

BOUNDED INFLUENCE ESTIMATORS FOR THE CENSORED REGRESSION MODEL

By

Franco Peracchi*

Department of Economics
University of California
Los Angeles, CA

July 1988

Working Paper No. 487

* This paper is based on a Chapter of my dissertation at Princeton University. I am very grateful to my advisors, Angus Deaton, Whitney Newey for kindly providing the data. An earlier version of this paper was presented at the 1988 Winer Symposium of the Econometric Society (Venice, Italy).

Abstract

This paper introduces a class of bounded-influence estimators for the normal censored regression 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, CLAD and SCLS estimators.

JEL Classification No. 211

Keywords: Bounded influence; Censored regression; Engel curves; Influential observations; Robustness.

1. Introduction

This paper introduces a class of robust estimators for the normal censored regression model. It is well known that the normal ML (or Tobit) estimator is very sensitive to small departures from the model assumptions, and therefore is not robust. Recently, several semi-parametric 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 some robustness properties, but they 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 assumed 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)] and they attain the best trade-off between efficiency and robustness. They can all be interpreted as weighted ML estimators, 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 non-negligible 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 normal censored regression model consistent with the data? Do more robust estimators lead to different conclusions than Tobit and why? What are the differences between semi-parametric 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. 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 sources of model failures, in particular outliers and influential observations.

The rest of the paper is organized as follows. Section 2 presents a variety of estimators for the censored regression model. Section 3 introduces the class of bounded-influence estimators. Section 4 contains the empirical results. Section 5 summarizes the conclusions.

2. Alternative 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 non-negative. A common statistical model for the relationship between y_n and x_n is the censored regression model

$$y_n = \max(0, x_n' \beta_0 + \sigma_0 r_n), \quad n = 1, \dots, N, \quad (1)$$

where r_n is an unobservable disturbance, assumed to be distributed independently 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. The model is often estimated under the assumption that the disturbances are independently and identically distributed (i.i.d.) as $N(0, 1)$. The resulting ML estimator is sometimes called the Tobit estimator. Table 1 gives the likelihood score function for this case. For convenience, the model has been reparameterized by putting $\alpha = \beta/\sigma$ and $\gamma = 1/\sigma$.

The expected value of the normal likelihood score for $\theta = (\alpha', \gamma)'$ is not zero in general, unless the disturbances in (1) are normal and homoskedastic, and so the Tobit estimator is not generally consistent when these distributional assumptions are violated. The bias of the Tobit estimator under heteroskedasticity and non-normality has been investigated in a few, simple cases [see e.g. Hurd (1979), Arabmazar and Schmidt (1981) and (1982), Goldberger (1983)]. A general finding is that it can be seriously biased, particularly when the scale parameter is unknown and the degree of censoring is high. Most studies, however, only consider the problem of estimating the mean of a population. Further, they restrict attention to symmetric distributions and do not investigate explicitly the

relationship between the bias and the tail behavior of the error distribution. This is unfortunate, because the bias depends also on the distribution of the regressors, and is likely to be more severe for non-symmetric or thick-tailed distributions.

Notice that ML estimators based on non-normal distributions need not suffer of the same problems as the Tobit estimator. For example, Table 1 also presents the likelihood score function for the case when the disturbances in (1) have a common Laplace (double-exponential) distribution. It is easy to verify that the conditional expectation of the Laplace likelihood score is equal to zero, hence the Laplace ML estimator is consistent, whenever the conditional distribution of the disturbances has median zero (homoskedasticity is not required) and the marginal distribution of the regressors gives no probability mass to points for which $x'\beta_0 < 0$.

The normality assumption may be tested in several ways. One 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. For example, Bera, Jarque and Lee (1984) nest the normal into the Pearson family of distributions, but other choices of nesting family are possible [see e.g. Ruud (1984)]. Although designed against a specific alternative, tests of this type have power against a variety of misspecification alternatives. A second 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 and y with the efficient estimator based on the

assumption of normal disturbances. Ruud (1984) compares the Probit and Tobit estimators of the normalized regression parameter α . Chesher, Lancaster and Irish (1985) compare alternative estimators of the information matrix for the normal model, as suggested by White (1982). All these tests can be interpreted as conditional moment tests [Newey (1985)]. They are not specifically designed to test normality and may lack power against certain alternatives. A third approach is to use graphical methods based on some non-parametric estimator of the error distribution, such as the Kaplan-Meier estimator [Chesher, Lancaster and Irish (1985)].

Several estimators that are consistent under weak distributional assumptions have been proposed in the literature. One approach involves joint estimation of the regression parameters and the distribution of the disturbances in (1) [see e.g. Buckley and James (1979), Duncan (1986), Fernandez (1986) and Horowitz (1986)]. All the proposed estimators require homoskedastic errors and their asymptotic distribution has not yet been established. Moreover, their computation is not a trivial problem. Another approach is to apply the method of moments to derive estimators that are consistent and asymptotically normal for a broad class of distributions. Here we consider two of these estimators.

Powell's (1984) censored least absolute deviation (CLAD) estimator is consistent and asymptotically normal provided that the conditional error distribution has median zero (homoskedasticity is not required). Estimating its asymptotic variance matrix, however, requires estimating the error density at the median. Powell's (1986) symmetrically censored least squares (SCLS) estimator is consistent and asymptotically normal under the somewhat stronger assumption that the conditional error distribution is symmetric about zero (again homoskedasticity is not

required). Both the CLAD and the SCLS estimators, as well as the Tobit and Laplace ML estimators, are M-estimators, that is, 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. The CLAD score function is equal to

$$\eta(z, \beta) = 1(x'\beta > 0) \text{ sign}(y - x'\beta) x,$$

where $1(A)$ denotes the indicator function of the event A . Notice that the CLAD score is similar but not equal to the Laplace likelihood score. The SCLS score function is equal to

$$\eta(z, \beta) = 1(x'\beta > 0) [\min(y, 2x'\beta) - x'\beta] x.$$

Interestingly, neither the CLAD nor the SCLS estimators require a knowledge of the scale parameter σ_0 .

Monte Carlo evidence [Paarsch (1984), Powell (1986)] indicates that both the CLAD and SCLS estimator 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 general conditions.

3. Optimal bounded-influence estimators for the Tobit model

It is important to investigate the behavior of a statistical procedure 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 the assumptions on which it is based. 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 measures 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.

In the econometric literature the notion of robustness is often used in a loose, heuristic way. Here we depart from this practice by adopting the definition proposed by Hampel (1971), which 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 $L_F(\hat{\theta}_N)$ 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 $F \rightarrow L_F(\hat{\theta}_N)$ is continuous at F with respect to the topology of weak

convergence.

In the case of M-estimators, qualitative robustness is easy to characterize. An estimator $\hat{\theta}_N$ can generally be represented, at least in large samples, as a functional of the empirical d.f. of the observations F_N , that is, $\hat{\theta}_N = \hat{\theta}(F_N)$. Let $F_{\epsilon, z}$ denote the mixture, with mixing probabilities equal to $1 - \epsilon$ and ϵ respectively, of the d.f. F and another d.f. with mass concentrated at the point z . Then the influence function (IF) of $\hat{\theta}_N$ at the d.f. F , denoted by $IF(z, \hat{\theta}, F)$, is defined to be the limit as $\epsilon \rightarrow 0$ of the difference $[\hat{\theta}(F_{\epsilon, z}) - \hat{\theta}(F)]/\epsilon$ [Hampel (1974)]¹. The IF can be used to approximate the asymptotic bias of $\hat{\theta}_N$, as an estimator of $\hat{\theta}(F)$, that may arise under small departures from the d.f. F [see e.g. Serfling (1980) and Fernholz (1983)]. If $\hat{\theta}_N$ is an M-estimator with associated score function $\eta(z, \theta)$, then its IF at the d.f. F is a non-singular linear transformation of $\eta(z, \hat{\theta}(F))$ [see e.g. Serfling (1980)]. Further, an M-estimator is qualitatively robust if and only if its IF is bounded and continuous [see e.g. Huber (1981)]. A natural quantitative measure of robustness is therefore provided by the sup-norm of an estimator's IF, called the estimator's sensitivity. An estimator with a finite sensitivity is called a bounded-influence estimator.

The Tobit estimator is clearly not a bounded-influence estimator. Indeed, one large disturbance or one gross-error in the data are enough to completely spoil the Tobit estimates. The Laplace ML, CLAD and SCLS estimators have some robustness properties, because all have a bounded IF when the regressors take values in a compact set. Even in this restrictive case, however, their sensitivity may be unacceptably high. Moreover, since their IF is not continuous, they can all be sensitive to rounding or grouping of the observations [see e.g. Hampel et al. (1986)].

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 model $F = (F_\theta: \theta \in \Theta, \Theta = \mathbb{R}^k \times (0, \infty))$, specified by the censored regression model (1) with normal disturbances². 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 under the normality assumption and cannot be improved upon simultaneously with respect to the criteria of asymptotic efficiency at the assumed model and robustness, that is, stability of performance under small departures from the model assumptions.

More precisely, let T_c be the class of M-estimators of θ_0 that are regular, i.e. possess an asymptotically normal distribution, and have a sensitivity that does not exceed a given bound c , i.e. $\sup_z \|IF(z, \hat{\theta}_N, F_\theta)\|_B \leq c$, where B is a given p.d. matrix³. Consider the problem of finding in T_c an estimator that is consistent and asymptotically efficient at the assumed model F according to a given asymptotic mean square error (MSE) criterion. 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. We shall restrict attention to the case when the sensitivity and the MSE criterion are both defined in the metric of the p.d. matrix B . Let the symmetric p.d. matrix $P(\theta)$ and the vector $a(\theta)$ be solutions to the equation system

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

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

where $s(z, \theta)$ is the likelihood score function of the assumed parametric model. We first give a necessary condition for $a(\theta)$ and $A(\theta)$ to exist. All proofs are presented in Appendix 1.

PROPOSITION 1: Suppose that $E_{\theta} \|s(z, \theta)\|$ exists and is positive. Then $a(\theta)$ and $P(\theta)$ exist only if $c \geq (\text{trace } B)/(E_{\theta} \|s(z, \theta)\|_B)$.

The next result characterizes the class of optimal bounded-influence estimators.

PROPOSITION 2: Assume that $a(\theta)$ and $A(\theta)$, implicitly defined by equations (2) and (3), exist for all θ in an open neighborhood of θ_0 . Let $\bar{\theta}_N$ be the regular 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

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

Then $\bar{\theta}_N$ minimizes $\text{trace } [B AV(\hat{\theta}_N, F_{\theta})]$ among all regular M-estimators $\hat{\theta}_N$ that satisfy $\sup_z \|IF(z, \hat{\theta}_N, F_{\theta})\|_B \leq c$.

Proposition 2 defines a whole class 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. $c = \infty$, $\bar{\theta}_N$ is the ML estimator for the given parametric model. The 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, $\tilde{\theta}_N$ is the Tobit analogue of the regression estimator of Hampel (1978) and Krasker (1980). When $B = AV(\tilde{\theta}_N, F_\theta)^{-1}$ 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 depends on the assumed model and ensures that $\tilde{\theta}_N$ is consistent for θ at F_θ . Geometrically, the likelihood score for one observation is shrunk to satisfy the bounded-influence constraint, and shifted to ensure consistency at the assumed model.

We now establish formally the asymptotic properties of $\tilde{\theta}_N$ when the assumed parametric model is the normal censored regression model. Since the score function (2) is not differentiable at the points where $\|P(\theta)^{-1} [s(z, \theta) - a(\theta)]\|_B = c$, standard techniques based on a Taylor's expansion cannot be applied. Asymptotic normality can however be established by verifying the sufficient conditions of Huber (1967).

PROPOSITION 3: 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 a compact set Θ_0 containing θ_0 in its interior, such that $P(\theta)$ is a p.d. matrix and $(a(\theta), P(\theta))$ solve (2)-(3) for all $\theta \in \Theta_0$. Then the optimal bounded-influence estimator $\tilde{\theta}_N$ exists and is consistent, and $N^{1/2}(\tilde{\theta}_N - \theta_0) \xrightarrow{d} N(0, P_0^{-1} Q_0 P_0^{-1})$, where $P_0 = (\partial/\partial\theta')$ $E_0 \eta(z, \theta_0)$ and $Q_0 = E_0 \eta(z, \theta_0) \eta(z, \theta_0)'$. A consistent estimate of the asymptotic variance matrix of $\tilde{\theta}_N$ is given by $\bar{P}_N^{-1} \bar{Q}_N \bar{P}_N^{-1}$, where $\bar{P}_N = N^{-1} \sum_{n=1}^N (\partial/\partial\theta') \eta(z_n, \tilde{\theta}_N)$ and $\bar{Q}_N = N^{-1} \sum_{n=1}^N \eta(z_n, \tilde{\theta}_N) \eta(z_n, \tilde{\theta}_N)'$.

The next result summarizes the robustness properties of $\tilde{\theta}_N$.

PROPOSITION 4: The optimal bounded-influence estimator $\tilde{\theta}_N$ is qualitatively robust at F . Moreover, $\sup_z \|\text{IF}(z, \tilde{\theta}_N, F)\|_B < \infty$ for all d.f. F and all B .

When the distribution of the regressors is unknown, the above results can be interpreted as conditional on the given set of regressors. When any of the assumptions of the normal censored regression model F is violated, $\tilde{\theta}_N$ will not generally be consistent. However, since the IF of $\tilde{\theta}_N$ is bounded, its asymptotic bias under any small departures from F will also be bounded. Thus, if the bias and the asymptotic variance of an estimator are combined together in some risk function, it will be possible to find a neighborhood of the assumed model F over which the optimal bounded-influence estimator has smallest asymptotic risk than the Tobit estimator [see e.g. Bickel (1984) and Hampel et al. (1986)].

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 have a bounded IF. Tests based on optimal bounded-influence estimators are robust and have, in addition, certain optimality properties.

The difference between Tobit and an optimal bounded-influence estimator $\tilde{\theta}_N$ can be used to construct a variety of specification tests of the type proposed, among others, by Hausman (1978). Tests of this type are

likely to be quite powerful, because the difference between the two estimators can be very large when the model is misspecified, but $\bar{\theta}_N$ will be only slightly less efficient than the Tobit estimator when the model is correctly specified.

The set of robust weights (4), computed for each observation in the sample, provide useful diagnostics for detecting outliers and influential observations. In the case of non-linear 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 $\bar{\theta}_N$ 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 $B = P(\theta)^2$. 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 sensitivity measure 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 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 denote by B10 the estimator based on a score function of the form $\eta(z, \theta) = w(z, \theta) [s(z, \theta) - a(\theta)]$, where $a(\theta) = E_{\theta} w(z, \theta) s(z, \theta)$ and the weight function $w(z, \theta)$ is given by

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

The B10 estimator is simple to compute because $w(z, \theta)$ does not depend on $a(\theta)$. It is easy to verify that the B10 estimator is consistent at the normal censored regression model, has a bounded IF and is asymptotically normal. Therefore, it should provide good starting values for one-step versions of the H-K and K-W estimators. The method of Bickel (1975) can be used to show that these one-step estimators are asymptotically equivalent to the fully iterated estimators. Table 2 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 ⁴. In this Section we analyze household budget data from the Sudan. A comparisons 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 ⁵. The original data set contains observations from

different regions of the Sudan but, 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 3 commodities with a non-negligible 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 :

$$\text{Working-Leser (WL):} \quad w_i = a_i + b_i \ln x$$

$$\text{Quadratic Working-Leser (QWL):} \quad w_i = a_i + b_i \ln x + d_i (\ln x)^2$$

$$\text{Linear expenditure system (LES):} \quad p_i q_i = a_i + b_i x$$

$$\text{Quadratic expenditure system (QES):} \quad p_i q_i = a_i + b_i x + d_i x^2$$

where w_i and $p_i q_i$ denote respectively the budget share and the total expenditure on the i -th commodity, and x denotes the total outlay. All Engel curves belong to the general class considered by Gorman (1981), and are all theory consistent in the sense that 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 (DRUR) 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 affect the whole set of parameters.

Definitions and summary statistics for the variables considered are presented in Table 3.

4.2 Preliminary tests of specification

For each commodity and functional form we first consider 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). The specification tests of Nelson (1981) and Ruud (1984) are also considered. The censored regression specification is tested 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). The joint hypothesis of conditional symmetry and censored regression specification is tested as in Newey (1987).

The Deaton-Irish test essentially compares a consistent estimate of $Pr(y > 0)$ with the efficient estimate based on the assumption of a normal censored regression model. The Lin-Schmidt test compares a consistent estimate of $E(xy|y > 0)$ with the efficient estimate for the normal censored regression model. Both tests are related to Nelson's specification test, and have power against a variety of alternatives,

including heteroskedasticity and non-normality.

All tests assume that the Engel curves are correctly specified. Therefore, they should all have power against misspecification arising from omitted variables or an incorrect functional form. All score test statistics are computed in an asymptotically equivalent form as N times the uncentered R^2 in the regression of a column of ones on the likelihood score for the unrestricted model (evaluated at the restricted estimates). 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 2 for the Bera-Jarque-Lee test, and to the number of regressors for all other tests. The Deaton-Irish statistic is asymptotically normal under the null hypothesis. As noted by Deaton and Irish (1984), if the statistic is negative and significantly different from zero, this is evidence against both Tobit and Cragg's Model I.

The various test statistics are presented in Table 4. The hypothesis of a Tobit model is strongly rejected in all cases, except the WL form for clothing. The results for the WL and LES forms are generally consistent with those for QWL and QES. 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 the 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.

Misspecification may also be detected by less formal procedures. Following Chesher, Lancaster and Irish (1985), we use the Tobit residuals

to compute the Kaplan-Meier estimate of the d.f. function of the errors. If the model is correctly specified the Kaplan-Meier estimate is consistent. Therefore, the plot of its inverse normal transform against the ordered Tobit residuals should be close to a 45-degree line. In agreement with our earlier findings, there is some evidence for normality only in the case of the WL form for clothing [Figures 1a and 1b]. However, even in this case, the Kaplan-Meier estimate has somewhat fatter tails than the normal. In all other cases, especially the expenditure equations, the Kaplan-Meier estimate looks often like the double-exponential distribution [see e.g. the case of the LES form for clothing in Figures 1c and 1d], which makes the CLAD estimator an interesting alternative to consider.

Thus, formal tests and graphical procedures all indicate that misspecification is likely to be present in most of the cases that we consider. However, it is hard to determine the exact nature of the misspecification, and in particular, whether it is due to failures of the censored regression specification or simply to failures of the normality assumption.

4.3 Point estimates and standard errors

Here we present the results obtained for 8 different estimators: the Tobit estimator, the 5 bounded-influence estimators discussed in Section 3, and Powell's CLAD and SCLS estimators. Details on the computation are given in Appendix 2.

For all estimators, the estimated standard errors are consistent under heteroskedasticity and non-normality. For the CLAD covariance estimates we consider different window widths for the non-parametric

estimator of the error density at the origin. The estimates are very sensitive to the degree of smoothing. We report results for three choices, corresponding respectively to setting $c_0 = 0.5, 1.0$ and 2.0 in equation (5.5) of Powell (1984).

Table 6 presents estimates of the income elasticity of demand evaluated at the median income. First consider the WL and LES forms. In the case of clothing estimates do not change much across models and estimation techniques, and are rather precise. In the case of transport the differences between Tobit and all other estimates tend to be large. Semi-parametric and bounded-influence estimates are generally close (with the exception of the H-K estimator, perhaps because of the numerical problems encountered in this case), but bounded-influence estimates tend to be more precise. In the case of tobacco again we find large differences between Tobit and all other estimators, but all estimates are very unprecise.

In the QWL case, estimated elasticities are close to the ones for the WL and LES forms. In the QES case, however, the Tobit estimates just blow up. For all goods, the Tobit estimates are at least twice as big as for the other specifications. The SCLS estimates are not reported because of divergence of the algorithm. The other estimates are also larger than for all previous specifications, but the increase is much less pronounced.

Our results are illustrated in Figure 2, that compares the shape of estimated Engel curves for transport in the case of a median household with 7 household members, 4 adults and 3 children, living in a rural area. The differences between Tobit and the other estimates are also very big in the case of tobacco, but are smaller in the case of clothing.

Specification tests based on the difference between Tobit and

bounded-influence estimates satisfy the conditions for powerful tests [see e.g. Ruud (1984)], namely a large difference between estimators under the alternative model, and a relatively efficient alternative estimator to ML. On the other hand, tests based on the difference between the Tobit and the CLAD or SCLS estimator satisfy the first but not the second conditions and should therefore be less powerful.

The specification test statistics are presented in Table 7. The test statistics are computed as N times the uncentered R^2 in the regression of a column of ones on the likelihood score and the influence function for the subset of regression parameters, both evaluated at the Tobit estimates. Under the null hypothesis of correct specification these statistics have an asymptotic χ^2 distribution with the number of degrees of freedom equal to the number of regression parameters. We find this artificial regression form more convenient than Hausman (1978) original form, because the covariance matrix for the difference between the two contrasts, even when constrained to be positive semi-definite, as suggested in Newey (1985), is typically singular.

Equality of the regression coefficients is typically rejected for clothing and transport, but not for tobacco. In the case of CLAD and SCLS rejection occurs less frequently, essentially because of the larger standard errors of these estimates ⁶.

4.4 Diagnostics for outliers and influential observations

A standard approach to identifying influential observations is based on deleting a subset of observations at a time and then comparing the resulting estimates with the ones obtained for the full sample. A subset

of observations is deemed to be influential if this difference is large. Usually only methods based on deleting a single observation are applied, because of the combinatorial problems arising with multiple deletion. This approach is straightforward for linear estimators [see e.g. Belsley, Kuh and Welsch (1980)], but even single deletion methods can be quite expensive for non-linear estimators, in particular when the number of observations is high. Another approach is to examine, for a given estimator, the norm of the IF for each observation [see e.g. Cook and Weisberg (1982) for the case of regression estimators]. An influential observation is one for which the norm of the IF is large. Asymptotically, this is equivalent to deleting one observation at a time and then computing the norm of the difference in the estimates with respect to the full sample. Neither approach is entirely satisfactory for outlier detection. On the one hand, it is well known that single deletion methods can fail to reveal multiple outliers [see e.g. Atkinson (1986)]. On the other hand, outliers may not be detected by methods that are based on estimators that are not robust.

As an illustration of the latter problem, consider the scatter of the relationship between log per-capita income and budget share on transport [Figure 3]. The symbol associated with each point in the scatter depends on the norm of the IF of the Tobit estimator, with the IF evaluated at the Tobit estimates. Two very influential points are clearly revealed. Notice that the norm of the IF is not very large for the points in the cluster on the top-right of the scatter. These points correspond to households with an exceptionally high expenditure share on transport (15% or more). By computing the Tobit estimates with and without these observations, it is easy to verify that it is their presence which explains why ML estimates

of elasticity are so large, especially in the QES case.

An alternative to the previous approaches is to consider the weights from bounded-influence estimation. An influential observation is now one which receives a small weight. Figure 4 is a graphical illustration of the use of these weights. The scatter of the log per-capita income and the budget share on transport is presented again, but now the symbol associated with each observation depends on the magnitude of the robust weights. Notice that the two very influential observations of Figure 3 are heavily downweighted, but so are now the points in the cluster on the top-right of the scatter. Figure 4 also illustrates how the different bounded-influence estimators work. In particular, the H-K estimator downweights more the former set of points, while the other bounded-influence estimators downweight more the latter.

5. Conclusions

This example demonstrates 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 semi-parametric and bounded-influence estimates tend to be close to each other, but the latter appear to be more precise and lead to tests that appear to be more powerful. It would be interesting to

verify these indications with a full scale Monte Carlo study.

In our view, using 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 difference with respect to Tobit estimates provides 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 ML if the assumed 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.

References

- Amemiya, T. (1985), *Advanced Econometrics*, Harvard University Press, Cambridge, Mass.
- Arabmazar and Schmidt (1981) "Further evidence on the robustness of the Tobit estimator to heteroskedasticity", *Journal of Econometrics*, 17, 253-258.
- Arabmazar and Schmidt (1982), "An investigation of the robustness of the Tobit estimator to non-normality", *Econometrica*, 50, 1055-1063.
- Atkinson, A.C. (1986), "Masking unmasked", *Biometrika*, 73, 533-541.
- Belsley, D.A., Kuh, E. and Welsch, R.E. (1980), *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*, Wiley, New York.
- Bera, A.K., Jarque, C.M. and Lee, L.F. (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.
- Blundell, R.W. and Meghir, C. (1987), "Bivariate alternatives to the univariate Tobit model", *Journal of Econometrics*, 33, 179-200.
- Chesher, A., Lancaster, T. and Irish, M. (1985), "On detecting the failure of distributional assumptions", *Annales de l'INSEE*, 59/60, 7- 45.
- Cook, R.D. and Weisberg, S. (1982), *Residuals and Influence in Regression*, Chapman and Hall, London.
- Cragg (1972), "Some statistical models for limited dependent variables with applications to the demand for durable goods", *Econometrica*, 39, 829-844.
- Deaton, A. and Irish, M. (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.

- Gorman (1980), "Some Engel curves", in A.S. Deaton (ed.), *Essays in Theory and Measurement of Consumer Behavior*, Cambridge University Press, Cambridge, U.K.
- 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, 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. (1967), "The behavior of maximum likelihood estimates under non-standard conditions", *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, 221-33.
- Hurd, M. (1979), "Estimation in truncated samples when there is heteroskedasticity", *Journal of Econometrics*, 11, 247-258.
- 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.
- Lin, T.F. and Schmidt, P. (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.
- Powell, J.L. (1984), "Least absolute 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.
- Ruud, P.A. (1984), "Tests of specification in econometrics", *Econometric Reviews*, 3, 211-242.
- Stefanski, L.A., Carroll, R.J. and Ruppert, D. (1986), "Optimally bounded score functions for generalized linear models with applications to logistic regression", *Biometrika*, 73, 413-424.
- White, H. (1982), "Maximum-likelihood estimation of misspecified models", *Econometrica*, 48, 817-838.

¹ Let $\hat{\theta}_N$ denote the value of the estimator for a given sample of N observations, and consider adding to this sample an additional observation z . Let $\hat{\theta}_{N+1}$ denote the value of the estimator for the new sample of $N + 1$ observations. Then, under mild regularity conditions, the IF of the estimator is equal to the limit in probability of the normalized difference $N (\hat{\theta}_{N+1} - \hat{\theta}_N)$.

² In what follows, E_0 denotes expectations taken with respect to the true d.f. F_0 , and E_θ denotes expectations taken with respect to F_θ .

³ $\|x\|_B = (x' B x)^{1/2}$ denotes the norm of a vector x in the metric of some p.d. matrix B . The Euclidean norm of x is simply denoted by $\|x\|$.

⁴ Cragg (1971) first pointed out that the censored (and truncated) regression model may not be a valid representation of demand, because it does not distinguish between the decision of purchasing a good and the decision of how much to purchase. See also Blundell and Meghir (1987).

⁵ The data set was 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.

⁶ Interestingly, equality of the whole parameter vector is almost always rejected in the case of bounded-influence estimators. This partly reflects the fact that ML estimates of the scale parameter are usually larger and less precise than bounded-influence estimates:

Appendix 1

PROOF OF PROPOSITION 1: See Peracchi (1987), Proposition 2.3.2.

PROOF OF PROPOSITION 2: See Peracchi (1987), Proposition 2.3.3. See also Hampel et al. (1986), Thm. 4.3.1, for the case when B is the identity matrix.

PROOF OF PROPOSITION 3: Under the stated assumptions the function $\eta: Z \times \Theta_0 \rightarrow \mathbb{R}^p$ exists, $\eta(\cdot, \theta)$ is measurable for each $\theta \in \Theta_0$ and $\eta(z, \cdot)$ is continuous on Θ_0 for each z . Since $\eta(z, \theta)$ is a bounded function, $\lambda(\theta) = E_0 \eta(z, \theta)$ exists for all $\theta \in \Theta_0$. Moreover, $\lambda(\theta)$ is continuous, and the equation $\lambda(\theta) = 0$ has a unique root at $\theta = \theta_0$ because the matrix $(\partial/\partial\theta') \lambda(\theta)$ is p.d. for all $\theta \in \Theta_0$, and is equal to $P(\theta)$ if the model is correctly specified. Therefore the function $S(\theta) = \|\lambda(\theta)\|^2$ is continuous on the compact set Θ_0 and attains a unique minimum of zero at $\theta = \theta_0$.

The estimator $\tilde{\theta}_N$ may equivalently be obtained by minimizing the function $S_N(\theta) = \|\lambda_N(\theta)\|^2$, where $\lambda_N(\theta) = N^{-1} \sum_{n=1}^N \eta(z_n, \theta)$. Clearly $S_N(\theta)$ is measurable and continuous. By the Weak Law of Large Numbers, $S_N(\theta)$ converges in probability to $S(\theta)$ for each fixed θ . We now verify that this convergence is uniform in $\theta \in \Theta_0$. Let $g(z, \theta)$ be defined on $Z \times \Theta_0$ by $g(z, \theta) = \eta(z, \theta) - \lambda(\theta)$. Then $g(\cdot, \theta)$ is measurable for each $\theta \in \Theta_0$ and $g(z, \cdot)$ is continuous on Θ_0 for each z . Moreover $E_0 g(z, \theta) = 0$, and $E_0 \sup_{\theta \in \Theta_0} \|g(z, \theta)\|$ is finite because $\eta(z, \theta)$ is a bounded function. It then follows from Theorem 4.2.1 of Amemiya (1985) that $\lambda_N(\theta)$ converges to $\lambda(\theta)$ in probability uniformly in $\theta \in \Theta_0$, and so the convergence of $S_N(\theta)$ to $S(\theta)$ is uniform in $\theta \in \Theta_0$. Consistency of $\tilde{\theta}_N$ then follows from Theorem 4.1.1 of Amemiya (1985).

We shall establish the asymptotic normality of $\tilde{\theta}_N$ by verifying the sufficient conditions of Huber (1967). These take care of the fact that $\eta(z, \theta)$ is not differentiable with respect to θ at points where $\|P(\theta)^{-1} [s(z, \theta) - a(\theta)]\|_B = c$. Measurability and separability of $\eta(\cdot, \theta)$ [Huber's Condition (N-1)] are easily verified. Condition (N-2) is satisfied because θ_0 is a root of $\lambda(\theta) = 0$. Since $\eta(z, \theta)$ is bounded, the trace of $E_0 \eta(z, \theta)$

$\eta(z, \theta)'$ is finite and so Condition (N-4) is also satisfied. Thus, we only have to verify Conditions (N-3), namely that there exist strictly positive numbers a_1, a_2, a_3 and d_0 such that

$$(i) \quad \|\lambda(\theta)\| \geq a_1 \|\theta - \theta_0\| \quad \text{for } \|\theta - \theta_0\| \leq d_0,$$

$$(ii) \quad E_0 u(z, \theta, d) \leq a_2 d \quad \text{for } \|\theta - \theta_0\| + d \leq d_0, \quad d \geq 0,$$

$$(iii) \quad E_0 u(z, \theta, d)^2 \leq a_3 d \quad \text{for } \|\theta - \theta_0\| + d \leq d_0, \quad d \geq 0,$$

where $u(z, \theta, d) = \sup_{\|\tau - \theta\| \leq d} \|\eta(z, \tau) - \eta(z, \theta)\|$. Part (i) is satisfied because $(\partial/\partial\theta') \lambda(\theta_0)$ is non-singular. Next notice that $\eta(z, \theta) \leq s(z, \theta) - a(\theta)$ for all z and $\theta \in \Theta_0$. Therefore, by the differentiability of $s(z, \cdot)$ and $a(\cdot)$,

$$\begin{aligned} \eta(z, \tau) - \eta(z, \theta) &\leq s(z, \tau) - s(z, \theta) - (a(\tau) - a(\theta)) \\ &\leq (\partial/\partial\theta') [s(z, \theta_0) - a(\theta_0)] (\tau - \theta) \end{aligned}$$

for all $\tau \in \Theta_0$ and $\theta \in \Theta_0$ such that $\|\tau - \theta\| \leq d$ and $\|\theta - \theta_0\| \leq d_0 - d$, where d is small and positive. The Lipschitz conditions (ii) and (iii) are easy to verify by taking norms and expectations, and using the fact that the Information matrix for the Tobit model is p.d. Since all conditions of Theorem 3 in Huber (1967) are satisfied, $\bar{\theta}_N$ is asymptotically normal with asymptotic variance matrix equal to $P_0^{-1} Q_0 P_0^{-1}$, where $P_0 = (\partial/\partial\theta') \lambda(\theta_0)$ and $Q_0 = E_0 \eta(z, \theta_0) \eta(z, \theta_0)'$. Finally, consistency of the proposed estimate of the asymptotic variance matrix of $\bar{\theta}_N$ follows immediately from Thm. 3.2. of White (1982).

PROOF OF PROPOSITION 4: The likelihood score function $s(\cdot, \theta)$ for the normal censored regression model is continuous. Thus, the weight function $w(\cdot, \theta)$ and therefore $IF(\cdot, \bar{\theta}_N, F_\theta)$ are also continuous. Since $IF(\cdot, \bar{\theta}_N, F_\theta)$ is bounded by construction, the conclusions of the Proposition follow immediately.

Appendix 2

Computation of the bounded-influence estimates proceeds as follows:

1. Start with $\theta^{(0)} = \hat{\theta}_{ML}$ and $a^{(0)} = 0$.
2. Choose $A^{(1)} = I_P$ for the B10 and B11 estimators, $A^{(1)} = J(\theta^{(0)})^{-1/2}$ for the B12 estimator, $A^{(1)} = P(\theta^{(0)})^{-1}$ for the H-K estimator and $A^{(1)} = Q(\theta^{(0)})^{-1/2}$ for the K-W estimator [$P(\theta^{(0)})$ is the solution to (11) for a given $a^{(0)}$, and $Q(\theta)$ is defined in Proposition 1].
3. Given $A^{(1)}$, compute $a^{(1)}$ as

$$a^{(1)} = \left[\sum_n E_{\Phi} \min \left\{ 1, \frac{c}{\|A^{(1)} [s_n(r, \theta^{(0)}) - b^{(0)}]\|} \right\} \right]^{-1} \\ \times \sum_n E_{\Phi} \min \left\{ 1, \frac{c}{\|A^{(1)} [s_n(r, \theta^{(0)}) - b^{(0)}]\|} \right\} s(r, \hat{\theta}^{(0)})$$

where, from Table 1,

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

$b^{(0)} = 0$ for the B11 estimator, and $b^{(0)} = a^{(0)}$ otherwise. The normal integrals are evaluated numerically by using the Gauss-Legendre subroutine in Quandt (1988).

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

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

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

5. Given $\theta^{(1)}$, compute $A^{(2)}$, $a^{(2)}$, $b^{(2)}$ and $\theta^{(2)}$ as in Step 2 to 4, 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%. 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 10 iterations of the outer loop. 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 $w(y, x, \beta) = 1(x'\beta) \min(|y - x'\beta|^{-1}, \epsilon^{-1})$, where ϵ is positive and small. The SCLS estimates are computed by the iterative LS algorithm mentioned in Powell (1986). The convergence criterion requires the maximal change in any of the parameter estimates to be less than 10^{-5} . CLAD estimates typically need more iterations to converge. In a few cases the limit of 100 iterations was reached without convergence.

Table 1

Likelihood score for the censored regression model ($r = \gamma y - x'\alpha$).

a) Normal disturbances

$$s(z, \theta) = \begin{pmatrix} [1(y > 0) r - 1(y = 0) \lambda(x'\alpha)] x \\ 1(y > 0) (\gamma y - \gamma^{-1}) \end{pmatrix}$$

where $\lambda(u) = \phi(u)/[1 - \Phi(u)]$.

b) Laplace disturbances

$$s(z, \theta) = \begin{pmatrix} [1(y > 0) \text{sign}(r) - 1(y = 0) \lambda(x'\alpha)] x \\ 1(y > 0) [\text{sign}(r) \cdot y - \gamma^{-1}] \end{pmatrix}$$

where $\lambda(u) = 1(u \geq 0) + 1(u < 0) [2 \exp(-u) - 1]^{-1}$.

Table 2

Bounded-influence estimators for the censored regression model.

All bounded-influence estimators in this paper are based on a score function of the form

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

where

Estimator	$b(\theta)$	$A(\theta)$	Metric on the IF
BIO	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 (Hampel-Krasker)	$a(\theta)$	$P(\theta)^{-1}$	I
K-W (Krasker-Welsch)	$a(\theta)$	$Q(\theta)^{-1/2}$	$P(\theta) Q(\theta)^{-1} P(\theta)$

Note: The $p \times 1$ vector $a(\theta)$ and the $p \times p$ matrix $P(\theta)$ are solutions to equations (2)-(3) in the text, and $Q(\theta) = E_{\theta} \eta(Z, \theta) \eta(Z, \theta)'$.

Table 3

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

SHXCLOTH: % share of total exp. on clothing and footwear.
 SHXTRANS: % share of total exp. on transport services and repairs.
 SHXTOBAC: % share of total exp. on tobacco products.
 XCLOTH : household expenditure on clothing and footwear.
 XTRANS : household expenditure on transport services and repairs.
 XTOBAC : household expenditure on tobacco products.
 LXPC : log of total expenditure per household member.
 LXPCSQ : square of LXPC.
 XPC : total expenditure per household member.
 XPCSQ : square of XPC.
 HHSIZE : number of household members.
 LT14 : household members less than 14 years old.

Variable	Min	Max	Median	MAD (1)
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
LXPC	7.97	11.68	9.41	0.31
LXPCSQ	63.45	136.48	88.54	5.77
XPC	2.88	118.46	12.20	3.73
XPCSQ	8.29	14032.43	148.89	84.30
HHSIZE	2	16	7	2
LT14	0	9	3	1

(1) Median absolute deviation from the median.

Table 4

Tests for normality, symmetry and Tobit specification.

Ruud: Ruud (1984), asymptotically χ^2_5 under H_0 .
 Nelson: Nelson (19881), asymptotically χ^2_5 under H_0 .
 BJL: Bera, Jarque and Lee (1984), asymptotically χ^2_2 under H_0 .
 L-S: Lin and Schmidt (1984), asymptotically χ^2_5 under H_0 .
 D-I: Deaton and Irish (1984), asymptotically $N(0,1)$ under H_0 .
 Newey: Newey (1987), asymptotically χ^2_5 under H_0 .

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

a: Asymptotic p-value greater than .05.

b: Asymptotic p-value between .01 and .05.

c: Not available.

Table 5

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 (.122)	1.23 (.475)	1.71 (2.48)	2.64 (.362)	5.85 (.745)	6.73 (2.67)	5.70 (26.7)	13.24 (1.77)	.620 (.660)	1.12 (.815)	.593 (13.4)	3.27 (1.67)
BIO	1.54 (.127)	1.60 (.207)	1.62 (2.18)	2.08 (.186)	3.10 (.383)	3.63 (.582)	3.46 (21.2)	4.60 (.377)	-.025 (.484)	.253 (.581)	-.073 (10.2)	1.42 (2.16)
BI1	1.54 (.128)	1.60 (.207)	1.62 (2.31)	2.08 (.186)	3.09 (.394)	3.63 (.582)	3.47 (22.8)	4.29 (.379)	-.027 (.484)	.254 (.583)	-.079 (10.2)	1.42 (2.15)
BI2	1.58 (.117)	1.62 (.199)	1.62 (3.33)	2.00 (.486)	2.97 (.317)	2.60 (.394)	3.05 (12.5)	3.09 (1.53)	-.022 (.471)	-.026 (.267)	-.119 (9.80)	.276 (1.41)
H-K	1.63 (.118)	1.54 (.318)	a	a	4.51 (.569)	6.99 (2.95)	a	a	.375 (.544)	.817 (.877)	a	a
K-W	1.59 (.119)	1.71 (.186)	1.63 (3.70)	2.09 (.763)	3.20 (.347)	3.30 (.459)	3.21 (12.7)	4.19 (1.09)	.031 (.495)	.314 (.654)	.022 (10.5)	1.59 (2.14)
SCLS	1.55 (.127)	.948 (.368)	1.66 (.124)	3.25 (.531)	3.68 (.895)	2.38 (1.40)	3.21 (.561)	b	-1.42 (.853)	-.084 (1.30)	-3.38 (3.05)	b
CLAD _c	1.61 (.082)	1.39 (.710)	1.61 (.099)	2.11 (.319)	3.13 (.532)	2.97 (2.41)	3.14 (2.17)	4.69 (.915)	.170 (.716)	0.010 (1.52)	.245 (.741)	.655 (.690)
	(.115)	(.227)	(.109)	(.484)	(.420)	(3.95)	(.429)	(.943)	(.628)	(2.01)	(.504)	(.982)
	(.113)	(.170)	(.114)	(.573)	(.502)	(1.85)	(.645)	(1.42)	(.794)	(.309)	(.667)	(.942)

a: Not computed.

b: Algorithm failed to converge.

c: Standard errors corresponding respectively to $c_0 = 0.5, 1.0$ and 2.0 in equation (5.5) of Powell (1984).

Table 6

Specification tests for the regression parameters.

	BIO	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	c	16.7	1.03 _a	3.64 _a
QES	30.1	30.1	56.4	c	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	c	43.5	32.3	21.9
QES	16.2 _b	17.0	56.6	c	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	c	12.0 _a	18.7	15.6 _b
QES	20.0	20.2	1.02 _a	c	7.90 _a	27.7	21.8

a: Asymptotic p-value greater than .05.

b: Asymptotic p-value between .01 and .05.

c: Not available.

FIGURE 1a

DEP VBL = SHXCLOTH, WORKING-LESER FORM
KAPLAN-MEIER ESTIMATE OF THE OF THE STANDARDIZED RESIDUALS

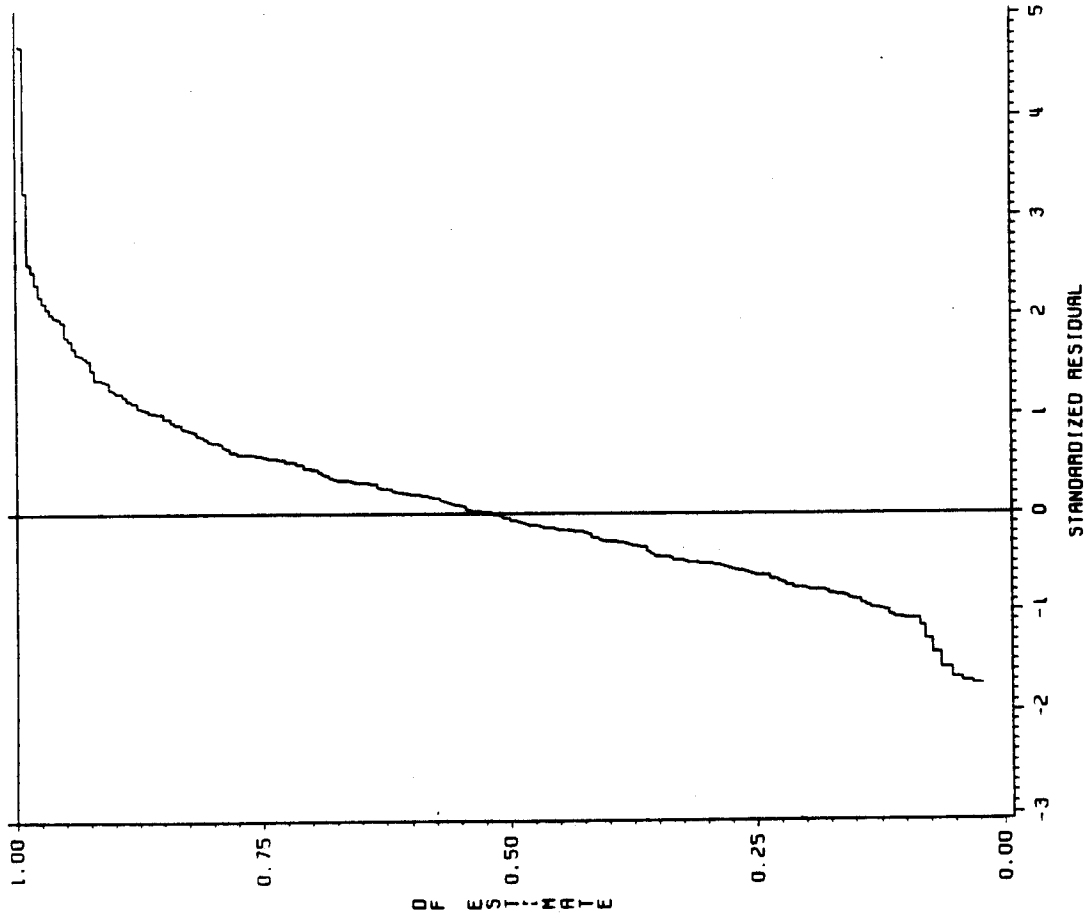


FIGURE 1b

DEP VBL = SHXCLOTH, WORKING-LESER FORM
INVERSE NORMAL TRANSFORMATION OF THE KAPLAN-MEIER ESTIMATE

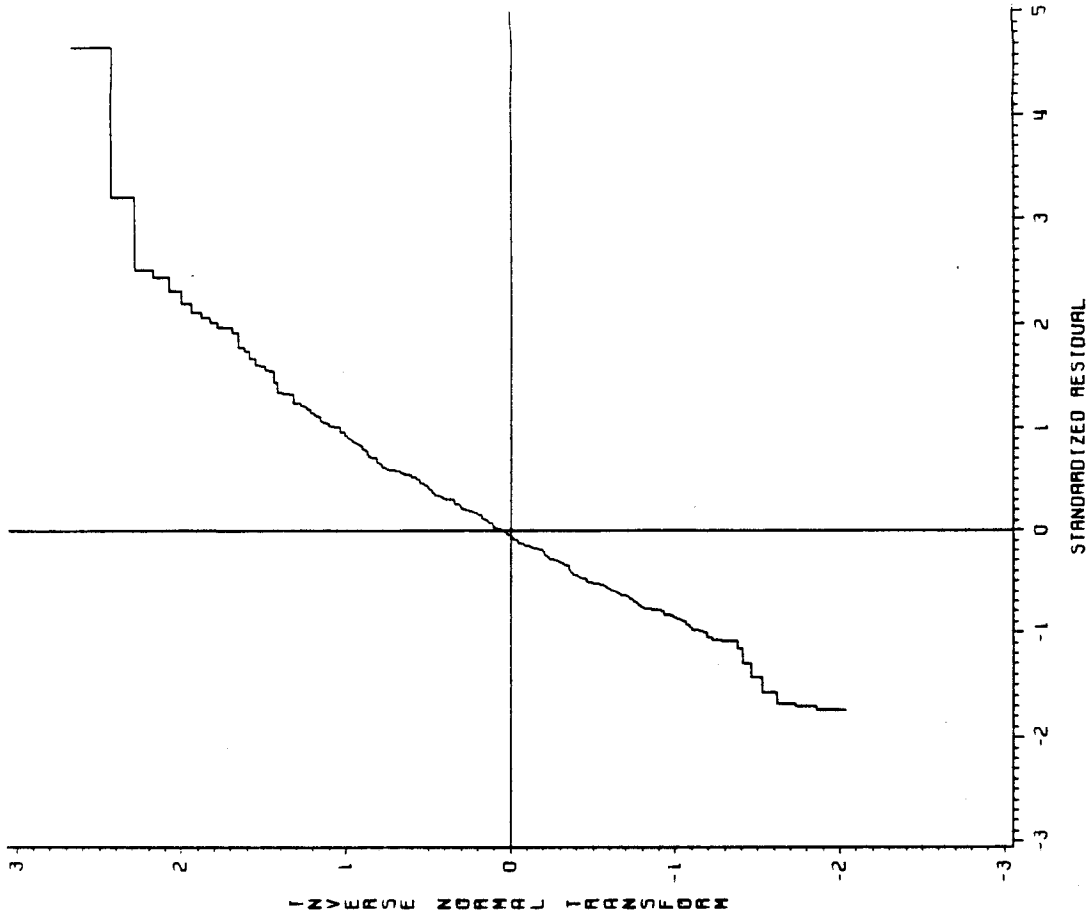


FIGURE 1C

DEP VBL = XCLOTH, LINEAR EXPENDITURE SYSTEM
KAPLAN-MEIER ESTIMATE OF THE DF OF THE STANDARDIZED RESIDUALS

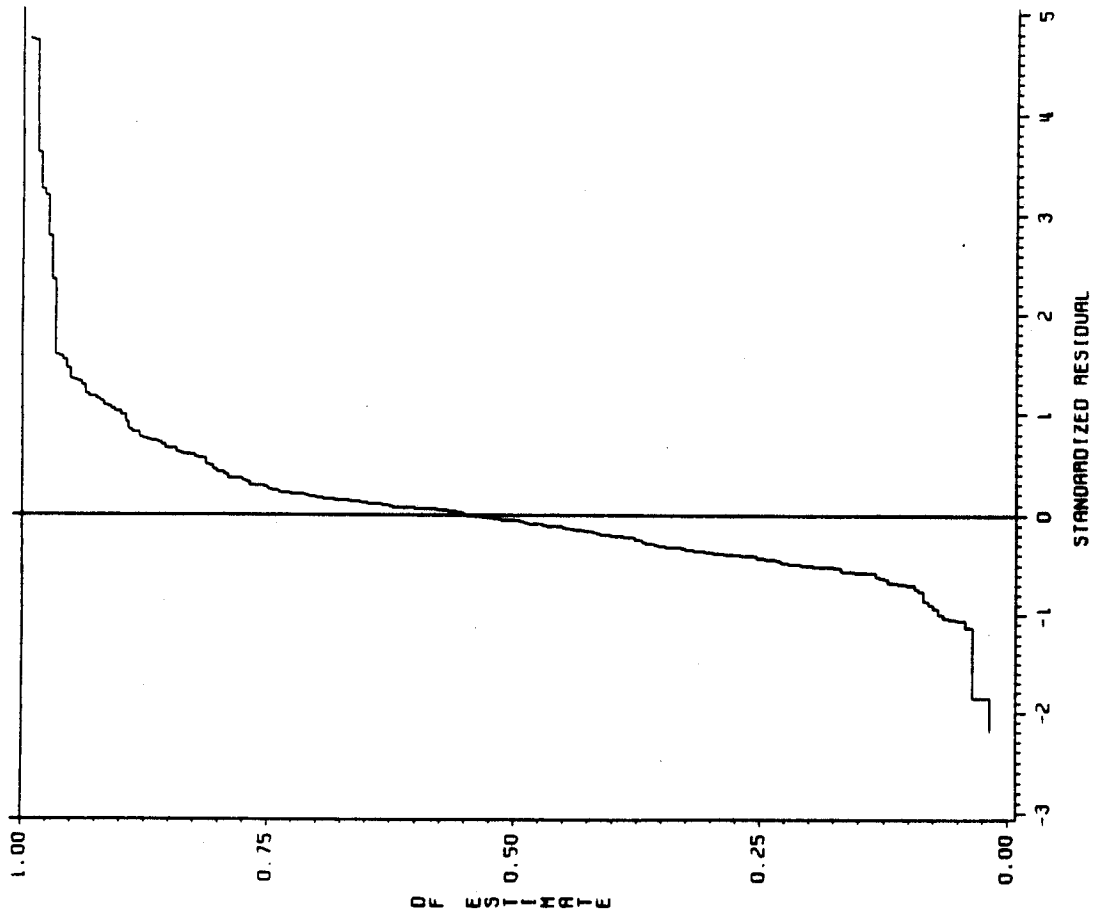


FIGURE 1d

DEP VBL = XCLOTH, LINEAR EXPENDITURE SYSTEM
INVERSE NORMAL TRANSFORMATION OF THE KAPLAN-MEIER ESTIMATE

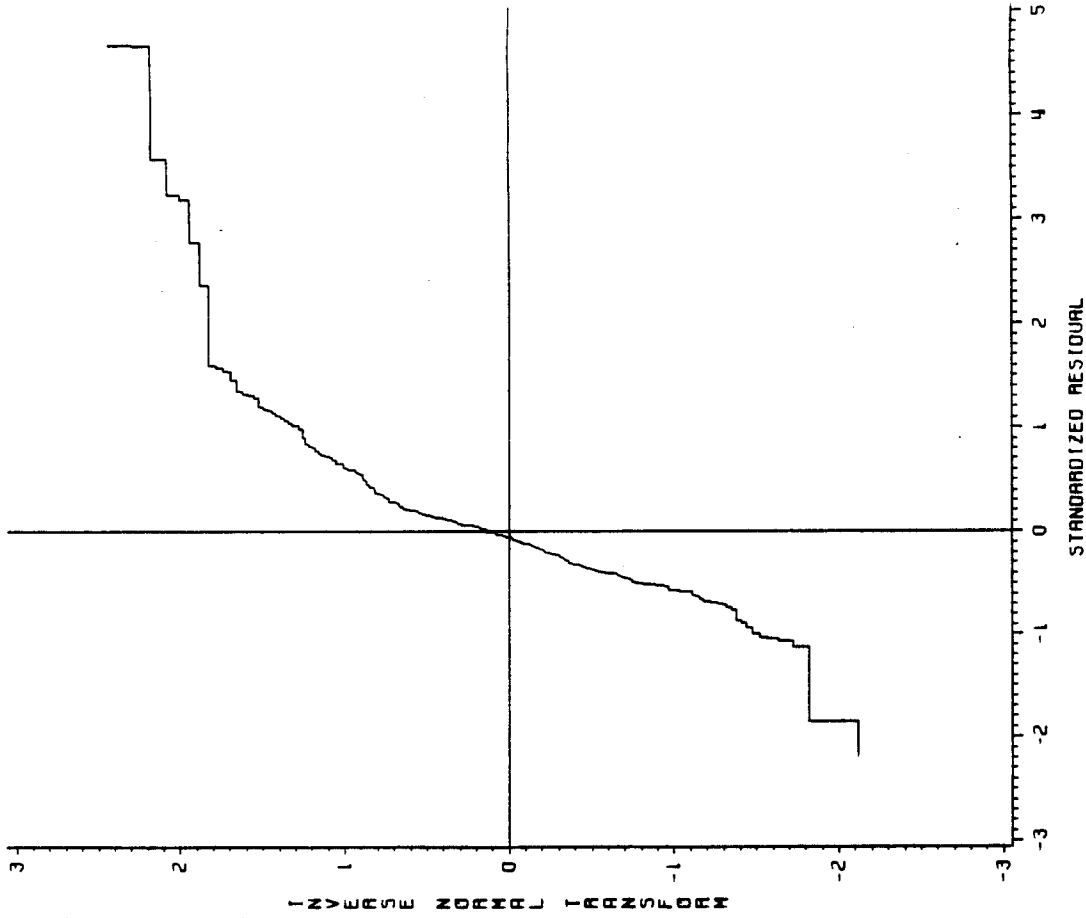


FIGURE 2a
 ESTIMATED ENGEL CURVES FOR TRANSPORT
 WORKING-LESER FORM
 LEGEND: ML=1, B11=2, B12=3, LA0=4

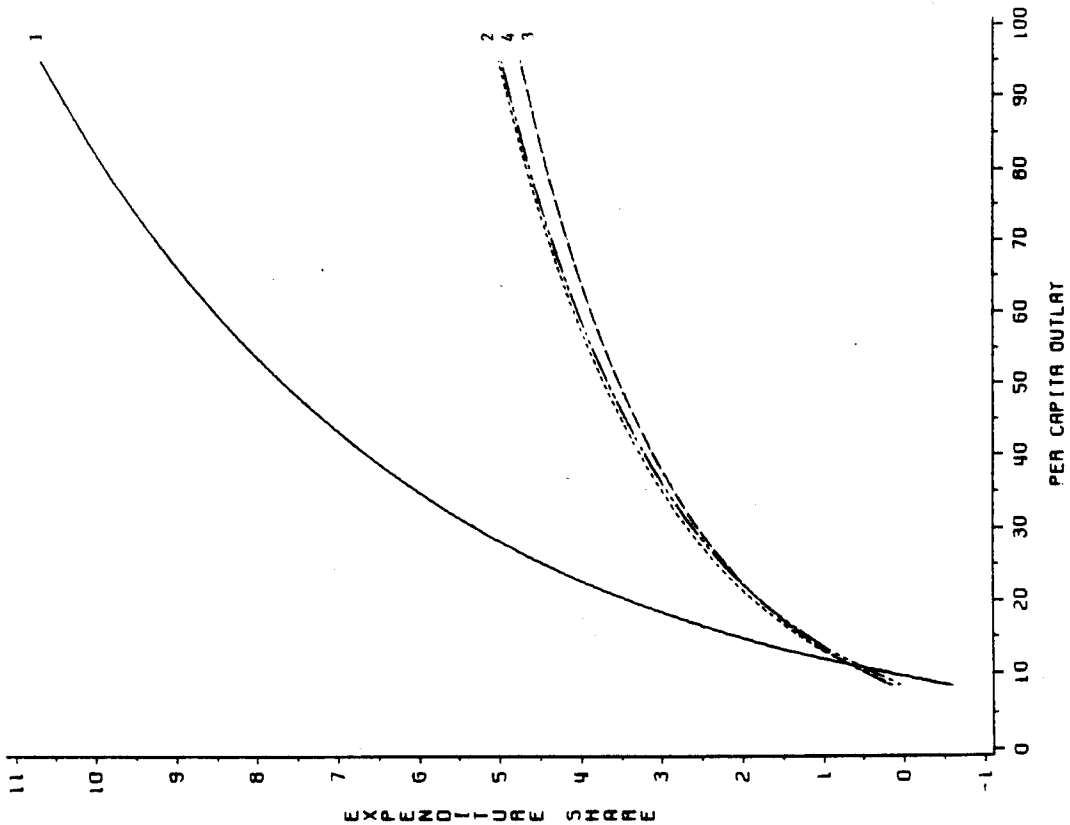


FIGURE 2b
 ESTIMATED ENGEL CURVES FOR TRANSPORT
 LINEAR EXPENDITURE SYSTEM
 LEGEND: ML=1, B11=2, B12=3, LA0=4

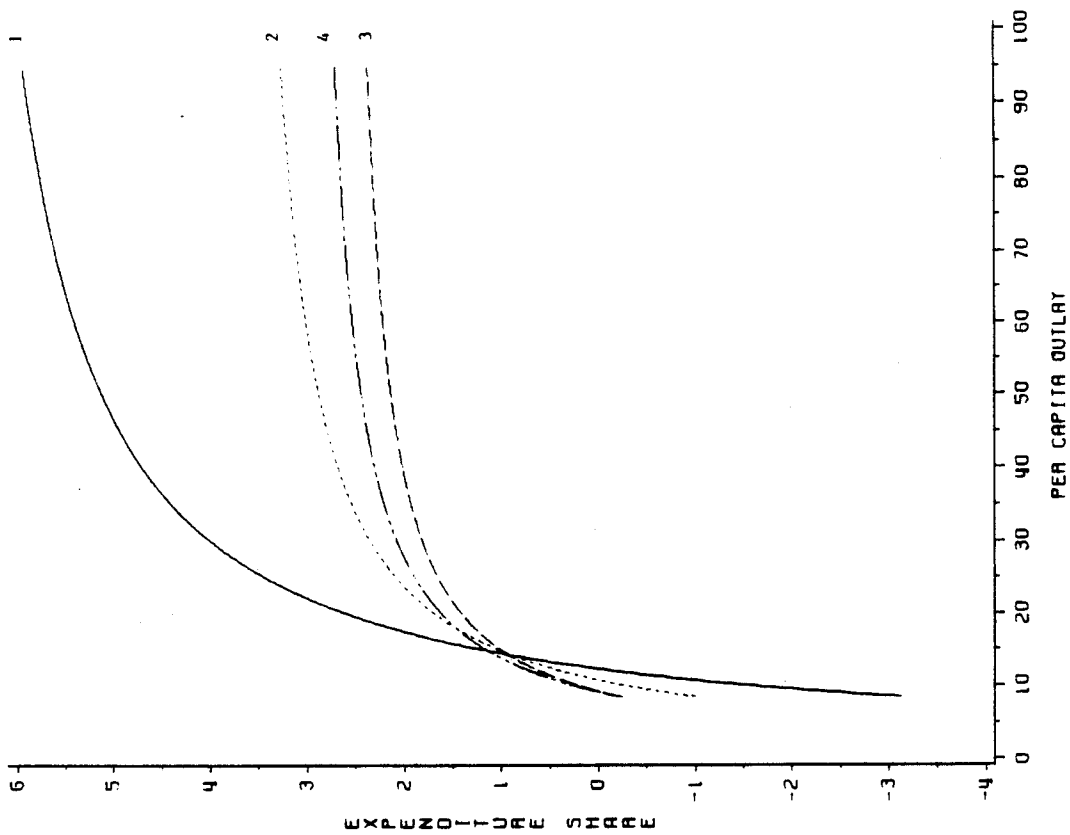


FIGURE 2c
 ESTIMATED ENGEL CURVES FOR TRANSPORT
 QUADRATIC WORKING-LESER FORM
 LEGEND: ML=1, B11=2, B12=3, LAD=4

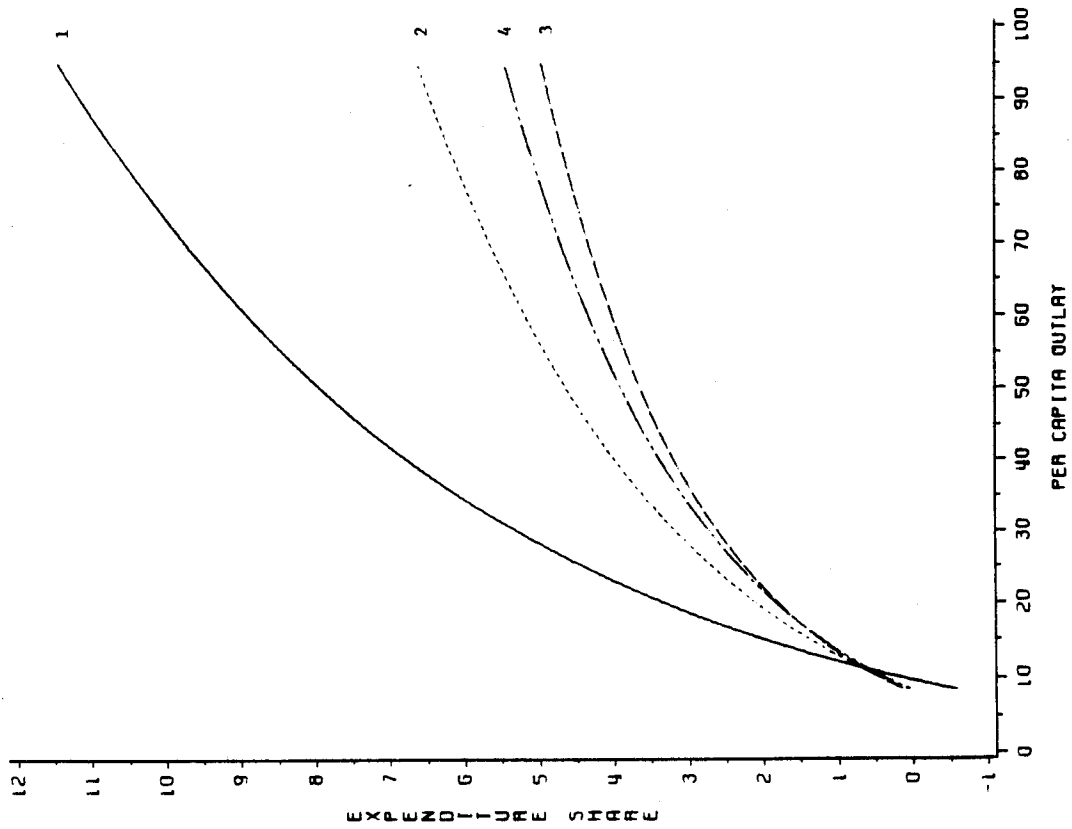


FIGURE 2d
 ESTIMATED ENGEL CURVES FOR TRANSPORT
 QUADRATIC EXPENDITURE SYSTEM
 LEGEND: ML=1, B11=2, B12=3, LAD=4

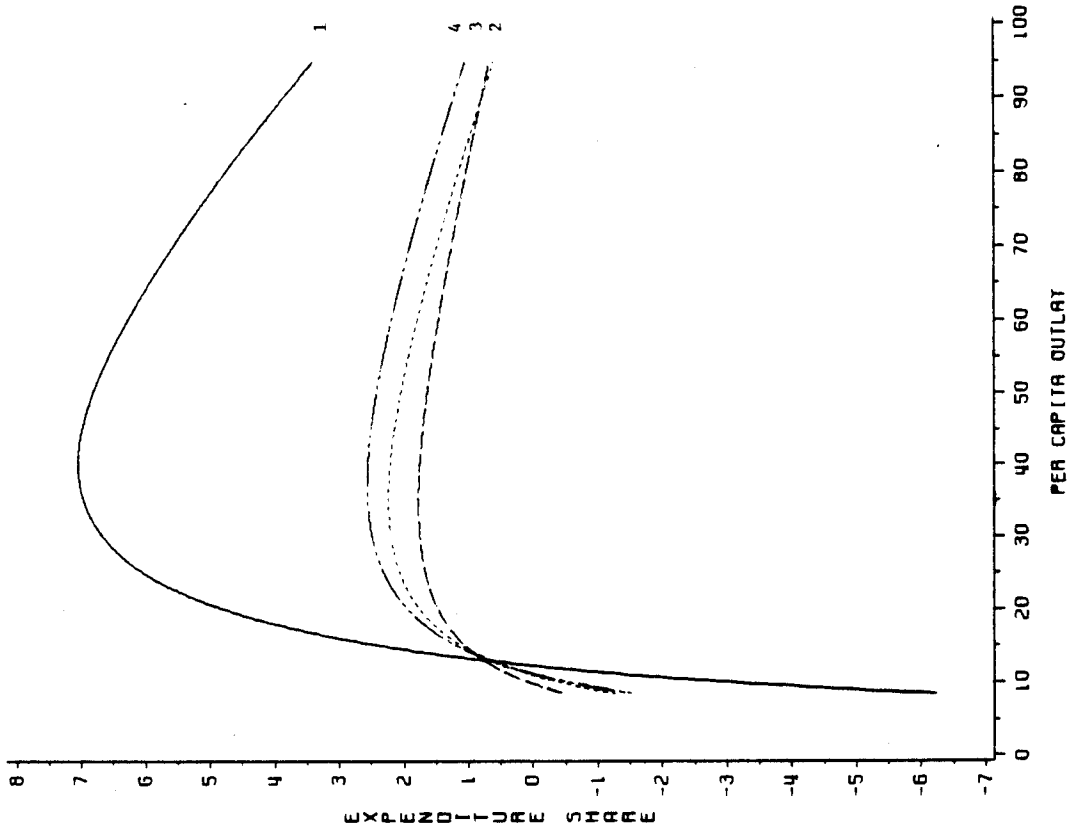
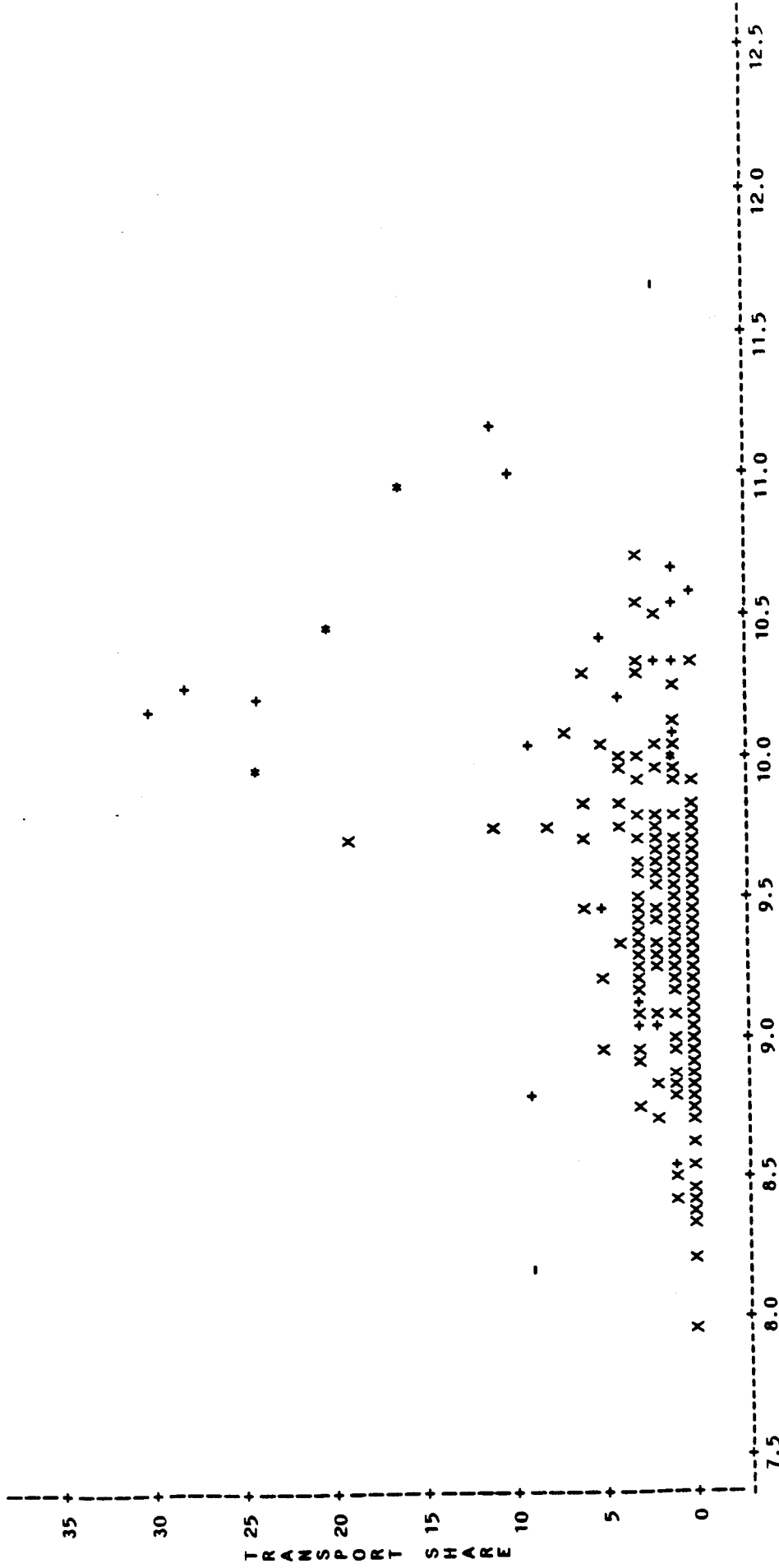


Figure 3
 PLOT OF TRANSPORT SHARE VS. PER CAPITA OUTLAY
 NORM OF TOBIT-ML INFLUENCE FUNCTION SUPERIMPOSED. WORKING-LESER FORM



SYMBOL	TNORM	SYMBOL	TNORM	SYMBOL	TNORM
XXXXXX	0.52839 -	*****	75.01079 -	=====	174.32067 -
++++++	25.35586 -	000000	124.66573 -	-----	223.97560 -
			174.32067		248.80307

Figure 4a

PLOT OF TRANSPORT SHARE VS. LOG PER-CAPITA OUTLAY
B11 WEIGHTS SUPERIMPOSED. WORKING-LESER FORM

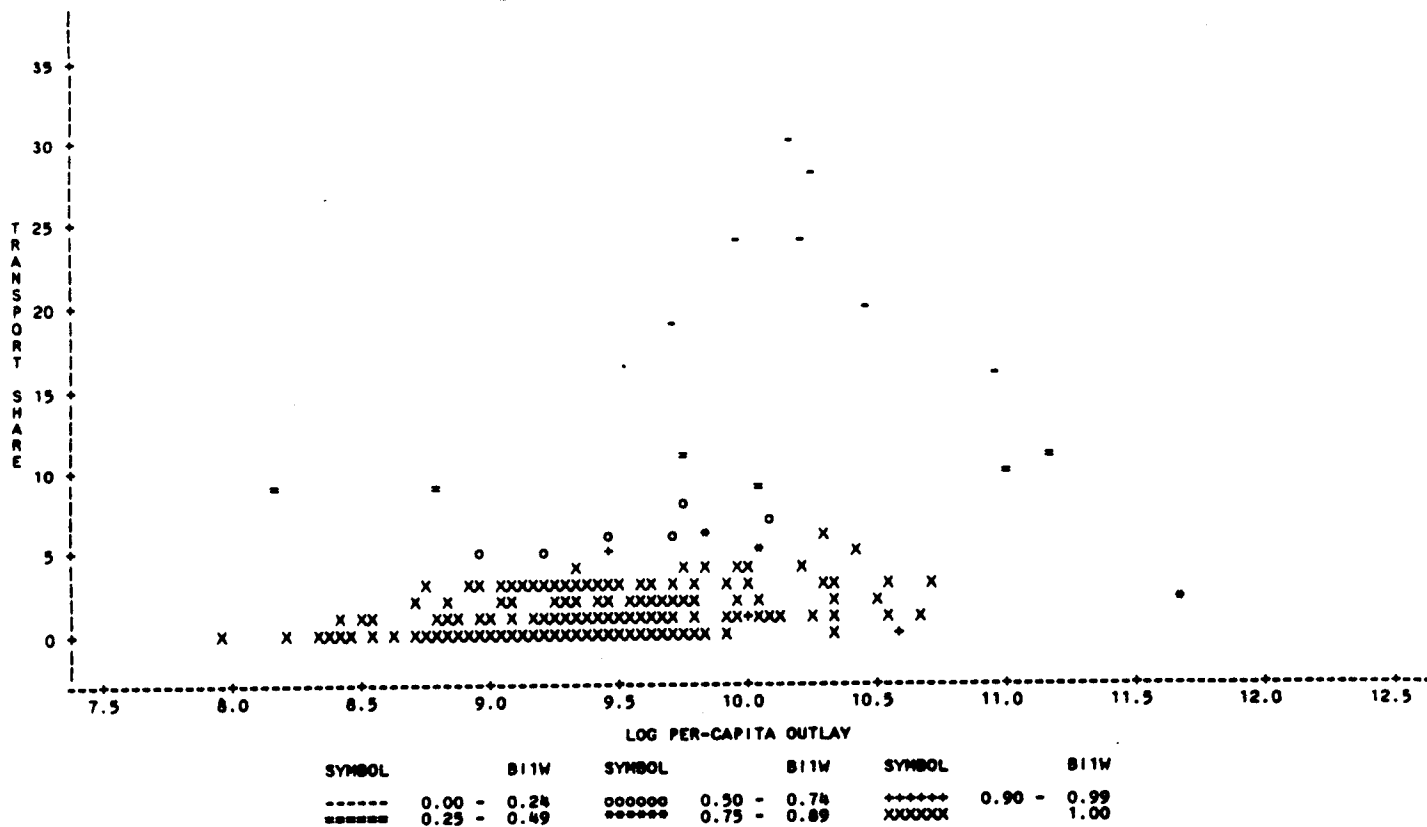
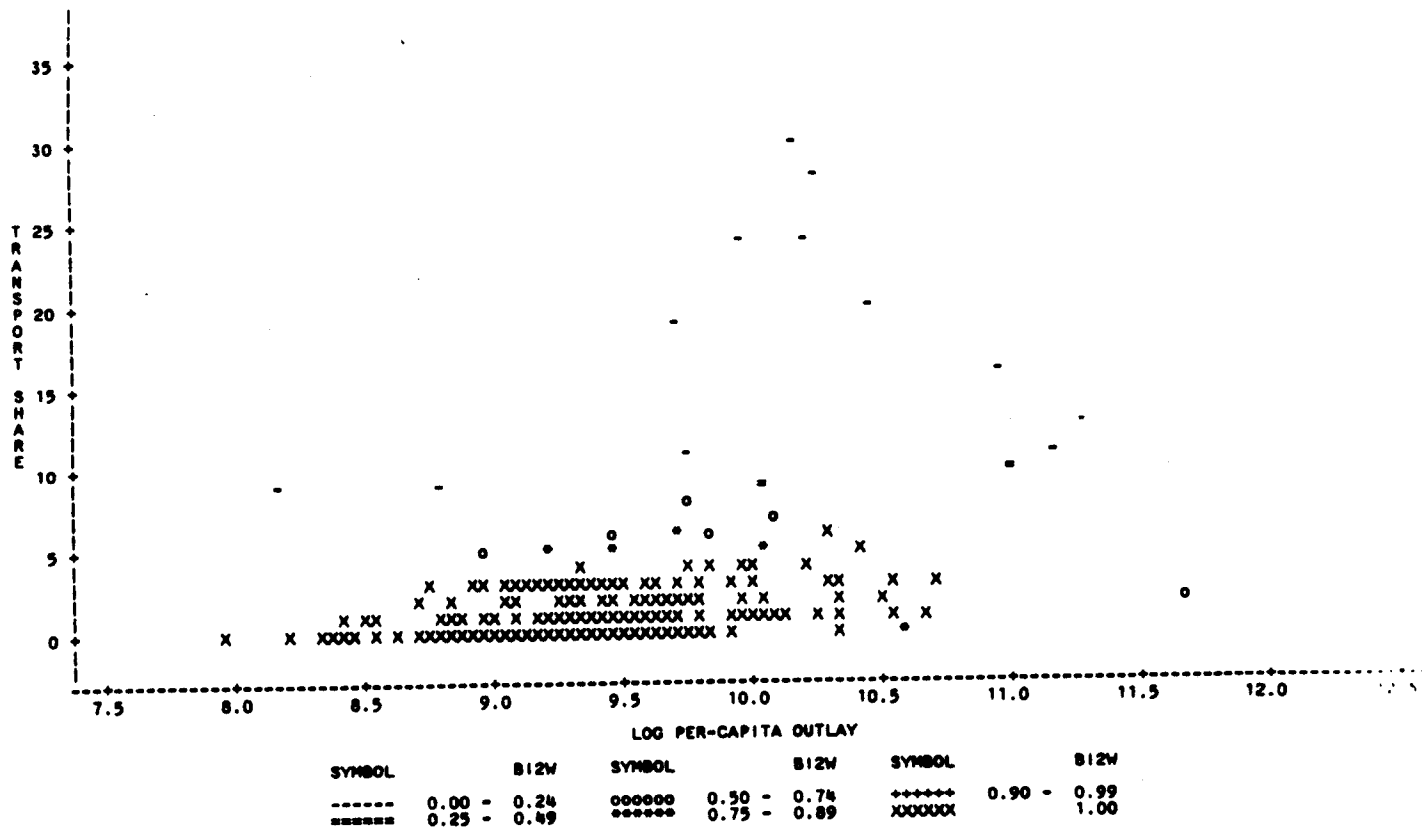


Figure 4b

PLOT OF TRANSPORT SHARE VS. LOG PER-CAPITA OUTLAY
B12 WEIGHTS SUPERIMPOSED. WORKING-LESER FORM



PLOT OF TRANSPORT SHARE VS. LOG PER-CAPITA OUTLAY
M-K WEIGHTS SUPERIMPOSED. WORKING-LESER FORM

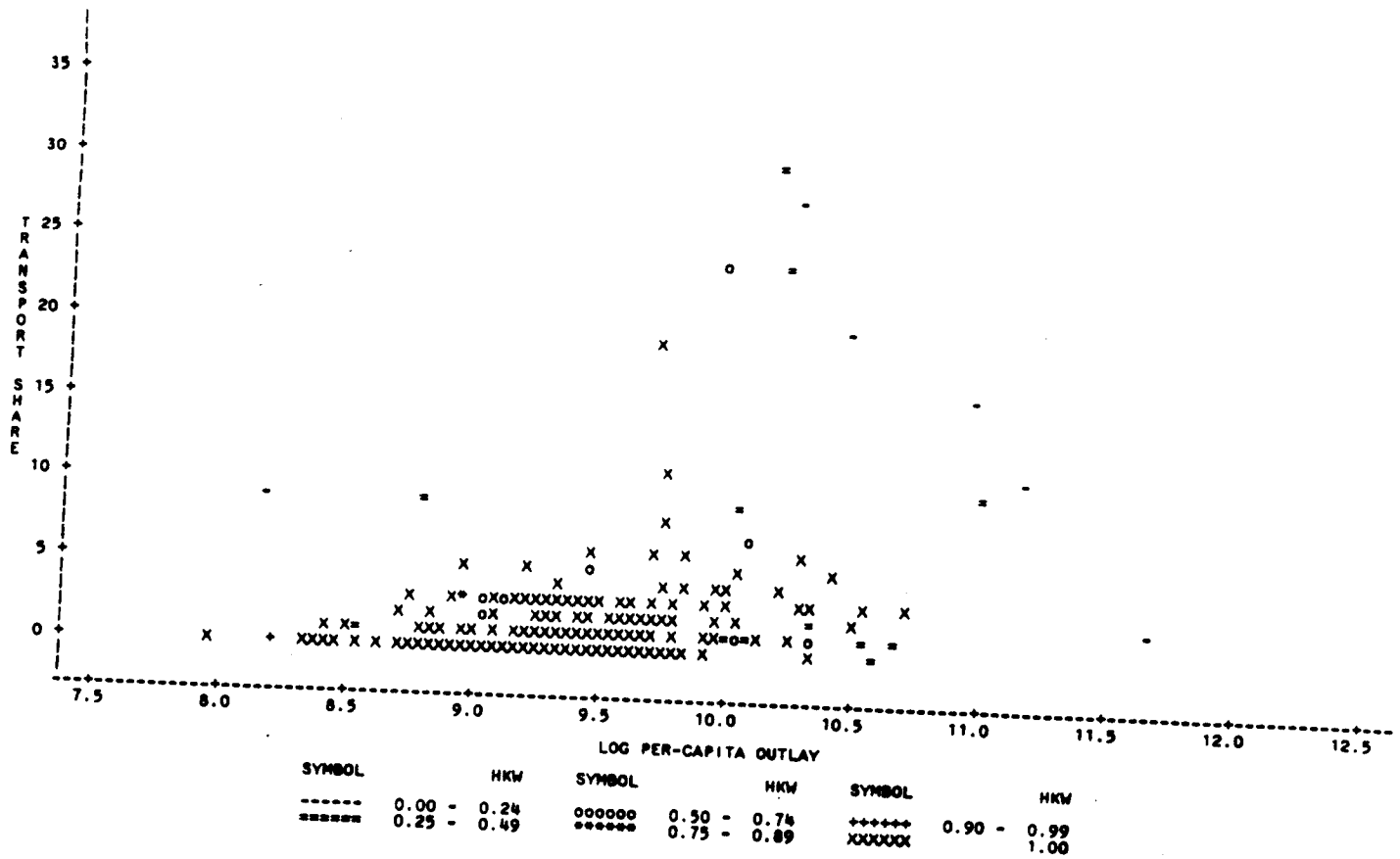


Figure 4d

PLOT OF TRANSPORT SHARE VS. LOG PER-CAPITA OUTLAY
K-W WEIGHTS SUPERIMPOSED. WORKING-LESER FORM
CONTOUR PLOT OF SHXTRANS*LXPC

