OXFORD BULLETIN OF ECONOMICS AND STATISTICS, 70, 4 (2008) 0305-9049
doi: 10.1111/j.1468-0084.2008.00506.x

PRACTITIONERS' CORNER

Practical Considerations for Choosing Between Tobit and SCLS or CLAD Estimators for Censored Regression Models with an Application to Charitable Giving*

MARK OTTONI WILHELM

Department of Economics, Indiana University–Purdue University at Indianapolis (IUPUI), 425 University Boulevard, Indianapolis, IN 46202, USA (e-mail: mowilhel@iupui.edu)

Abstract

Practical considerations for choosing between Tobit, symmetrically censored least squares (SCLS) and censored least absolute deviations (CLAD) estimators are offered. Practical considerations deal with when a Hausman test is better than a conditional moment test for judging the severity of a misspecification, the need to bootstrap the sampling distributions of the Hausman tests, what to look for in a graphical examination of the residuals and the limited value of SCLS. The practical considerations are applied to a model of the intergenerational transmission of charitable giving using new data from the Panel Study of Income Dynamics (PSID). The paper shows how to use relative distribution methods to calculate CLAD-based marginal effects on the observable dependent variable.

I. Introduction

Alternatives to Tobit – symmetrically censored least squares (SCLS) and censored least absolute deviations (CLAD) – are being used more frequently to estimate

*I am grateful to seminar participants at IUPUI, Adriaan Kalwij, and an anonymous referee for helpful comments. The usual disclaimer applies. I thank David Drukker for kindly sharing the National Longitudinal Survey of Older Women data. I am grateful for support from NICHD Grant Number 1 R01 HD04645633-01. A generous grant from Atlantic Philanthropies to the Center on Philanthropy at Indiana University financed the collection of philanthropy data in the PSID that motivated this research.

JEL Classification numbers: C12, C14, C15, C24, C52.

censored regression models (see Chay and Honoré, 1998; Chay and Powell, 2001). SCLS and CLAD are attractive because, unlike Tobit, they are robust against departures of errors from homoskedasticity and normality. A natural question is then how should one conduct a thorough specification evaluation to choose between Tobit and SCLS or CLAD estimators?

One approach is to use all three estimators and informally compare the point estimates to see if the differences between them are economically important – that is, large in terms of the economic implications of the differences or large in percentage terms relative to each other. An alternative approach is to use formal specification testing: estimate Tobit and then conduct conditional moment tests to detect departures of the errors from homoskedasticity and normality. Neither approach allows you to evaluate whether the difference between point estimates caused by departures from homoskedasticity and normality is 'statistically important' – that is, whether the difference between point estimates is large enough relative to the standard errors to be statistically significant. However, a third approach – the Hausman test – is intended to detect misspecifications that are serious enough to cause a large difference between point estimates, as Hausman (1978, p. 1253) explained in the motivation to his original article: 'Hopefully, this procedure will lead to powerful tests in important cases because the misspecification is apt to be serious only when the two estimates differ substantially'.

A thorough specification evaluation requires all three approaches, but only the conditional moment test for normality has been subject to thorough Monte Carlo investigation. It is now well known that the normality test is oversized, and in most experiments the oversize is dramatic (cf. Ericson and Hansen, 1999; Skeels and Vella, 1999; Drukker, 2002). However, approximating the sampling distribution with a parametric bootstrap delivers a correctly sized test with moderate-to-good power (Drukker, 2002). The conditional moment test for heteroskedasticity is also oversized (in Skeels and Vella the oversize is negligible; in Ericson and Hansen the oversize is dramatic), but no studies have analysed if a parametric bootstrap can deliver a correctly sized test. Furthermore, these papers do not consider whether statistically significant differences in the tests are accompanied by economically important differences in point estimators. Economically important differences in point estimators were studied by early researchers (Paarsch, 1984; Powell, 1986) with Monte Carlo experiments based on (only) one, synthetically created (rather than real-world) independent variable, an experimental setting fine for an initial investigation but with little resemblance to the setting in which most applied work is done. The early studies did not consider whether differences in point estimators are statistically significant.

Indeed, very little is known about the statistical significance of differences in point estimators caused by departures from homoskedasticity and normality. To my knowledge, only Ericson and Hansen (1999) have investigated the Hausman test for Tobit vs. SCLS. They found the Hausman test to be only slightly oversized. Ericson and Hansen also found the Hausman test to have weak power in detecting heteroskedasticity and non-normal errors, but recall that a Hausman test is not intended to have power against heteroskedasticity and non-normal misspecifications in general, rather only in misspecifications serious in Hausman's sense of causing 'the two estimates [to] differ substantially'. Because differences in point estimators are not reported in Ericson and Hansen's article, it may be that the Hausman test's 'weakness' is due to the heteroskedasticity and non-normality specified in the experiments not being serious enough to have caused large differences in the point estimators. Finally, again to my knowledge, there has not been a Monte Carlo investigation of the Hausman test for Tobit vs. CLAD.

This study extends this previous research by conducting a Monte Carlo investigation of the Hausman tests for Tobit vs. SCLS and for Tobit vs. CLAD while simultaneously investigating the other two approaches – informal comparison of point estimates and conditional moment tests. Therefore, the Monte Carlo results provide some practical guidance about how to conduct thorough specification evaluation using all three approaches. Then the paper illustrates the use of the approaches plus the graphical analysis of residuals in an application: estimating the intergenerational transmission of charitable giving using new data from the *Center on Philanthropy Panel Study* (Wilhelm *et al.*, 2001), a module recently added to the *Panel Study of Income Dynamics*. Indeed, it was this application that motivated the Monte Carlo investigations performed in this paper.

In addition to the first Monte Carlo results for the Hausman test for Tobit vs. CLAD, the paper's most important results are the considerations drawn from the Monte Carlo experiments that offer practical guidance for researchers doing applied work with censored regression models. In the experiments the conditional moment tests generally have reasonable to very good power to detect departures from homoskedasticity and normality (after all, the tests are asymptotically locally most powerful, as pointed out by Pagan and Vella, 1989). But conditional moment test rejections of homoskedasticity or normality – even with large values of the test statistics – do not necessarily imply 'serious' misspecification in the sense that Hausman used the term to motivate his original test. In large samples ($N = 2{,}500$) the Monte Carlo experiments suggest that the Hausman tests are reasonably powerful when the two estimates differ substantially in percentage terms. Hence, when working with a large sample a Hausman test's failure to reject in a project's baseline model provides justification for using Tobit to generate the bulk of the project's point estimates. However, the baseline model should nevertheless be subjected to conditional moment tests of homoskedasticity and normality. If either of these reject, the heteroskedasticity and non-normality of the residuals should be examined graphically to see if residuals suggest either severe heteroskedasticity or asymmetry. The examination of residuals is specifically for severe heteroskedasticity or severe asymmetry because the Monte Carlo experiments suggest that it is possible for these types of misspecifications to cause important biases in the Tobit point estimator that are not statistically significant in the Hausman sense. In the intergenerational transmission of charitable giving application I examine the asymmetry of the CLAD residuals using relative distribution techniques (see Handcock and Morris, 1999; relative distribution techniques are not well known among economists). The advantage of relative distribution techniques is that they make the severity of any asymmetry easy to judge, and if the asymmetry is judged to be severe the relative distribution suggests an easy way to generate asymmetric errors for the estimation of marginal effects.

In small samples ($N = 610$ in the experiments) the Hausman tests lack power even when the percentage difference between the two estimates is large. Hence, when working with a small sample the researcher must rely on the conditional moment tests and his/her judgements about the economic importance of any differences between the Tobit and SCLS or CLAD point estimates.

There are several other results. First, the Monte Carlo experiments demonstrate that the Tobit–SCLS and Tobit–CLAD Hausman tests can be dramatically oversized if conducted with asymptotic approximations to the sampling distributions. Note that the dramatic oversize result found for the Tobit–SCLS test is in contrast to the (only) slight oversize found by Ericson and Hansen (1999). Second, approximating the Hausman tests' sampling distributions with a parametric bootstrap using a modest number of bootstrap replications delivers an asymptotic refinement and correctly sized tests. Third, an error in the literature's formulation of the Tobit–SCLS test (Newey, 1987) is corrected. Without the correction, application of the test formulated by Newey is dramatically oversized (and the parametric bootstrap cannot correct the problem).

Fourth, the oversize problem in the conditional moment–homoskedasticity test documented by Skeels and Vella (1999) and Ericson and Hansen (1999) can be resolved when a parametric bootstrap is used to approximate the sampling distribution (this result is not all that surprising in the light of Drukker's 2002 study of the normality test). In addition, the Monte Carlo experiments suggest that the bootstrapped homoskedasticity test has power against heteroskedasticity when the heteroskedasticity is substantial or when the sample size is large, but not when the heteroskedasticity is more modest and the sample size is small.

Fifth, the Monte Carlo experiments are fairly discouraging as regards the use of SCLS. SCLS is consistent in the face of heteroskedasticity, but the experiments indicate that the SCLS small sample bias can sometimes be worse than Tobit's even when the sample size is not all that 'small' ($N = 610$). Neither Tobit nor SCLS is consistent in the face of asymmetric errors, but in all the experiments with asymmetric errors the SCLS bias was as bad as Tobit's or worse, even in large samples.

Sixth, the experiments illustrate the importance of building data generating processes with real-world data when conducting Monte Carlo experiments: the evidence of dramatic oversize in the Hausman test for Tobit vs. SCLS would have been missed if an experiment with a synthetic data generating process was the only one run.

Finally, the paper contains a new derivation of the Hausman test for Tobit vs. CLAD. The Hausman test for Tobit vs. CLAD was first derived by Horowitz and Neumann (1987), but unfortunately their paper is not well known (the only application of their test of which I am aware is by themselves in Horowitz and Newman, 1989). In addition to fixing ideas, the new derivation of the Tobit–CLAD test in the present paper has the advantage of making transparent the similarity between it and Newey's (1987) derivation of the Tobit–SCLS test.

II. Hausman tests for Tobit vs. CLAD and SCLS

2.1. Tobit vs. CLAD

Consider the censored regression model:

$$\begin{aligned} y_i^* &= Z_i\delta_0 + u_i \quad i = 1, \ldots, N \\ y_i &= \begin{cases} y_i^* & \text{if } y_i^* > 0 \\ 0 & \text{otherwise} \end{cases} \end{aligned} \tag{1}$$

where y_i^* is an unobserved latent variable, y_i is observed, Z_i is a $1 \times q$ vector of exogenous variables, and δ_0 is the $q \times 1$ parameter vector to be estimated. We assume that the errors u_i are independently and identically distributed (i.i.d.), but that their distribution is unknown.

If the u_i are $N(0, \sigma_0^2)$ – normal and homoskedastic – then the Tobit estimator $\hat{\delta}_{\mathrm{T}}$ is consistent and efficient. However, if either assumption does not hold, then $\hat{\delta}_{\mathrm{T}}$ is inconsistent. Pagan and Vella (1989) describe conditional moment tests of both assumptions. The advantage of conditional moment tests is that they can be implemented after Tobit estimation without estimating more complicated models. Should these tests reject either or both assumptions, one option is to use the SCLS estimator (Powell, 1986); SCLS is consistent if the u_i are heteroskedastic and non-normal as long as they are symmetrically distributed. Alternatively, the CLAD estimator (Powell, 1984) is consistent even if the errors are asymmetrically distributed as long as the median $(u_i | Z_i) = 0$. Of course, the other estimators require that $E(u_i | Z_i) = 0$.

Although the conditional moment tests alert us to specification problems, our interest also lies in whether these problems are serious in the sense Hausman (1978) suggested – serious enough to distort the Tobit estimator of δ_0 from the CLAD or SCLS estimators. To fix ideas, I begin with a derivation of the Tobit–CLAD test and then discuss the Tobit–SCLS test.[1]

The desired test statistic is:

$$h = N(\hat{\delta}_{\mathrm{C}} - \hat{\delta}_{\mathrm{T}})' \hat{V}^{-1} (\hat{\delta}_{\mathrm{C}} - \hat{\delta}_{\mathrm{T}}) \tag{2}$$

where $\hat{\delta}_{\mathrm{C}}$ is the CLAD estimator and $\hat{V}$ is an estimate of the covariance matrix of $N^{1/2}(\hat{\delta}_{\mathrm{C}} - \hat{\delta}_{\mathrm{T}})$. Under the null hypothesis of normal, homoskedastic errors the test statistic h is asymptotically chi-square with degrees of freedom q. Now form $\hat{V}$ as an estimate of the covariance matrix of two method-of-moments estimators as follows.

Assuming normality and homoskedasticity the Tobit estimator is consistent and asymptotically normal under regularity conditions (Amemiya, 1985, p. 363). In particular:

$$N^{1/2}(\hat{\delta}_{\mathrm{T}} - \delta_0) = [I_q, 0] \hat{J}^{-1} N^{-1/2} \sum_{i=1}^{N} s_i(\theta_0) + o_p(1) \tag{3}$$

[1] Because the conditional moment tests are well known, I do not discuss their derivation. See Pagan and Vella (1989) or Greene (2003, pp. 772ff).

where $\theta_0 \equiv (\delta_0', \sigma_0^2)'$ is a $(q+1) \times 1$ parameter vector, $s_i(\theta_0)$ is the $(q+1) \times 1$ score vector of the Tobit log-likelihood, $[I_q, 0]$ is a $q \times (q+1)$ matrix where the last column contains zeros, and $\hat{J}$ is an estimate of the $(q+1) \times (q+1)$ Tobit information matrix. I use the outer product gradient estimate of the information matrix:

$$\hat{J} = N^{-1} \sum_{i=1}^{N} s_i(\theta_0) s_i(\theta_0)'.$$

The CLAD estimator is (strongly) consistent and asymptotically normal under the regularity conditions described by Powell (1984). In particular:

$$N^{1/2}(\hat{\delta}_C - \delta_0) = \hat{H}^{-1} N^{-1/2} \sum_{i=1}^{N} g_i(\delta_0) + o_p(1) \tag{4}$$

where

$$\hat{H} = 2 \frac{1}{N h_N} \sum_{i=1}^{N} 1(Z_i \hat{\delta}_C > 0) 1(0 \le \hat{u}_i \le h_N) Z_i' Z_i \tag{5}$$

$$g_i(\delta) = 1(Z_i \delta > 0) \mathrm{sign}[\min(y_i, 2Z_i\delta) - Z_i\delta] Z_i' \tag{6}$$

and

$$\hat{u}_i = y_i - Z_i \hat{\delta}_C. \tag{7}$$

The $\hat{u}_i$ are the CLAD residuals and the $1(Z_i\delta > 0)\mathrm{sign}[\min(y_i, 2Z_i\delta) - Z_i\delta]$ are the symmetrically trimmed residuals (Pagan and Vella, 1989, p. s37).[2]

Subtract equation (3) from equation (4) to get $N^{1/2}(\hat{\delta}_C - \hat{\delta}_T)$ whose covariance matrix is (under independence of the u_i conditional on Z_i and the previously mentioned regularity conditions):

$$\hat{V} = \hat{A} \hat{C} \hat{A}' \tag{8}$$

where:

$$\hat{A} = [\hat{H}^{-1}, -[I_q, 0] \hat{J}^{-1}] \tag{9}$$

$$\hat{C} = N^{-1} [G, S]' [G, S] \tag{10}$$

and G is the $N \times q$ matrix of CLAD moments and S is the $N \times (q+1)$ matrix of Tobit scores (e.g. the ith rows are $g_i(\hat{\delta}_C)'$ and $s_i(\hat{\theta}_T)'$, where $\hat{\theta}_T$ is the Tobit estimator of θ_0). Forming $\hat{V}$ in this way guarantees that it is positive semi-definite.

[2]As written, the expression in equation (5) for $\hat{H}$ makes the Hausman test robust to heteroskedastic errors under the alternative. Instead, if it is assumed that the u_i are homoskedastic then the estimate of $f_u(0|Z)$ – i.e. $1(0 \le \hat{u}_i \le h_N)$ – can be factored out of the summation in equation (5). I use Silverman's (1986) suggestion for the bandwidth: $h_N = 0.9N^{-1/5}\min$[standard deviation, interquartile range/1.34].

2.2. Revisiting the Hausman test for Tobit vs. SCLS

The above derivation of the Hausman test for Tobit vs. CLAD follows Newey's (1987) derivation of the Hausman test for Tobit vs. SCLS. In Newey's derivation:

$$N^{1/2}(\hat{\delta}_{\mathrm{S}} - \delta_0) = \hat{H}^{-1} N^{-1/2} \sum_{i=1}^{N} g_i(\delta_0) + o_p(1) \tag{11}$$

where $\hat{\delta}_{\mathrm{S}}$ is the SCLS estimator,

$$g_i(\delta) = 1(Z_i\delta > 0)[\min(y_i, 2Z_i\delta) - Z_i\delta]Z_i'$$

and

$$\hat{H} = \sum_{i=1}^{N} 1(Z_i\hat{\delta}_S > 0)Z_i'Z_i/N. \tag{12}$$

However, equation (12) is not correct. First, the expression for $\hat{H}$ in equation (12) does not follow from Powell's (1986) derivation of SCLS covariance matrices. Instead, Powell's derivation leads to:

$$\hat{H} = \sum_{i=1}^{N} 1(0 < y_i < 2Z_i\hat{\delta}_S)Z_i'Z_i/N. \tag{13}$$

Second, in Monte Carlo experiments the Hausman test statistic constructed using equation (12) is dramatically oversized even when the parametric bootstrap is used to approximate the sampling distribution (for details, see section 4.1, paragraph 3).

Therefore, I construct the Hausman test statistic for Tobit vs. SCLS as in Newey (1987) except using equation (13) instead of equation (12) to define $\hat{H}$. As will be seen below, the test statistic constructed using equation (13) is correctly sized when the parametric bootstrap is used.

2.3. Bootstrapping

Although the Hausman test statistic h comparing Tobit with CLAD or Tobit with SCLS is asymptotically chi-square under the normal, homoskedastic null, Ericson and Hansen (1999) found that the Tobit–SCLS Hausman test was slightly oversized when using the asymptotic approximation to the sampling distribution. Compared with the asymptotic distribution, the bootstrap of a pivotal statistic offers a better approximation of the actual sampling distribution (Horowitz, 2001). Accordingly, I investigate the parametric bootstrap of the Hausman test statistic h.

In the parametric bootstrap, bootstrapped errors are generated under the null hypothesis that the errors in equation (1) are normal and homoskedastic. Specifically, the parametric bootstrap for the Tobit–CLAD test is implemented as follows (the Tobit–SCLS bootstrap is obviously similar):

Step i. With the sample data $\{(y_i, Z_i), i=1, \ldots, N\}$ get the Tobit estimates $\hat{\delta}_{\mathrm{T}}$ and $\hat{\sigma}^2_{\mathrm{T}}$ and the CLAD estimates $\hat{\delta}_{\mathrm{C}}$; calculate the value of the test statistic h.
Step ii. Generate bootstrapped errors $\{u_{i,b}, i=1, \ldots, N\}$ using $N(0, \hat{\sigma}^2_{\mathrm{T}})$.
Step iii. Generate the bootstrapped latent variable $y^*_{i,b}$ and the bootstrapped observable $y_{i,b}$ using Z_i, $\hat{\delta}_{\mathrm{T}}$, and the $u_{i,b}$ in equation (1).
Step iv. Get Tobit and CLAD estimates using $y_{i,b}$ and Z_i; form the bootstrapped test statistic h_b.
Step v. Repeat steps (ii)–(iv) B times; $\{h_b, b=1, \ldots, B\}$ is the bootstrap approximation of the sampling distribution of h (under the null).
Step vi. Conduct the Hausman test by comparing the value of the test statistic h from step (i) with critical values from $\{h_b, b=1, \ldots, B\}$.

Steps (i)–(vi) produce a test with asymptotic refinement because h is asymptotically pivotal [$h \sim \chi^2(q)$ under H_0]. An alternative procedure would be to generate a bootstrap estimate of V (called $\hat{V}_{\mathrm{boot}}$), form an alternative test statistic $h_{\mathrm{alternative}} = N(\hat{\delta}_{\mathrm{C}} - \hat{\delta}_{\mathrm{T}})'\hat{V}^{-1}_{\mathrm{boot}}(\hat{\delta}_{\mathrm{C}} - \hat{\delta}_{\mathrm{T}})$, and compare $h_{\mathrm{alternative}}$ with $\chi^2(q)$ critical values. However, because the estimator of V is non-pivotal, the alternative procedure would produce a test without refinement. A test with refinement can also be obtained by implementing steps (i)–(vi) but using the empirical distribution function of the data rather than the parametric bootstrap; however, the numerical accuracy of the parametric bootstrap is likely superior (see Horowitz, 2001).

III. Experimental design

The Monte Carlo experiments use three sets of independent variables. The first set contains six synthetic random variables correlated in the following way:

$$\begin{aligned}
&z_{i1} = \upsilon_{i1}; \quad z_{i2} = 0.3z_{i1} + \upsilon_{i2}; \quad z_{i3} = 0.15z_{i1} + 0.15z_{i2} + \upsilon_{i3};\\
&z_{i4} = 0.1z_{i1} + 0.1z_{i2} + 0.1z_{i3} + \upsilon_{i4}; \quad z_{i5} = -0.1z_{i1} - 0.1z_{i2} - 0.1z_{i3} - 0.1z_{i4} + \upsilon_{i5};\\
&z_{i6} = -0.075z_{i1} - 0.075z_{i2} - 0.075z_{i3} - 0.075z_{i4} + 0.075z_{i5} + \upsilon_{i6} \qquad (14)
\end{aligned}$$

where the υ_{ij} are $N(0, 1)$. The latent variable is:

$$y^*_i = z_{i1} + z_{i2} + z_{i3} + z_{i4} + z_{i5} + z_{i6} + u_i. \qquad (15)$$

The second set of independent variables are from the *National Longitudinal Survey of Older Women* data on labour supply originally analysed by Moffitt (1984); the labour supply data are used to build the data generating process:

$$\begin{aligned}
y^*_i = {} & 39.992 - 0.01084\mathrm{nwi}_i + 2.0173\mathrm{educ}_i - 6.3468\mathrm{clt6}_i - 9.3085\mathrm{ms}_i\\
& - 1.1148\mathrm{age}_i + 1.7784\mathrm{race}_i + u_i \qquad (16)
\end{aligned}$$

where y^*_i is latent hours worked per week and the independent variables are annual non-wage income, years of education, number of children under 6 years old, marital status (1 if married), age and race (1 if white). The parameter values come from

Tobit estimation of the labour supply function matching equation (16) with the 610 observations available in the data.

The third set of independent variables comes from the *Center on Philanthropy Panel Study* data on charitable giving (see Wilhelm, 2006, 2007):

$$y_i^* = -16 + 0.968 \log \text{income}_i + 0.249\text{educ}_i + 0.079\text{kids}_i + 0.785\text{ms}_i + 0.043\text{age}_i - 0.226\text{race}_i + u_i \tag{17}$$

where y_i^* is the latent log of charitable giving per year and the independent variables are selected to match those in equation (16) as closely as possible, except that the log of total family income replaces equation (16)'s linear non-wage income. Once more, the parameter values come from Tobit estimation of the giving function matching equation (17), but this time using the 4,834 observations available in the data.[3]

Design (16) generates results that can be compared with those of Skeels and Vella (1999) and Drukker (2002), both of whom used this data generating process to study conditional moment tests. The design provides a correlation pattern in the independent variables from a real-world application, hence a pattern similar to that which researchers may actually encounter. This turns out to be important. Interest in real-world correlation patterns also motivates the use of design (17). In addition, the larger sample N available with the giving data permits experimental study of the results' sensitivity with respect to N. The results from the real-world designs can be compared with those of the synthetic design (14)–(15), similar to designs typically used in Monte Carlo studies. Design (14)–(15) also facilitates experimental study of the results' sensitivity with respect to increasing the number of independent variables while holding constant the percentage of observations being censored. As the number of independent variables is increased, the correlation pattern among the independent variables in equation (14) generates an increasingly complex $Z'Z$. Although this mimics what typically happens in applied work, design (14)–(15) can be easily modified so that the independent variables are not correlated with each other – then the effect of increasing the number of independent variables can be studied while holding constant the complexity of $Z'Z$.

In the test size experiments (the experiments conducted to compare the actual size of the tests with the nominal size), the u_i are specified to be normal, homoskedastic random variables with zero mean and variances 4, 1162.6 and 4.937 for the respective designs [the latter two are the Tobit estimates corresponding to the parameter values used in equations (16) and (17)]. For experiments considering departures from homoskedasticity, the error variance is a linear function of $Z_i\delta_0$, similar to the departures considered by Powell (1986). Specifically, the observations are ordered from smallest to largest according to values of $Z_i\delta_0$ and then designs where the ratio of variances $\text{var}(u_N)/\text{var}(u_1)$ is 3, 15, 1/3 and 1/15 are examined. The departures from

[3]To make the percentage of observations censored in the giving model similar to the percentage censored in the labour supply model (50%), the constant term from the Tobit estimation of the giving model (-14.382) is replaced with -16 in the Monte Carlo experiments.

normality are those considered by Skeels and Vella (1999) and Drukker (2002) – a Cauchy distribution and a t-distribution with five degrees of freedom (symmetric distributions) and a chi-squared distribution with one degree of freedom, a chi-squared distribution with five degrees of freedom and a mixture of chi-squares (asymmetric distributions). The chi-squared distributions are re-centred to have zero mean (e.g. $\chi_1^2 - 1$). The chi-squared mixture $[0.4(\chi_1^2 - 1) + 0.4(\chi_{25}^2 - 25)]$ is bimodal, though not dramatically enough to be readily apparent from a visual inspection of the errors. Each of the departures from normality is rescaled to have the same variance as the original design to which it is applied [e.g. rescaled to have variance 4.937 for design (17)].

Each Monte Carlo experiment consists of $M = 2{,}000$ replications. Each replication performs a parametric bootstrap with $B = 100$ bootstrap replications. Several experiments were performed with $B = 250$ bootstrap replications; the additional bootstrap replications delivered no discernible improvements in the accuracy of the bootstrapped Hausman–SCLS test and only very modest improvements in the bootstrapped Hausman–CLAD tests. Experiments with $B = 500$ bootstrap replications did not improve the accuracy of either test beyond that achieved with $B = 250$ replications. Because increasing the bootstrap replications above 100 substantially increases the computational burden of the experiment (a typical $M = 2{,}000$, $B = 100$ experiment takes 2 weeks on a 1.7 GHz Pentium IV computer with 2 GB of RAM), the modest (at best) accuracy improvements do not seem worth the additional computational cost of running the experiments.[4]

IV. Results

4.1. Test size experiments

Table 1 presents the actual test sizes of the conditional moment and Hausman tests applied to the three designs (14)–(17), with sample $N = 610$. The rejection decisions are based on the asymptotic (top panel) and bootstrapped (bottom panel) approximations to the test statistics' sampling distributions. When rejection decisions are based on the asymptotic approximation, the results in the first six columns indicate that the normality and homoskedasticity tests are dramatically oversized, replicating previous results (Ericson and Hansen, 1999; Drukker, 2002). At first, the Hausman–SCLS test does not appear to be oversized – when applied to the synthetic data the actual test sizes are 0.104, 0.061 and 0.012 – but the test is dramatically oversized when applied to either the labour or giving designs. This result illustrates the importance of going beyond synthetic designs to check test performance with independent variables

[4]Although an $M = 2{,}000$ experiment takes 2 weeks, a single Monte Carlo replication with $B = 100$ takes only about 10 minutes. I say more about computation time in section 4.3. I use STATA programs written by Jolliffe, Krushelnytskyy and Semykina (2001) and Moreira (see Chay and Powell, 2001) to estimate the CLAD and SCLS models.

TABLE 1

Conditional moment and Hausman tests: actual test sizes using asymptotic and bootstrapped distributions

Test:	*Conditional moment*						*Hausman*					
	Normality			*Homoskedasticity*			*SCLS*			*CLAD*		
Nominal test size:	*0.10*	*0.05*	*0.01*	*0.10*	*0.05*	*0.01*	*0.10*	*0.05*	*0.01*	*0.10*	*0.05*	*0.01*
Asymptotic distribution												
Design												
Synthetic (equations 14, 15)	0.219	0.163	0.083	0.179	0.108	0.027	0.104	0.061	0.012	0.232	0.118	0.048
Labour (equation 16)	0.210	0.157	0.088	0.188	0.108	0.034	0.402	0.285	0.196	0.240	0.092	0.044
Giving (equation 17)	0.188	0.131	0.064	0.185	0.116	0.028	0.260	0.198	0.108	0.315	0.239	0.137
Bootstrapped distribution												
Design												
Synthetic (equations 14, 15)	0.109	0.055	0.014	0.110	0.059	0.014	0.116	0.053	0.014	0.112	0.062	0.016
Labour (equation 16)	0.110	0.067	0.010	0.112	0.062	0.014	0.097	0.050	0.012	0.082	0.041	0.008
Giving (equation 17)	0.097	0.049	0.012	0.104	0.052	0.010	0.104	0.057	0.009	0.099	0.052	0.009

Note: $N = 610$.

drawn from real-world applications. The Hausman–CLAD test is dramatically oversized in all three designs.[5]

Rejection decisions based on the bootstrapped approximations are appropriately sized in all but a few cases: the normality test using the labour design at 5% (actual test size is 0.067) and the Hausman–CLAD test using the labour design at 10% (actual test size is 0.082).[6] Even in these cases the actual test sizes are much closer to the nominals than when using the asymptotic distribution. Moreover, the bootstrapped tests become appropriately sized when the sample N is increased to $N = 1{,}000$ (not shown), whereas the asymptotic test remains oversized (see Figure 3; to be discussed below).

However, the bootstrapped Hausman–SCLS test remains dramatically oversized when conducted with Newey's (1987) formulation of $\hat{H}$ equation (12). For example, when conducting the test with equation (12) using the synthetic design, the asymptotic approximation produces actual test sizes 0.500, 0.393 and 0.228, and the bootstrap approximation produces actual test sizes 0.499, 0.377 and 0.221. Hence, the bootstrap approximation cannot correct the oversize problem in the Hausman–SCLS test conducted with equation (12).

In the light of the evidence that the asymptotic distribution can produce substantial over-rejection of a true null hypothesis, in what practical situations might this be problematic? Figures 1–3 present simulation evidence that the asymptotic distribution's over-rejections generally (although not always) increase:

(i) with the number of independent variables,
(ii) as censoring increases, and
(iii) at small N.

Figure 1 plots actual test sizes for the conditional moment and Hausman tests when the tests are conducted nominally at 5% using the synthetic design ($N = 100$) and beginning with a single independent variable (z_{i1}) and successively adding the next variable until all six are included. As the number of independent variables increases, the over-rejections become more frequent. The over-rejections increase more in the conditional moment tests, less so in the Hausman–SCLS test, but not at all in the Hausman–CLAD test. Regenerating Figure 1 with design (14)–(15), modified so that the independent variables are not correlated with each other, produces similar results (not shown) – hence, the more frequent over-rejections are due to the increased number of independent variables, and not the increased complexity of $Z'Z$.

Figure 2 plots actual test sizes from nominally 5% tests using the giving design ($N = 610$) but changing the constant term to adjust the amount of censoring from

[5]The Hausman tests are based on the slope coefficients only – the tests do not take account of differences between estimates of constant terms, though including the constant terms in the tests makes no qualitative difference in the results.

[6]In these cases the hypothesis that the actual size equals the nominal size can be rejected at 1%. Four other cases can reject this hypothesis at 5%: the homoskedasticity test using the labour design at 5%, the Hausman–SCLS test using the synthetic design at 10%, and the Hausman–CLAD test using the synthetic design at 5% and 1%.

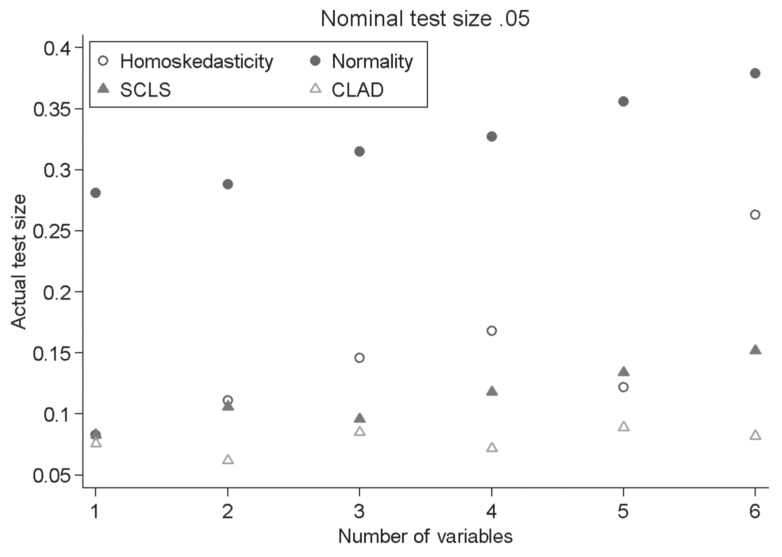

Figure 1. Actual test size and number of variables. Data in the experiments were generated with the synthetic design. $N = 100$

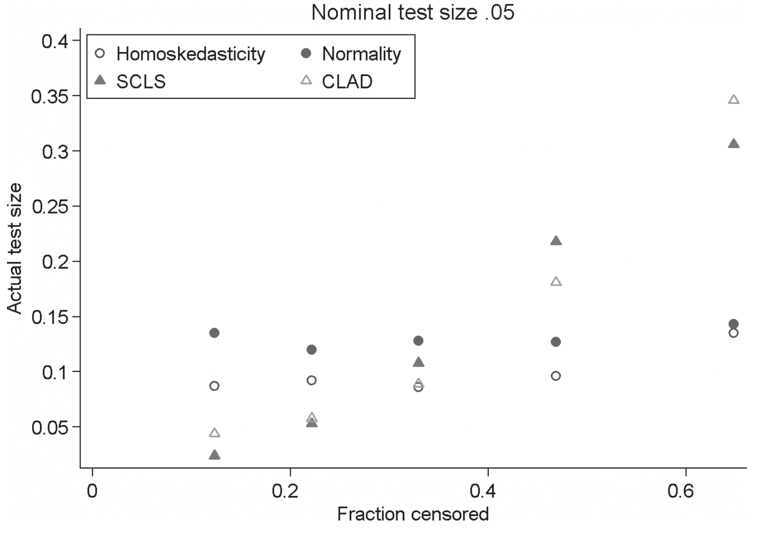

Figure 2. Actual test size and censoring. Data in the experiments were generated with the giving design. $N = 610$

about 12% to 65%. As the censoring increases the over-rejections in the Hausman tests become much more frequent; the conditional moment tests do not seem to be sensitive to increased censoring.[7] Figure 3 uses the giving design again (with the constant term returned to -16 as in equation 17) and increases the sample N from

[7]There are minor differences between the Figure 2 results at 0.47 censoring (constant term set at -16) and the Table 1 row 3 results. The two sets of results come from different Monte Carlo experiments.

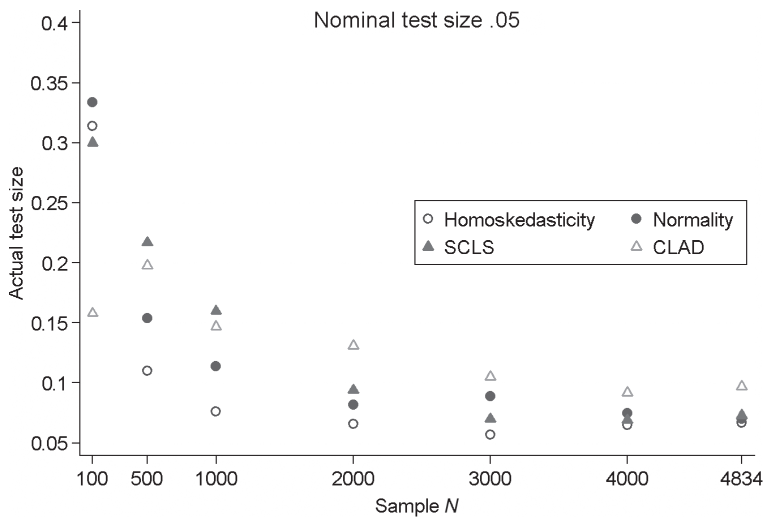

Figure 3. Actual test size and sample N. Data in the experiments were generated with the giving design

$N = 100$ to $N = 4{,}834$ (the maximum available in the *Center Panel*). Over-rejections become less frequent as the sample N increases, but even with large samples ($N = 2{,}000$ to 3,000) the tests remain oversized. Even with a very large sample ($N = 4{,}834$) the Hausman–CLAD test remains noticeably oversized.

4.2. Departures from homoskedasticity and normality

Table 2 presents the rejection frequencies for bootstrapped tests conducted using the giving design with various departures from homoskedasticity and normality. Columns (1)–(3) present the means and standard errors (in parentheses) of Tobit, SCLS and CLAD estimates of the coefficient on log income (true parameter = 0.968) to provide an indication of how seriously the various departures bias the Tobit estimates and how well the alternative estimation techniques perform. Columns (4) and (5) present the rejection frequencies from bootstrapped conditional moment tests for homoskedasticity and normality. Finally, columns (6) and (7) contain the rejection frequencies for the bootstrapped Hausman tests; these tests are based on all the slope coefficients (not just the log income coefficients). All the tests are conducted at 5% significance and the experiments are run at a small ($N = 610$) and large ($N = 2{,}500$) sample size.

The experiment in row (1a) shows the effects of heteroskedasticity increasing three times over the range of the data in a sample of size $N = 610$. The Tobit estimator of the income elasticity is biased too high, but only by about 7%. The finite sample bias in SCLS is yet higher, even though SCLS is consistent in the face of heteroskedasticity. There is negligible bias in the CLAD estimator. The rejection frequency of the conditional moment–homoskedasticity test is 0.172, showing weak power to detect the specification error. The Hausman tests seldom reject, a result

TABLE 2

Departures from homoskedasticity and normality

		Coefficient on log income (true coefficient = 0.968)			Rejection frequency at $\alpha = 0.05$			
					Conditional moment		Hausman	
Specification error	*N*	*Tobit (1)*	*SCLS (2)*	*CLAD (3)*	*Homo. (4)*	*Normal (5)*	*SCLS (6)*	*CLAD (7)*
Heteroskedasticity								
(1a) Increasing: three times	610	1.038	1.093	0.977	0.172	0.077	0.050	0.052
		(0.155)	(0.328)	(0.263)				
(1b)	2,500	1.026	0.985	0.964	0.484	0.192	0.074	0.100
		(0.074)	(0.132)	(0.127)				
(2a) Increasing: 15 times	610	1.105	1.118	0.976	0.488	0.126	0.068	0.068
		(0.161)	(0.380)	(0.274)				
(2b)	2,500	1.091	0.988	0.968	0.982	0.560	0.163	0.201
		(0.079)	(0.147)	(0.137)				
(3a) Decreasing: three times	610	0.902	1.026	0.951	0.144	0.054	0.064	0.052
		(0.144)	(0.240)	(0.222)				
(3b)	2,500	0.909	0.980	0.972	0.608	0.229	0.057	0.055
		(0.068)	(0.105)	(0.111)				
(4a) Decreasing: 15 times	610	0.806	0.999	0.943	0.857	0.221	0.067	0.076
		(0.133)	(0.189)	(0.194)				
(4b)	2,500	0.826	0.975	0.965	1.000	0.927	0.296	0.205
		(0.065)	(0.086)	(0.095)				
Non-normality								
(5a) Cauchy	610	1.320	0.971	0.968	0.991	1.000	0.392	0.337
		(0.297)	(0.048)	(0.017)				
(5b)	2,500	1.370	0.969	0.968	0.996	1.000	0.862	0.888
		(0.269)	(0.015)	(0.004)				
(6a) t_5	610	1.015	1.045	0.964	0.217	0.519	0.038	0.027
		(0.157)	(0.253)	(0.201)				
(6b)	2,500	1.009	0.984	0.974	0.340	1.000	0.059	0.054
		(0.075)	(0.104)	(0.098)				
(7a) χ^2_1	610	1.221	1.293	0.964	0.980	1.000	0.010	0.031
		(0.234)	(0.308)	(0.197)				
(7b)	2,500	1.195	1.198	0.969	1.000	1.000	0.012	0.307
		(0.115)	(0.117)	(0.092)				
(8a) χ^2_5	610	1.065	1.219	0.958	0.479	0.984	0.043	0.037
		(0.203)	(0.388)	(0.265)				
(8b)	2,500	1.058	1.100	0.969	0.988	1.000	0.042	0.059
		(0.096)	(0.146)	(0.129)				
(9a) $0.4\,\chi^2_1 + 0.6\,\chi^2_{25}$	610	1.005	1.136	0.967	0.110	0.305	0.046	0.038
		(0.170)	(0.349)	(0.270)				
(9b)	2,500	1.003	1.028	0.966	0.396	0.995	0.054	0.054
		(0.079)	(0.129)	(0.120)				

Notes: The first three columns contain the mean of the estimates of the log *income* coefficient across $M = 2{,}000$ simulations; standard errors of the estimates are in parentheses. Columns 4 and 5 contain the rejection frequencies for the conditional moment tests of homoskedasticity and normality. Columns 6 and 7 contain the rejection frequencies for the Hausman tests of Tobit vs. SCLS and CLAD. For each simulation the specification tests are conducted at significance level 0.05 and with bootstrapped sampling distributions.

that is not at all surprising given that the standard errors of the estimators are much larger than the amount of bias (0.155 for the Tobit and 0.263 for the CLAD, using the income elasticity to illustrate). Rerunning the experiment with $N = 2{,}500$ (row 1b) shows improved performance of the SCLS estimator, more power in the conditional moment–homoskedasticity test, but only slightly higher frequencies of rejection by the Hausman tests.

In row 2 the heteroskedasticity is worse – increasing 15 times over the range of the data – and the conditional moment–homoskedasticity test is better able to detect it. Despite the substantially worse heteroskedasticity, the Tobit estimator of the income elasticity is biased only 13% to 14% too high. The magnitude of bias is smaller than the estimator's standard errors in the smaller sample, and consequently the Hausman tests seldom reject. At $N = 2{,}500$ the Hausman tests reject somewhat more frequently but still in only 16% to 20% of the simulations; the standard errors of the point estimators are smaller but still about the same size as the magnitude of the bias. A similar pattern of results appears in the experiments, looking at decreasing heteroskedasticity (rows 3 and 4).

Among all the departures from homoskedasticity and normality in the experiments, the Cauchy errors – symmetric but with extremely long tails – in row 5 cause the most serious Tobit bias.[8] However, both SCLS and CLAD estimators are robust against Cauchy errors. The conditional moment test has very good power in detecting the non-normality. The Hausman tests reject in 34% to 39% of the simulations when $N = 610$ and in 86% to 89% when $N = 2{,}500$.

With the less extreme long tails of the t_5 errors, bias in the Tobit estimator is very slight. Again, the finite sample SCLS bias is somewhat worse than the Tobit bias when the sample is small. The conditional moment–normality test has reasonable power when $N = 610$ and very good power when $N = 2{,}500$. The Hausman tests seldom reject.

The remaining error distributions in rows 7–9 are asymmetric and SCLS is no longer consistent. In these experiments the SCLS biases are as bad as the Tobits or worse (hence it makes little sense to examine the Hausman–SCLS test results), while CLAD is unbiased. In row 7 the χ^2_1 errors cause large biases, and the Hausman–CLAD test rejects in 31% of the simulations with the larger sample, but the test seldom rejects in the smaller sample. Tobit estimators are only slightly biased when the errors are χ^2_5 or the chi-square mixture, and not surprisingly the Hausman–CLAD test seldom rejects. In all these experiments the conditional moment–normality test has very good power, except in the small-sample experiment with the chi-squared mixture.

How would a thorough specification evaluation using all three approaches play out in each of these misspecifications? Let us begin with the large-sample experiments. The largest Tobit bias is caused by the extremely long-tailed Cauchy errors (row 5b): the income elasticity estimator is 42% too high. Of course, whether a bias

[8]This is similar to Chay and Honoré's (1998) application in which 'abnormally long tails in the log-earnings distribution is the major source of misspecification in the [Tobit] estimates of the black-white earnings gap' (p. 24).

this size is economically important depends upon the application, but I assume that in most applications this amount of bias would be considered serious and economically important. The conditional moment–normality test detects the departure from normality and the Hausman tests seem to meet Hausman's goal of being powerful when the misspecification has serious effects.

At the other extreme, many of the misspecifications – the three-times-increasing (or decreasing) heteroskedasticity, t_5 errors, the χ^2_5 errors, and the chi-square mixture – cause relatively small biases: from 4% to 9% in the estimator for income elasticity. Again, whether this size of the bias is economically important depends upon the application, but it would seem reasonable to assume that in many applications a 4% to 9% bias would be considered neither serious nor economically important. The conditional moment tests reject often in these cases – 48% to 61% in the heteroskedasticity simulations and nearly always in the non-normality simulations. In contrast, Hausman tests reject in only 4% to 10% of the simulations, again meeting Hausman's goal: in these simulations the misspecifications do not have serious effects and the Hausman tests do not reject.

There is an area of ambiguity in between these two extremes. In the χ^2_1 misspecification the Tobit estimator for income elasticity is biased 23% too high. Readers may have different opinions as to whether this size of bias is serious, but (my opinion) I would be concerned about it. The conditional moment–normality test always rejects, but the Hausman–CLAD test rejects in only 31% of the replications. The 15-times-increasing (or decreasing) heteroskedasticity is similarly ambiguous: 13% to 15% bias in the Tobit income elasticity estimator, the conditional moment–homoskedasticity test nearly always rejects, but the Hausman tests reject in 30% of the replications or less. In a sense, this is how you would expect the Hausman test to perform: there is some ambiguity about whether the misspecifications have serious effects and the Hausman tests have some, though not a lot of power. Still, this kind of ambiguity reminds us of the potential danger in relying on the Hausman tests alone – in isolation from conditional moment tests and an informal comparison of point estimates – when evaluating a specification. More importantly for practical guidance, the simulations suggest that this kind of ambiguity is likely to arise when the misspecification is driven by strongly asymmetric errors, or perhaps by strong heteroskedasticity. To guard against this, when the Hausman–CLAD test fails to reject, but one of the conditional moment tests does reject and an informal comparison of the Tobit and CLAD point estimates suggests an economically important difference, a graphical examination of the residuals should be done to detect whether strongly asymmetric or heteroskedastic errors are the likely cause of the Hausman test's failure to reject. I illustrate such a graphical examination of residuals in the application below.

To summarize the large-sample experiments: the conditional moment tests have power to detect departures from homoskedasticity and normality even when the departures do not have serious effects on the point estimates. The Hausman tests have power, but only when the departures have a serious effect on the point estimates.

In the small-sample experiments the Hausman tests add little to what can be learned from the conditional moment tests and an informal comparison of point estimates. Even in the Cauchy $N = 610$ experiment where the bias is most severe, the Hausman tests reject in only 34% to 39% of the simulations. In the χ^2_1, $N = 610$ experiment the Hausman–CLAD test almost never rejects. For practical guidance, the small-sample experiments suggest reliance on conditional moment tests and an informal comparison of point estimates for specification evaluation.

4.3. Application: intergenerational transmission of charitable giving

Table 3 contains estimates of the intergenerational transmission of charitable giving. Once more the data are from the *Center Panel* but now the sample is restricted to adult children whose parents are still respondents in the PSID ($N = 2{,}384$). The dependent variable is the adult children's charitable giving (log). The independent variables are the parent's giving (log), the average income of the child over the past 5 years (log), the child's education, religious affiliation and marital status, and whether the child resides in the south. The first column contains the Tobit estimates; they indicate that the elasticity of children's giving with respect to parent's giving is 0.2 and the children's own income elasticity is 1.453. SCLS and CLAD estimates of these two parameters are similar to the Tobits (0.167 and 1.320; 0.175 and 1.346). The other point estimates are qualitatively similar across the three estimation methods. Most of the SCLS estimates in column 2 are closer to zero than their Tobit counterparts and most of the CLAD estimates in column 3 move back away from zero (back toward the Tobit estimates). Roughly speaking, the differences between the Tobit and SCLS/CLAD estimates are on the order of one standard error.

Despite the qualitative similarity of the point estimates across the three methods, the conditional moment tests for homoskedasticity and normality both show rejections with large values of the test statistics. Are the biases caused by departures from homoskedasticity and normality serious as judged by the Hausman tests? No. For both SCLS and CLAD, the Hausman test cannot reject the equality of the estimates with Tobit. Despite the strong conditional moment test rejections, the qualitative similarity of the point estimates and the failure-to-reject result of the Hausman tests justify carrying out the bulk of the estimation work in this application with Tobit.[9]

Before settling on this conclusion some judgement must be made about whether strongly asymmetric errors or strong heteroskedasticity could cause ambiguity in the Hausman test results. Starting with heteroskedasticity, Figure 4 plots the second moment residuals (minus one) from the Tobit model against the predicted values $Z_i\hat{\delta}_\mathrm{T}$ and fits a locally smooth regression to the residuals (see Chesher and Irish, 1987). The regression fit indicates that the error variance declines as the predicted value

[9] In some applications, interest may be focused on one estimate and the Hausman test can be conducted accordingly. For example, in this application, interest may be focused on the parent giving estimate: the Hausman test statistic for the Tobit–CLAD difference in this one estimate is 1.1 (P-value $= 0.302$).

TABLE 3

Intergenerational transmission of charitable giving: estimates and specification tests

Variables	*Tobit*	*SCLS*	*CLAD*
log parent's charitable giving	0.200	0.167	0.175
	(0.025)	(0.027)	(0.028)
log income	1.453	1.320	1.346
	(0.087)	(0.097)	(0.090)
Less than high school	−0.622	−0.655	−1.217
	(0.188)	(0.637)	(0.507)
Some college	0.478	0.387	0.398
	(0.118)	(0.110)	(0.129)
College	0.703	0.629	0.577
	(0.144)	(0.117)	(0.130)
Post-college	1.105	0.933	1.028
	(0.172)	(0.140)	(0.162)
Catholic	0.639	0.433	0.588
	(0.171)	(0.177)	(0.206)
Protestant	1.257	0.983	1.009
	(0.154)	(0.151)	(0.190)
Jewish	1.001	0.768	0.737
	(0.307)	(0.205)	(0.247)
Other religious affiliation	0.837	0.596	0.661
	(0.194)	(0.191)	(0.225)
Married	0.767	0.702	0.818
	(0.117)	(0.123)	(0.125)
South	0.113	0.064	0.028
	(0.105)	(0.097)	(0.111)
Constant	−16.996	−14.734	−14.999
	(0.909)	(1.002)	(0.961)
Conditional moment test			
Homoskedasticity	127.3	—	—
	[0.000]*		
Normality	200.1	—	—
	[0.000]*		
Hausman test	—	13.9	14.8
		[0.311]	[0.236]

Notes: The dependent variable is the log of children's charitable giving. Standard errors in parentheses (500 bootstrap replications used to calculate the SCLS and CLAD standard errors). Bootstrapped P-values in square brackets (based on 500 bootstrap replications of the test statistics). $N = 2{,}384$.

*Value of the test statistic is much larger than the largest bootstrapped value.

increases.[10] The decline is only 3.5 times over the range of the data, and the Table 2 experiments suggest that that amount of heteroskedasticity is not likely to cause much bias. Indeed, if declining heteroskedasticity causes serious bias and generates

[10]The rising 'curve' is not a regression fit, but rather the estimated expected values of the second moment residuals for the censored observations.

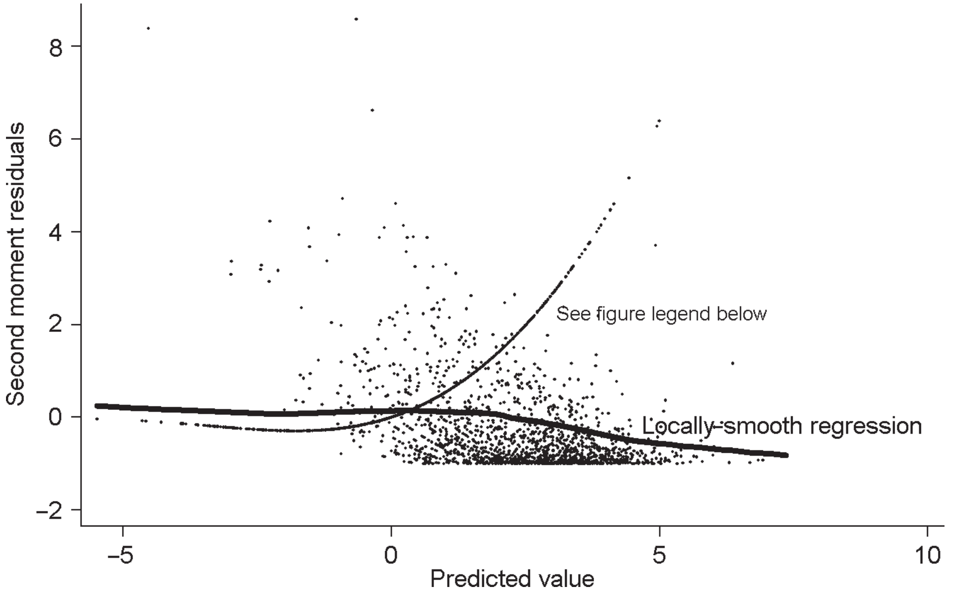

Figure 4. Second moment residuals and predicted values. The rising 'curve' is not a regression fit, but rather the points representing the estimated expected values of the second moment residuals for the censored observations

ambiguity in the Hausman tests, Table 2's declining heteroskedasticity experiments would lead us to expect SCLS estimates larger than the Tobit estimates, but that is not the case in Table 3.

That leaves asymmetry. Figure 5 plots a kernel density estimate using the uncensored residuals from the CLAD specification (Chesher, Lancaster and Irish, 1985). The density is certainly non-normal, but not extremely asymmetric. This is easier to see using the histogram of the residuals relative to the normal distribution (Handcock and Morris, 1999) shown in Figure 6. The relative histogram defines bins based on the normal distribution and then places the uncensored CLAD residuals within those bins; if the residuals were normal, the relative histogram would be uniform. The relative histogram shows more residuals (than a normal distribution) in both the lowest decile and highest decile, and more residuals in both the fourth and seventh/eighth deciles; hence, the departures from normality are roughly symmetric on both sides of the median. Overall the relative distribution is not drastically different from a uniform histogram (chi-squared tests show that only the difference at the lowest decile is statistically significant). This evidence, along with the closeness of the SCLS estimates to the CLAD estimates, suggests that asymmetric errors are not a large problem in this application, and do not cause ambiguity in the Hausman tests.

Although the informal comparison of point estimates and the Hausman tests both suggest that the departures from homoskedasticity and normality detected by the conditional moment tests do not cause serious bias in the point estimates, it does not necessarily follow that the marginal effects on the observable y calculated in the Tobit

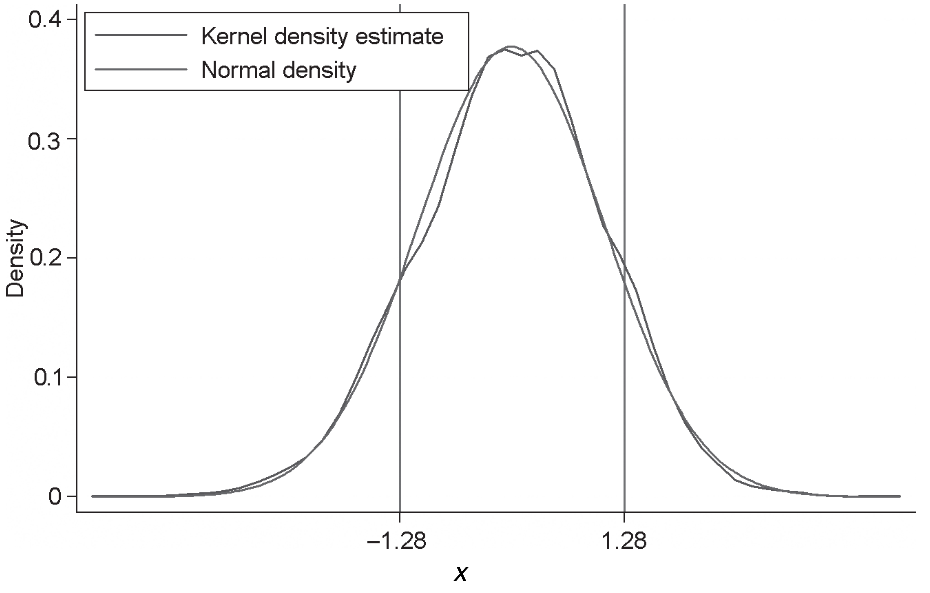

Figure 5. CLAD uncensored residuals

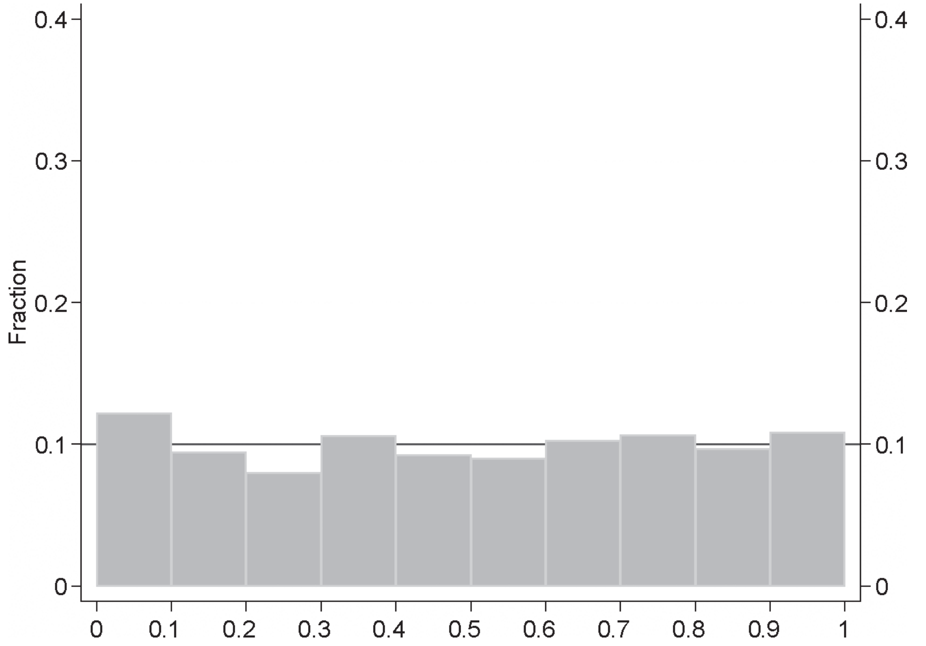

Figure 6. CLAD uncensored residuals relative to normal

model based on the homoskedastic-normal assumption should be accepted without additional scrutiny. One might judge that the heteroskedasticity and non-normality visually described in Figures 4 and 6 seem neither heteroskedastic enough nor non-normal enough to qualitatively affect the marginal effects on the observable *y*

TABLE 4

Intergenerational transmission of charitable giving: marginal effects of parent charitable giving

Marginal effects	*Tobit*	*CLAD*
$\partial E[y]/\partial x$	0.156	0.145
$\partial P[y>0]/\partial x$	0.028	0.017
$\partial E[y \mid y>0]/\partial x$	0.113	0.118

Notes: Marginal effects calculated using the point estimates from Table 3 columns 1 and 3. The CLAD marginal effects are based on a simulation of the random errors so that the errors mimic the heteroskedasticity pattern in Figure 4 and the distributional shape in Figure 6. The simulation is based on 10,000 draws of the random errors for each of the 2,384 observations.

relative to a Tobit approximation, but to my knowledge there has been no previous experience making such judgements. We can check this judgement by using the visual descriptions of the departures in Figures 4 and 6 along with the CLAD estimates from Table 3 to simulate the marginal effect of an increase in parent giving on the observable y, the probability $y>0$, and the observable y conditional on $y>0$. That is, simulate the random errors so that they mimic the empirical distribution in Figures 4 and 6. Table 4 contains the results: column 1 contains the marginal effects calculated from the Tobit model and column 2 contains the simulation – the two sets of marginal effects are very close. The simulated marginal effects on the observable y and the probability $y>0$ are slightly smaller than their Tobit counterparts, while the calculations of the marginal effect on the observable y conditional on $y>0$ are virtually identical in Tobit and CLAD.

Despite the strong rejections by the conditional moment tests, the graphical examination of the residuals and the similarity of the Tobit and CLAD-based marginal effects back up the Hausman tests' statistical justification for using Tobit. At this point different applied researchers would make different choices as to how to proceed. The justification for using Tobit notwithstanding, some researchers nevertheless would choose to use CLAD. Others would choose to carry out any remaining estimation work on this model using Tobit. Obviously, if there are only one or two remaining models to estimate most researchers would just use CLAD and simulate the marginal effects as described in the previous paragraph. Conversely, if there are numerous sensitivity checks to be conducted on the baseline model, many researchers would be inclined to carry out those checks using Tobit (and they would have a statistical justification for so doing). If there are only one or two remaining models to estimate the computational advantage of Tobit is negligible, but as the number of sensitivity checks to be conducted increases, so does Tobit's computational advantage.[11]

[11]The CLAD estimates in Table 3 took 11.5 minutes to calculate on the computer described at the end of section 3 (most of this time is due to bootstrapping the CLAD standard errors). The Hausman Tobit–CLAD test in Table 3 took 15 minutes to calculate (with $B=500$). SCLS is computationally less intensive than CLAD: the SCLS estimates in Table 3 took 3 minutes and the Tobit–SCLS Hausman test took 4 minutes. The conditional moment tests in Table 3 took 1 minute each.

V. Conclusion

It has become common practice in applied work to estimate a censored regression model with SCLS and CLAD in addition to Tobit, informally compare the three sets of point estimates to judge the economic importance of any differences, and conduct conditional moment tests of the homoskedasticity and normality assumptions upon which the Tobit estimator's consistency rests. The present paper uses Monte Carlo experiments to draw additional practical considerations for a thorough specification evaluation of censored regression models.

First, conditional moment test rejection of homoskedasticity or normality does not necessarily imply serious misspecification of the Tobit estimator. In large samples ($N = 2{,}500$ in the paper's experiments) the Hausman test can be used to test for serious misspecification, but in small samples ($N = 610$ in the experiments) it cannot. In a large sample if the misspecification detected with a conditional moment test is judged to be not serious according to the Hausman test, the researcher has a statistical justification for carrying out the bulk of a project's estimation work with Tobit. Second, when conducting the conditional moment and Hausman tests, rejection decisions should be based on the bootstrapped sampling distribution to avoid over-sized tests. Third, when a conditional moment test rejects either homoskedasticity or normality but the Hausman test fails to reject, the residuals should be graphically examined. The graphical examination can be done using a locally smooth regression for the second moment residuals (to reveal the heteroskedasticity pattern) and the relative distribution of the uncensored CLAD residuals (to reveal the non-normality pattern). The graphical examination is to judge the severity of the heteroskedasticity and asymmetry of the errors because the Monte Carlo experiments suggest that it is against severe heteroskedasticity and asymmetry that the Hausman test may have less power. The graphs can also be used to simulate the empirical distribution of the errors that in turn can be used to calculate marginal effects from the CLAD estimates. Finally, the Monte Carlo experiments suggest little value in using SCLS. When heteroskedasticity or non-normality is thought to cause serious bias in the Tobit estimator, CLAD is the better alternative.

Final Manuscript Received: February 2008

References

Amemiya, T. (1985). *Advanced Econometrics*, Harvard University Press, Cambridge, MA.

Chay, K. Y. and Honoré, B. E. (1998). 'Estimation of semiparametric censored regression models: an application to changes in Black-White earnings inequality during the 1960s', *Journal of Human Resources*, Vol. 33, pp. 4–38.

Chay, K. Y. and Powell, J. L. (2001). 'Semiparametric censored regression models', *Journal of Economic Perspectives*, Vol. 15, pp. 29–42.

Chesher, A. and Irish, M. (1987). 'Residual analysis in the grouped and censored normal linear model', *Journal of Econometrics*, Vol. 34, pp. 33–61.

Chesher, A., Lancaster, T. and Irish, M. (1985). 'On detecting the failure of distributional assumptions', *Annales de l'Insee*, Vol. 59/60, pp. 7–44.
Drukker, D. M. (2002). 'Bootstrapping a conditional moment test for normality after Tobit estimation', *The Stata Journal*, Vol. 2, pp. 125–139.
Ericson, P. and Hansen, J. (1999). 'A note on the performance of simple specification tests for the Tobit model', *Oxford Bulletin of Economics and Statistics*, Vol. 61, pp. 121–127.
Greene, W. H. (2003). *Econometric Analysis*, 5th edn, Prentice-Hall, Upper Saddle River, NJ.
Handcock, M. S. and Morris, M. (1999). *Relative Distribution Methods in the Social Sciences*, Springer-Verlag, New York.
Hausman, J. A. (1978). 'Specification tests in econometrics', *Econometrica*, Vol. 46, pp. 1251–1271.
Horowitz, J. L. (2001). 'The bootstrap', in Heckman J. J. and Leamer E. (eds), *Handbook of Econometrics*, Vol. 5, Elsevier, Amsterdam.
Horowitz, J. L. and Neumann, G. R. (1987). 'Semiparametric estimation of employment duration models', *Econometric Reviews*, Vol. 6, pp. 5–40.
Horowitz, J. L. and Neumann, G. R. (1989). 'Semiparametric testing in censored regression models, parametric and semiparametric methods', *Journal of Applied Econometrics*, Vol. 4, pp. s61–s86.
Jolliffe, D., Krushelnytskyy, B. and Semykina, A. (2001). 'Censored least absolute deviations estimator: CLAD', *STATA Technical Bulletin*, Vol. 10, pp. 240–244.
Moffitt, R. (1984). 'Estimation of a joint wage-hours labor supply model', *Journal of Labor Economics*, Vol. 2, pp. 550–566.
Newey, W. (1987). 'Specification tests for distributional assumptions in the Tobit model', *Journal of Econometrics*, Vol. 34, pp. 125–146.
Pagan, A. and Vella, F. (1989). 'Diagnostic tests for models based on individual data: a survey', *Journal of Applied Econometrics*, Vol. 4, pp. s29–s59.
Paarsch, H. J. (1984). 'A Monte Carlo comparison of estimators for censored regression models', *Journal of Econometrics*, Vol. 24, pp. 197–213.
Powell, J. L. (1984). 'Least absolute deviations estimation for the censored regression model', *Journal of Econometrics*, Vol. 25, pp. 303–325.
Powell, J. L. (1986). 'Symmetrically trimmed least squares estimation for Tobit models', *Econometrica*, Vol. 54, pp. 1435–1460.
Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*, Chapman and Hall/CRC, London.
Skeels, C. L. and Vella, F. (1999). 'A Monte Carlo investigation of the sampling behavior of conditional moment tests in Tobit and probit models', *Journal of Econometrics*, Vol. 92, pp. 275–294.
Wilhelm, M. O. (2006). 'New data on charitable giving in the PSID', *Economics Letters*, Vol. 92, pp. 26–31.
Wilhelm, M. O. (2007). 'The quality and comparability of survey data on charitable giving', *Nonprofit and Voluntary Sector Quarterly*, Vol. 36, pp. 65–84.
Wilhelm, M. O., Brown, E., Rooney, P. M. and Steinberg, R. (2001). *The Center on Philanthropy Panel Study* [machine-readable data file]/Director and Principle Investigator, Mark O. Wilhelm; Co-Principle Investigators, Eleanor Brown, Patrick M. Rooney, and Richard Steinberg; Sponsored by Atlantic Philanthropies. In the *Panel Study of Income Dynamics* Wave XXXII [machine-readable data file]/Director and Principle Investigator, Frank P. Stafford; Associate Director and Principle Investigator, Robert F. Schoeni; Co-Principle Investigators, Jacquelynne S. Eccles, Katherine McGonagle, Wei-Jun Jean Yeung, and Robert B. Wallace. Ann Arbor, MI: Institute for Social Research, The University of Michigan.