Equivariant, Monotonic, 50% Breakdown Estimators

GILBERT W. BASSETT, JR.*

The median is equivariant, monotonic, and has 50% breakdown. No other estimator has all three properties.

KEY WORDS: Least median of squares; Median.

INTRODUCTION

Given a sample $\mathbf{x} = (x_1, \ldots, x_n)$, let the order statistics be denoted by $x_{(1)}, \ldots, x_{(n)}$, $n = 2p + q$, where q is 0 or 1 as the sample size is even or odd. The median interval is $[x_{(p+q)}, x_{(p+1)}]$. This interval is a point when n is odd and defines the median estimate. When n is an even number, the median will be defined as $1/2(x_{(p+q)} + x_{(p+1)})$. [To make the median estimate equivariant for an even number of observations ($q = 0$), it is necessary to define the median as the midpoint of the median interval.] Alternatively, the median can be identified by its least absolute deviations property: $s(m) = \inf s(b)$, where $s(b) = \Sigma|x_i - b|$ and m is any point in the median interval.

Three easily verified properties of the median are as follows. First, the median is equivariant. If observations are shifted and rescaled, then the median will be similarly shifted and rescaled. Second, the median is monotonic. If an observation increases, then the median will either stay constant or increase—it cannot decrease. Third, the median has the 50% breakdown property (see Donoho and Huber 1983; Hampel, Ronchetti, Rousseeuw, and Stahel 1986). This means that the median estimate stays bounded when 50% or fewer of the observations become unbounded.

Examples of other estimates with two of the three median properties are presented below. The theorem following the examples shows that no other estimate has all three properties, and hence the median is the equivariant, monotonic, 50% breakdown estimator.

This characterization of the median can be used to introduce students to the monotonicity and 50% breakdown concepts. The simple proof can be used to introduce the close connection between 50% breakdown and an exact fit property satisfied by any such estimate. Finally, the result can be used to discuss trade-offs between the monotonicity and 50% breakdown properties.

PROPERTIES

Equivariance. An estimate $T(\mathbf{x})$ is equivariant if $T(\delta\mathbf{x} + b) = \delta T(\mathbf{x}) + b$. This says that the estimate will be translated by b and scaled by δ if the data are similarly transformed. The median and many other estimators satisfy this property.

Monotonicity. $T(\mathbf{x})$ is monotonic if $T(\mathbf{x}) \geq T(\mathbf{x}')$, $\mathbf{x} \geq \mathbf{x}'$, where the vector inequality is read componentwise. This says that the estimate will not decrease in value when an observation is increased. Monotonic estimates besides the median include the mean and linear combinations (with nonnegative weights) of order statistics. For a recent use of monotonicity, see He, Jureckova, Koenker, and Portnoy (1990).

50% Breakdown. For a given $\mathbf{x}$, let $\mathbf{x}^{(s)}$ denote a vector of contaminated observations. The contamination replaces s of the original observations with arbitrary (typically unbounded) values. $T(\mathbf{x})$ has 50% breakdown if $\sup|T(\mathbf{x}) - T(\mathbf{x}^{(p+q)})| = \infty$ and $\sup|T(\mathbf{x}) - T(\mathbf{x}^{(s)})| < \infty$, for $s < p + q$; the suprema are taken with respect to any $\mathbf{x}^{(s)}$ with the specified amount of contamination. This means that the estimate remains bounded when 50% ($p - q$ out of the $2p + q$) of the observations are replaced by arbitrary values. This is desirable because the estimator then fits the majority of the data without being influenced by discrepant observations.

Examples

The mean is equivariant and monotonic, but it does not have the 50% breakdown property. Indeed, one arbitrary observation is enough to move the mean arbitrarily far from its original position and its breakdown is zero.

The least median of squares (LMS, see Rousseeuw 1984) for the location submodel is the midpoint of the shortest half of the sample, where "shortest half" means the half-sample with the smallest range. This estimate is similar to the shorth, the mean of the observations contained in the shortest half, which was considered in Andrews et al. (1972). The LMS estimate is equivariant and has 50% breakdown, but it is not monotonic.

To see this, consider the five observations (0, 0, 1, 1, x_5). The LMS is $(1 + x_5/2)$ for $x_5 \in [1, 2]$, but it decreases to .5 when x_5 increases past 2. A similar lack of monotonicity holds for the shorth or any other estimate in the shortest half-range of the sample.

Other 50% breakdown estimates can be constructed by using different measures to identify the "shortest" half-sample (see Butler 1982 for asymptotic properties of these estimators). For one example, let T^* denote the median of the half-sample with the smallest sum of absolute deviations. This estimate has 50% breakdown, it is equivariant, and it is not the same as the median, LMS, or shorth. An illustration is presented in Table 1. It shows how T^* is computed and how it compares with the other estimates. Notice that T^* identifies a different shortest

*Gilbert W. Bassett, Jr., is Associate Professor of Economics, Department of Economics, University of Illinois at Chicago, Box 4348, Chicago, IL 60680. The author would like to thank Peter Rousseeuw and a referee for their helpful comments.

Table 1. 50% Breakdown Estimators

Observation	X_i	Half-subset Observations[a]	Range	Least absolute deviations[b]
1	0	1, 2, 3, 4, 5	5.00 = (5 − 0)[c]	9.80 = (2.5 + 2.4 + 0 + 2.4 + 2.5)
2	.1	2, 3, 4, 5, 6	7.45 = (7.55 − 2.4)	9.95 = (4.8 + 2.4 + 0 + 1 + 2.65)
3	2.5	3, 4, 5, 6, 7	5.05 = (7.55 − 2.5)	7.70 = (2.5 + .1 + 0 + 2.55 + 2.55)
4	4.9	4, 5, 6, 7, 8	5.01 = (9.91 − 4.9)	7.56 = (2.65 + 2.55 + 0 + 0 + 2.36)
5	5.0	5, 6, 7, 8, 9	5.10 = (10.1 − 5)	7.46 = (2.55 + 0 + 0 + 2.36 + 2.55)[c]
6	7.55			
7	7.55			
8	9.91			
9	10.1			
Median	5			
Least median of squares	2.5			
Shorth	2.5 (0 + .1 + 2.5 + 4.9 + 5)/5			
T^*	7.55			

[a]The other half-subsets (e.g., 1, 2, 3, 4, 6) have larger half-subset values.
[b]The sum of absolute deviations from the half-subset median—for example, the median of the first half-subset is 2.5 and 9.80 is the value of the sum of absolute errors on the first half-subset.
[c]Shortest half-subset.

half-subset than the LMS estimate. The estimate is equivariant and has 50% breakdown, but as implied by the following theorem, it cannot be monotonic.

Remark. The 50% breakdown estimates identify a cluster of data that has the shortest length. This makes the estimates behave like modes and explains why monotonicity and 50% breakdown are almost inconsistent with one another. It is easy to construct instances in which the mode is not a monotonic function of the mass of a distribution; moving some mass from around the mode to a larger value moves the mode in the opposite direction. The mode-like feature of 50% breakdown estimates means that with the exception of the median, they cannot satisfy the monotonicity property.

Theorem. The median is the only equivariant, monotonic estimator with 50% breakdown.

Proof. Let T denote an equivariant estimate with 50% breakdown. Any such estimate must satisfy the so-called exact-fit property; that is, if half (or more) of the sample values are equal to k, then the estimate also must be k (see Rousseeuw and Leroy 1987, p. 123). By equivariance k can be taken to be zero, so

$$T(0, \ldots, 0, x_{p+2}, \ldots, x_n) = 0. \tag{1}$$

To see why this must hold, suppose to the contrary, that

$$\sigma T(0, \ldots, 0, x_{p+2}, \ldots, x_n) = \sigma A, \qquad A \neq 0,$$

or, by equivariance,

$$T(0, \ldots, 0, \sigma x_{p+2}, \ldots, \sigma x_n) = \sigma A.$$

The right side of the equality goes to $+\infty$ by letting $|\sigma|$ go to $+\infty$, but this is impossible because the 50% breakdown estimate on the left side with only $[n - (p + 2) + 1] = p + q - 1$ arbitrary arguments must be bounded. Hence any 50% breakdown estimate must satisfy the exact-fit property.

Suppose now that T is also monotone. Let the x_i be in increasing order and use equivariance so that

$$\begin{aligned} T(x_{(1)}, \ldots, x_{(n)}) - x_{(p+1)} &= T(x_{(1)} - x_{(p+1)}, \ldots, x_{(n)} - x_{(p+1)}) \\ &\leq T(0, \ldots, 0, x_{(p+2)} - x_{(p+1)}, \ldots, x_{(n)} - x_{(p+1)}) \\ &= 0. \end{aligned}$$

The inequality follows from monotonicity ($x_{(i)} - x_{(p+1)} \leq 0$ for the first $p + 1$ terms), and the last step is the exact-fit property. In a similar fashion

$$\begin{aligned} T(x_{(1)}, \ldots, x_{(n)}) - x_{(p+q)} &= T(x_{(1)} - x_{(p+q)}, \ldots, x_{(n)} - x_{(p+q)}) \\ &\geq T(x_{(1)} - x_{(p+q)}, \ldots, x_{(p+q-1)} - x_{(p+q)}, 0, \ldots, 0) \\ &= 0. \end{aligned}$$

Combining gives

$$x_{(p+q)} \leq T \leq x_{(p+1)}. \tag{2}$$

This proves the theorem because (a) when $q = 0$ the median is the equivariant estimate in the median interval and (b) when $q = 1$ the inequality (2) already says that T is the median.

Remark. The proof can be easily modified to characterize equivariant and monotonic estimators with $100r$% breakdown: any such estimate must be in between the r and $(1 - r)$ sample quantiles. The proof is left as an exercise. [To prove the result, use the exact-fit property for $100r$% breakdown estimators: If 100 $(1 - r)$% of the sample values are equal to k, then the estimate is equal to k. Complete the proof by mimicking the inequalities used in the $r = 1/2$ version of the theorem.]

CONCLUSION

Monotonicity and 50% breakdown both seem desirable. A nonmonotonic estimator can move in the opposite direction of a perturbed observation, and at least at first glance, this does not seem a good feature for a location estimator. The 50% breakdown property is useful because it protects against discrepant observations. The desirable properties, however, are almost inconsistent with one another. With the single exception of the median, no equivariant estimator can both be monotonic and have 50% breakdown.

[*Received April 1989. Revised January 1990.*]

REFERENCES

Andrews, D. G., Bickel, P. J., Hampel, F. R., Huber, P. J., Rogers, W. H., and Tukey, J. W. (1972), *Robust Estimates of Location: Survey and Advances*, Princeton, NJ: Princeton University Press.

Butler, R. W. (1982), "Nonparametric Interval and Point Prediction Using Data Trimming by a Grubbs-Type Outlier Rule," *Annals of Statistics*, 10, 197–204.

Donoho, D. L., and Huber, P. J. (1983), "The Notion of Breakdown Point." In *A Festschrift for Erich L. Lehmann*, eds. P. J. Bickel, K. A. Doksum, and J. L. Hodges, Jr. Belmont, CA: Wadsworth, pp. 157–184.

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

He, X., Jureckova, J., Koenker, R., and Portnoy, S. (1990), "Tail Behavior of Regression Estimators and Their Breakdown Points," *Econometrica*, 58, 1195–1214.

Rousseeuw, P. J. (1984), "Least Median of Squares Regression," *Journal of the American Statistical Association*, 79, 871–880.

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

A Simple Characterization of Seemingly Unrelated Regressions Models in Which OLS Is BLUE

ROBERT BARTELS and DENZIL G. FIEBIG*

Recent contributions to the discussion about the conditions under which ordinary least squares in the seemingly unrelated regressions (SUR) model is the best linear unbiased estimator suggest a characterization of SUR models that is convenient for checking whether these conditions are satisfied. Standard pedagogic examples for SUR's, as well as more complex cases, are easily derived from our characterization.

KEY WORDS: Equivalence classes; Necessary and sufficient conditions; Translog system.

1. INTRODUCTION

Although the conditions under which ordinary least squares (OLS) in the linear regression model is the best linear unbiased estimator (BLUE) are now well known (see, e.g., Puntanen and Styan 1989), the formulations of these conditions generally fall outside the intuition of most students. For the seemingly unrelated regressions (SUR) model, one particular form of the necessary and sufficient condition under which OLS is BLUE has recently been discussed by Baltagi (1988) and Baksalary and Trenkler (1989). In this article we use this condition to provide a new characterization of the SUR model, in which the cases where OLS is BLUE become quite transparent. This characterization is of pedagogic value to intermediate-level students.

The SUR model proposed by Zellner (1962) consists of a set of standard linear regression equations in which the errors are correlated across equations. Suppose that there are M equations in the set, then for $i = 1, \ldots, M$ we can write the individual equations in data matrix form as

$$y_i = X_i\beta_i + u_i \quad \text{with } E(u_i) = 0, \quad V(u_i) = \sigma_{ii}I. \quad (1)$$

The special feature of the SUR model is that for any pair of equations i and j, the errors may be correlated; that is, $E(u_iu_j') = \sigma_{ij}I$. As a result, efficiency gains over OLS estimation of the individual equations may be possible by treating the equations as a system and using generalized least squares (GLS) estimation. By stacking the y_i, β_i, and u_i vectors and creating a new design matrix with the X_i as blocks along the diagonal, we can write the M equations in system form as (see, e.g., Judge, Griffiths, Hill, Lütkepohl, and Lee, 1985, p. 467)

$$y = X\beta + u \qquad \text{with } E(u) = 0,$$
$$V(u) = \Omega = \Sigma \otimes I, \quad (2)$$

where $y = (y_1', y_2', \ldots, y_M')'$, $u = (u_1', u_2', \ldots, u_M')'$, $\beta = (\beta_1', \beta_2', \ldots, \beta_M')'$, $X = \text{diag}(X_1, X_2, \ldots, X_M)$, and $\Sigma = (\sigma_{ij})$ is the $M \times M$ matrix of variances and covariances among the equations. The symbol $\otimes$ denotes the Kronecker matrix product. Since in general the variance–covariance matrix for the model is not of the form σ^2I, OLS will generally not be BLUE, so for efficient estimation GLS is required. There are some well-known examples, however, where OLS *is* BLUE; in particular, this is the case (a) if all the covariances between the

*Robert Bartels and Denzil G. Fiebig are Senior Lecturers, Department of Econometrics, University of Sydney, New South Wales 2006, Australia. This article was written while Robert Bartels held a Visiting Professorship at Bonn University. Financial assistance from the von Humboldt Foundation is gratefully acknowledged.