VATT-KESKUSTELUALOITTEITA
VATT-DISCUSSION PAPERS

16

NONLINEARITY AND HETEROSCEDASTICY IN TOBIT MODELS: Score tests for misspecification with one degree of freedom

Ilpo Suoniemi

Government Institute for Economic Research
Helsinki 1992

ISBN 951-561-020-6
ISSN 0788-5016

Valtion taloudellinen tutkimuskeskus
Hämeentie 3, 00530 Helsinki

Valtion painatuskeskus
Pasilan VALTIMO
Helsinki 1992

Ilpo Suoniemi

SUONIEMI, ILPO: NONLINEARITY AND HETEROSCEDASTICITY IN TOBIT MODELS: SCORE TESTS FOR MISSPECIFICATION WITH ONE DEGREE OF FREEDOM. Helsinki: VATT, Valtion taloudellinen tutkimuskeskus, 1992. (C, ISSN 0788-5016, No 16). ISBN 951-561-020-6.

ABSTRACT: In contrast with regression models Tobit models are not robust against specification errors such as heteroscedasticity and non-normality. There is thus a need for specification tests which have power against a wide range of alternatives and are easy to implement. In the paper the transformation family introduced by MacKinnon and Magee is applied in the Tobit framework to derive a score test for misspecification with one degree of freedom. The test statistic is found to be sensitive for misspecification in the first three conditional moments of the positive observations. The test is compared with the RESET-test as well as with score tests which also have one degree of freedom and test for linearity of the mean and the type of heteroscedasticity related to the mean. An empirical example is presented in which the derived test clearly indicated excess skewness in the residuals and simultaneously accounts for the results given by the other two score tests. It is recommended that the test statistic should be routinely used in model diagnostics.

KEY WORDS: Tobit models, misspecification, functional form

SUONIEMI, ILPO: NONLINEARITY AND HETEROSCEDASTICITY IN TOBIT MODELS: SCORE TESTS FOR MISSPECIFICATION WITH ONE DEGREE OF FREEDOM. Helsinki: VATT, Valtion taloudellinen tutkimuskeskus, 1992. (C, ISSN 0788-5016, No 16). ISBN 951-561-020-6.

TIIVISTELMÄ: Tavallisista regressiomalleista poiketen Tobit mallit eivät ole robusteja spesifikaatiovirheille kuten jäännösten heteroskedastisuus ja ei-normaalisuus. Tästä syystä tarvitaan spesifikaatiotestejä, joilla on voimaa useiden vaihtoehtojen suhteen ja joita on helppo soveltaa. Soveltamalla MacKinnonin ja Mageen esittämää muunnosperhettä johdetaan yhden vapausasteen pisteytystesti (score test) Tobit mallin spesifikaatiolle. Testisuure osoittautuu herkäksi positiivisten havaintojen kolmen ensimmäisen ehdollisen momentin väärinspesifioinnille. Testiä verrataan sekä RESET-testiin että yhden vapausasteen pisteystesteihin, jotka testaavat odotusarvon lineaarisuutta ja sellaista heteroskedastisuutta, joka riippuu odotusarvon koosta. Työssä esitetään empiirinen esimerkki, jossa edellinen testi viittaa selvästi residuaalien liialliseen vinouteen ja samalla paljastaa jälkimmäisten pisteystestien tulokset. Testitunnuslukua suositellaan rutiinikäyttöön osana mallin diagnostista tarkastelua.

AVAINSANAT: Tobit mallit, väärinspesifiointi, funktiomuoto

CONTENTS

1. INTRODUCTION

Tobit models and their multivariate generalizations involving several latent or partially observed normally distributed random variables are widely used in applied econometric work. These models are usually estimated by solving numerically the nonlinear estimating equations given by the principle of maximum likelihood. Therefore, investigators often display a natural reluctance to test the specification of the model with the thoroughness and vigor comparable to the common practice in the normal regression framework. In contrast it is well known, however, that Tobit models are not robust against specification errors such as heteroscedasticity and violation of the normality assumption. There is thus a need for specification tests of Tobit models which have power against a wide range of alternatives and are easy to implement and inexpensive to compute.

In this context it seems natural to consider the use of score, or LM, tests because they require estimates only under the null hypothesis and can often be computed by means of artificial linear regressions. In this paper a misspecification test is presented which is derived by considering a possible misspecification in the transformation applied to the observed values of the dependent variable in the Tobit model.

In the statistical analysis transformations of the dependent variable are used to obtain three objectives. The first of these is to normalize the random variable in question. The second objective is to stabilize its variance. A classic example of a simultaneously normalizing and variance stabilizing transformation concerns the sample correlation coefficient of a bivariate normal distribution, where the transformation $\tanh^{-1}$ is used. In regression models the third objective is to obtain linearity of the conditional mean for data which have been conveniently transformed.

In econometrics the family of monotonic power transformations is widely used to obtain the above objectives (Box and

Cox, 1964). However, the Box-Cox transformation cannot be applied to variables which can have zero or negative values. Therefore it cannot be used in limited dependent variable models, eg. the Tobit model, where a normally distributed latent variable is observed only at above zero values and otherwise a limit value of zero is observed.[1]

MacKinnon and Magee (1990) have proposed a family of transformations which can sensibly be applied to variables that can take any values. It is interesting to apply their method of analyzing ordinary regression models to Tobit models. MacKinnon and Magee derive score tests for the null hypothesis that the dependent variable has not been transformed against the alternative that a transformation of this family has been applied to it. These tests, which do not require that the exact form of the transformation is specified and are thus interpretable as implicit misspecification tests in the sense of Hausman (1978), are in this paper extended to the Tobit model where the underlying latent variable is similarly transformed. The misspecification test is based on the score under the null and has one degree of freedom.

In analogy with the results that MacKinnon and Magee derive in an ordinary regression model, the score test presented for Tobit models can be seen as testing simultaneously for three restrictions that affect conditional moments. The first of these, which is closely related to the well-known RESET test of Ramsey (1969), is that there is no correlation between the squared conditional mean of the latent variable and the model residuals of positive observations. In this case one has to correct the residuals for their expected values which are generally non-zero in the ordinary Tobit model.

---

[1] This problem could be overcome by adding a positive constant to the dependent variable prior to the application of the power transformation. Alternatively the family of modulus transformations proposed by John and Draper (1980) for obtaining approximate normality from symmetric long-tailed distributions could be considered.

The second conditional moment restriction is that there is no such heteroscedasticity which results in nonzero correlation between the conditional mean and squared residuals in the Tobit model, after the heteroscedasticity induced by the ordinary Tobit model is allowed for.

The third restriction is that the Tobit model residuals have a third moment of zero, again after the positive skewness induced by the ordinary Tobit model is allowed for. In case of misspecification in a Tobit model, the estimated residuals of positive observations would be expected to suffer from problems which would affect their first three moments in a way not allowed by the ordinary Tobit model.

The above implicit test for misspecification based on the MacKinnon-Magee family of transformations has some power against a wide range of alternative models. This is attributable to the way in which information on the non-linear model of the mean is confounded with information concerning the distribution of disturbances. The lack of fit detected may be due to the mean model, or the disturbance model, or both. This feature is common to all transformation families, eg. Box-Cox, that affect only the values of the dependent variable. To cater for the above situation two score tests are considered in the remainder of the paper. These two tests specialize in testing linearity of the mean, and heteroscedasticity, respectively[2]. The relative success of these three tests in rejecting the null would suggest whether modifications in modelling the mean, or heteroscedasticity, or the type of non-normality which affects the third conditional moment of the data would be necessary.

Heteroscedasticity causes the parameter estimates from a Tobit model to be inconsistent. It is thus a more severe problem than in the case of ordinary regression. In additi-

---

[2] The heteroscedasticity test is a special (one degree of freedom) case of the score test introduced to the regression framework by Breusch and Pagan (1979) and to Tobit models by Lee and Maddala (1985). The one degree of freedom form of the test is based on an idea due to Anscombe (1961).

on, because Tobit models are usually estimated using cross-sectional data it is a problem likely to be encountered quite often. Therefore a score test based on a specific form of heteroscedasticity proposed by Anscombe (1961) is discussed as an additional diagnostic tool in testing heteroscedasticity.

Finally the use of these tests is illustrated by applying them to an empirical example of estimating a demand function for alcoholic beverages using data from the 1981 Finnish Family Expenditure Survey. In this particular case the implicit misspecification test based on the idea introduced by MacKinnon and Magee (1990) is quite successful in narrowing the type of misspecification present due to excess skewness in the dependent variable.

2. THE MACKINNON-MAGEE TRANSFORMATION IN TOBIT MODELS

The following family of transformations introduced by MacKinnon and Magee (1990) is considered in this paper

$$g(\gamma y)/\gamma, \tag{1}$$

where the function g is monotonic and satisfies the following properties:

$$g(0) = 0, \tag{2}$$

$$g'(0) = 1, \tag{3}$$

$$g''(0) \neq 0. \tag{4}$$

If one allows for γ to vary in a suitable way, the above transformation (1) is homogeneous of degree one in y. This property of scale invariance is not shared by the Box-Cox transformation. In addition transformation (1) is applicable to variables y that can have negative values. These features make it interesting to examine the properties of (1) in Tobit models. Property (4) is needed because otherwise the partial derivative of the loglikelihood function w.r.t. γ

would be zero at the point $\gamma = 0$. This, however, rules out skew-symmetric functions g (for detailed discussion, see MacKinnon and Magee, 1990).

In the sequel one considers the following Tobit model, where the observed variable y^* is generated by

$$y^* = \max\{0, y\}, \tag{5}$$

and the latent variable y which is partially observed is generated by

$$r(\gamma, y) \equiv g(\gamma y)/\gamma \sim N(X^T b, \sigma^2). \tag{6}$$

Here X is a column vector of k independent variables and b is a vector of parameters to be estimated together with a dispersion parameter σ^2.

One important feature of (2) in Tobit models is that the probability of a limit observation, i.e. $P(y^* = 0)$ is independent of the transformation (1), and the transformation affects only the distribution of nonlimit observations, i.e. $y > 0$.[3] In particular the straightforward use of an implicitely defined Probit model would consistently estimate the parameter vector β defined below in equation (8).

In the following we consider the score test for the null hypothesis that $\gamma = 0$. Under this null $r(0, y)$ reduces to y and the model (5) reduces to the ordinary Tobit model. This is easily verified by taking the appropriate limit of $r(\gamma, y)$ and using $g'(0) = 1$.

In the paper Olsen's reparametrization of the Tobit model is used, giving new parameters β and h by

---

[3] Considering measurement errors in y, one may hypothesize that the limit observations, i.e. zeros, are measured more accurately than nonlimit observations. In addition the limit observations are typically less informative on the possible form of misspecification and the resulting loss of power due to the form of representation adopted in the paper is probably not too severe, especially so if the number of limit observations is small compared to the total sample size.

$$h = 1/\sigma, \tag{7}$$

$$\beta = \frac{1}{\sigma} b. \tag{8}$$

The cumulative distribution function of the latent variable y is given by

$$F(y) \equiv P(\underline{y} \le y) = P(r(\gamma,\underline{y}) \le r(\gamma,y)) = P\left(\frac{r(\gamma,\underline{y})-X^T b}{\sigma} \le \frac{r(\gamma,y)-X^T b}{\sigma}\right)$$

$$= \Phi\left(\frac{g(\gamma y)}{\gamma\sigma} - \frac{X^T b}{\sigma}\right) = \Phi\left(\frac{hg(\gamma y)}{\gamma} - X^T\beta\right). \tag{9}$$

The corresponding density function is given by

$$f(y) \equiv hg'(\gamma y)\phi\left(\frac{hg(\gamma y)}{\gamma} - X^T\beta\right). \tag{10}$$

Here Φ and ϕ are, respectively the cumulative distribution and density function of a standardized normal distribution.

Example

Consider the transformation

$$g = \kappa(\Phi^{-1} \circ F), \tag{11}$$

where κ is a scaling constant to be determined later and F is a given standardized cumulative distribution function, with a median of zero, i.e. $F(0) = \frac{1}{2}$. In this case condition (2) is satisfied and (3) holds if

$$g'(0) = \kappa \frac{f(0)}{\phi(g(0)/\kappa)} = \kappa \frac{f(0)}{\phi(0)} = 1, \tag{12}$$

where $f = dF$. In order to satisfy (12) one must set $\kappa = [\sqrt{2\pi}\ f(0)]^{-1}$. Similarly

$$g''(0) = \kappa \frac{f'(0)}{\phi(0)} + \frac{1}{\kappa} g(0)[g'(0)]^2 = \kappa \frac{f'(0)}{\phi(0)}, \tag{13}$$

This expression reveals that in order to have $g''(0) \neq 0$ one must have $f'(0) \neq 0$, i.e. the distribution F must be such that the mode is not equal to the median which is zero.
In this case the cumulative distribution function of the latent variable y is given by

$$\begin{aligned} P(\underline{y} \leq y) &= P\left(\frac{r(\gamma, \underline{y}) - X^T b}{\sigma} \leq \frac{r(\gamma, y) - X^T b}{\sigma}\right) \\ &= \Phi\left(\frac{g(\gamma y)}{\gamma \sigma} - \frac{X_T b}{\sigma}\right) = \Phi\left(\frac{\kappa h}{\gamma} \Phi^{-1}(F(\gamma y)) - X^T \beta\right). \end{aligned} \tag{14}$$

In other words, one has taken a random variable distributed according to F up to a scale factor γ and transformed it to get a normally distributed random variable, with mean $X_k^T b$, where $b = (\kappa/\gamma\sigma)\beta$, and deviation $\gamma\sigma/\kappa$. The corresponding density function is in this case given by

$$\kappa h f(\gamma y) \phi(\Phi^{-1}(F(\gamma y)) \phi\left(\frac{\kappa h}{\gamma} \Phi^{-1}(F(\gamma y)) - X^T \beta\right). \tag{15}$$

The example shows that setting $\gamma = \kappa/\sigma$ ensures that the score test derived below has some power if one considers the alternative hypotheses characterized by a distribution similar to F in the above example.

3. SCORE TEST FOR H_0: $\gamma = 0$

The loglikelihood of an individual observation, $y_k^* = I_k y_k$, X_k, $k = 1, \ldots, n$, where I_k is an indicator function for the event $\{y_k > 0\}$, is given by (leaving out the inessential constant, $-(1/2)\log(2\pi)$)

$$\mathcal{L}_k \equiv \mathcal{L}(\gamma, h, \beta | I_k y_k, X_k) = I_k \left[\log(h) + \log(g'(\gamma y_k)) - \frac{1}{2}\left(\frac{hg(\gamma y_k)}{\gamma} - X_k^T\beta\right)^2\right]$$
$$+ (1 - I_k)\log(1 - \Phi(X_k^T\beta)). \tag{16}$$

The derivative of $\mathcal{L}_k$ w.r.t γ is

$$\frac{\partial \mathcal{L}_k}{\partial \gamma} = I_k \left[\frac{g''(\gamma y_k) y_k}{g'(\gamma y_k)} - \left(\frac{hg(\gamma y_k)}{\gamma} - X_k^T\beta\right)\left(\frac{hg'(\gamma y_k) y_k}{\gamma} - \frac{hg(\gamma y_k)}{\gamma^2}\right)\right]. \tag{17}$$

Applying l'Hopital's rule to the term in the last brackets implies

$$\lim_{\gamma \to 0} \frac{hg'(\gamma y_k) y_k}{\gamma} - \frac{hg(\gamma y_k)}{\gamma^2} = \frac{h}{2} g''(0) y_k^2, \tag{18}$$

giving

$$\left[\frac{\partial \mathcal{L}_k}{\partial \gamma}\right]_{\gamma = 0} = g''(0) I_k \left[y_k - \frac{h y_k^2}{2}\left(h y_k - X_k^T\beta\right)\right]. \tag{19}$$

The other derivatives of the loglikelihood function under $\gamma = 0$ are the same as in the ordinary Tobit model:

$$\left[\frac{\partial \mathcal{L}_k}{\partial h}\right]_{\gamma = 0} = I_k \left[\frac{1}{h} - (h y_k - X_k^T\beta) y_k\right], \tag{20}$$

$$\left[\frac{\partial \mathcal{L}_k}{\partial \beta}\right]_{\gamma=0} = I_k\,(hy_k - X_k^T\beta)X_k - (1-I_k)\,\frac{\phi(X_k^T\beta)}{1-\Phi(X_k^T\beta)}X_k. \tag{21}$$

Under H_0 the maximum likelihood estimates of β and h in the ordinary Tobit model satisfy

$$\left[\frac{\partial \mathcal{L}}{\partial \beta}\right]_{\gamma=0} \equiv \sum_{k=1}^{n}\left[\frac{\partial \mathcal{L}_k}{\partial \beta}\right]_{\gamma=0} = 0 = \sum_{k=1}^{n}\left[\frac{\partial \mathcal{L}_k}{\partial h}\right]_{\gamma=0} \equiv \left[\frac{\partial \mathcal{L}}{\partial h}\right]_{\gamma=0}. \tag{22}$$

A score test for $\gamma = 0$ can always be interpreted as a test for

$$\text{plim}_{n\uparrow\infty}\left[\frac{1}{n}\frac{\partial \mathcal{L}}{\partial \gamma}\right]_{\gamma=0} = 0. \tag{23}$$

Denoting the residuals of positive observations by $I_k u_k = hI_k y_k - I_k X_k^T\beta$, one can easily show using the properties of truncated normal variables that the first three moments are

$$\mathscr{E}(I_k u_k) = \phi(X_k^T\beta), \tag{24 i}$$

$$\mathscr{E}(I_k u_k^2) = \Phi(X_k^T\beta) - X_k^T\beta\phi(X_k^T\beta), \tag{24 ii}$$

$$\mathscr{E}(I_k u_k^3) = (X_k^T\beta)^2\phi(X_k^T\beta) + 2\phi(X_k^T\beta), \tag{24 iii}$$

and the higher moments are give by the recursion formula

$$\mathscr{E}(I_k u_k^m) = (-X_k^T\beta)^{m-1}\phi(X_k^T\beta) + (m-1)\mathscr{E}(I_k u_k^{m-2}), \quad m \geq 2. \tag{24 iv}$$

One can write

$$\frac{1}{n}\left[\frac{\partial \mathcal{L}}{\partial \gamma}\right]_{\gamma=0} = \left(\frac{-g''(0)\sigma}{n}\right)\sum_{k=1}^{n} I_k\left(\frac{1}{2}u_k(X_k^T\beta + u_k)^2 - (X_k^T\beta + u_k)\right) \tag{25}$$

$$= \left(-g''(0)\sigma\right)\left(\sum_{k=1}^{n}\frac{1}{2n}I_k(X_k^T\beta)^2\left(u_k - \frac{\phi(X_k^T\beta)}{\Phi(X_k^T\beta)}\right)\right. \qquad (26\ i)$$

$$+ \sum_{k=1}^{n}\frac{I_k(X_k^T\beta)}{n}\left(u_k^2 - \left(1 - X_k^T\beta\frac{\phi(X_k^T\beta)}{\Phi(X_k^T\beta)}\right)\right) \qquad (26\ ii)$$

$$\left.+ \sum_{k=1}^{n}\frac{I_k}{2n}\left(u_k^3 - (X_k^T\beta)^2\frac{\phi(X_k^T\beta)}{\Phi(X_k^T\beta)} - 2u_k\right)\right). \qquad (26\ iii)$$

Using the properties (24 i-iii), one easily shows that each of terms in (26 i-iii) have a probability limit of zero, as $n \uparrow \infty$.

If one tests an ordinary Tobit model against an alternative defined by (6), one tests simultaneously for three restrictions that affect conditional moments. The first of these, (26 i), is that there is no correlation between $(X_k{}^T\beta)^2$ and u_k when the latter is corrected for its expectation under condition $y_k > 0$. This is what a popular RESET test by Ramsey (1969) tests for. In Tobit models this form of the RESET test tests for $\delta = 0$ in a 'HECKIT-regression' using only the positive observations when the regression equation is given by

$$hy_k = X_k^T\beta + \delta(X_k^T\tilde{\beta})^2 + \frac{\phi(X_k^T\tilde{\beta})}{\Phi(X_k^T\tilde{\beta})}, \qquad (27)$$

where β is the maximum likelihood estimate of β in the ordinary Tobit model.

The second restriction, (26 ii), is that there is no such heteroscedasticity which results in nonzero correlation between $(X_k^T\beta)$ and squared residuals in Tobit model, after the heteroscedasticity induced by the ordinary Tobit model is allowed for.

The third restriction, (26 iii), is that the Tobit model residuals have a third moment of zero, again after the positive skewness induced by the Tobit model is allowed for. Therefore, if the data were generated by (6) with $\gamma \neq 0$, and one estimated an ordinary Tobit model, the estimated residuals of positive observations would be expected to suffer from problems which would affect their first three conditional moments in a way not allowed by the ordinary Tobit model. These observations, made originally by MacKinnon and Magee (1990) in ordinary regression models, carry over to the Tobit model.

The RESET test will have some power against $\gamma \neq 0$ in (6) because of (26 i). If the fit of the Tobit model is very good by which it is meant that σ is small relative to the variation of $(X_k^T b)$, then any violation to the condition that the probability limits of (26 ii) and (26 iii) are zero, will contribute relatively little to the score test. If the fit of the Tobit model is poor, however, RESET test will have low power compared to the score test against the above specific alternative hypothesis (6). This follows from the observation that as $\sigma \downarrow 0$ then typically $(X_k^T\beta)^2 \uparrow \infty$, and the first term (26 i) will dominate over the other two terms (26 ii) and (26 iii).

The easiest way to calculate the score test for $\gamma = 0$ is to replace the information matrix of the parameter estimates with its finite sample approximation by the "Outer Product of the Gradient," or OPG,

$$\text{plim}_{n\uparrow\infty}\left[\left[\frac{\partial\mathcal{L}}{\partial\theta}\right]\left[\frac{\partial\mathcal{L}}{\partial\theta}\right]^T\right]_{\gamma=0} = \mathcal{E}\left[-\frac{\partial^2\mathcal{L}}{\partial\theta\partial\theta^T}\right]_{\gamma=0} \equiv \mathfrak{I}. \tag{28}$$

The OPG test statistic, popularized by Godfrey and Wickens (1981), can be computed as n minus the sum of squared residuals, or nR^2, from the artificial linear regression

$$\iota = \left[\frac{\partial \mathcal{L}}{\partial \boldsymbol{\theta}}\right] c + \text{remainder}, \tag{29}$$

where ι is an n-vector of ones and the regressor matrix is an n $\times$ p matrix of the derivatives of the loglikelihood of each of the observations evaluated at the restricted ordinary Tobit model estimates. Here n = k + 2, and the corresponding components are given in the formulae (19) -(21). Note that the actual value of g"(0) has no effect on the test statistic and might as well be omitted.

Alternatively one can calculate the second derivatives of the loglikelihood

$$\left[\frac{\partial^2 \mathcal{L}_k}{\partial \gamma \partial h}\right]_{\gamma = 0} = -\frac{g''(0)}{2} I_k\, y_k^2 (2h y_k - X_k^T \beta), \tag{30}$$

$$\left[\frac{\partial^2 \mathcal{L}_k}{\partial \gamma \partial \beta}\right]_{\gamma = 0} = \frac{g''(0)\, I_k\, h y_k^2}{2} X_k . \tag{31}$$

Partitioning the information matrix as

$$\mathfrak{I} = \begin{bmatrix} \iota_{\gamma\gamma} & \mathfrak{I}_{\gamma 2} \\ \mathfrak{I}_{2\gamma} & \mathfrak{I}_{22} \end{bmatrix}, \tag{32}$$

One can write the score statistic LM

$$LM = \frac{\left[\frac{\partial \mathcal{L}}{\partial \gamma}\right]^2_{\gamma = 0}}{\iota_{\gamma\gamma} - \mathfrak{I}_{\gamma 2}\mathfrak{I}_{22}^{-1}\mathfrak{I}_{2\gamma}}$$

$$= \frac{\left(\sum_{k=1}^{n} I_k \left(\frac{1}{2} u_k (X_k^T\beta + u_k)^2 - (X_k^T\beta + u_k)\right)\right)^2}{v_{\gamma\gamma} - 1_\gamma^T V 1_\gamma}, \tag{33}$$

where

$$v_{\gamma\gamma} = \frac{h^2 \iota_{\gamma\gamma}}{g''(0)^2} = \sum_{k=1}^{n} \mathscr{E}\left[I_k \left(\frac{1}{2} u_k (X_k^T\beta + u_k)^2 - (X_k^T\beta + u_k)\right)^2\right], \qquad (34\ i)$$

$$l_\gamma = \frac{h}{g''(0)} \mathfrak{S}_{2\gamma} = \frac{1}{2}\left[\begin{array}{l} \sum_{k=1}^{n} \mathscr{E}\left(I_k\ h y_k^2 (2hy_k - X_k^T\beta)\right) \\ -\sum_{k=1}^{n} \mathscr{E}\left(I_k (hy_k)^2 X_k\right) \end{array}\right], \qquad (34\ ii)$$

and V is the estimated covariance matrix of the maximum likelihood estimators of $(h, \beta^T)^T$ in the ordinary Tobit model. The formulae for $v_{\gamma\gamma}$ and l_γ are given in the appendix. They involve the moments of positive observations up to the fourth degree and are straightforward, albeit relatively tedious to derive.

The above score test based on the MacKinnon-Magee family of transformations has some power against a wide range of alternative models. This is attributable to the way in which information on the non-linear model of the mean is confouded with information concerning the distribution of disturbances. The lack of fit detected may be due to the mean model, or the disturbance model, or both. This feature of model (1) is common to all transformation families, eg. Box-Cox, that affect only the values of the dependent variable. To cater for the above situation two score tests are considered below. These two tests specialize in testing linearity of the mean (26 i), and heteroscedasticity (26 ii), respectively. If any of these tests rejects the null substantially more emphatically than the score test for $\gamma = 0$ in (6), that would suggest that (6) is not the appropriate model and modifications in modelling the mean, or heteroscedasticity would be necessary.

4. SCORE TEST FOR THE LINEARITY OF THE MEAN[4]

One may consider instead of (6) the class of models

$$hy_k = \frac{g(\delta X_k^T\beta)}{\delta} + \epsilon_k , \qquad (35)$$

and test for $\delta = 0$ in (35).

The loglikelihood of an individual observation, $y_k^* = I_k y_k$, X_k, $k = 1, \ldots, n$, where I_k is an indicator function for the event $\{y_k > 0\}$, is given by (without the inessential constant $-(1/2)\log(2\pi)$)

$$\mathcal{L}_k \equiv \mathcal{L}(\delta, h, \beta | I_k y_k, X_k) = I_k \left[\log(h) - \frac{1}{2}\left(hy_k - \frac{g(\delta X_k^T\beta)}{\delta}\right)^2 \right] + (1-I_k)\log\left(1 - \Phi\left(\frac{g(\delta X_k^T\beta)}{\delta}\right)\right). \qquad (36)$$

Using (18) the derivative of $\mathcal{L}_k$ w.r.t δ, under $\delta = 0$ is given by

$$\left[\frac{\partial \mathcal{L}_k}{\partial \delta}\right]_{\delta = 0} = \frac{g''(0)(X_k^T\beta)^2}{2}\left[I_k \left(hy_k - X_k^T\beta\right) - (1-I_k)\left(\frac{\phi(X_k^T\beta)}{1 - \Phi(X_k^T\beta)}\right)\right]. \qquad (37)$$

The other derivatives of the loglikelihood function under $\gamma = 0$ are the same as in the ordinary Tobit model (20)-(21).

The score test for $\delta = 0$ is easily seen to be asymptotically

---

[4] In the rest of the paper the transformations considered are defined to affect also the limit observations, i.e. the zeros, in the Tobit model. The resulting discrepancy with the earlier part of the paper seems, however, to have limited practical consequences, when these tests are calculated in typical empirical examples (see footnote 1 and the example in section 6).

equivalent to a form of RESET test in which one includes an additional explanatory variable $(X_k^T\beta)^2$ in a second stage Tobit model which uses all the observations, also the limit observations, in contrast to (27). The score test may in this case also be seen as a form of regression test, where one tests for correlation between $(X_k^T\beta)^2$ and the heteroscedastic 'residuals' of in the ordinary Tobit model[5]. The zero mean residuals e_k are defined by

$$e_k = I_k e_{1k} + (1-I_k) e_{0k}, \text{ with} \tag{38}$$

$$e_{1k} = y_k - X_k^T\beta; \quad e_{0k} = -\frac{\phi(X_k^T\beta)}{1-\Phi(X_k^T\beta)}. \tag{39}$$

In analogy with (25) one can write

$$\frac{1}{n}\left[\frac{\partial \mathcal{L}}{\partial \delta}\right]_{\delta=0} = \left(\frac{g''(0)}{2n}\right)\sum_{k=1}^{n}(X_k^T\beta)^2\left(I_k u_k + (1-I_k)\left(\frac{\phi(X_k^T\beta)}{1-\Phi(X_k^T\beta)}\right)\right) \tag{40}$$

$$= \left(\frac{g''(0)}{2}\right)\left(\sum_{k=1}^{n}\frac{1}{n}I_k(X_k^T\beta)^2\left(u_k - \frac{\phi(X_k^T\beta)}{\Phi(X_k^T\beta)}\right)\right. \tag{41}$$

$$\left. + \sum_{k=1}^{n}\frac{1}{n}I_k(X_k^T\beta)^2\left(\frac{\phi(X_k^T\beta)}{\Phi(X_k^T\beta)(1-\Phi(X_k^T\beta))}\right) - \sum_{k=1}^{n}\frac{1}{n}(X_k^T\beta)^2\left(\frac{\phi(X_k^T\beta)}{1-\Phi(X_k^T\beta)}\right)\right).$$

Using (24 i) one easily shows that, under $\delta = 0$ both terms in (41) have a probability limit of zero as $n \uparrow \infty$.

---

[5] For this definition residuals in Tobit models, see Lee and Maddala (1985).

5. SCORE TESTS FOR HETEROSCEDASTICITY IN TOBIT MODELS

Consider a Tobit model, where (6) is replaced by

$$hy = X_k^T\beta + g'(\delta X^T\beta)\epsilon, \tag{42}$$

where g is a function which satisfies properties (2)-(4), and $\epsilon \sim N(0,1)$. In the above model the variance of the latent variable y depends on its mean through the function g'. In the sequel a score test for the hypothesis $\delta = 0$ is considered.

The motivation for basing the test on model (42) is that heteroscedasticity is often related to values of some important explanatory variables in the model. These variables correlate with the fit, $X_k^T\beta$, and the test for $\delta = 0$ is likely to have some power even in the case where heteroscedasticity is related to the values of a single important explanatory variable. The above idea was proposed originally for regression models by Anscombe (1961).

The loglikelihood of an individual observation, $y_k^* = I_k y_k$, X_k, $k = 1, \ldots, n$, where I_k is an indicator function for the event $\{y_k > 0\}$, is given by (without the inessential constant $-(1/2)\log(2\pi)$)

$$\mathcal{L}_k \equiv \mathcal{L}(\delta, h, \beta \mid I_k y_k, X_k)$$

$$= I_k\left[\log(h) - \log(g'(\delta X_k^T\beta)) - \frac{1}{2[g'(\delta X_k^T\beta)]^2}\left(hy_k - X_k^T\beta\right)^2\right]$$

$$+ (1-I_k)\log\left(1 - \Phi\left(\frac{X_k^T\beta}{g'(\delta X_k^T\beta)}\right)\right). \tag{43}$$

The derivative of $\mathcal{L}_k$ w.r.t δ, under $\delta = 0$ is given by

$$\left[\frac{\partial \mathcal{L}_k}{\partial \delta}\right]_{\delta=0} = g''(0)X_k^T\beta\left[I_k\left((hy_k - X_k^T\beta)^2 - 1\right) + (1-I_k)\left(\frac{\phi(X_k^T\beta)X_k^T\beta}{1-\Phi(X_k^T\beta)}\right)\right], \quad (44)$$

The other derivatives of the loglikelihood function under H_0: $\delta = 0$ are the same as in the ordinary Tobit model (20)-(21).

In this case the score test may again be seen as a form of regression test, where one tests for zero correlation between $X_k{}^T\beta$ and the squared 'residuals' in the ordinary Tobit model. These squared residuals $e_k{}^2$ are defined by

$$e_k^2 = I_k e_{1k}^2 + (1-I_k)e_{0k}^2, \text{ with} \quad (45)$$

$$e_{1k}^2 = u_k^2 - 1; \quad e_{0k}^2 = \frac{\phi(X_k^T\beta)X_k^T\beta}{1-\Phi(X_k^T\beta)}. \quad (46)$$

In analogy with (25) one can write

$$\frac{1}{n}\left[\frac{\partial \mathcal{L}}{\partial \delta}\right]_{\delta=0} = \left(\frac{g''(0)}{n}\right)\sum_{k=1}^{n} X_k^T\beta\left(I_k(u_k^2-1) + (1-I_k)\left(\frac{X_k^T\beta\,\phi(X_k^T\beta)}{1-\Phi(X_k^T\beta)}\right)\right) \quad (47)$$

$$= g''(0)\left(\sum_{k=1}^{n}\frac{1}{n}I_k X_k^T\beta\left(u_k^2 - \left(1 - \frac{X_k^T\beta\,\phi(X_k^T\beta)}{\Phi(X_k^T\beta)}\right)\right)\right. \quad (48)$$

$$\left. - \sum_{k=1}^{n}\frac{1}{n}I_k(X_k^T\beta)^2\left(\frac{\phi(X_k^T\beta)}{\Phi(X_k^T\beta)(1-\Phi(X_k^T\beta))}\right) + \sum_{k=1}^{n}\frac{1}{n}(X_k^T\beta)^2\left(\frac{\phi(X_k^T\beta)}{1-\Phi(X_k^T\beta)}\right)\right).$$

Using (24 i) one easily shows that, under $\delta = 0$ both terms in (48) have a probability limit of zero, as $n \uparrow \infty$.

Possible functional forms for g, which define locally equivalent alternative models, include the exponential function, e^x, giving

$$g'(\delta X^T\beta) = \exp(\delta X^T\beta), \tag{49}$$

considered in the Tobit models by Lee and Maddala (1985), or the affine function, $1 + (1/\alpha)x$, giving

$$g'(\delta X^T\beta) = \frac{1}{\alpha}(\alpha + \delta X^T\beta). \tag{50}$$

In the affine case a positive constant α, $\alpha > 0$, has been added to the formula of the conditional variance to guarantee that $\alpha + X^T\beta$ stays positive over the sample range. This is done in order to ensure that the negative values in the conditional mean do not produce odd behaviour in the conditional variance.[6] A similar augmentation with a translation term α is used below, where an additional and related example of a score test for heteroscedasticity is briefly discussed.

Consider a Tobit model, where (6) is replaced by

$$hy = X_k^T\beta + h^{-\delta}(\alpha + X^T\beta)^\delta\epsilon, \tag{51}$$

where $\epsilon \sim N(0,1)$. In the above model the variance of the latent variable y depends on its mean through the parameter δ. A fixed constant α is chosen in such a way as to guarantee that $\alpha + X_k{}^T\beta$ stays positive over the whole sample. Defining

$$g'(\delta) = h^{-\delta}(\alpha + X^T\beta)^\delta, \tag{52}$$

gives

$$g''(0) = \log(\alpha + X^T\beta) - \log(h). \tag{53}$$

---

[6] A problem in using the various score tests for $\delta = 0$ with a specific transformation g' in mind is that the curvature of the likelihood function at the null hypothesis may be very different from that at the maximum likelihood estimate of δ. This may result in poor power properties of the score test in comparison with the corresponding likelihood ratio test.

Forming the loglikelihood of an individual observation, $y_k^* = I_k y_k$, X_k, $k = 1, \ldots, n$, where I_k is an indicator function for the event $\{y_k > 0\}$, one gets the derivative of $\mathcal{L}_k$ w.r.t δ, under $\delta = 0$,

$$\left[\frac{\partial \mathcal{L}_k}{\partial \delta}\right]_{\delta=0} = s_k \left[I_k \left((hy_k - X_k^T\beta)^2 - 1 \right) + (1 - I_k) \left(\frac{\phi(X_k^T\beta) X_k^T\beta}{1 - \Phi(X_k^T\beta)} \right) \right], \qquad (54)$$

where $s_k = \log(\alpha + X_k^T\beta) - \log(h) = \log(a + X_k^T b)$, say. The other derivatives of the loglikelihood function under $\delta = 0$ are the same as in the ordinary Tobit model (20)-(21).
Using the Taylor expansion of $(a + X_k^T b)^\delta$ w.r.t. δ one has an approksimation for a small δ

$$(a + X_k^T b)^\delta \approx 1 + \delta \log(a + X_k^T b) = 1 + \delta s_k. \qquad (55)$$

Similarly with the previous case the score test may also be seen as a form of regression test, where one tests for $\delta = 0$ using approximation (55), i.e. for zero correlation between s_k and the squared 'residuals' in the ordinary Tobit model. These squared residuals e_k^2 are defined by eqs. (45)-(46).

6. AN EMPIRICAL EXAMPLE

The implicit misspecification test derived above was applied to an empirical example analyzed earlier by the author (Suoniemi, 1990). Finnish data from the 1981 Household Expenditure Survey were used to estimate a demand function in budget share form for household consumption of alcoholic beverages. In the analysis an ALIDS-functional form (Deaton & Muellbauer, 1980) was augmented with demographic variables and possible nonlinearity in the Engel-curves was allowed. There were twenty-four regressors and 7295 observations, 55.1 per cent of which reported no consumption of alcoholic beverages. In the original analysis Tobit models and their extensions with latent unobservable variables were used. The Tobit model chosen as 'the preferred specification' was subjected to the score tests for misspecification derived earlier in the paper. Results are given in Table 1.

TABLE 1

PERFORMANCE OF THE TOBIT MODEL FOR ALCOHOL CONSUMPTION

Test	Score	Test statistic χ^2-form	t-form
LM(ff)		0.367	-0.605
LM(h1)		3.082	-1.753
LM(h2)		3.521	-1.873
LM(γ)	-81.382	137.461	-11.815
LM_1	-0.285	0.413	
LM_2	8.481	4.431	
LM_3	-89.577	143.000	

The score test statistics are asymptotically $\chi^2(1)$ when the underlying Tobit model is actually generating the data. First row gives the test statistic for misspecification in the functional form of the mean, with score given in eq. (37). The second and third row give the test statictics for heteroscedasticity, with the scores given in eqs (44) and (54), respectively. The fourth row gives the implicit test for misspecification given in eq. (33). In their calculation the outer product forms were used, and corresponding asymptotically equivalent statistics based on Student's t-distribution are given in the last column. The values LM_1, LM_2, and LM_3 correspond to the decomposition of the score, given in eqs. (26 i-iii), respectively, and indicate the contributions of the misspecification of the mean, heteroscedasticity and skewness in the implied residual distribution corrected for censoring.

The highly significant test statistics LM(γ) clearly indicates that the model is misspecified. Interestingly enough this rejection seems not have been due to the misspecification of the mean (row one) and is only mildly affected by heteroscedasticity (rows two or three). In contrast the test statistic LM(γ) suggests that positive skewness in the distribution is the main underlying problem.

An additional interesting feature of the example is that the decomposition of the score (first column in Table 1) gives practically the same information as the score-tests taylored to test separately for the misspecification of the mean (row one) and such heteroscedasticity which is related to the mean (rows two and three).

In the original analysis of the data the problem of excess skewness was detected by less systematic methods.[7] An intuititively interpretable solution to the problem was formulated using a model with two populations labelled as "normal consumers" and "heavy drinkers", with heavy drinkers consuming a fixed amount of alcohol while the behaviour of the normal consumers is governed by a demand function.[8] The type of drinkers is not observed but instead it is assumed that heavy drinkers are uniformly mixed within the normal population. The proportion of heavy drinkers in the population, the mean of their consumption and the parameters affecting the demand function of the normal consumers were estimated by applying a Tobit-type model to a mixture of two normally distributed distributions. The model was found to be able to account for the distributional features in the data.

---

[7] Note that the demand function was estimated in a budget share form and that the test statistics for heteroscedasticity which are based on a Student's t-approximation indicate that the conditional variance is inversely related to the fit $X^T\beta$.

[8] This is a common response to rejection of the null when one uses tests that are interpreted as implicit misspecification tests. One should first try to modify the model specification to account for the poor properties of the model revealed by testing rather than directly adopting the implicit alternative. The latter less appealing alternative would in the above case lead to searching for a suitable normalizing transformation of the data.

7. CONCLUSION

In the paper the transformation family introduced by MacKinnon and Magee has been applied in the Tobit framework to derive a score test for misspecification with one degree of freedom. The test statistic is found to be sensitive for misspecification in the first three conditional moments of the positive observations in the Tobit model. The test is compared with the well known RESET-test as well as with score tests which also have one degree of freedom and test for linearity of the mean and the type of heteroscedasticity related to the mean of the latent variable in the model. Finally an empirical example has been presented in which the derived test clearly indicated excess skewness in the Tobit residuals and simultaneously accounts for the results given by the other two score tests mentioned above. Because the test statistic is easy to implement it is recommended that it should be routinely calculated and used in model diagnostics in applications based on the Tobit model.

Robust estimation methods are best seen as a complementary rather than an alternative tool to diagnostic methods such as the test developed in the paper. First since robust estimation methods assume symmetric error distributions, e.g. (Powell, 1986), tests for possible skewness in the data should precede the application of such methods. Second diagnostic methods may result in the recognition of important phenomena that might otherwise have gone undetected. This was demonstrated in the empirical example where the outlying observations labelled as "heavy drinkers" may indicate cases under which the consumption process works differently. Indentification of phenomena like this may in fact have at least equal scientific importance than the analysis of the bulk of data.

REFERENCES

Anscombe, F.J., "Examination of Residuals," **Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability**, 4 (1961), 1-36.

Box, G.E.P. and D.R. Cox, "An Analysis of Transformations," **Journal of the Royal Statistical Society, Series B**, 26 (1964), 211-252.

Breusch, T.S. and A.R. Pagan, "A Simple Test for Heteroscedasticity and Random Coefficient Variation," **Econometrica**, 47 (1979), 1287-1294.

Deaton, A.S. and J. Muellbauer, "An almost ideal demand system," **American Economic Review**, 70 (1980), 312-326.

Godfrey, L.G. and M.R. Wickens, "Testing Linear and Log-linear Regressions for Functional Form," **Review of Economic Studies**, 48 (1981), 487-496.

Hausman, J., "Specification Tests in Econometrics," **Econometrica**, 46 (1978), 1251-1271.

John, J.A. and N.R. Draper, "An alternative family of power transformations," **Applied Statistics**, 29 (1980), 190-197.

MacKinnon, J.G. and L. Magee, "Transforming the Dependent Variable in Regression Models," **International Economic Review**, 31 (1990), 315-339.

Lee, L-F. and G.S. Maddala, "The Common Structure of Tests for Selectivity Bias, Serial Correlation, Heteroscedasticity and Non-normality in the Tobit Model," **International Economic Review**, 26 (1985), 1-20.

Powell, J.L., "Symmetrically Trimmed Least Squares Estimation for Tobit Models," **Econometrica**, 54 (1986), 1435-1460.

Ramsey, J., "Tests for Specification Errors in Classical Linear Least Squares Regression Analysis," **Journal of the Royal Statistical Society, Series B**, 31 (1969), 350-371.

Suoniemi, I. "Analysing Demand for Alcoholic Beverages Using Household Data I: A model for household consumption using data from the 1981 Household Expenditure Survey" (in Finnish), The Finnish State Alcohol Monopoly, Economic Research and Planning Unit: Research Report 17 (1990).

APPENDIX

Using repeatedly the equations for the conditional moments of positive observations in the Tobit model (24) (i)-(iv), one can write

$$v_{\gamma\gamma} = \frac{h^2 \iota_{\gamma\gamma}}{g''(0)^2} = \sum_{k=1}^{n} \mathscr{E}\left[I_k \left(\frac{1}{2} u_k (X_k^T\beta + u_k)^2 - (X_k^T\beta + u_k)\right)^2\right]$$

$$= \frac{1}{4}\sum_{k=1}^{n} \left[\left((X_k^T\beta)^4 + 10(X_k^T\beta)^2 + 7\right)\Phi(X_k^T\beta) + \left((X_k^T\beta)^3 + 9X_k^T\beta\right)\phi(X_k^T\beta)\right], \quad \textbf{(A 1)}$$

similarly,

$$l_\gamma = \frac{h}{g''(0)} \mathfrak{S}_{2\gamma} = \frac{1}{2}\left[\begin{array}{c} \sum_{k=1}^{n} \mathscr{E}\left(I_k\, h y_k^2 (2hy_k - X_k^T\beta)\right) \\ -\sum_{k=1}^{n} \mathscr{E}\left(I_k (hy_k)^2 X_k\right) \end{array}\right]$$

(A 2)

$$= \frac{1}{2}\left[\begin{array}{l} \sigma \sum_{k=1}^{n} \left[\left((X_k^T\beta)^3 + 5X_k^T\beta\right)\Phi(X_k^T\beta) + \left((X_k^T\beta)^2 + 4\right)\phi(X_k^T\beta)\right] \\ -\sum_{k=1}^{n} \left[\left((X_k^T\beta)^2 + 1\right)\Phi(X_k^T\beta) + (X_k^T\beta)\phi(X_k^T\beta)\right]X_k \end{array}\right].$$