Applied Economics, 2003, **35**, 1163–1169

Estimation of a Tobit model with unknown censoring threshold

THOMAS W. ZUEHLKE

Department of Economics, Florida State University, Tallahassee, Florida 32306, USA
E-mail: tzuehlke@mailer.fsu.edu

Conventional wisdom suggests that only the estimated intercept is affected by imposition of a zero censoring threshold on a Tobit model. This is true for Heckman–Lee estimation. For maximum likelihood (ML) estimation, however, it is only true if the censoring threshold is known and is subtracted from the dependent variable. Failure to properly transform the dependent variable prior to ML estimation of a zero threshold Tobit model will generally bias the coefficient estimates. A long neglected topic is ML estimation of a Tobit model with common, but unknown, censoring threshold. This paper shows that the ML estimator of the censoring threshold is the minimum order statistic from the observed subsample, and that existing software for estimation of a zero-threshold Tobit model is easily adapted to include estimation of the censoring threshold.

I. INTRODUCTION

Conventional wisdom suggests that only the estimated intercept is affected by imposition of a zero censoring threshold on a Tobit model.[1] This is true for Heckman–Lee (HL) estimation. For maximum likelihood (ML) estimation, however, it is only true if the censoring threshold is known and is subtracted from the dependent variable. Failure to properly transform the dependent variable prior to ML estimation of a zero threshold Tobit model will generally bias the coefficient estimates. A long neglected topic is ML estimation of a Tobit model with common, but unknown, censoring threshold. This paper shows that the ML estimator of the censoring threshold is the minimum order statistic from the observed subsample, and that existing software for estimation of a zero-threshold Tobit model is easily adapted to include estimation of the censoring threshold.

II. THE TOBIT MODEL

The model considered in this paper has been classified as a 'Type I Tobit' model by Amemiya (1985). The form of the latent regression is:

$$Y_i^* = \alpha + \mathbf{X_i}\beta + \sigma\varepsilon_i$$

where $\varepsilon_i \sim$ i.i.d.N(0, 1). The dependent variable Y_i^* is not observed. Instead, a censoring indicator, J_i, is observed where

$$J_i = 1 \quad \text{if} \quad Y_i^* \geqslant \delta$$

and

$$J_i = 0 \quad \text{if} \quad Y_i^* < \delta$$

The parameter δ is a common censoring threshold.[2] The observed dependent variable, denoted Y_i, equals Y_i^* if $J_i = 1$.

[1] Examples are numerous. Amemiya (1985, 363) denotes the censoring limit by y_0, and states that a zero censoring limit can be imposed 'without essentially changing the model, whether y_0 is known or unknown, because y_0 can be absorbed into the constant term of the regression'. Maddala (1983, 159) specifies the Tobit model with zero censoring limit in his Equation (6.1), and states that 'the Tobit model can be specified as in Equation (6.1) without loss of generality'. Finally, Greene (2000, 906) states that he will 'assume that the censoring point is zero, although this is only a convenient normalization'.
[2] As presented here, an observation is censored if it is strictly less than δ. In what follows, the only result that changes if an observation is censored when it is less than or equal to δ, is that the maximum likelihood estimator of δ corresponds to the supremum of the likelihood function rather than the maximum.

Applied Economics ISSN 0960–3107 print/ISSN 1466–4305 online © 2003 Taylor & Francis Ltd
http://www.tandf.co.uk/journals
DOI: 10.1080/0003684032000081339

The value of Y_i is missing (although commonly coded as zero) if $J_i = 0$. The vector of regressors, $\mathbf{X_i}$, is observed regardless of any censoring of the dependent variable.

If δ is known, then the model may be expressed in equivalent form as a zero-threshold model. Subtracting δ from both sides of the latent regression gives

$$y_i^* = \gamma + \mathbf{X_i}\beta + \sigma\varepsilon_i$$

where $y_i^* = Y_i^* - \delta$ and $\gamma = \alpha - \delta$. The censoring indicator, J_i, is determined as

$$J_i = 1 \quad \text{if} \quad y_i^* \geqslant 0$$

and

$$J_i = 0 \quad \text{if} \quad y_i^* < 0$$

The observed dependent variable, y_i, equals y_i^* if $J_i = 1$. The value of y_i is missing if $J_i = 0$.

Heckman (1976) and Lee (1976) provide a particularly simple estimator for the Tobit model. In the first stage, a Probit model is estimated using only the qualitative information in the observed values of J_i. The log-likelihood function of the Probit model is

$$\ln L(\alpha, \beta, \sigma, \delta) = \sum_{i=1}^{n}\left\{J_i \ln\left[1 - \Phi\left(\frac{\delta - \alpha - \mathbf{X_i}\beta}{\sigma}\right)\right] + (1 - J_i)\ln\Phi\left(\frac{\delta - \alpha - \mathbf{X_i}\beta}{\sigma}\right)\right\}$$

where $\Phi()$ denotes the standard normal distribution function. Without quantitative information, only the standardized parameters $\lambda = (\alpha - \delta)/\sigma$ and $\pi = \beta/\sigma$ are identified. The identification conditions $\sigma = 1$ and $\delta = 0$ may be imposed on the Probit model without loss of generality.[3] The first stage estimates of λ and π are then used to construct an auxiliary regressor based on the conditional expectation $E(\varepsilon_i \mid J_i = 1)$. In the second stage, ordinary least squares (OLS) is applied to the subsample regression function

$$Y_i = \alpha + \mathbf{X_i}\beta + \sigma\left[\frac{\phi(-\hat{\lambda} - \mathbf{X_i}\hat{\pi})}{1 - \Phi(-\hat{\lambda} - \mathbf{X_i}\hat{\pi})}\right] + \eta_i$$

where $\phi()$ denotes the standard normal density function. The zero threshold form of the subsample regression function is obtained by subtracting δ from both sides of the equation. Since the second stage regression is estimated by OLS, the only effect of subtracting δ from the dependent variable is to reduce the estimated intercept accordingly. The estimate of β, and the maximized value of the likelihood function from the second stage regression are unchanged. This is probably why $\delta = 0$ is often mistaken as an identification condition for the Tobit model.

Assuming that δ is known, the log-likelihood function of the Tobit model is

$$\ln L(\alpha, \beta, \sigma) = \sum_{i=1}^{n}\left\{J_i\left[-\ln(\sigma) + \ln\phi\left(\frac{Y_i - \alpha - \mathbf{X_i}\beta}{\sigma}\right)\right] + (1 - J_i)\ln\Phi\left(\frac{\delta - \alpha - \mathbf{X_i}\beta}{\sigma}\right)\right\}$$

The zero threshold form of the log-likelihood function is

$$\ln L(\gamma, \beta, \sigma) = \sum_{i=1}^{n}\left\{J_i\left[-\ln(\sigma) + \ln\phi\left(\frac{y_i - \gamma - \mathbf{X_i}\beta}{\sigma}\right)\right] + (1 - J_i)\ln\Phi\left(\frac{-\gamma - \mathbf{X_i}\beta}{\sigma}\right)\right\}$$

It is important to note that the zero threshold log-likelihood function is expressed in terms of the transformed dependent variable, $y_i = Y_i - \delta$, not the original dependent variable, Y_i. If the dependent variable y_i is properly constructed, then maximization of $\ln L(\gamma, \beta, \sigma)$ is equivalent to maximization of $\ln L(\alpha, \beta, \sigma)$. The estimates of β and σ are identical, and the estimates of γ and α are related as $\hat{\alpha} = \hat{\gamma} + \delta$. When the censoring threshold is known, either model can be estimated, provided that the proper dependent variable is constructed.

This does not imply that the restriction $\delta = 0$ is an identification condition for the Tobit model. Failure to construct the transformed dependent variable, $y_i = Y_i - \delta$, when estimating a zero threshold Tobit model will constrain the maximized value of the likelihood function, and will result in biased coefficient estimates. The intuition behind this bias is simple. When expressed as a zero threshold model, the intercept of the regression model is reduced by δ. If the dependent variable is not properly transformed, then this reduction in mean is not reflected in the empirical distribution of Y_i, and any decrease in the intercept degrades the fit of the observed subsample. Since the Tobit ML estimates are the solution to a set of simultaneous nonlinear implicit functions, the effects of a failure to properly transform the dependent variable will spill over to the estimates of β and σ. In contrast, the HL estimator of β is invariant to the choice of δ, because it fails to impose the functional relationship between the parameters in the first and second stages of estimation. Of course, this is also the source of the inefficiency of the HL estimator.

III. ML ESTIMATION OF THE CENSORING THRESHOLD

This section considers the case where δ is unknown.[4] The maximization problem is complicated by the fact that the

[3] In contrast, when the quantitative information in the observed subsample of Y_i is included, the parameters α, β, δ, and σ are all identified.

[4] To my knowledge, estimation of the Tobit model when δ is unknown has not been considered in the literature. Of course, there have been many generalizations of the Tobit model that relax the assumption of a common censoring threshold. These models are more appropriate in many applications than the standard Tobit model. Nevertheless, the standard Tobit model is still widely used.

sample space of the observed dependent variable, Y_i, is a function of the unknown parameter δ.[5] The likelihood function above, $\ln L(\alpha, \beta, \sigma)$, is only valid for values of δ that keep all observed values of Y_i in the sample space. Specifically, it is required that $\delta \leqslant Y_i$ for all observations in the observed subsample. If $\delta > Y_i$ for any observation, there is a contradiction; the observed value of Y_i should have been censored. This restriction may be simplified to $\delta \leqslant Y_1$, where Y_1 denotes the minimum order statistic from the observed subsample. The minimum order statistic for the observed subsample provides an upper bound for the censoring threshold.

The log-likelihood function of the Tobit model may be generalized to include this restriction on the sample space through the use of an indicator function. Specifically, let $\Gamma(\delta) = 1$ if $\delta \leqslant Y_1$ and $\Gamma(\delta) = 0$ if $\delta > Y_1$. If $\Gamma(\delta) = 0$, a value of δ has been chosen that is inconsistent with the observed data. Using this notation, the log-likelihood function may be written as

$$\ln L(\alpha, \beta, \sigma, \delta) = \sum_{i=1}^{n} \left\{ J_i \left[-\ln(\sigma) + \ln \phi\left(\frac{Y_i - \alpha - \mathbf{X_i}\beta}{\sigma}\right) + \ln \Gamma(\delta) \right] + (1 - J_i) \ln \Phi\left(\frac{\delta - \alpha - \mathbf{X_i}\beta}{\sigma}\right) \right\}$$

This generalization simply recognizes that the probability of (Y_i, J_i) pairs outside the sample space is zero.[6]

Where it exists, the score equation for δ is

$$\frac{\partial \ln L(\alpha, \beta, \sigma, \delta)}{\partial \delta} = \sum_{i=1}^{n} (1 - J_i)\sigma^{-1} \phi\left(\frac{\delta - \alpha - \mathbf{X_i}\beta}{\sigma}\right) \times \left[\Phi\left(\frac{\delta - \alpha - \mathbf{X_i}\beta}{\sigma}\right) \right]^{-1}$$

which is always positive. The log-likelihood function is increasing in δ, for any $\delta < Y_1$. The log-likelihood function is discontinuous at $\delta = Y_1$, however. For any $\delta > Y_1$, the value of $\Gamma(\delta)$ is zero, and the log-likelihood function falls precipitously; essentially to minus infinity. The maximum to the likelihood function over δ occurs at Y_1. The ML estimator, $\hat{\delta}$, is the minimum order statistic from the observed subsample of Y.[7]

Substituting $\hat{\delta} = Y_1$ into $\ln L(\alpha, \beta, \sigma, \delta)$ gives the concentrated log-likelihood function

$$\ln L(\alpha, \beta, \sigma, \hat{\delta}) = \sum_{i=1}^{n} \left\{ J_i \left[-\ln(\sigma) + \ln \phi\left(\frac{Y_i - \alpha - \mathbf{X_i}\beta}{\sigma}\right) \right] + (1 - J_i) \ln \Phi\left(\frac{\hat{\delta} - \alpha - \mathbf{X_i}\beta}{\sigma}\right) \right\}$$

Letting $y_i = Y_i - \hat{\delta}$ and $\gamma = \alpha - \hat{\delta}$, the concentrated log-likelihood function may be written as

$$\ln L(\gamma, \beta, \sigma, \hat{\delta}) = \sum_{i=1}^{n} \left\{ J_i \left[-\ln(\sigma) + \ln \phi\left(\frac{y_i - \gamma - \mathbf{X_i}\beta}{\sigma}\right) \right] + (1 - J_i) \ln \Phi\left(\frac{-\gamma - \mathbf{X_i}\beta}{\sigma}\right) \right\}$$

This is just the log-likelihood function of the zero threshold Tobit model, where the ML estimator $\hat{\delta}$ is subtracted from the dependent variable. Consequently, software for estimation of a zero threshold Tobit model is easily adapted to the case of an unknown censoring threshold.[8] Conventional test statistics for the Tobit model are valid conditional on $\hat{\delta}$.

IV. DATA GENERATION

The purpose of the Monte Carlo portion of this study is to examine how imposition of a zero censoring threshold affects the estimate of β when the true censoring threshold is positive. In order to focus on this topic, the structure of the model is kept as simple as possible. The model contains an intercept, α, and a single regressor, X, with coefficient, β. The regressor, X, is a random draw of independent standard normals, and is fixed in repeated sampling. The assumption of a zero mean and unit variance for the regressor involves no loss in generality. In practice, standardizing a regressor will simply scale the coefficient estimate without affecting the precision of the estimate or the fit of the model. The disturbance, ε, is a sequence of independent standard normals that are statistically independent of X. Given this structure, the unconditional mean of the latent dependent variable Y^* is controlled by the single parameter α.

[5] The sample space of the latent dependent variable, y_i^*, is the set of real numbers, while the sample space of the observed dependent variable, y_i, is the set of nonnegative real numbers.

[6] Current econometric software is written for the case of known δ, and does not control for the discontinuity in the likelihood function induced by the restricted sample space of Y. For example, if the transformed dependent variable, $y_i = Y_i - \delta$, is constructed using an invalid value of δ (one exceeding Y_1), and zero threshold software is used to maximize $\ln L(\gamma, \beta, \sigma)$, incorrect values for the likelihood function will be reported without any error message or warning. Furthermore, for invalid δ, the reported value of the log-likelihood function will differ from one software package to the next, depending on how the censored subsample is determined. (See footnote 8.)

[7] The problem is similar to that of finding the ML estimator of θ for a random sample of uniform random variables on the interval $(0, \theta)$. The ML estimator for this problem is the largest order statistic.

[8] A practical problem arises with software packages that determine the censored subsample from the numerical value of the dependent variable, rather than from a separate binary variable that indicates censoring. If observations for which $y_i \leqslant 0$ are treated as censored, then the observation corresponding to Y_1 will be switched from the observed to the censored subsample after construction of y_i. The simplest solution to this problem is to choose as the estimate of δ a number 'slightly' smaller than Y_1.

Two factors affecting the performance of Tobit estimators are the fit of the latent regression and the degree of censoring. Nelson (1984) shows that the variance of the Tobit estimator is affected proportionally by a change in σ. When comparing the *relative* performance of the estimator, the choice of σ is arbitrary. Changes in σ affect the absolute, but not the relative scale of the variances. If the normalization $\sigma = (1-\beta^2)^{1/2}$ is adopted, then the fit of the latent regression is controlled by the single parameter, β. For this choice of σ, the slope coefficient, β, corresponds to the correlation coefficient between Y^* and X. The parameters δ and α affect the probability of censoring only through the difference $\delta - \alpha$. Constructing α as $\alpha = \delta - \theta$, where θ is the critical value of the standard normal distribution function that gives the desired degree of censoring, results in equivalent variation in α and δ that leaves the probability of censoring unchanged. This allows an examination of the effect of an increase in δ on the performance of the zero threshold Tobit model, without confounding variation in the degree of censoring.

A brief summary of the Monte Carlo process is as follows:

(1) The regressor, X, is drawn. It is fixed across repetitions.
(2) The disturbance ε is drawn. Given the regressor, X, and the parameters β, δ, and θ, the values of Y and J are computed.
(3) Coefficient estimates are obtained for each of the estimators.
(4) Steps 2 and 3 are repeated on successive repetitions, and sample moments for the estimators are compiled across repetitions.

The data generation process was carefully structured in order to limit intra-experiment random variation.[9] Each estimator is applied to the same sequence of data sets for any given parameter combination, (β, δ, θ). This will limit random variation in comparisons across estimators. In addition, the same sequence of independent standard normal errors, ε, will be used to construct the data sequence (Y, J) required for each distinct parameter combination.[10] This will limit random variation in comparisons across parameter values.

V. RESULTS

Zero threshold software may be used to obtain ML estimates of the Tobit model, whether δ is known or unknown. The essential step is construction of the proper dependent variable. If δ is known, the user must construct the transformed dependent variable, $y_i = Y_i - \delta$. This estimator will be referred to as the KT-MLE (known threshold). If δ is unknown, the user must construct $y_i = Y_i - Y_1$. This will be called the MOS-MLE (minimum order statistic). Use of the original dependent variable, Y_i, in conjunction with zero threshold software will be called the ZT-MLE (zero threshold). For nonzero δ, this estimator is generally biased.

Section II showed that the HL estimator of β is invariant to the value of δ. The MOS-MLE and KT-MLE are also invariant to variation in δ, provided the degree of censoring is held constant. Recall that the degree of censoring is held constant in the face of an increase in δ, by imposing a symmetric increase in α that leaves the difference $\delta - \alpha$ unchanged. Since an increase in α results in one-to-one increases in Y_i and Y_1, simultaneous increases in α and δ will leave the transformed dependent variables, $Y_i - Y_1$ and $Y_i - \delta$, unchanged. Only the ZT-MLE will be affected by variation in δ.

The results of this section are based on 500 repetitions of samples of size 100. Figure 1 plots the mean bias of the estimate, as a percentage of the true β, against the true value of the censoring threshold, δ. Figure 2 plots the root mean square error of the estimate, again as a percentage of the true β, against the censoring threshold. The mean values appear as a solid line, while a 95% asymptotic confidence interval is given by the surrounding dotted lines. Results are shown for censored subsamples of 25%, 50% and 75%. In both figures, the explanatory power of the latent regression is 50%.[11]

Figure 1 shows that the mean bias of the ZT-MLE is increasing in δ and increasing in the degree of censoring. The mean bias of the ZT-MLE is linear in δ for any given degree of censoring. For example, with a censored subsample of 50%, a one standard deviation increase in δ results in a 75% increase in mean bias.[12] As noted earlier, the MOS-MLE and HL estimates of β are invariant to changes in δ. Since the mean bias of both of these estimators is statistically zero, results are only reported for the case of censored subsamples of 25%. Finally, the mean bias of the KT-MLE (for all values of δ) is given by the vertical intercept of the ZT-MLE. The mean bias of the KT-MLE is also statistically zero.

Figure 2 reports the root mean square error (RMSE) of the estimators as a function of δ. The RMSE of the ZT-MLE is increasing in δ and increasing in the degree of censoring. For δ greater than 0.5, the RMSE is essentially linear in δ for any given degree of censoring. For a censored subsample of 50%, a one standard deviation increase in δ results in approximately a 90% increase in RMSE.

[9] See Hendry (1984).
[10] The independent standard normals were obtained with the algorithm of Forsythe *et al.* (1977).
[11] The results obtained when the explanatory power was increased to 75% or decreased to 25% were almost indistinguishable from those presented in Figs 1 and 2. The mean values of each specification generally fell within the envelope formed by the 95% confidence intervals.
[12] Examination of the log-likelihood function shows that it is the size of δ relative to σ that determines its empirical significance.

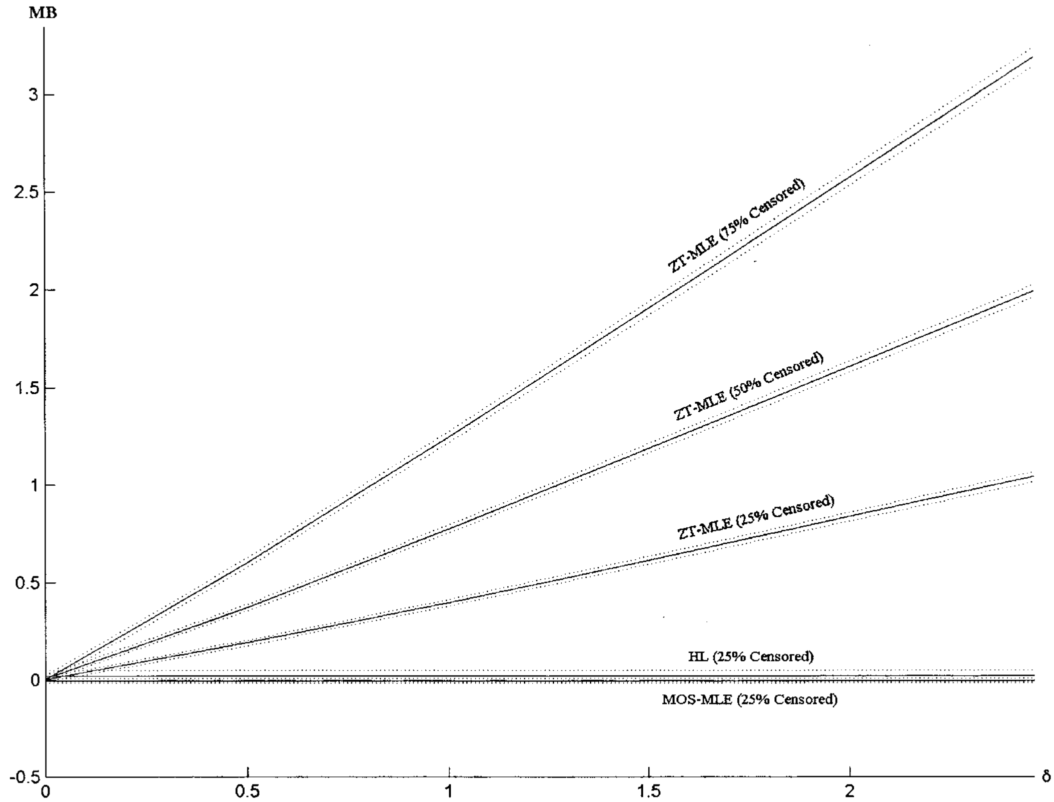

Fig. 1. *Mean bias (%) as a function of* δ (n = *100, repetitions* = *500,* $R^2 = 0.50$)

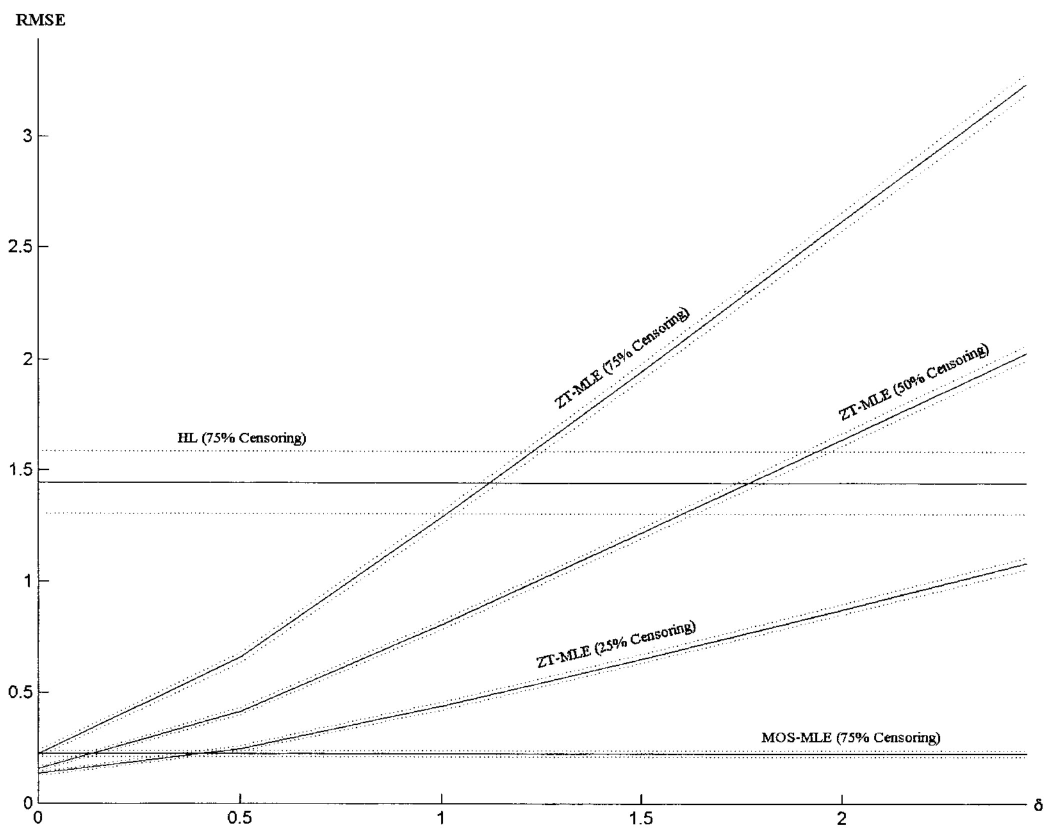

Fig. 2. *Root mean square error (%) as a function of* δ (n = *100, repetitions* = *500,* $R^2 = 0.50$)

Results for the HL and MOS-MLE are only reported for censored subsamples of 75%. This figure represents a 'worst case' scenario for the HL estimator. As would be expected, the MOS-MLE has a MSE advantage on the HL estimator. As the degree of censoring is reduced, the RMSE of both of these estimators falls. The size of the MSE advantage diminishes as the degree of censoring is reduced.

Despite its bias, the ZT-MLE has a MSE advantage over the HL estimator for sufficiently small δ. With censored subsamples of 75%, the RMSE of the ZT-MLE is lower than that of the HL estimator for δ less than about 1.1 standard deviations. As the degree of censoring falls, the RMSE of the HL estimator falls. For censored subsamples of 25% (not shown), the RMSE of the ZT-MLE is lower than that of the HL estimator for δ less than about 0.75 standard deviations. This is just a reflection of the relative inefficiency of the HL estimator.

Finally, the RMSE of the KT-MLE (for all values of δ) is given by the vertical intercept of the ZT-MLE. The RMSE of the KT-MLE and MOS-MLE are virtually identical for like degrees of censoring. Since imposing a valid constraint can only improve the precision of the estimates, the KT-MLE is expected to have a MSE advantage over the MOS-MLE. The results of Fig. 2 show that this MSE advantage is trivial for the sample sizes considered here. Even for observed subsamples averaging 25 observations (where 75% of 100 observations are censored), the precision of the MOS estimator of δ is sufficient to yield almost indistinguishable results.

While the ZT-MLE of β is biased upwards, its corresponding *t*-statistic is biased downward. Figure 3 depicts the relationship between the censoring threshold, δ, and the mean value of the *t*-statistic obtained under the null hypothesis $\beta = 0$, for different degrees of censoring and explanatory power. First, the degree of censoring is fixed at 75%, and the degree of explanatory power, represented by R^2, is decreased from 75%, to 50%, to 25%. In each case, the mean value of the *t*-statistic is decreasing in δ, and this relationship is shifted downward by a decrease in explanatory power. For sufficiently large values of δ and sufficiently small values of R^2, the outcome of the test is altered by use of the ZT-MLE.

Figure 3 also illustrates the effects of variation in the degree of censoring. The degree of explanatory power is fixed at 25%, and the degree of censoring is increased from 25%, to 50%, to 75%. The mean value of the *t*-statistic is again decreasing in δ, and the relationship is shifted

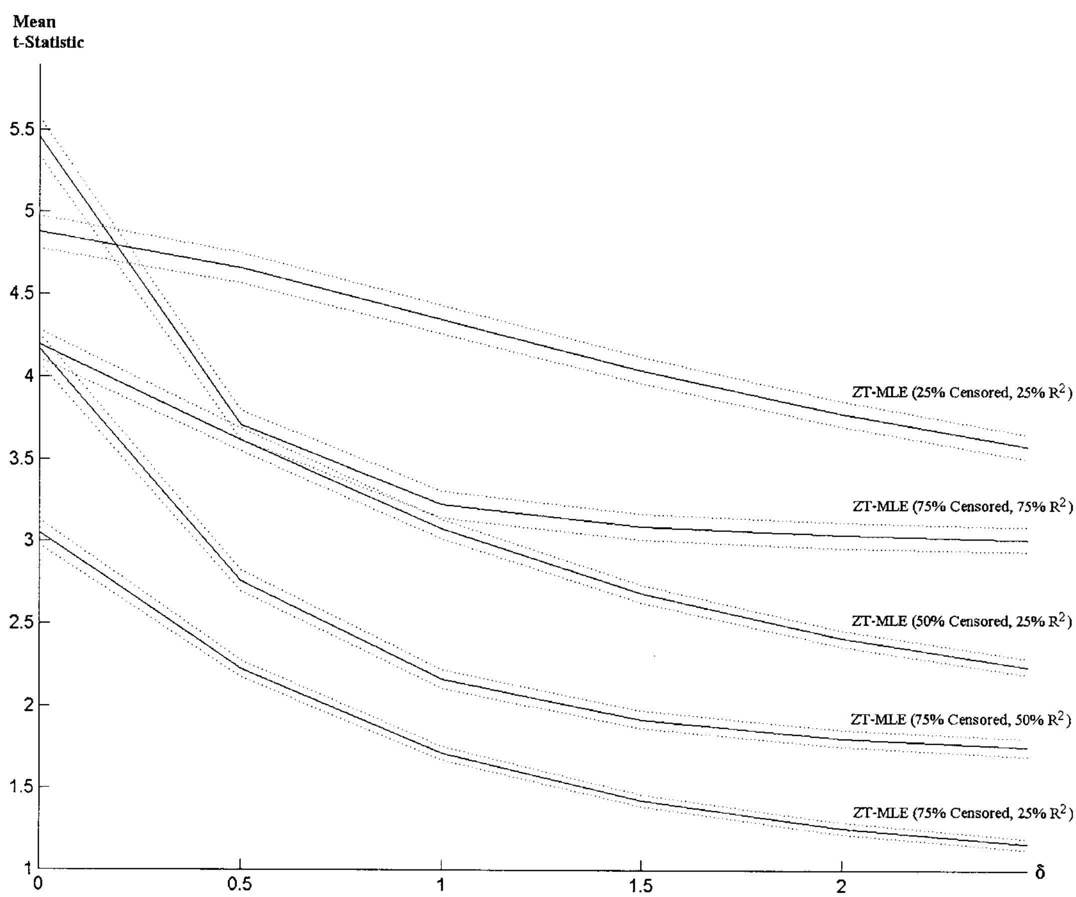

Fig. 3. *Mean value of* t*-statistic* ($H_o: \beta = 0$) *as a function of* δ (n = *100, repetitions* = *500*)

downward by an increase in the degree of censoring. For sufficiently large values of δ and a sufficiently high degree of censoring, the outcome of the test is altered by use of the ZT-MLE.

VI. CONCLUSION

This paper shows that proper treatment of the censoring threshold when estimating a Tobit model is no trivial matter. If the censoring threshold is known, it must be subtracted from the dependent variable prior to ML estimation of a zero-threshold model. Failure to do so will result in an upward bias in the coefficient estimates, and a downward bias in conventional t-statistics. The ML estimate of a common censoring threshold is shown to be the minimum order statistic from the observed subsample. Existing software for estimation of the zero-threshold Tobit model is easily adapted to this case, by simply subtracting the ML estimate of the censoring threshold from the dependent variable prior to use. The MSE of this generalized ML estimator is found to rival that of the conventional ML estimator (where the censoring threshold is known), even for relatively small observed subsamples. Whether the censoring threshold is known or unknown, failure to properly transform the dependent variable mis-specifies the Tobit model and results in a substantial increase in MSE.

REFERENCES

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

Forsythe, G. E., Malcolm, M. A. and Moler, C. B. (1977) *Computer Methods for Mathematical Computations*, Prentice-Hall, Englewood Cliffs, New Jersey.

Greene, W. H. (2000) *Econometric Analysis, Fourth Edition*, Prentice-Hall, Englewood Cliffs, New Jersey.

Heckman, J. J. (1976) The common structure of statistical models of truncation, sample selection, and limited dependent variables and a simple estimator for such models, *Annals of Economic and Social Measurement*, **5** (Fall), 475–92.

Hendry, D. F. (1984) Monte Carlo experimentation in econometrics, in *Handbook of Econometrics* **2**, (Eds) Z. Griliches and M. D. Intriligator, Elsevier Science Publishers, Amsterdam, pp. 937–76.

Lee, L.-F. (1976) *Estimation of limited dependent variable models by two-stage methods*, Doctoral Dissertation, University of Rochester.

Maddala, G. S. (1983) *Limited-Dependent and Qualitative Variables in Econometrics*, Cambridge University Press, Cambridge.

Nelson, F. D. (1984) Efficiency of the two-step estimator for models with endogenous sample selection, *Journal of Econometrics*, **24** (January/February), 181–96.