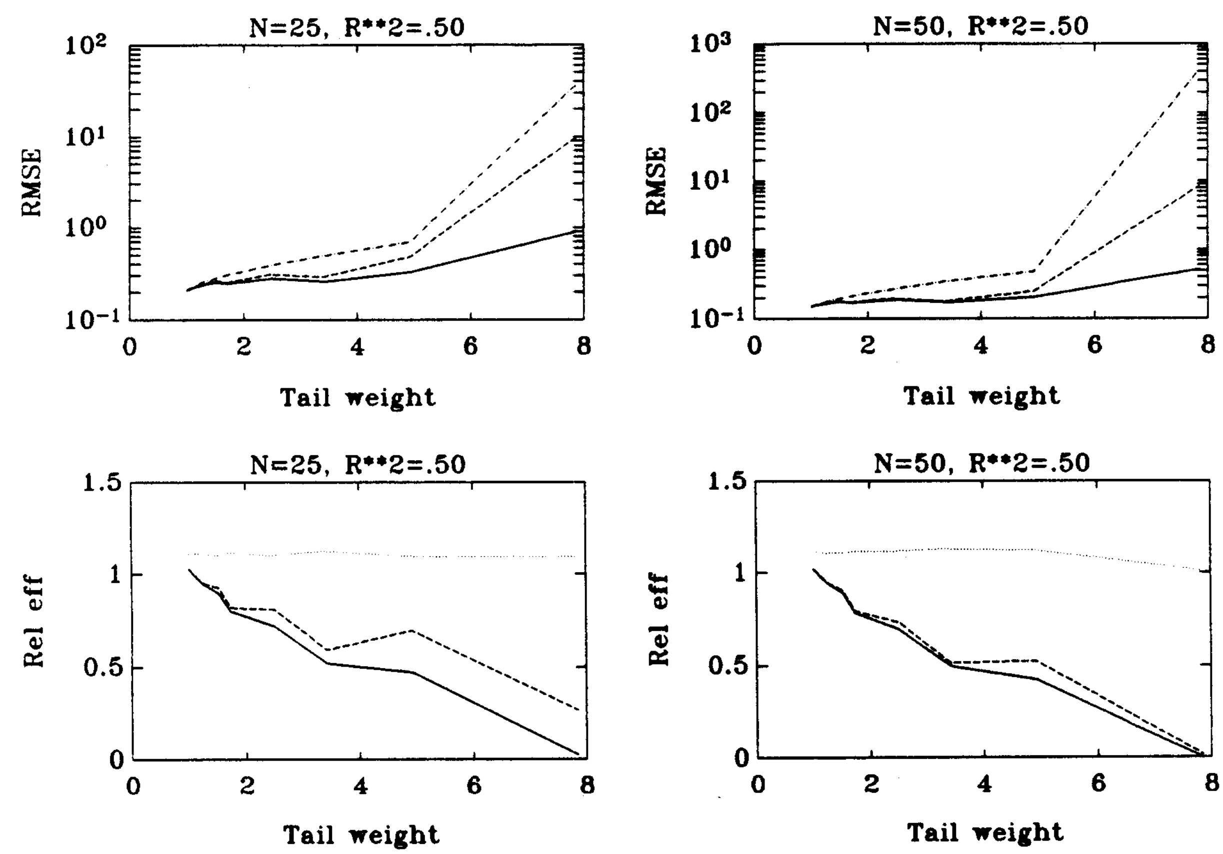
Table 1

Measures of tail length.

Distributions	Tail length
$N(0,1)$	1.00
$CN(.05,3)$	1.24
$CN(.10,3)$	1.53
$CN(.05,5)$	1.73
$CN(.10,5)$	2.50
$CN(.05,10)$	3.43
$CN(.10,10)$	4.93
Slash	7.85

Figure 1

Monte Carlo RMSE and relative efficiency of alternative estimators of the slope parameter in the first equation.

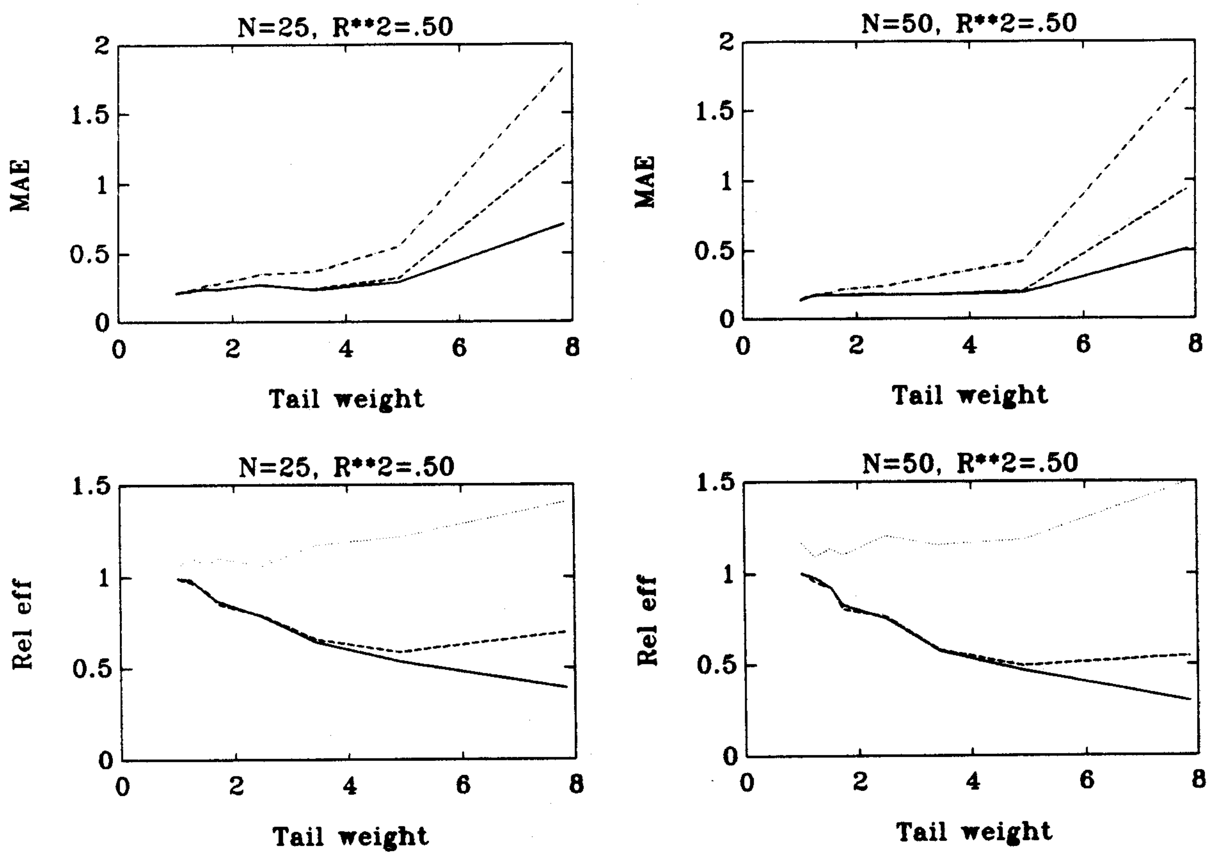


Legend:

· · · · · · · · ·	*Single-equation LS*
—·—·—·—	*Zellner's feasible GLS*
————————	*Huber-type*
— — — — —	*BI1*

Figure 2

Monte Carlo MAE and relative efficiency of alternative estimators of the slope parameter in the first equation.

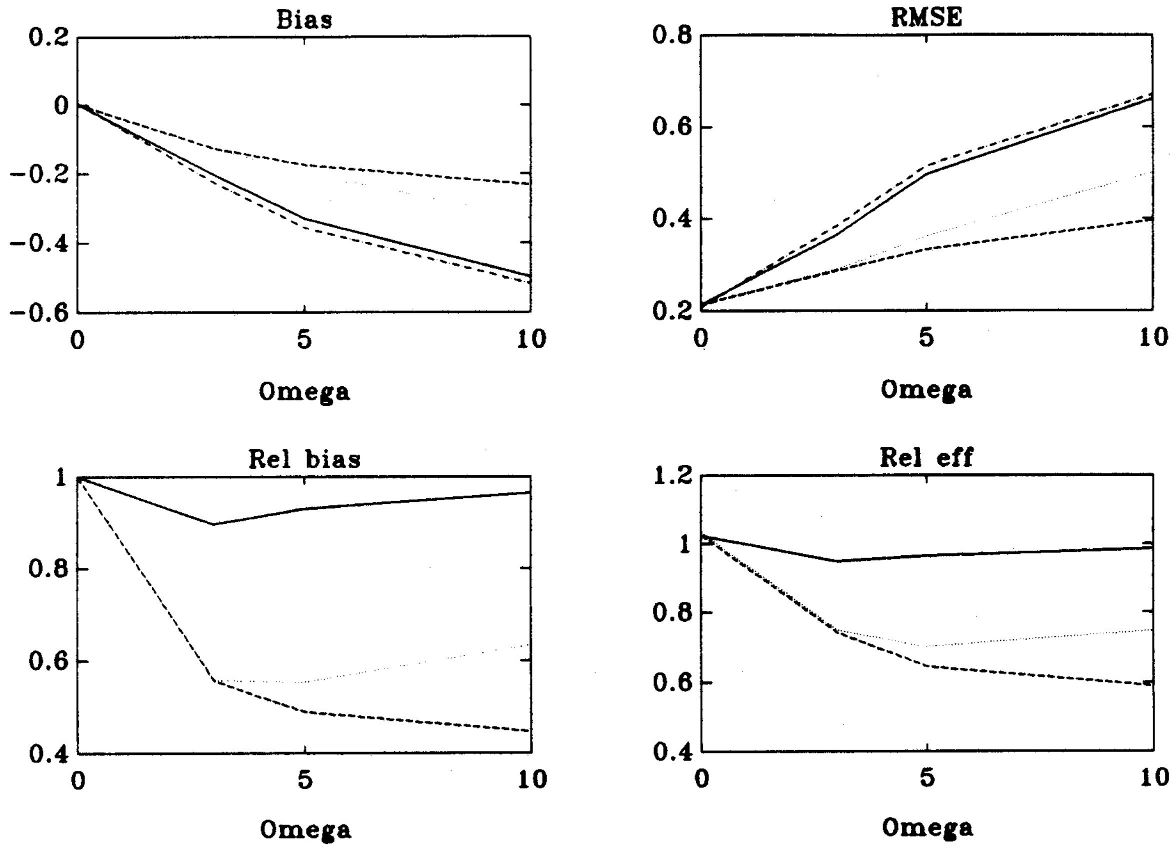


Legend:

....	*Single-equation LS*
-.-.-.-.-	*Zellner's feasible GLS*
————	*Huber-type*
- - - - -	*BI1*

Figure 3

Monte Carlo bias and RMSE of alternative estimators
of the slope parameter in the first equation.
Gross-error model: $\delta = 0$, $\pi = .05$, $N = 25$.



Legend:

-·-·-·-·- *Zellner's feasible GLS*

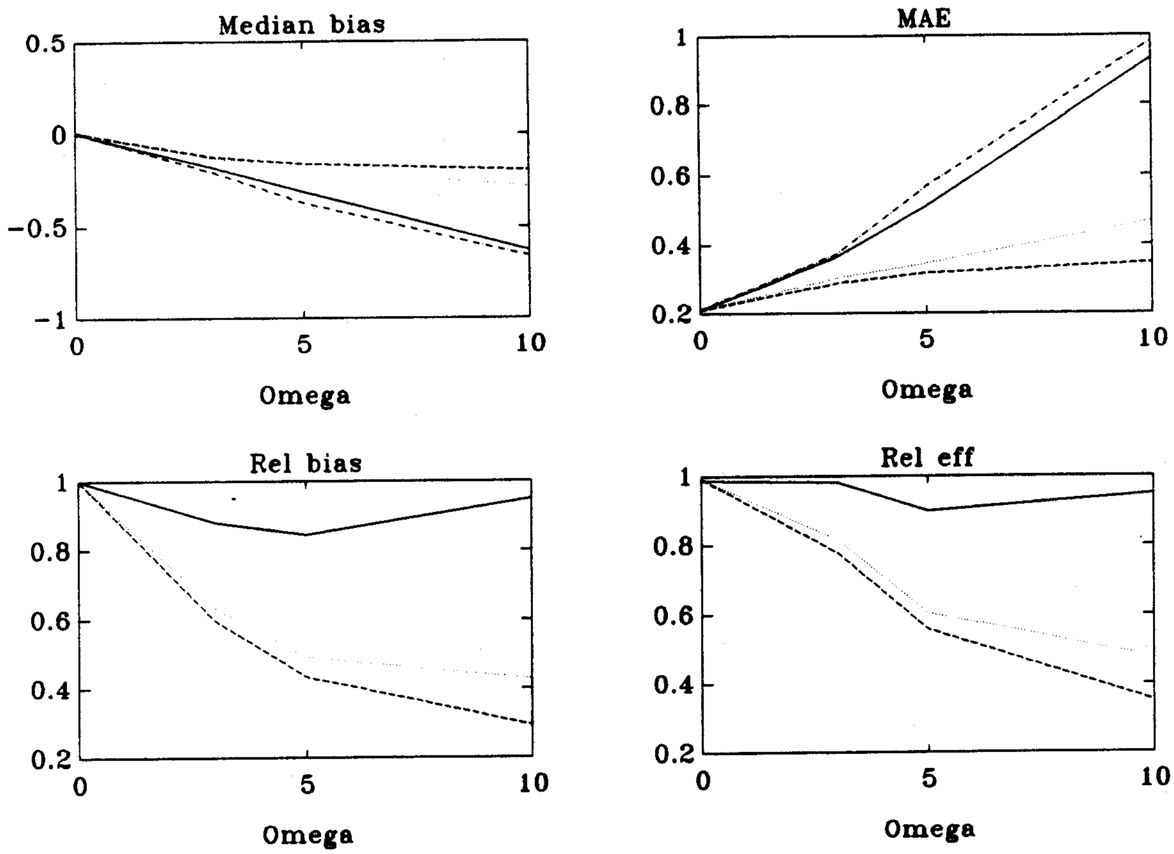
———— *Huber-type*

- - - - - *BI1*

········· *BI2*

Figure 4

Monte Carlo median bias and MAE of alternative estimators of the slope parameter in the first equation.
Gross-error model: $\delta = 0$, $\pi = .05$, $N = 25$.

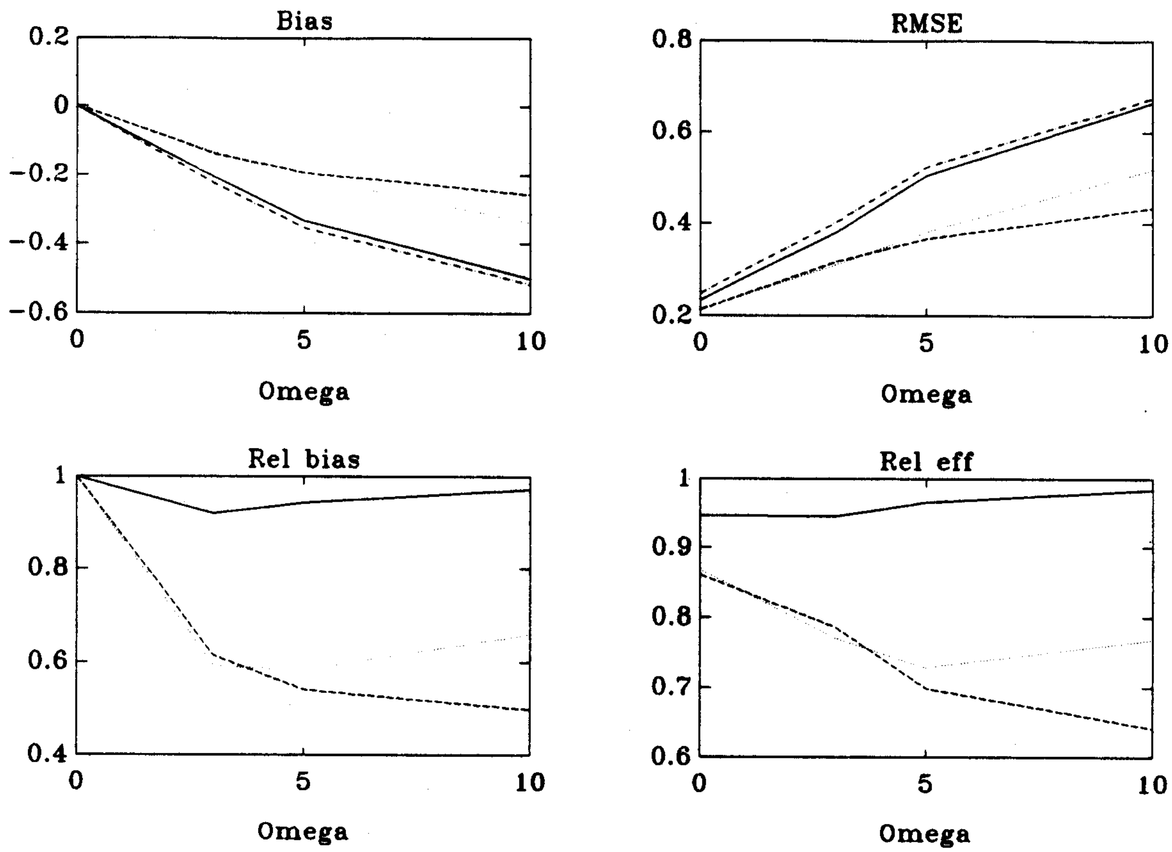


Legend:

−·−·−·−·−	*Zellner's feasible GLS*
————	*Huber-type*
− − − − −	*BI1*
· · · · · · · · ·	*BI2*

Figure 5

Monte Carlo bias and RMSE of alternative estimators
of the slope parameter in the first equation.
Gross-error model: $\delta = 0$, $\pi = .10$, $N = 25$.



Legend:

—·—·—·—·— *Zellner's feasible GLS*

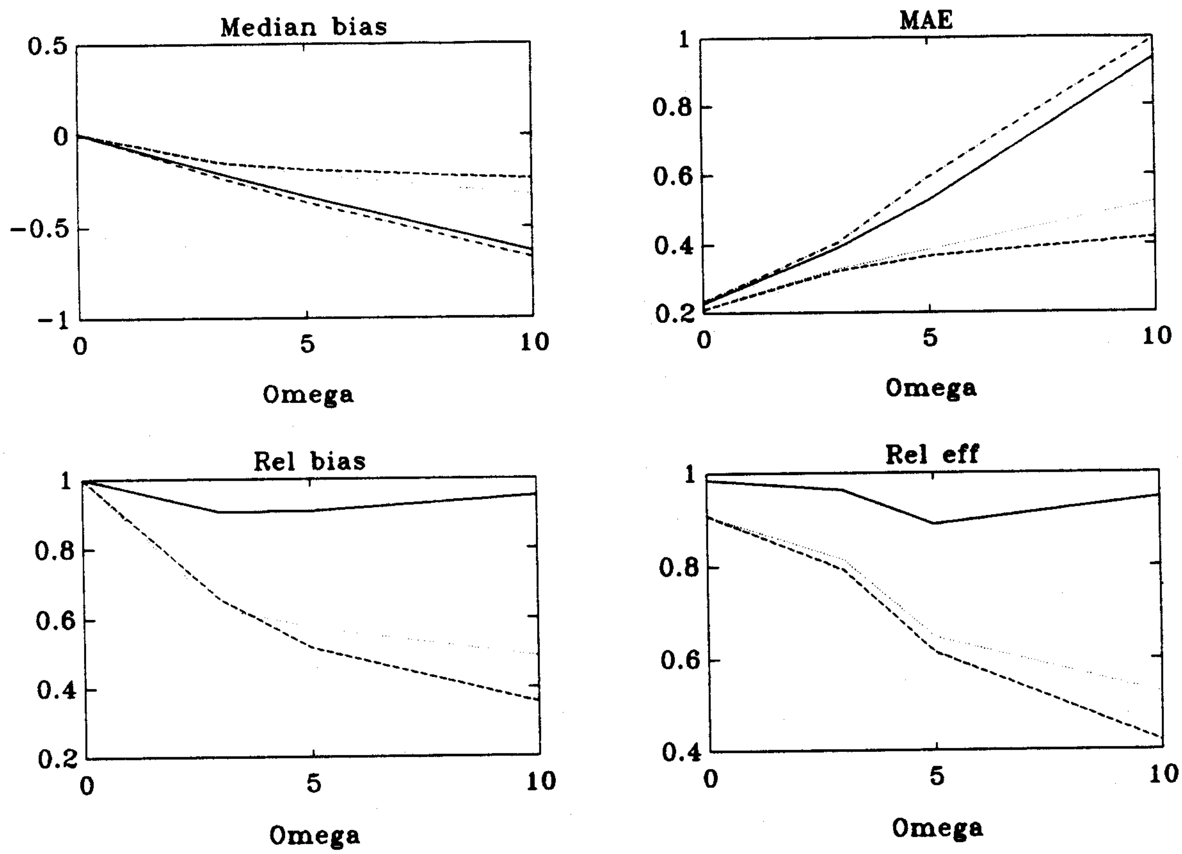
——————— *Huber-type*

– – – – – *BI1*

· · · · · · · · · *BI2*

Figure 6

Monte Carlo median bias and MAE of alternative estimators of the slope parameter in the first equation. Gross-error model: $\delta = 0$, $\pi = .10$, $N = 25$.

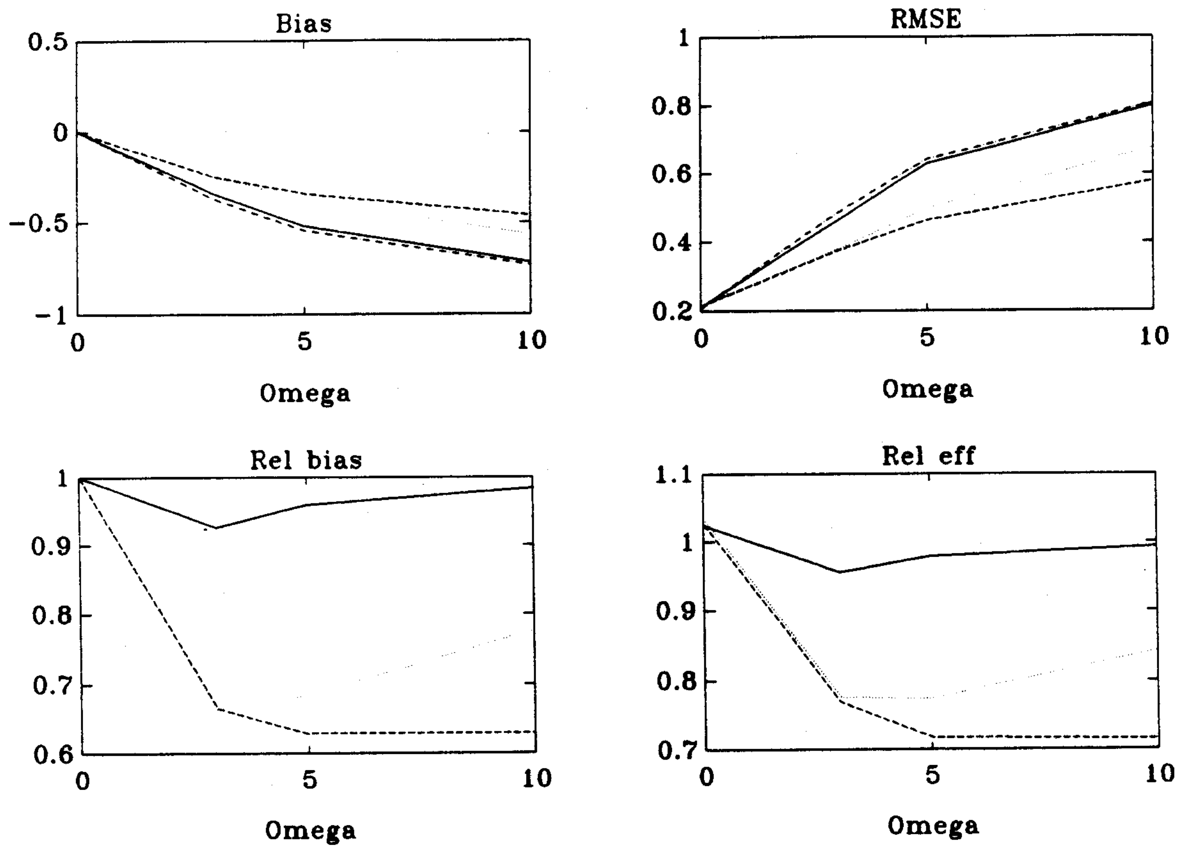


Legend:

—·—·—·—	*Zellner's feasible GLS*
————	*Huber-type*
– – – – –	*BI1*
·········	*BI2*

Figure 7

Monte Carlo bias and RMSE of alternative estimators
of the slope parameter in the first equation.
Gross-error model: $\delta = 3$, $\pi = .05$, $N = 25$.

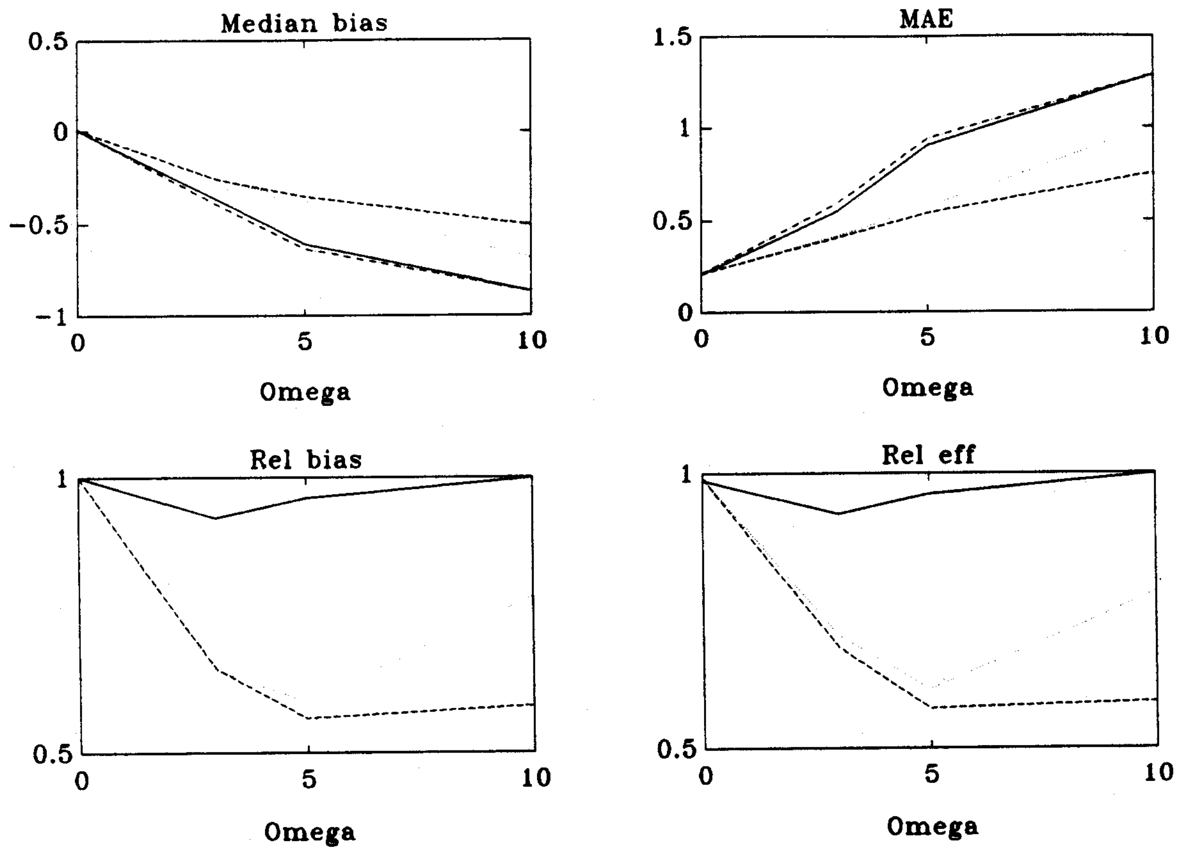


Legend:

—·—·—·—	*Zellner's feasible GLS*
————	*Huber-type*
– – – – –	*BI1*
·········	*BI2*

Figure 8

Monte Carlo median bias and MAE of alternative estimators
of the slope parameter in the first equation.
Gross-error model: $\delta = 3$, $\pi = .05$, $N = 25$.



Legend:

-.-.-.-.- *Zellner's feasible GLS*

——— *Huber-type*

- - - - - *BI1*

.... *BI2*

Footnotes

[1] 'Vec' denotes the operator that stacks the columns of a matrix in a single column vector. Strictly speaking, we should consider the q(q + 1)/2-vector of distinct elements of the matrix Σ_0, but this would only complicate the notation.

[2] This implies that the variance of the regressors is equal to $\xi^2/3$.

[3] The program has been written using GAUSS, Version 1.49b. Computations have been carried out on an IBM PS/2 Model 80.

[4] $CN(\pi,\sigma)$ denotes a contaminated normal distribution with contamination proportion equal to π and standard deviation of the contaminating distribution equal to σ. The $CN(.05,5)$ and $CN(.10,5)$ distributions have approximately the same tail weight as the t_2 and t_3 distributions. The 'slash' distribution is the ratio of a $N(0,1)$ and a $U(0,1)$ distribution. The tail behavior of this distribution is less extreme than the Cauchy, for which $\tau(F) = 9.22$.