Journal of Econometrics 44 (1990) 107–126. North-Holland

BOUNDED-INFLUENCE ESTIMATORS FOR THE TOBIT MODEL*

Franco PERACCHI

University of California, Los Angeles, CA 90024, USA

This paper introduces a class of bounded-influence estimators for the Gaussian censored-regression or Tobit model. These estimators can be interpreted as weighted ML estimators, with weights chosen to attain the best trade-off between efficiency and robustness. An empirical example illustrates the feasibility and usefulness of these estimators, as well as their performance vis-à-vis the Tobit ML, CLAD, and SCLS estimators.

1. Introduction

This paper introduces a class of robust estimators for the Gaussian censored-regression (or Tobit) model. It is well known that the Tobit ML estimator is not robust, being very sensitive to small departures from the model assumptions. Recently, several semiparametric estimators have been proposed in the literature. Examples include Powell's (1984) censored least absolute deviation (CLAD) estimator and Powell's (1986) symmetrically censored least squares (SCLS) estimator. These estimators have certain robustness properties, but can be very inefficient, for they disregard entirely the information contained in the parametric assumptions. The estimators presented in this paper provide a compromise between efficiency and robustness, for they make use of parametric assumptions, thereby attaining high efficiency at the Tobit model, but are robust in Hampel's (1971) sense, that is, their probability distribution changes only little under small changes in the underlying probability distribution of the observations. These estimators will be referred to as 'optimal bounded-influence estimators', for they have a bounded-influence function [Hampel (1974)], which ensures protection against the effects of small departures from the Tobit assumptions, and attain the best trade-off between efficiency and robustness. They can all be interpreted as weighted ML estima-

*This paper is based on a chapter of my dissertation at Princeton University. I am very grateful to my advisors, Angus Deaton, Whitney Newey, and Elvezio Ronchetti, for their generous guidance. I also thank Anne Case for kindly providing the data. Earlier versions of this paper were presented at the 1988 Winter Symposium of the Econometric Society (Venice, Italy) and the 3rd Congress of the European Economic Association.

tors, the exact form of the weights depending on the particular choice of the efficiency and robustness criteria.

So far, bounded-influence estimation has been largely confined to the linear-regression model [see, e.g., Hampel (1978), Krasker (1980) Krasker and Welsch (1982), and Hampel et al. (1986)]. A notable exception is the bounded-influence estimator proposed by Stefanski, Carroll, and Ruppert (1986) for the Logit model. This paper provides another example of the feasibility and usefulness of bounded-influence estimation outside the context of the linear-regression model. As an illustration, we estimate Engel curves using Sudanese household budget data with a nonnegligible fraction of reported zero expenditures. We compare several bounded-influence estimators with the Tobit, CLAD, and SCLS estimators. In particular, we address the following questions: Is the Tobit model consistent with the data? Do more robust estimators lead to different conclusions than Tobit and why? What are the differences between semiparametric and bounded-influence estimators? What diagnostic information is provided by the various methods.

Our findings may be summarized as follows. The joint hypothesis of normality and censored-regression specification is often at odds with the data. The Tobit estimates are very sensitive to a few extreme observations and look way off in some cases. Bounded-influence, CLAD, and SCLS estimates are close to each other and look more reliable. Bounded-influence estimates, however, appear to be more precise than CLAD and SCLS. Finally, bounded-influence weights provide useful diagnostic information for identifying potential sources of model failure, in particular outliers and influential observations.

The rest of the paper is organized as follows. Section 2 briefly discusses parametric and semiparametric estimation of the censored-regression model. Section 3 introduces the class of bounded-influence estimators for the Tobit model. Section 4 contains the empirical results. Section 5 offers some conclusions.

2. Parametric and semi-parametric estimators for censored regression

Let $z_n = (y_n, x_n')'$ be a vector of observations on $k+1$ variables, with y_n restricted to be nonnegative. A common statistical model for the relationship between y_n and x_n is the censored-regression model

$$y_n = \max\{0, x_n'\beta_0 + \sigma_0 r_n\}, \qquad n = 1, \ldots, N, \tag{1}$$

where r_n is an unobservable disturbance assumed to be independent of x_n, $\beta_0 \in \mathbb{R}^k$ is a vector of unknown regression parameters, and $\sigma_0 \in (0, \infty)$ is an unknown scale parameter. When the disturbances are assumed to be independently and identically distributed (i.i.d.) as N(0, 1), model (1) is known as the Tobit model and the resulting ML estimator as the Tobit estimator.

In general, the expected value of the likelihood score for the Tobit model is equal to zero only when the disturbances in (1) are Gaussian and homoskedastic, and so the Tobit estimator is generally inconsistent when these distributional assumptions are violated. The available Monte Carlo evidence [see, e.g., Arabmazar and Schmidt (1981, 1982) and Goldberger (1983)] shows that the bias of the Tobit estimator under heteroskedasticity or nonnormality can be quite serious, particularly when the scale parameter is unknown and the degree of censoring is high. This lack of robustness of the Tobit estimator motivates two lines of research in the literature. The first is concerned with testing the normality assumption. The second with consistent estimation of the regression parameter β_0 under weak distributional assumptions.

The normality assumption may be tested in several ways. One approach is to use graphical methods based on some nonparametric estimator of the error distribution, such as the Kaplan–Meier estimator [Chesher, Lancaster, and Irish (1985)]. A more formal approach is to nest the normal distribution in a larger parametric family and then construct a standard score test of the restrictions implied by normality [see, for example, Bera, Jarque and Lee (1984), and Ruud (1984)]. Although designed against one specific alternative, tests of this type have power against a variety of misspecification alternatives. A third approach is to construct general specification tests based on the comparison of two estimators that are both consistent at the assumed model, but have different probability limits when the model is misspecified. The various tests differ in the choice of what estimators to compare. For example, Nelson (1981) compares a consistent estimator of the covariance of x_n and y_n with the efficient estimator based on the assumption of Gaussian disturbances, while Ruud (1984) compares the Probit and Tobit estimators of the normalized parameter β_0/σ_0. All these tests can be interpreted as conditional moment tests [Newey (1985)]. They are not specifically designed to test the normality assumption and may lack power against certain alternatives.

Knowledge of the parametric form of the error distribution is not necessary for consistent estimation of β_0. For example, if the errors in (1) are only assumed to be i.i.d., β_0 can be estimated consistently jointly with the unknown error distribution [see, e.g., Buckley and James (1979), Duncan (1986), Fernandez (1986) and Horowitz (1986)]. Little is known, however, about the sampling distribution of the proposed estimators. Another possibility is to impose restrictions on the conditional distribution of the disturbances that are weaker than a parametric assumption but still sufficient to construct $\sqrt{N}$-consistent and asymptotically normal estimators of β_0. Two such estimators are Powell's (1984) censored least absolute deviation (CLAD) estimator and Powell's (1986) symmetrically censored least squares (SCLS) estimator. The CLAD estimator, defined as

$$\hat{\beta}_{\text{CLAD}} = \underset{\beta}{\operatorname{argmin}} \sum_{n=1}^{N} |y_n - \max\{0, x_n'\beta\}|,$$

is consistent under the assumption that the conditional error distribution has median zero (homoskedasticity is not required). Estimating the CLAD asymptotic variance matrix, however, involves nonparametric estimation of the error density at the median. The SCLS estimator, defined as

$$\hat{\beta}_{\mathrm{SCLS}} = \underset{\beta}{\operatorname{argmin}} \sum_{n=1}^{N} \Big[(y_n - \max\{y_n/2, x_n'\beta\})^2 + 1(y_n > 2x_n'\beta)\big[(y_n/2)^2 - (\max\{0, x_n'\beta\})^2\big]\Big],$$

where 1(.) denotes the usual indicator function, is consistent under the somewhat stronger assumption that the conditional error distribution is symmetric about zero (again homoskedasticity is not necessary). Neither estimator requires a knowledge of the scale parameter σ_0. Monte Carlo evidence [Paarsch (1984), Powell (1986)] indicates that both estimators can be very inefficient relative to the ML estimator based on a correctly specified model. This raises the question of whether too much information is ignored in order to attain consistency under very weak distributional assumptions.

3. Bounded-influence estimators for the Tobit model

The behavior of a statistical procedure should be investigated not just at the assumed model, but also under small departures from the model assumptions. Some kind of 'stability of behavior' is necessary, because the assumed model need not be exactly true, no matter how weak its assumptions are. In the case of model (1), for example, normality and independence between the errors and the regressors may fail because of a few gross-errors in the data, and linearity of the underlying regression relationship may fail for extreme values of the regressors. Further, it may be difficult to assess the exact nature of the model misspecification, and therefore it may not be clear what corrective actions are appropriate. In this kind of situations, which appear to arise frequently in empirical work, it may be sensible to consider statistical procedures that are reasonably efficient when the model is correctly specified, while being robust, that is, not too sensitive to small violations of the model assumptions.

3.1. Qualitative and B-robustness

A rigorous definition of robustness has been proposed by Hampel (1971), and formalizes the notion that an estimator $\hat{\theta}_N$, indexed by the sample size N, is robust if small changes in the probability distribution of the observations have only small effects on the probability distribution of $\hat{\theta}_N$. More precisely, let $\mathscr{L}_F(\hat{\theta}_N)$ denote the distribution function (d.f.) of $\hat{\theta}_N$ when F is the underlying d.f. of the observations. Then, the sequence of estimators $\{\hat{\theta}_N\}$ is called qualitatively robust at the d.f. F if, for large enough N, the mapping

$F \to \mathscr{L}_F(\hat{\theta}_N)$ is continuous at F with respect to the topology of weak convergence.

A simple characterization of qualitative robustness can be obtained in the case of M- (or generalized ML) estimators, defined as roots of 'estimating equations' of the form

$$\sum_{n=1}^{N} \eta(z_n, \theta) = 0,$$

where the function $\eta(z, \theta)$ is called the score function associated with the given estimator. Tobit, CLAD, and SCLS are all members of this class. The Tobit estimator corresponds to choosing $\eta(z, \theta) = s(z, \theta)$, where $s(z, \theta)$ denotes the Tobit likelihood score and $\theta = (\beta', \sigma)'$. The CLAD and SCLS estimators correspond to choosing, respectively, $\eta(z, \beta) = 1(x'\beta > 0)[\text{sign}(y - x'\beta)]x$ and $\eta(z, \beta) = 1(x'\beta > 0)[\min(y - x'\beta, x'\beta)]x$.

An M-estimator $\hat{\theta}_N$ can generally be represented, at least in large samples, as a function $\hat{\theta}(\cdot)$ defined on a suitable space of d.f.'s, that is, $\hat{\theta}_N$ is asymptotically equivalent to $\hat{\theta}(F_N)$, where F_N denotes the empirical d.f. of the observations. Let $F_{\varepsilon, z}$ denote the mixture, with mixing probabilities $1 - \varepsilon$ and ε, of the d.f. F and another d.f. with mass concentrated at the point z. The influence function (IF) of $\hat{\theta}$ at F, denoted by $IF(z, \hat{\theta}, F)$, is defined to be the limit, as $\varepsilon \to 0$, of the normalized difference $[\hat{\theta}(F_{\varepsilon, z}) - \hat{\theta}(F)]/\varepsilon$ [Hampel (1974)]. The IF provides an approximation to the asymptotic bias of $\hat{\theta}_N$, as an estimator of $\hat{\theta}(F)$, under a small contamination of the d.f. F by a point mass distribution centered at z. The sup-norm of an estimator's IF, called the estimator's sensitivity, provides a natural quantitative measure of robustness. An estimator with a finite sensitivity is called a bounded-influence or B-(bias-) robust estimator. For an M-estimator with a continuous IF, B-robustness and qualitative robustness are equivalent [see, e.g., Hampel et al. (1986)]. Further, since the IF and the score function of an M-estimator are related by a nonsingular linear transformation [see, e.g., Serfling (1980)], B-robustness is equivalent to the score function being bounded, and qualitative robustness to the score function being bounded and continuous.

The Tobit estimator is not qualitatively robust because the Tobit score is not bounded. The CLAD and SCLS score functions are bounded only if the regressors take values in a bounded set. Even in this restrictive case, however, their sensitivity can be unacceptably high. Further, their score function is not continuous, and so both can be sensitive to rounding or grouping of the observations [see, e.g., Hampel et al. (1986)].

3.2. *Optimal bounded-influence estimators for the Tobit model*

In this section we shall assume that the observations $\{z_n\}$ are i.i.d. and their common d.f. F_0 belongs to the parametric family $\mathscr{F} = \{F_\theta \colon \theta \in \Theta, \Theta = \mathbb{R}^k \times$

$(0, \infty)]$, specified by the Tobit model (1). Thus $F_0 = F_{\theta_0}$ for some $\theta_0 \in \Theta$. We shall exploit this information to construct estimators of θ_0 that are consistent at the Tobit model and cannot be improved upon simultaneously with respect to the criteria of asymptotic efficiency at the assumed model and robustness to small departures from the model assumptions. More precisely, let $\mathscr{T}_c$ denote the class of M-estimators of θ_0 that are consistent and asymptotically normal at the Tobit model and have a sensitivity that does not exceed a given bound c, i.e., $\sup_z \|IF(z, \hat{\theta}, F_\theta)\|_B \leq c$, where $\|x\|_B$ denotes the norm of the vector x in the metric of the p.d. matrix B. Within the class $\mathscr{T}_c$, we seek an estimator that is efficient at the Tobit model according to an asymptotic mean square error (MSE) criterion of the form $\mathrm{MSE}(\hat{\theta}, F_\theta, Q) = \mathrm{trace}[Q\, \mathrm{AV}(\hat{\theta}, F_\theta)]$, where Q is some p.d. matrix and $\mathrm{AV}(\hat{\theta}, F_\theta)$ denotes the asymptotic variance matrix of $\hat{\theta}$ at the d.f. F_θ. Such an estimator, called an optimal bounded-influence estimator, is qualitatively robust if, in addition, its score function is continuous.

Peracchi (1987) provides conditions for the existence of an optimal bounded-influence estimator and characterizes its score function in the case of a general parametric model and arbitrary metrics for the sensitivity and the MSE criterion. The form of the optimal bounded-influence estimator simplifies considerably when the sensitivity and the MSE criterion are both defined in the same metric, that is, $Q = B$. By specializing his results to the Tobit case we obtain the following:

Proposition 1. Suppose that Amemiya's (1973) *conditions for consistency and asymptotic normality of the Tobit estimator are satisfied, and assume that there exists a pair of continuously differentiable functions* $(a(\cdot), P(\cdot))$, *defined on an open set* Θ_0 *containing* θ_0, *such that* $P(\theta)$ *is a p.d. matrix and* $(a(\theta), P(\theta))$ *solve the system of equations*

$$\mathrm{E}_\theta \min\{1, c/\|P^{-1}[s(z,\theta) - a]\|_B\}[s(z,\theta) - a] = 0, \tag{2}$$

$$\mathrm{E}_\theta \min\{1, c/\|P^{-1}[s(z,\theta) - a]\|_B\}[s(z,\theta) - a]s(z,\theta)' - P = 0, \tag{3}$$

where E_θ *denotes expectations taken with respect to* F_θ. *Let* $\tilde{\theta}_N = \tilde{\theta}(F_N)$ *be the* M-*estimator of* θ_0 *based on the score function* $\eta(z,\theta) = w(z,\theta)[s(z,\theta) - a(\theta)]$, *where the function* $w(z,\theta)$ *is defined by*

$$w(z,\theta) = \min\left\{1, \frac{c}{\|P(\theta)^{-1}[s(z,\theta) - a(\theta)]\|_B}\right\}. \tag{4}$$

Then, for any $\theta \in \Theta_0$, $\mathrm{MSE}(\tilde{\theta}, F_\theta, B) \leq \mathrm{MSE}(\hat{\theta}, F_\theta, B)$ *for all estimators* $\hat{\theta}$ *in* $\mathscr{T}_c$. *Moreover,* $N^{1/2}(\tilde{\theta}_N - \theta_0) \xrightarrow{d} N(0, P_0^{-1}Q_0P_0'^{-1})$, *where* $P_0 = (\partial/\partial\theta')\mathrm{E}_0\eta(z,\theta_0)$, $Q_0 = \mathrm{E}_0\eta(z,\theta_0)\eta(z,\theta_0)'$ *and* E_0 *denotes expectations taken with respect to* F_0. *A consistent estimate of the asymptotic variance matrix of* $\tilde{\theta}_N$

is given by $\tilde{P}_N^{-1}\tilde{Q}_N\tilde{P}_N'^{-1}$, *where* $\tilde{P}_N = N^{-1}\Sigma_{n=1}^N(\partial/\partial\theta')\eta(z_n,\tilde{\theta}_N)$ *and* $\tilde{Q}_N = N^{-1}\Sigma_{n=1}^N\eta(z_n,\tilde{\theta}_N)\eta(z_n,\tilde{\theta}_N)'$.

The existence of $\tilde{\theta}_N$ depends on the existence of a solution to the equation system (2)–(3). A necessary condition is the following:

Proposition 2. Suppose that $E_\theta\|s(z,\theta)\|$ *exists and is positive. Then* $a(\theta)$ *and* $P(\theta)$ *exist only if* $c \geq (\text{trace } B)/(E_\theta\|s(z,\theta)\|_B)$.

Proposition 1 defines a whole family of estimators, indexed by the choice of the matrix B and the sensitivity bound c. Clearly, when the bounded-influence constraint is not binding, i.e., $c = \infty, \tilde{\theta}$ is the Tobit estimator. An optimal bounded-influence estimator can be interpreted as a weighted ML estimator, where the weight function $w(z,\theta)$ depends on the matrix B. When B is equal to the identity matrix, $\tilde{\theta}$ is the Tobit analogue of the regression estimator of Hampel (1978) and Krasker (1980). When $B = \text{AV}(\tilde{\theta}, F_\theta)^{-1}$ we obtain the analogue of the regression estimator of Krasker and Welsch (1982). other choices of B will be discussed later. The vector $a(\theta)$ is a bias correction term that ensures consistency of $\tilde{\theta}$ at the Tobit model. Geometrically, the Tobit score for one observation is censored to satisfy the bounded-influence constraint, and shifted to guarantee consistency at the assumed model. Since the Tobit score is continuous, it is clear that $\tilde{\theta}$ is qualitatively robust.

When the distribution of the regressors is unknown, the above results should be interpreted as conditional on the given set of regressors. When any of the assumptions of the Tobit model is violated, $\tilde{\theta}$ is not generally consistent. However, since the *IF* of $\tilde{\theta}$ is bounded, combining the bias and the asymptotic variance of an estimator in some risk function, it is possible to find a neighborhood of the assumed model over which the optimal bounded-influence estimator has smallest asymptotic risk than the Tobit estimator.

It can be shown that tests inherit the efficiency and robustness properties of the estimators on which they are based [see, e.g., Peracchi (1987)]. In particular, tests based on bounded-influence estimators are robust, that is, their level and power are relatively stable under small departures from the model assumptions. This property is not shared by tests based on estimators that do not possess a bounded *IF*. Tests based on optimal bounded-influence estimators are robust and have, in addition, certain optimality properties in terms of asymptotic power.

The difference between Tobit and an optimal bounded-influence estimator $\tilde{\theta}$ can be used to construct a variety of specification tests of the type proposed, among others, by Hausman (1978). This type of tests are likely to be quite powerful, because the difference between the two estimators can be very large when the model is misspecified, but $\tilde{\theta}$ will be only slightly less efficient than the Tobit estimator when the model is correct.

The robust weights (4), computed for each observation in the sample, provide useful diagnostics for detecting outliers and influential observations. In the case of nonlinear estimators, the use of these weights represent an alternative to methods based on deleting a subset of observations at a time and then comparing the resulting estimates with the ones obtained from the full sample [see, e.g., Belsley, Kuh, and Welsch (1980) for the linear-regression case]. Since the robust weights are jointly computed with the parameter estimates, they require no additional calculation. Further, they are easy to interpret, because of the weighted ML nature of an optimal bounded-influence estimator.

The computation of $\tilde{\theta}$ may be quite expensive, but considerable simplifications can be obtained by exploiting the arbitrariness of the metric in which the sup-norm of the *IF* is defined. Here we propose two possibilities. The first is to choose $B = P(\theta)^2$, where $P(\theta)$ was defined in Proposition 1. Although not very natural, this metric is convenient from the point of view of computation, since it eliminates the need of solving for the matrix P at each iteration. However, the resulting estimator is not invariant under a reparameterization of the model. One choice that leads to invariance is $B = P(\theta)J(\theta)^{-1}P(\theta)$, where $J(\theta)$ is the information matrix associated with the parametric model F_θ. The resulting weight function, which is also computationally simple, rescales the recentered likelihood score whenever its norm, in the metric of the inverse information matrix, is greater than the given bound c. The estimators based on these two choices of weight function will be denoted by *BI1* and *BI2*, respectively. We shall also consider the estimator based on a score function of the form $\eta(z,\theta) = w(z,\theta)[s(z,\theta) - a(\theta)]$, where the weight function $w(z,\theta)$ is given by

$$w(z,\theta) = \min\{1, c/\|s(z,\theta)\|\},$$

Table 1

Bounded-influence estimators for the censored regression model. All bounded-influence estimators in this paper are based on a score function of the form

$$\eta(z,\theta) = \min\left\{1, \frac{c}{\|A(\theta)[s(z,\theta) - b(\theta)]\|}\right\}[s(z,\theta) - a(\theta)].^{a}$$

Estimator[b]	$b(\theta)$	$A(\theta)$	Metric on the *IF*
BI0	0	I	—
BI1	$a(\theta)$	I	$P(\theta)^2$
BI2	$a(\theta)$	$J(\theta)^{-1/2}$	$P(\theta)J(\theta)^{-1}P(\theta)$
H-K	$a(\theta)$	$P(\theta)^{-1}$	I
K-W	$a(\theta)$	$Q(\theta)^{-1/2}$	$P(\theta)Q(\theta)^{-1}P(\theta)$

[a] The $p \times 1$ vector $a(\theta)$ and the $p \times p$ matrix $P(\theta)$ are solutions to eqs. (2)–(3) in the text, and $Q(\theta) = \mathrm{E}_\theta \eta(z,\theta)\eta(z,\theta)'$.

[b] H-K = Hampel-Krasker type, K-W = Krasker-Welsch type.

and $a(\theta) = E_\theta w(z,\theta)s(z,\theta)$. This estimator, denoted by *BIO*, does not have any optimality property but is simple to compute because $w(z,\theta)$ does not depend on $a(\theta)$. Since the *BIO* estimator has a bounded *IF* and can be shown to be consistent at the Tobit model, it should provide good starting values for one-step versions of an optimal-bounded influence estimator. The method of Bickel (1975) can then be used to show that these one-step estimators are asymptotically equivalent to the fully iterated estimators. Table 1 summarizes the score function for each of the estimators introduced in this section.

4. Empirical application

The censored-regression model is frequently used to analyze the income–expenditure relationship when household budget data contain a significant fraction of reported zero expenditures. In this section we analyze household budget data from the Sudan. A comparison between the Tobit, CLAD, SCLS, and several bounded-influence estimators is carried out.

4.1. The data and the fitted models

The data are taken from the 1978–80 Household Income and Expenditure Survey of the Sudan, and were chosen as an example of the type of low-quality data that are often used by economists. The data are contaminated in various ways, including misreporting by individual households, coding and punching errors, data manipulation at the editing stage, etc., but the actual amount of contamination is unknown. To keep the model as simple as possible, we only consider the subset of 268 observations from the Nile region. We estimate Engel curves for three commodities with a nonnegligible fraction of reported zero expenditures, namely clothing and footwear ('clothing'), transport services and repairs ('transport'), and tobacco products ('tobacco'). The degree of censoring differs for the various commodities and is equal to 8.2% for clothing, 23.9% for transport, and 31.7% for tobacco.

We consider a number of popular models of Engel curves for an individual commodity i:

$$w_i = a_i + b_i \ln x \qquad \text{[Working-Leser (WL)]},$$

$$w_i = a_i + b_i \ln x + d_i (\ln x)^2 \qquad \text{[Quadratic Working-Leser (QWL)]},$$

$$p_i q_i = a_i + b_i x \qquad \text{[Linear Expenditure System (LES)]},$$

$$p_i q_i = a_i + b_i x + d_i x^2 \qquad \text{[Quadratic Expenditure System (QES)]},$$

where w_i and $p_i q_i$ denote, respectively, the budget share and the total expenditure on the ith commodity, and x denotes the total outlay. These Engel curves are all theory-consistent, that is, each of them can be derived by Shephard's Lemma from some nice cost function. QWL and QES may be interpreted as second-order approximations, based, respectively, on powers of $\ln x$ and of x, to an arbitrary Engel curve.

Demographic and area effects are introduced in the analysis by expressing income in per capita terms, and by assuming that for each model the intercept a_i depends linearly on a number of household characteristics, including the household size, a household composition effect (number of household members less than 14 years old), and an area dummy with a value of one for households living in rural areas and zero for households living in urban areas. This specification may be restrictive, because demographic and area effects may in principle affect the whole set of parameters.

Definitions and summary statistics for all variables considered are presented in table 2.

Table 2

Definition and summary statistics for the variables in the data set.

Variable[a]	Min	Max	Median	MAD[b]
SHXCLOTH	0.00	22.35	4.0	92.24
SHXTRANS	0.00	30.43	0.97	0.97
SHXTOBAC	0.00	12.27	0.52	0.52
XCLOTH	0.00	52.63	4.00	2.73
XTRANS	0.00	73.53	0.94	0.94
XTOBAC	0.00	19.44	0.34	0.34
XPC	2.88	118.46	12.20	3.73
XPCSQ	8.29	14032.43	148.89	84.30
LXPC	7.97	11.68	9.41	0.31
LXPCSQ	63.45	136.48	88.54	5.77
HHSIZE	2	16	7	2
*LT*14	0	9	3	1

[a]*SHXCLOTH* = % share of total expenditure on clothing and footwear,
SHXTRANS = % share of total expenditure on transport services and repairs,
SHXTOBAC = % share of total expenditure on tobacco products,
XCLOTH = household expenditure on clothing and footwear,
XTRANS = household expenditure on transport services and repairs,
XTOBAC = household expenditure on tobacco products,
XPC = total expenditure per household member,
XPCSQ = square of XPC,
LXPC = 1000 * log *XPC*,
LXPCSQ = square of *LXPC*,
HHSIZE = number of household members,
*LT*14 = household members less than 14 years old.

[b]Median absolute deviation from the median.

4.2. Preliminary tests of specification

For each commodity and functional form we first present a number of tests for normality, conditional symmetry of the error distribution, and censored-regression specification. The normality assumption is tested against the general Pearson family using the score test of Bera, Jarque, and Lee (1984) and the specification tests of Nelson (1981) and Ruud (1984). As Cragg (1971) first pointed out, the censored-regression model may not provide an appropriate representation of demand behavior because it does not distinguish between the decision to purchase a good and the decision of how much to purchase. We test the censored-regression specification against Cragg's (1971) Model I by using the score test of Deaton and Irish (1984), and against Cragg's Model II by using the score test of Lin and Schmidt (1984). Both are tests of the joint hypothesis of normality of the error distribution and censored-regression specification. The joint hypothesis of conditional symmetry and censored-

Table 3

Tests for normality, symmetry, and Tobit specification.

	Tests[a]					
	Ruud	Nelson	BJL	L-S	D-I	Newey
			Clothing			
WL	14.1[c]	4.01[b]	4.22[b]	14.1[c]	1.68[b]	6.88[b]
LES	93.2	71.5	36.6	94.6	4.06	19.6
QWL	15.9[c]	7.21[b]	4.78[b]	16.0[c]	1.69[b]	8.73[b]
QES	99.0	127.0	20.2	100.6	0.95[b]	9.14[b]
			Transport			
WL	110.3	40.6	60.8	111.8	10.2	16.4
LES	161.5	80.7	37.8	162.5	12.7	12.0[c]
QWL	118.9	61.0	59.6	119.4	10.6	15.5
QES	171.5	184.3	70.0	178.0	11.9	n/a
			Tobacco			
WL	48.9	11.1[c]	45.8	49.0	6.74	6.39[b]
LES	70.7	23.8	61.5	71.1	7.22	6.84[b]
QWL	50.2	17.3	46.7	50.2	6.80	3.94[b]
QES	75.5	82.0	54.5	77.4	6.37	n/a

[a]Ruud = Ruud (1984),
Nelson = Nelson (1981),
BJL = Bera, Jarque, and Lee (1984),
L-S = Lin and Schmidt 1984),
D-I = Deaton and Irish (1984),
Newey = Newey (1987).
[b]Not significant at the 5% level.
[c]Significant at the 5% but not at the 1% level.

regression specification is tested as in Newey (1987). Since all these tests assume that the Engel curves are correctly specified, they should also have power against misspecification arising from omitted variables or an incorrect functional form. Under the null hypothesis, all test statistics except the Deaton–Irish statistic have an asymptotic χ^2 distribution. The number of degrees of freedom is equal to two for the Bera–Jarque–Lee test, and to the number of regressors for all other tests. The Deaton–Irish statistic has an asymptotic N(0, 1) distribution under the null hypothesis. If the statistic is positive and significantly different from zero, this is evidence against both Tobit and Cragg's Model I.

The various test statistics are reported in table 4. The hypothesis of a Tobit model is rejected in all cases, except the WL form for clothing. Nelson's test tends to reject less frequently than the others. This may be a consequence of its low power, as suggested by Ruud (1984). The Ruud and Lin–Schmidt statistics are very close and always lead to rejection. The Deaton–Irish statistic is always positive, which indicates rejection of the Tobit model but, interestingly, not in the direction of Cragg's Model I. The results for the conditional symmetry hypothesis are mixed, with rejections in the case of transport and the LES form for clothing. All these results indicate that misspecification is likely to be present in most cases. However, it is hard to determine its exact nature, and, in particular, whether it is due to failures of the censored-regression specification, misspecification of the Engel curves, or simply failure of the assumption of Gaussian disturbances.

4.3. Estimation results

We now present the results obtained by estimating the various models using eight different methods: Tobit, the five bounded-influence estimators discussed in section 3, and Powell's CLAD and SCLS estimators. Details on the computations are given in the appendix. The estimated standard errors are consistent under heteroskedasticity and nonnormality. CLAD standard errors are reported for different choices of the bandwidth for the nonparametric estimation of the error density at the median.

Table 4 reports estimates of the income elasticity of demand evaluated at the median income. In the case of clothing, the estimates for the WL, LES, and QWL specifications tend to be clustered around 1.6. Estimated elasticities for the QES specification are higher, particularly in the case of Tobit and SCLS. For a given specification of the Engel curve, CLAD and bounded-influence estimates look very similar.

In the case of transport, estimated elasticities vary significantly depending on the choice of functional form and estimation method. In particular, Tobit gives high point estimates, especially in the QES case. For the WL, LES, and QWL specifications, the Tobit estimates are about 6. The H-K estimates also

Table 4

Income elasticity of demand evaluated at the median (asymptotic standard errors in parentheses).

	Clothing				Transport				Tobacco			
	Wl	LES	QWL	QES	WL	LES	QWL	QES	WL	LES	QWL	QES
Tobit	1.61	1.23	1.71	2.64	5.85	6.73	5.70	13.24	0.620	1.12	0.593	3.27
	(0.122)	(0.475)	(0.248)	(0.362)	(0.745)	(2.67)	(2.67)	(1.77)	(0.660)	(0.815)	(1.34)	(1.67)
BI0	1.54	1.60	1.62	2.08	3.10	3.63	3.46	4.60	−0.025	0.253	−0.73	1.42
	(0.127)	(0.207)	(0.218)	(0.186)	(0.383)	(0.582)	(0.212)	(0.377)	(0.484)	(0.581)	(1.02)	(2.16)
BI1	1.54	1.60	1.62	2.08	3.09	3.63	3.47	4.29	−0.027	0.254	−0.079	1.42
	(0.128)	(0.207)	(0.231)	(0.186)	(0.394)	(0.582)	(0.228)	(0.379)	(0.484)	(0.583)	(1.02)	(2.15)
BI2	1.58	1.62	1.62	2.00	2.97	2.60	3.05	3.09	−0.022	−0.026	−0.119	0.276
	(0.117)	(0.199)	(0.333)	(0.486)	(0.317)	(0.394)	(1.25)	(1.53)	(0.471)	(0.267)	(0.980)	(1.41)
H-K	1.63	1.54	[a]	[a]	4.51	6.99	[a]	[a]	0.375	0.817	[a]	[a]
	(0.118)	(0.318)			(0.569)	(2.95)			(0.544)	(0.877)		
K-W	1.59	1.71	1.63	2.09	3.20	3.30	3.21	4.19	0.031	0.314	0.022	1.59
	(0.119)	(0.186)	(0.370)	(0.763)	(0.347)	(0.459)	(1.27)	(1.09)	(0.495)	(0.654)	(1.05)	(2.14)
SCLS	1.55	0.948	1.66	3.25	3.68	2.38	3.21	[a]	−1.42	−0.084	−3.38	[a]
	(0.127)	(0.368)	(0.124)	(0.531)	(0.895)	(1.40)	(0.561)		(0.853)	(1.30)	(3.05)	
CLAD[b]	1.61	1.39	1.61	2.11	3.13	2.97	3.14	4.69	0.170	0.010	0.245	0.655
	(0.082)	(0.710)	(0.099)	(0.319)	(0.532)	(2.41)	(2.17)	(0.915)	(0.716)	(1.52)	(0.741)	(0.690)
	(0.115)	(0.227)	(0.109)	(0.484)	(0.420)	(3.95)	(0.429)	(0.943)	(0.628)	(2.01)	(0.504)	(0.982)
	(0.113)	(0.170)	(0.114)	(0.573)	(0.502)	(1.85)	(0.645)	(1.42)	(0.794)	(0.309)	(0.667)	(0.942)

[a]Algorithm failed to converge.
[b]Standard errors corresponding respectively to $c_0 = 0.5$, 1.0, and 2.0 in eq. (5.5) of Powell (1984).

Table 5

Tests of equality between Tobit and the other estimates of the regression parameters.

BI0	BI1	BI2	H-K	K-W	SCLS	CLAD	
			Clothing				
WL	20.0	19.3	18.3	3.82[a]	15.8	2.02[a]	1.55[a]
LES	98.8	98.9	111.9	19.5	105.1	7.14[a]	19.8
QWL	7.86[a]	7.81[a]	20.4	n/a	16.7	1.03[a]	3.64[a]
QES	30.1	30.1	56.4	n/a	44.7	27.7	18.9
			Transport				
WL	253.9	257.0	228.4	24.5	123.1	23.6	20.7
LES	171.2	175.4	164.5	100.0	86.2	38.4	13.9[b]
QWL	101.3	120.1	203.6	n/a	43.5	32.3	21.9
QES	16.2[b]	17.0	56.6	n/a	13.0[b]	54.4	32.1
			Tobacco				
WL	7.19[a]	7.51[a]	8.98[a]	16.8	13.0[b]	16.9	15.2
LES	7.58[a]	7.69[a]	2.61[a]	24.8	8.13[a]	26.7	30.8
QWL	19.4	19.8	14.3[b]	n/a	12.0[a]	18.7	15.6[b]
QES	20.0	20.2	1.02[a]	n/a	7.90	27.7	21.8

[a]Not significant at the 5% level.
[b]Significant at the 5% but not at the 1% level.

tend to be high, whereas all other estimates tend to fall between 3.0 and 3.5. In the QES case, Tobit gives an estimate of 13.2, which is at least 3 times larger than all other estimates.

For both clothing and transport, elasticities appear to be estimated very precisely. However, SCLS and CLAD standard errors tend to be larger than the bounded-influence ones. CLAD standard errors are also sensitive to the choice of the bandwith for the estimation of the error density at the median.

In the case of tobacco, Tobit again leads to high point estimates, particularly in the QES case. All other point estimates are close to zero or negative. However, elasticities are now estimated very imprecisely in all cases.

Table 5 reports tests of significance for the difference between Tobit and the other estimates of the regression parameter β_0. If the Tobit model is correctly specified, these differences should be small and the tests statistics should have an asymptotic χ^2 distribution with number of degrees of freedom equal to the number of regression parameters. Equality of the regression coefficients is typically rejected for clothing and transport, but not for tobacco, because of the high imprecision of the estimates. In the case of CLAD and SCLS, rejection occurs less frequently, because of the larger standard errors of these estimates.

These results show that using semiparametric (CLAD or SCLS) or bounded-influence estimators can lead to significant differences with respect to

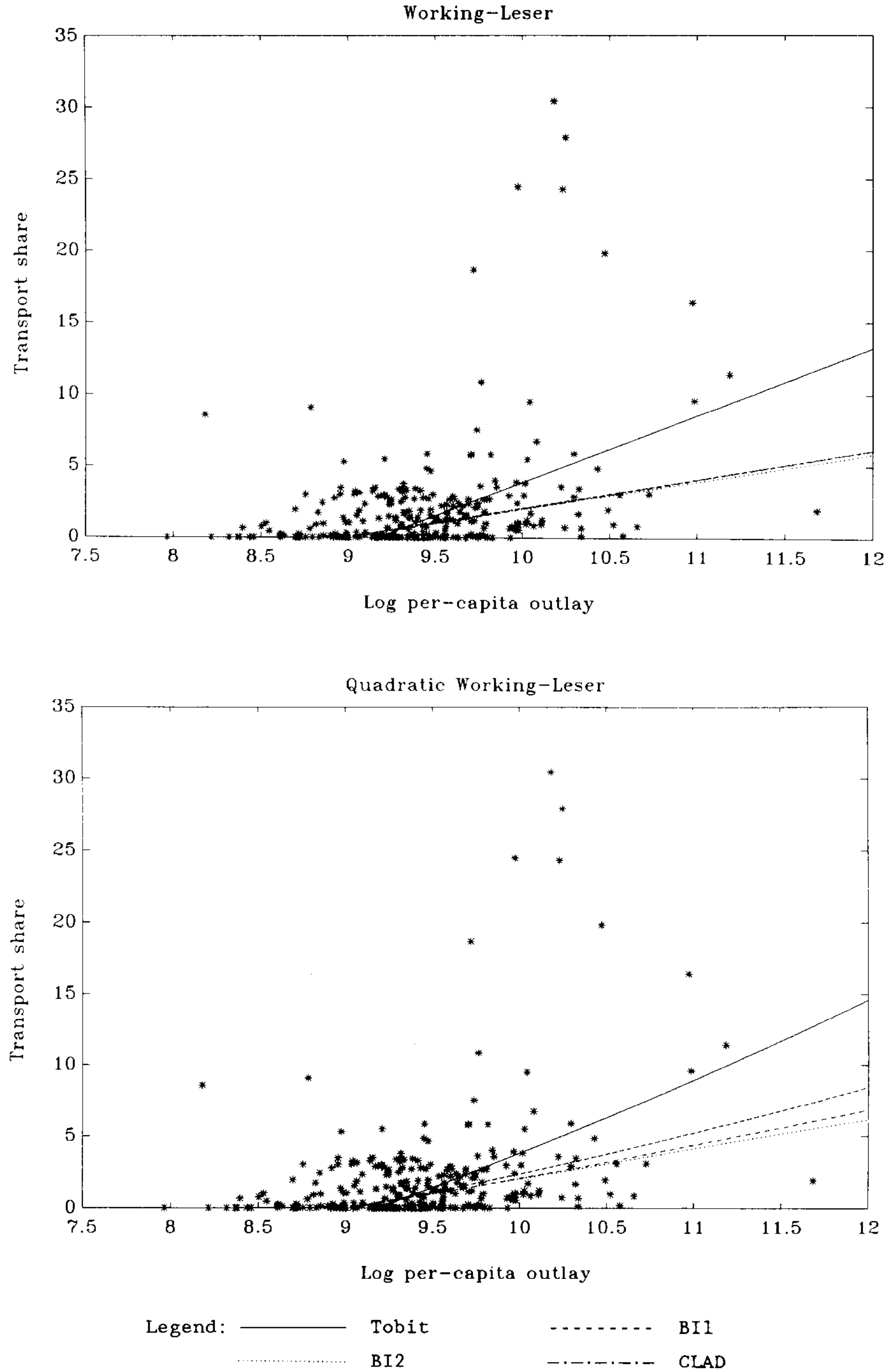

Fig. 1a. Scatter of log per-capita outlay and transport share; estimated Engel curves superimposed.

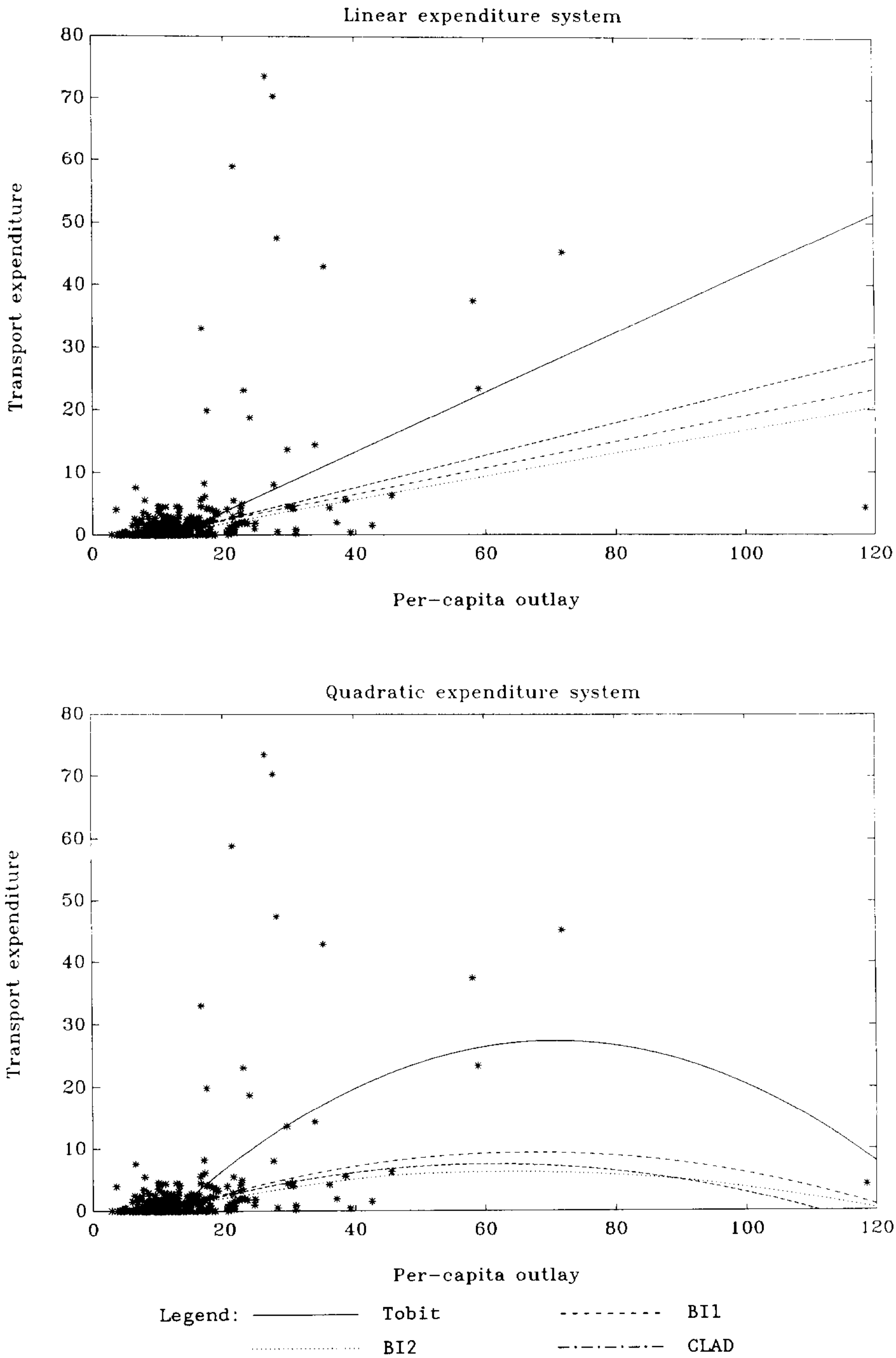

Fig. 1b. Scatter of per-capita outlay and transport expenditure; estimated Engel curves superimposed.

Tobit. Semiparametric and bounded-influence estimates tend to be close, but bounded-influence estimates appear to be more precise. Moreover, the CLAD standard errors appear to be sensitive to the choice of bandwidth. Among the bounded-influence estimators, the relatively simple BI0, BI1, and BI2 estimators proposed in section 3.2 appear to behave very well.

To gain some insight into the reason for the differences between Tobit and the other estimates, it is useful to inspect the scatter of the data. For simplicity, we only consider the demand for transport. Fig. 1a superimposes the estimated Engel curves for Tobit, BI1, BI2, and CLAD to the scatter of the observations on log per-capita outlay and transport share. Fig. 1b does the same for the scatter of the observations on per-capita outlay and transport expenditure. The Engel curves are evaluated for a median household with seven household members, four adults and three children, living in a rural area. In all cases, Tobit Engel curves lie well above the other estimates. This is due to the influence of a small cluster of data points (less than 3% of the total sample) characterized by medium–high income and a budget share on transport that exceeds 15%. The influence exercized by this cluster is particularly strong for the quadratic specifications, QWL and QES. Inspection of the robust weights (4) shows that this cluster of points is heavily downweighted by the bounded-influence estimators. In addition, these estimators heavily downweigh the outlier on the bottom right of the scatter. The leverage exercized by this point on the Tobit estimates is not very strong in the WL, LES, and QWL cases, but is responsible for the high curvature of the estimated income–expenditure relationship in the QES case. The fact that, for this sample, outliers tend to be concentrated in the center of the income distribution explains why bounded-influence and semiparametric estimates are similar. Both methods tend to downweigh the influence of these points, but only bounded-influence estimators are explicitly designed to deal with outliers that are also high leverage points.

5. Conclusions

The results of section 4 demonstrate the feasibility of bounded-influence estimation outside the context of the linear-regression model. It shows that Engel curves estimated from the same set of censored data can differ significantly depending on the choice of the estimation technique. In particular, the Tobit estimates can differ significantly from other more robust estimates as a consequence of the presence of only a small fraction of extreme observations. We found that semiparametric and bounded-influence estimates tend to be close to each other, but the latter appear to be more precise. It would be interesting to verify these indications with a full-scale Monte Carlo study.

In our view, the class of bounded-influence estimators discussed in this paper offers several advantages. First, it ensures protection against the negative effects, on both estimation and inference, of small departures from the assumed parametric model, while maintaining high efficiency if the assumed model is correctly specified. Second, the differences with respect to the Tobit estimates provide the basis for specification tests that have power against a variety of alternatives. Third, the weights from bounded-influence estimation provide useful diagnostics for detecting outliers and influential observations. The price one has to pay by using these estimators is a loss of efficiency with respect to the Tobit estimator if the Tobit model is indeed correct. However, and this is yet another advantage, the investigator can choose the efficiency loss that he/she is willing to tolerate.

Appendix

Computation of the bounded-influence estimates proceeds as follows:

(1) Start with $\theta^{(0)} = \hat{\theta}_{\mathrm{ML}}$, $A^{(0)} = I_p$ and $a^{(0)} = b^{(0)} = 0$.

(2) For the BI0 and BI1 estimators put $A^{(1)} = I_P$, for the BI2 estimator put $A^{(1)} = J(\theta^{(0)})^{-1/2}$, for the H-K estimator put $A^{(1)} = P(\theta^{(0)})^{-1}$ where

$$P(\theta^{(0)}) = \frac{1}{N} \sum_{n=1}^{N} \min\left\{1, \frac{c}{\|A^{(0)}[s(z_n, \theta^{(0)}) - b^{(0)}]\|}\right\} \times \left[s(z_n, \hat{\theta}^{(0)}) - a^{(0)}\right] s(z_n, \hat{\theta}^{(0)})',$$

for the K-W estimator put $A^{(1)} = Q(\theta^{(0)})^{-1/2}$ where

$$Q(\theta^{(0)}) = \frac{1}{N} \sum_{n=1}^{N} \min\left\{1, \frac{c}{\|A^{(0)}[s(z_n, \theta^{(0)}) - b^{(0)}]\|}\right\}^2 \times \left[s(z_n, \hat{\theta}^{(0)}) - a^{(0)}\right]\left[s(z_n, \hat{\theta}^{(0)}) - a^{(0)}\right],$$

(3) Given $A^{(1)}$, compute $a^{(1)}$ as

$$a^{(1)} = \left[\sum_{n=1}^{N} \mathrm{E}_{\Phi} \min\left\{1, \frac{c}{\|A^{(1)}[\sigma_n(r, \theta^{(0)}) - b^{(0)}]\|}\right\}\right]^{-1} \times \sum_{n=1}^{N} \mathrm{E}_{\Phi} \min\left\{1, \frac{c}{\|A^{(1)}[\sigma_n(r, \theta^{(0)}) - b^{(0)}]\|}\right\} \sigma_n(r, \hat{\theta}^{(0)}),$$

where

$$\sigma_n(r, \theta) = \begin{pmatrix} [1(y_n > 0) r - 1(y_n = 0)\lambda(x_n'\alpha)] x_n \\ 1(y_n > 0)(\gamma^{-1} - r y_n) \end{pmatrix},$$

and $\alpha = \beta/\sigma$ and $\gamma = 1/\sigma$. The normal integrals are evaluated numerically using the Gauss–Legendre subroutine in Quandt (1988).

(4) Put $b^{(1)} = 0$ for the BI0 estimator and $b^{(1)} = a^{(1)}$ otherwise.

(5) Given $A^{(1)}$, $a^{(1)}$, and $b^{(1)}$, compute $\theta^{(1)}$ by solving

$$\sum_{n=1}^{N} \min\left\{1, \frac{c}{\|A^{(1)}[s(z_n, \theta) - b^{(1)}]\|}\right\}[s(z_n, \theta) - a^{(1)}] = 0.$$

This is done by using the Newton–Raphson algorithm NEWRAP in GQOPT.

(6) Given $\theta^{(1)}$, compute $A^{(2)}$, $a^{(2)}$, $b^{(2)}$, and $\theta^{(2)}$ as in step (2) to (5) and iterate. Convergence of this algorithm is not guaranteed.

The sensitivity bound c is chosen so as to obtain an average weight of about 95%, that is, $N^{-1}\Sigma_{n=1}^{N} w(z_n, \tilde{\theta}_n) \simeq 0.95$, where $w(z, \theta)$ is given by (4). When $c = \infty$, all bounded-influence estimators that we consider are the same as the ML estimator, with an average weight equal to unity. Thus, our choice of the sensitivity bound may be interpreted as resulting in an efficiency loss of about 5% when the Tobit model is indeed correct. The % of downweighted observations varies depending on the specification and, to a lesser extent, the type of estimator. Typically is between 10% and 15% for the WL and QWL forms, and is somewhat lower for LES and QES. In the latter case, however, the value of the minimum weight is much smaller, which indicates the presence of highly influential observations.

The convergence criterion requires the maximal change in any of the parameter estimates to be less than 10^{-4}. Convergence is typically attained after 5 to ten iterations. We had numerical problems with the H-K estimator, in particular for the QWL and QES specifications, and we do not report results for these two cases. For the other bounded-influence estimators, sometimes the algorithm cycled between two values very close to each other. In these cases convergence was always reached by weakening the tolerance to 10^{-3}.

The CLAD estimates are computed by iteratively reweighted LS with weight function given by $w(y, x, \beta) = 1(x'\beta)\min\{|y - x'\beta|^{-1}, \varepsilon^{-1}\}$, where ε is positive and small. The SCLS estimates are computed by the iterative LS algorithm suggested in Powell (1986). The convergence criterion requires the maximal change in any of the parameter estimates to be less than 10^{-5}. SCLS estimates typically need more iterations to converge. In a few cases the limit of 100 iterations was reached without convergence.

References

Amemiya, T., 1973, Regression analysis when the dependent variable is truncated normal, Econometrica 42, 759–762.

Arabmazar, A. and P. Schmidt, 1981, Further evidence on the robustness of the Tobit estimator to heteroskedasticity, Journal of Econometrics 17, 253–258.

Arabmazar, A. and P. Schmidt, 1982, An investigation of the robustness of the Tobit estimator to non-normality, Econometrica 50, 1055–1063.

Belsley, D.A., E. Kuh, and R.E. Welsch, 1980, Regression Diagnostics: Identifying influential data and sources of collinearity (Wiley, New York, NY).

Bera, A.K., C.M. Jarque, and L.F. Lee, 1984, Testing for the normality assumption in limited dependent variable models, International Economic Review 25, 563–578.

Bickel, P.J., 1975, One-step Huber estimates in the linear model, Journal of the American Statistical Association 70, 428–434.

Chesher, A., T. Lancaster, and M. Irish, 1985, On detecting the failure of distributional assumptions, Annales de l'INSEE 59/60, 7–45.

Cragg, J.G., 1972, Some statistical models for limited dependent variables with applications to the demand for durable goods, Econometrica 39, 829–844.

Deaton, A. and M. Irish, 1984, Statistical models for zero expenditures in household budgets, Journal of Public Economics 23, 59–80.

Goldberger, A., 1983, Abnormal selection bias, in: S. Karlin, T. Amemiya, and L.A. Goodman, eds., Studies in econometrics, time series and multivariate statistics (Academic Press, New York, NY).

Hampel, F.R., 1971, A general qualitative definition of robustness, Annals of Mathematical Statistics 42, 1887–1896.

Hampel, F.R., 1974, The influence curve 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, DC).

Hampel, F.R., E. Ronchetti, P.J. Rousseeuw, and W.A. Stahel, 1986, Robust statistics: The approach based on influence functions (Wiley, New York, NY).

Hausman, J.A., 1978, Specification tests in econometrics, Econometrica 46, 1251–1272.

Krasker, W.S., 1980, Estimation in linear regression models with disparate data points, Econometrica 48, 1333–1346.

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

Lin, T.F. and P. Schmidt, 1984, A test of the Tobit specification against an alternative suggested by Cragg, Review of Economics and Statistics 66, 174–177.

Nelson, F.D., 1981, A test for misspecification in the censored normal model, Econometrica 49, 1317–1329.

Newey, W.K., 1985, Maximum likelihood specification testing and conditional moment tests, Econometrica 53, 1047–1069.

Newey, W.K., 1987, Specification tests for distributional assumptions in the Tobit model, Journal of Econometrics 34, 125–145.

Paarsch, H.J., 1984, A Monte Carlo comparison of estimators for censored regression models, Journal of Econometrics 24, 197–213.

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

Powell, J.L., 1984, Least aboslute deviations estimation for the censored regression model, Journal of Econometrics 25, 303–325.

Powell, J.L., 1986, Symmetrically trimmed least squares estimation for Tobit models, Econometrica 54, 1435–1460.

Quandt, R.E., 1988, Econometrics of disequilibrium (Blackwell, New York, NY).

Ruud, P.A., 1984, Tests of specification in econometrics, Econometric Review 3, 211–242.

Stefanski, L.A., R.J. Carroll, and D. Ruppert, 1986, Optimally bounded score functions for generalized linear models with applications to logistic regression, Bometrika 73, 413–424