Journal of Applied Statistics

ISSN: 0266-4763 (Print) 1360-0532 (Online) Journal homepage: https://www.tandfonline.com/loi/cjas20

Testing for heteroskedasticity in the tobit and probit models

Darryl Holden

To cite this article: Darryl Holden (2011) Testing for heteroskedasticity in the tobit and probit models, Journal of Applied Statistics, 38:4, 735-744, DOI: 10.1080/02664760903563684

To link to this article: https://doi.org/10.1080/02664760903563684

Journal of Applied Statistics
Vol. 38, No. 4, April 2011, 735–744

Testing for heteroskedasticity in the tobit and probit models

Darryl Holden*

Department of Economics, University of Strathclyde, Glasgow, UK

(*Received 26 June 2009; final version received 7 November 2009*)

Non-constant variance across observations (heteroskedasticity) results in the maximum likelihood estimators of tobit and probit model parameters being inconsistent. Some of the available tests for constant variance across observations (homoskedasticity) are discussed and examined in a small Monte Carlo experiment.

Keywords: tobit; probit; heteroskedasticity; score test; Monte Carlo

1. Introduction

The tobit and probit models both assume that the variable y_i^* in

$$y_i^* = \boldsymbol{X}_i'\boldsymbol{\beta} + u_i \tag{1}$$

is not fully observed. In the tobit framework the variable $y_i = \max(0, y_i^*)$ is observed, so that y_i^* is observed if, and only if, it assumes a positive value. In the probit framework the observed variable is

$$d_i = \begin{cases} 0 & \text{if and only if } y_i^* \leq 0 \\ 1 & \text{if and only if } y_i^* > 0, \end{cases}$$

so that only the sign of y_i^* is observed. The maximum likelihood estimation of $\boldsymbol{\beta}$ and σ^2 in the tobit model and of $\boldsymbol{\beta}/\sigma$ in the probit model, based on an assumption that the u_i are independent $N(0, \sigma^2)$, are standard options in software packages intended for econometric work. However, these estimators are inconsistent if $\text{var}(u_i)$ in fact varies across i rather than being constant.[1] This suggests that an assessment of the validity of the assumption that $\text{var}(u_i)$ is constant across i, that is for homoskedasticity of u_i, might be regarded as important in work using the tobit or probit models. A number of appropriate tests in fact exist. Section 2 presents some tests for the tobit model and Section 3 presents some tests for the probit model. The tests which are considered are interpreted as different varieties of Lagrange multiplier (score) test. Section 4 presents the results of a small scale Monte Carlo experiment and Section 5 presents some concluding comments.

*Email: d.holden@strath.ac.uk

ISSN 0266-4763 print/ISSN 1360-0532 online

DOI: 10.1080/02664760903563684
http://www.informaworld.com

2. Tests for the tobit model

Following [3] we assume that the u_i in Equation (1) are independent $N(0, \sigma_i^2)$ with

$$\sigma_i^2 = \mathrm{var}(u_i) = f(\alpha_0 + \boldsymbol{\alpha}_1' \boldsymbol{W}_i) \tag{2}$$

and derive score tests of $H_0 : \boldsymbol{\alpha}_1 = \mathbf{0}$ since $\boldsymbol{\alpha}_1 = \mathbf{0}$ implies $\mathrm{var}(u_i) = \sigma_i^2 = f(\alpha_0)$ which is constant across i so that the u_i are homoskedastic. In Equation (2) the function f is required to always return a positive value but is otherwise unspecified.

A score test for $H_0 : \boldsymbol{\alpha}_1 = \mathbf{0}$ will be of the form

$$\ell_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}})' \mathbf{Q}^{-1} \ell_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}), \tag{3}$$

where $\ell_{\boldsymbol{\theta}}$ is the vector of first derivatives of the heteroskedastic tobit log-likelihood function, $\hat{\boldsymbol{\theta}}$ is the maximum likelihood estimator when the null hypothesis is imposed in estimation, and $\mathbf{Q}$, which is required to consistently estimate $-\mathrm{E}(\ell_{\boldsymbol{\theta},\boldsymbol{\theta}'})$, where $\ell_{\boldsymbol{\theta},\boldsymbol{\theta}'}$ is the matrix of second derivatives of the heteroskedastic tobit log-likelihood function, can be chosen in several ways. The various choices for $\mathbf{Q}$ considered below each give a statistic of the form in Equation (3) with an asymptotic χ_p^2 distribution under the null hypothesis, which is rejected for large values of the statistic. The degree of freedom p is the number of elements of $\boldsymbol{\alpha}_1$.

The heteroskedastic tobit log-likelihood function is

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \left[(1 - d_i) \ln(1 - \Phi_i) + d_i \ln \left[\frac{\phi(a_i)}{\sigma_i} \right] \right],$$

where Φ_i is Φ, the cumulative distribution function of the standard normal, at $\delta_i = \boldsymbol{X}_i' \boldsymbol{\beta} / \sigma_i$, a_i is $(y_i - \boldsymbol{X}_i' \boldsymbol{\beta}) / \sigma_i$, and ϕ is the density function of the standard normal. The first derivatives of $\ell(\boldsymbol{\theta})$, which are the components of $\ell_{\boldsymbol{\theta}}$, are

$$\frac{\partial \ell}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{n} \frac{(d_i a_i - (1 - d_i)\lambda_i)}{\sigma_i} \boldsymbol{X}_i, \tag{4}$$

where λ_i is $\phi_i / (1 - \Phi_i)$ and ϕ_i is ϕ at δ_i,

$$\frac{\partial \ell}{\partial \alpha_0} = \sum_{i=1}^{n} \frac{f_i'(d_i(a_i^2 - 1) + (1 - d_i)\lambda_i \delta_i)}{2\sigma_i^2}, \tag{5}$$

where f' is the first derivative of f, and

$$\frac{\partial \ell}{\partial \boldsymbol{\alpha}_1} = \sum_{i=1}^{n} \frac{f_i'(d_i(a_i^2 - 1) + (1 - d_i)\lambda_i \delta_i)}{2\sigma_i^2} \boldsymbol{W}_i. \tag{6}$$

The estimator $\hat{\boldsymbol{\theta}}$ is obtained by imposing the null hypothesis on the expressions in Equations (4) and (5), equating to $\mathbf{0}$, and solving for $\boldsymbol{\beta}$ and $\sigma^2 = f(\alpha_0)$.

There are several possible $\mathbf{Q}$ which be can be used in Equation (3). One choice is $\mathbf{Q}_1 = (-\mathrm{E}(\ell_{\boldsymbol{\theta},\boldsymbol{\theta}'}))_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$, where the information matrix equality, $-\mathrm{E}(\ell_{\boldsymbol{\theta},\boldsymbol{\theta}'}) = \mathrm{E}(\ell_{\boldsymbol{\theta}} \ell_{\boldsymbol{\theta}}')$, means that it is not necessary to obtain the second derivatives of $\ell(\boldsymbol{\theta})$ in order to obtain $\mathbf{Q}_1$. If we let $\underline{\mathbf{Q}}$ denote

$-\mathrm{E}(\ell_{\boldsymbol{\theta},\boldsymbol{\theta}'})$ it is possible to show that the blocks of

$$\underline{\mathbf{Q}} = \begin{bmatrix} \underline{\mathbf{Q}}_{\boldsymbol{\beta},\boldsymbol{\beta}'} & \underline{\mathbf{Q}}_{\boldsymbol{\beta},\alpha_0} & \underline{\mathbf{Q}}_{\boldsymbol{\beta},\boldsymbol{\alpha}_1'} \\ & \underline{\mathbf{Q}}_{\alpha_0,\alpha_0} & \underline{\mathbf{Q}}_{\alpha_0,\boldsymbol{\alpha}_1'} \\ & & \underline{\mathbf{Q}}_{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_1'} \end{bmatrix}$$

are

$$\underline{\mathbf{Q}}_{\boldsymbol{\beta},\boldsymbol{\beta}'} = \sum_{i=1}^{n} \frac{(\Phi_i + \phi_i(\lambda_i - \delta_i))}{\sigma_i^2} \boldsymbol{X}_i \boldsymbol{X}_i', \tag{7}$$

and

$$\underline{\mathbf{Q}}_{\boldsymbol{\beta},\alpha_0} = \sum_{i=1}^{n} \frac{f_i' s_i}{2\sigma_i^3} \boldsymbol{X}_i, \quad \underline{\mathbf{Q}}_{\boldsymbol{\beta},\boldsymbol{\alpha}_1'} = \sum_{i=1}^{n} \frac{f_i' s_i}{2\sigma_i^3} \boldsymbol{X}_i \boldsymbol{W}_i', \tag{8}$$

where s_i is $\phi_i(\delta_i^2 - \lambda_i \delta_i + 1)$, and

$$\underline{\mathbf{Q}}_{\alpha_0,\alpha_0} = \sum_{i=1}^{n} \frac{(f_i')^2 t_i}{4\sigma_i^4}, \quad \underline{\mathbf{Q}}_{\alpha_0,\boldsymbol{\alpha}_1'} = \sum_{i=1}^{n} \frac{(f_i')^2 t_i}{4\sigma_i^4} \boldsymbol{W}_i', \quad \underline{\mathbf{Q}}_{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_1'} = \sum_{i=1}^{n} \frac{(f_i')^2 t_i}{4\sigma_i^4} \boldsymbol{W}_i \boldsymbol{W}_i', \tag{9}$$

where t_i is $2\Phi_i + \phi_i \delta_i(\lambda_i \delta_i - \delta_i^2 - 1)$. $\mathbf{Q}_1$ is obtained by imposing the null hypothesis on the expressions in Equations (7)–(9), and estimating remaining unknowns using $\hat{\boldsymbol{\theta}}$. As various appearances of $f'(\hat{\alpha}_0)$ in the implied version of Equation (3) cancel, the $\mathbf{Q} = \mathbf{Q}_1$ statistic can be described as being invariant to the choice of f in Equation (2). The $\mathbf{Q} = \mathbf{Q}_1$ based statistic seems to have originally appeared in Jarque [13], where the derivation is based on specifying $\mathrm{var}(u_i)$ as $\alpha_0 + \boldsymbol{\alpha}_1' \boldsymbol{W}_i$. Jarque's statistic can also be derived from the results presented in [14]. The specification in Equation (2) is adopted by Lee and Maddala [16], who clearly have $\mathbf{Q} = \mathbf{Q}_1$ in mind for Equation (3), though it would be difficult to determine the required calculations on the basis of the information they present. The invariance to the choice of f is noted in [16] and by Godfrey [8], who refers to [14] and [13] for details of obtaining the $\mathbf{Q}$ under discussion. Orme [18] shows how the current version of Equation (3) can be obtained from an artificial 'double length' regression. None of the above papers present any Monte Carlo evidence.

An alternative to $\mathbf{Q}_1$ is $\mathbf{Q}_2 = (-\ell_{\boldsymbol{\theta},\boldsymbol{\theta}'})_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$ although obtaining the second derivatives of $\ell(\boldsymbol{\theta})$ is not an especially pleasing prospect. These second derivatives will involve f' and f'', the first and second derivatives of f, and (3) based on $\mathbf{Q}_2$ will require a choice of f to be made. An asymptotically valid alternative is to drop the terms in f'' in $\ell_{\boldsymbol{\theta},\boldsymbol{\theta}'}$. If $\ell^*_{\boldsymbol{\theta},\boldsymbol{\theta}'}$ denotes $\ell_{\boldsymbol{\theta},\boldsymbol{\theta}'}$ minus the terms in f'' then the statistic will be Equation (3) with $\mathbf{Q}$ set at $-\ell^*_{\boldsymbol{\theta},\boldsymbol{\theta}'}$ after imposing the null hypothesis and estimating remaining unknowns using $\hat{\boldsymbol{\theta}}$. This statistic will be invariant to the choice of the function f and can be shown to be equal to the statistic presented in Table 3 of [5] in a form involving generalised, or moment, residuals. However, the statistic has the disadvantage of not automatically being positive. If the option of making a choice for f is pursued then the exponential function might appeal if only because $f = f' = f''$ ought to simplify the required calculations. In fact, the resulting statistic can be related to that presented in Table 3 of [5]: it requires replacing $\hat{e}^{(4)} - (\hat{e}^{(2)})^2 - 6\hat{e}^{(2)}$ with $\hat{e}^{(4)} - (\hat{e}^{(2)})^2 - 4\hat{e}^{(2)}$ in the lower right-hand matrix in block B of that table.[2]

An alternative to $\mathbf{Q}_1$ and $\mathbf{Q}_2$ is $\mathbf{Q}_3 = (\sum_{i=1}^{n} \ell_{\boldsymbol{\theta}}(i)\ell_{\boldsymbol{\theta}}(i)')_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$, where $\ell_{\boldsymbol{\theta}}(i)$ denotes the contribution to $\ell_{\boldsymbol{\theta}}$ of observation i and can be deduced from Equations (4)–(6). $\mathbf{Q}_3$ produces a computationally appealing version of Equation (3) since the implied statistic is just n times the uncentred R^2 from a regression of 1 on $\ell_{\boldsymbol{\theta}}(i)$ after imposing the null hypothesis and using $\hat{\boldsymbol{\theta}}$ to estimate remaining unknowns.[3] This statistic appears in [5] in a form involving generalised residuals, and in [19,21] as a conditional moment test. It appears in [9] both as a conditional moment test and as a (supposedly different[4]) Lagrange multiplier test. Pagan and Vella [19] stress the computational simplicity of the $\mathbf{Q}_3$-based test and the intuitive appeal of $\ell_{\boldsymbol{\theta}} = \sum_{i=1}^{n} \ell_{\boldsymbol{\theta}}(i)$ at $\hat{\boldsymbol{\theta}}$ as the basis for a test of the assumption of constant variance in response to 'the sorry state of affairs' wherein estimated tobit models are not typically subjected to diagnostic testing. Skeels and Vella [21] provide some Monte Carlo results regarding the 'outer product of the gradient' statistic under discussion (their OPG) using a reduced form labour supply equation and a data set for 610 individuals originally used in [17]. The finding that OPG rejects a true null too frequently is a fairly standard conclusion in studies of the finite sample behaviour of asymptotic test procedures. An alternative statistic considered in [21] (their AD), is found to reject a true null too infrequently.[5] The AD statistic for the tobit model, $\mathrm{AD_t}$ say, is derived as a conditional moment test but can be presented in the form of Equation (3) if that expression is first written as

$$\ell_{\boldsymbol{\alpha}_1}(\hat{\boldsymbol{\theta}})'(\mathbf{Q}_{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_1'} - \mathbf{Q}_{\boldsymbol{\alpha}_1,\boldsymbol{\gamma}'}(\mathbf{Q}_{\boldsymbol{\gamma},\boldsymbol{\gamma}'})^{-1}\mathbf{Q}_{\boldsymbol{\gamma},\boldsymbol{\alpha}_1'})^{-1}\ell_{\boldsymbol{\alpha}_1}(\hat{\boldsymbol{\theta}}),$$

where $\boldsymbol{\gamma} = [\boldsymbol{\beta}' \ \ \alpha_0]'$. $\mathrm{AD_t}$ is then based on the estimation of $\mathbf{Q}_{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_1'} - \mathbf{Q}_{\boldsymbol{\alpha}_1,\boldsymbol{\gamma}'}(\mathbf{Q}_{\boldsymbol{\gamma},\boldsymbol{\gamma}'})^{-1}\mathbf{Q}_{\boldsymbol{\gamma},\boldsymbol{\alpha}_1'}$ as

$$\mathbf{Q}^{[3]}_{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_1'} - \mathbf{Q}^{[3]}_{\boldsymbol{\alpha}_1,\boldsymbol{\gamma}'}(\mathbf{Q}^{[2]}_{\boldsymbol{\gamma},\boldsymbol{\gamma}'})^{-1}\mathbf{Q}^{[*]}_{\boldsymbol{\gamma},\boldsymbol{\alpha}_1'} - \mathbf{Q}^{[*]}_{\boldsymbol{\alpha}_1,\boldsymbol{\gamma}'}(\mathbf{Q}^{[2]}_{\boldsymbol{\gamma},\boldsymbol{\gamma}'})^{-1}\mathbf{Q}^{[3]}_{\boldsymbol{\gamma},\boldsymbol{\alpha}_1'} + \mathbf{Q}^{[*]}_{\boldsymbol{\alpha}_1,\boldsymbol{\gamma}'}(\mathbf{Q}^{[2]}_{\boldsymbol{\gamma},\boldsymbol{\gamma}'})^{-1}\mathbf{Q}^{[3]}_{\boldsymbol{\gamma},\boldsymbol{\gamma}'}(\mathbf{Q}^{[2]}_{\boldsymbol{\gamma},\boldsymbol{\gamma}'})^{-1}\mathbf{Q}^{[*]}_{\boldsymbol{\gamma},\boldsymbol{\alpha}_1'}, \tag{10}$$

where the $^{[2]}$ and $^{[3]}$ indicate terms are estimated using $\mathbf{Q}_2$ or $\mathbf{Q}_3$, at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$. The matrix $\mathbf{Q}^{[*]}_{\boldsymbol{\gamma},\boldsymbol{\alpha}_1'}$ is obtained by estimating $\mathbf{Q}_{\boldsymbol{\beta},\boldsymbol{\alpha}_1'}$ using $\mathbf{Q}_2$ and $\underline{\mathbf{Q}}_{\alpha_0,\boldsymbol{\alpha}_1'}$ unbiasedly.[6] An advantage of $\mathrm{AD_t}$ is computational ease since the statistic can be calculated using the output from an artificial regression.[7] Despite Skeels and Vella [21] concluding that there is no evidence in their results to suggest $\mathrm{AD_t}$ is an improvement on OPG, which is the $\mathbf{Q}_3$-based version of Equation (3) recall, it is included in the Monte Carlo experiments of Section 4.

3. Tests for the probit model

The probit framework maximum likelihood estimator is typically interpreted as estimating $\boldsymbol{\gamma} = \boldsymbol{\beta}/\sigma$ in the equation which follows from dividing through Equation (1) by σ. The resulting expression includes $u_i^{\mathrm{p}} = u_i/\sigma$ which has unit variance under the null, and it follows that we need to write

$$\sigma_i^2 = \mathrm{var}(u_i^{\mathrm{p}}) = f(f^{-1}(1) + \boldsymbol{\alpha}_1'\boldsymbol{W}_i), \tag{11}$$

rather than Equation (2), when we work with the probit framework. The null hypothesis $H_0\colon \boldsymbol{\alpha}_1 = \mathbf{0}$ then gives $\sigma_i^2 = f(f^{-1}(1)) = 1$. A score test for $H_0\colon \boldsymbol{\alpha}_1 = \mathbf{0}$ will still be of the form in Equation (3) if ℓ is now the heteroskedastic probit log-likelihood function given by

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n}((1-d_i)\ln(1-\Phi_i) + d_i \ln(\Phi_i)),$$

where Φ_i is Φ at $\delta_i = \boldsymbol{X}_i'\boldsymbol{\gamma}/\sigma_i$. The first derivatives are

$$\frac{\partial \ell}{\partial \boldsymbol{\gamma}} = \sum_{i=1}^{n} \frac{\phi_i(d_i - \Phi_i)}{\sigma_i \Phi_i(1-\Phi_i)}\boldsymbol{X}_i \tag{12}$$

and

$$\frac{\partial \ell}{\partial \boldsymbol{\alpha}_1} = -\sum_{i=1}^{n} \frac{(d_i - \Phi_i)\phi_i\delta_i f_i'}{2\sigma_i^2\Phi_i(1-\Phi_i)}\boldsymbol{W}_i, \tag{13}$$

where ϕ_i is ϕ at δ_i. The estimator $\hat{\boldsymbol{\theta}}$ is obtained by imposing the null hypothesis on $\partial\ell/\partial\boldsymbol{\gamma}$ in Equation (12), equating to $\mathbf{0}$ and solving for $\boldsymbol{\gamma}$.

The choices of $\mathbf{Q}$ in the probit framework mirror those considered in the tobit framework so that we first consider $\mathbf{Q}_1 = (-\mathrm{E}(\ell_{\boldsymbol{\theta},\boldsymbol{\theta}'}))_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$. Again using $\underline{\mathbf{Q}}$ to denote $-\mathrm{E}(\ell_{\boldsymbol{\theta},\boldsymbol{\theta}'})$ it is possible to show that the blocks of

$$\underline{\mathbf{Q}} = \begin{bmatrix} \underline{\mathbf{Q}}_{\boldsymbol{\gamma},\boldsymbol{\gamma}'} & \underline{\mathbf{Q}}_{\boldsymbol{\gamma},\boldsymbol{\alpha}_1'} \\ & \underline{\mathbf{Q}}_{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_1'} \end{bmatrix}$$

are

$$\underline{\mathbf{Q}}_{\boldsymbol{\gamma},\boldsymbol{\gamma}'} = \sum_{i=1}^{n} p_i \boldsymbol{X}_i \boldsymbol{X}_i', \quad \underline{\mathbf{Q}}_{\boldsymbol{\gamma},\boldsymbol{\alpha}_1'} = -\sum_{i=1}^{n} \frac{\delta_i f_i'}{2\sigma_i} p_i \boldsymbol{X}_i \boldsymbol{W}_i', \quad \underline{\mathbf{Q}}_{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_1'} = \sum_{i=1}^{n} \left[\frac{\delta_i f_i'}{2\sigma_i}\right]^2 p_i \boldsymbol{W}_i \boldsymbol{W}_i', \quad (14)$$

where p_i is $\phi_i^2/(\sigma_i^2 \Phi_i(1-\Phi_i))$. Imposing the null hypothesis on the expressions in (12)–(14) and evaluating remaining unknowns at $\hat{\boldsymbol{\theta}}$ gives a version of Equation (3) in which the various appearances of $f'(f^{-1}(1))$ cancel so that the statistic is invariant to the choice of f in Equation (11). The $\mathbf{Q}_1$-based statistic under discussion is in fact identical to the LM_2 statistic of [6], where the derivation assumes f in Equation (11) is the exponential function. Davidson and MacKinnon [6] show that LM_2 can be obtained as the explained sum of squares from an artificial regression. Two other statistics associated with the same artificial regression are judged inferior to LM_2 on the basis of the Monte Carlo experiments performed in [6].[8] Godfrey [8] presents the statistic in a way which involves reference to Equation (11).

As in the tobit model a version of Equation (3) can be based on $\mathbf{Q}_2 = (-\ell_{\boldsymbol{\theta},\boldsymbol{\theta}'})_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$, where the second derivatives will be quite complicated and will involve f''. Again as in the tobit model, ignoring the terms involving f'' in $\ell_{\boldsymbol{\theta},\boldsymbol{\theta}'}$ leads to a statistic which is identical to that featured in Table 3 of [5] and which has the disadvantage of being able to take on negative values. The statistic based on $-\ell_{\boldsymbol{\theta},\boldsymbol{\theta}'}$ with f chosen to be the exponential function can be obtained from Table 3 of [5] using the same modifications as outlined above in the tobit framework.

The probit statistic based on $\mathbf{Q}_3 = (\sum_{i=1}^{n} \ell_{\boldsymbol{\theta}}(i)\ell_{\boldsymbol{\theta}}(i)')_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$ is in fact the LM_1 statistic of [6], where the derivation is based on the exponential function for f in Equation (11). In the Monte Carlo experiments performed by Davidson and MacKinnon [6] the statistic rejects a true null hypothesis too frequently. The same statistic appears in [5] in a form involving generalised residuals, and in [19,21] as a conditional moment test. The AD statistic of [21], AD_p say, is derived as a conditional moment statistic but can be presented in the form (3) if that expression is firstly written as

$$\ell_{\boldsymbol{\alpha}_1}(\hat{\boldsymbol{\theta}})'(\mathbf{Q}_{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_1'} - \mathbf{Q}_{\boldsymbol{\alpha}_1,\boldsymbol{\gamma}'}(\mathbf{Q}_{\boldsymbol{\gamma},\boldsymbol{\gamma}'})^{-1}\mathbf{Q}_{\boldsymbol{\gamma},\boldsymbol{\alpha}_1'})^{-1}\ell_{\boldsymbol{\alpha}_1}(\hat{\boldsymbol{\theta}}),$$

where $\mathbf{Q}_{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_1'} - \mathbf{Q}_{\boldsymbol{\alpha}_1,\boldsymbol{\gamma}'}(\mathbf{Q}_{\boldsymbol{\gamma},\boldsymbol{\gamma}'})^{-1}\mathbf{Q}_{\boldsymbol{\gamma},\boldsymbol{\alpha}_1'}$ is estimated as

$$\mathbf{Q}^{[3]}_{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_1'} - \mathbf{Q}^{[3]}_{\boldsymbol{\alpha}_1,\boldsymbol{\gamma}'}(\mathbf{Q}^{[2]}_{\boldsymbol{\gamma},\boldsymbol{\gamma}'})^{-1}\mathbf{Q}^{[2]}_{\boldsymbol{\gamma},\boldsymbol{\alpha}_1'} - \mathbf{Q}^{[2]}_{\boldsymbol{\alpha}_1,\boldsymbol{\gamma}'}(\mathbf{Q}^{[2]}_{\boldsymbol{\gamma},\boldsymbol{\gamma}'})^{-1}\mathbf{Q}^{[3]}_{\boldsymbol{\gamma},\boldsymbol{\alpha}_1'} + \mathbf{Q}^{[3]}_{\boldsymbol{\alpha}_1,\boldsymbol{\gamma}'}(\mathbf{Q}^{[3]}_{\boldsymbol{\gamma},\boldsymbol{\gamma}'})^{-1}\mathbf{Q}^{[2]}_{\boldsymbol{\gamma},\boldsymbol{\gamma}'}(\mathbf{Q}^{[3]}_{\boldsymbol{\gamma},\boldsymbol{\gamma}'})^{-1}\mathbf{Q}^{[3]}_{\boldsymbol{\gamma},\boldsymbol{\alpha}_1'}, \quad (15)$$

where the $^{[2]}$ and $^{[3]}$ indicate terms are estimated using $\mathbf{Q}_2$ or $\mathbf{Q}_3$, at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$. As with AD_t it is possible to calculate AD_p using the output from an artificial regression.

4. Monte Carlo evidence

We consider the results of 2500 replications for each of 45 experiments involving the model

$$y_i^* = \beta_0 + \beta_1 X_i + u_i.$$

Experiments differ by the value of n, the number of observations, β_0, and a, a parameter determining the degree of heteroskedasticity, as explained below.[9] The n considered are 150, 300, and 600. The values of β_0 considered are -1.28, -0.68, 0, 0.68, and 1.28. Varying the value of β_0 allows for different numbers of limit observations, where a limit observation corresponds to $y_i^* \leq 0$ and $y_i = 0$ in the tobit model. A larger β_0 lowers the likelihood of limit observations. Intuitively the more limit observations there are the less additional information is available for the tobit estimator compared with the probit estimator. The X_i are based on a random generation of 150 U$(-1, 1)$ values. Twofold (fourfold) repetition is needed for the $n = 300$ ($n = 600$) simulations. The X_i are fixed across replications. The u_i are generated as $N(0, \sigma_i^2)$ where

$$\sigma_i = \exp(a \times (X_i + 1)^2),$$

where $a = 0$ implies $\sigma_i = 1$ and homoskedasticity and $a = 0.2$ and 0.4 imply heteroskedasticity. The various tests are based on the $\boldsymbol{W}_i$ vector being specified as the scalar X_i. The tests considered are the $\mathbf{Q}_1$-based test for the tobit and probit models ($T_{1\mathrm{t}}$ and $T_{1\mathrm{p}}$, respectively), the $\mathbf{Q}_3$-based test for the tobit and probit models ($T_{3\mathrm{t}}$ and $T_{3\mathrm{p}}$, respectively), AD_t, AD_p, and T_{lat}, a test based on assuming the y_i^* are available.

The $a = 0$ rejection percentages when testing is done at the 5% significance level are in Table 1. The most noticeable feature is the overrejection of the null by $T_{3\mathrm{p}}$ and $T_{3\mathrm{t}}$. The degree of overrejection decreases with n but is still noticeable for $T_{3\mathrm{t}}$ ($T_{3\mathrm{p}}$) with negative (large absolute)

Table 1. Rejection percentages when $a = 0$.

β_0 equals	-1.28	-0.68	0	0.68	1.28
$n = 150$					
$T_{1\mathrm{p}}$	3.76	5.08	5.24	5.48	4.24
$T_{3\mathrm{p}}$	21.48	9.52	6.52	8.16	16.24
AD_p	13.64	8.00	5.80	6.60	10.32
$T_{1\mathrm{t}}$	3.60	5.04	4.68	4.84	4.40
$T_{3\mathrm{t}}$	17.12	11.24	7.52	7.32	6.36
AD_t	5.24	5.32	4.36	4.64	3.80
T_{lat}	4.96	4.72	4.76	4.96	4.84
$n = 300$					
$T_{1\mathrm{p}}$	4.72	4.72	5.04	5.32	4.92
$T_{3\mathrm{p}}$	11.84	6.80	5.20	6.56	10.88
AD_p	9.00	6.32	4.96	6.00	8.60
$T_{1\mathrm{t}}$	4.68	4.56	5.00	4.92	4.52
$T_{3\mathrm{t}}$	13.08	7.92	6.52	5.76	6.00
AD_t	5.52	4.92	4.96	4.60	4.88
T_{lat}	5.28	4.36	4.16	4.48	4.96
$n = 600$					
$T_{1\mathrm{p}}$	5.12	5.52	5.32	5.12	5.96
$T_{3\mathrm{p}}$	8.60	6.12	5.20	5.60	9.16
AD_p	7.56	5.96	5.08	5.32	7.72
$T_{1\mathrm{t}}$	4.36	5.64	4.80	4.88	4.20
$T_{3\mathrm{t}}$	9.48	7.12	5.76	5.52	5.20
AD_t	5.96	5.56	4.88	4.52	4.28
T_{lat}	4.80	5.00	4.60	4.72	4.84

β_0 at $n = 600$. For each n the overrejection displayed by T_{3t} tends to decrease as β_0 increases. For each n the overrejection displayed by T_{3p} is minimised at $\beta_0 = 0$. For $n = 300$ and 600 the rejection rate of T_{3p} seems symmetric about $\beta_0 = 0$. The contrasting behaviour of AD_p and AD_t is interesting. AD_p behaves in a broadly similar manner to T_{3p}, though with a reduced degree of overrejection in all cases, whereas the rejection rate of AD_t is acceptably close to 5%. The same comment seems to be applicable as regards T_{1p}, T_{1t}, and T_{lat}.

Table 2 presents averages and variances for the $n = 150$ experiments.[10] The values for T_{3p} and T_{3t} are again the most noteworthy. For both statistics, for all β_0 the average exceeds one and the variance exceeds two, on occasions by a striking amount. The closest correspondence to an average of one and a variance of two is at $\beta_0 = 0$ ($\beta_0 = 1.28$) for T_{3p} (T_{3t}). The contrast between AD_p and AD_t is also worth noting. AD_p tends to be too large and too variable whereas AD_t tends to assume the correct value but to vary insufficiently.

The $a = 0.2$ rejection percentages when testing is done at the 5% significance level are in Table 3. There is some uniformity in the results in that (i) in all 45 experiments the rejection rate of T_{lat} is never exceeded, (ii) in 44 of the 45 experiments the three lowest rejection rates are displayed by T_{1p}, T_{3p}, and AD_p[11], so that (iii) in 44 of the 45 experiments the power of the tobit statistics T_{1t}, T_{3t}, and AD_t lies between the power of T_{lat} and that of T_{1p}, T_{3p}, and AD_p, and (iv) aside from T_{3p} when $\beta_0 = -1.28$ the rejection frequency is always non-decreasing in n. The rejection frequency displayed by T_{1t}, T_{3t}, and AD_t never decreases as β_0 increases. When β_0 is negative there is a suggestion of greater power for AD_t than T_{1t}. It is more difficult to arrive at general conclusions regarding the probit statistics. We might concentrate on T_{1p}, given the performance of the various statistics under the null, and note the tendency for the power of this statistic to increase with β_0. For negative β_0 the rejection rates of T_{1p} are low both in absolute terms and in comparison to, for example, T_{1t}. For $n = 150$ and $n = 300$ the behaviour of AD_p seems almost the exact opposite to that of T_{1t}.

The $a = 0.4$ rejection percentages when testing is done at the 5% significance level are in Table 4. To a large extent the comments which are appropriate are a repeat of those associated with Table 3. However, it is worth noting the curious behaviour of AD_p when β_0 is positive, with the exception of $\beta_0 = 0.68$ and $n = 600$. In these five cases we find the increase in a from $a = 0.2$ in Table 3 to $a = 0.4$ in Table 4 leads to a reduced rate of rejection of the null hypothesis. Otherwise the rejection rate increases when $a = 0.2$ in Table 3 is replaced by $a = 0.4$ in Table 4.

Table 2. Average values and variances of the statistics when $a = 0$ and $n = 150$.

β_0 equals	-1.28	-0.68	0	0.68	1.28
(a) Averages					
T_{1p}	1.003	0.990	1.016	0.995	1.011
T_{3p}	2.810	1.374	1.103	1.231	2.228
AD_p	1.674	1.208	1.051	1.105	1.417
T_{1t}	0.893	0.973	0.991	0.969	0.957
T_{3t}	2.321	1.522	1.248	1.177	1.119
AD_t	1.128	1.054	0.989	1.000	0.952
T_{lat}	1.004	1.004	1.008	0.990	0.984
(b) Variances					
T_{1p}	1.751	1.905	2.056	1.941	1.828
T_{3p}	25.058	5.775	2.616	3.862	16.683
AD_p	5.357	3.481	2.227	2.611	3.673
T_{1t}	1.862	2.283	2.166	1.866	2.059
T_{3t}	15.724	5.538	3.164	2.724	2.578
AD_t	1.967	1.831	1.555	1.711	1.575
T_{lat}	1.946	1.951	1.863	1.827	2.005

Table 3. Rejection percentages when $a = 0.2$.

β_0 equals	-1.28	-0.68	0	0.68	1.28
$n = 150$					
T_{1p}	5.28	8.92	13.72	17.48	16.72
T_{3p}	24.24	17.44	15.12	16.28	15.12
AD_p	17.96	14.88	11.32	5.72	1.60
T_{1t}	13.20	33.04	62.04	83.52	91.88
T_{3t}	45.48	57.40	75.84	89.12	94.00
AD_t	21.48	38.64	61.68	81.52	89.68
T_{lat}	97.04	96.76	97.28	97.08	97.32
$n = 300$					
T_{1p}	8.24	13.52	19.40	27.04	28.84
T_{3p}	22.04	18.92	20.76	26.04	26.52
AD_p	20.04	18.20	17.92	17.44	6.80
T_{1t}	29.92	62.28	90.84	98.92	99.80
T_{3t}	55.56	76.96	94.32	99.16	99.92
AD_t	40.96	67.04	91.08	98.60	99.68
T_{lat}	100.00	100.00	99.96	100.00	100.00
$n = 600$					
T_{1p}	12.64	21.08	36.88	46.64	53.32
T_{3p}	21.96	25.40	37.48	45.28	50.80
AD_p	20.80	24.64	36.16	41.04	31.04
T_{1t}	58.08	91.88	99.84	100.00	100.00
T_{3t}	73.52	95.24	99.88	100.00	100.00
AD_t	66.08	93.60	99.88	100.00	100.00
T_{lat}	100.00	100.00	100.00	100.00	100.00

Table 4. Rejection percentages when $a = 0.4$.

β_0 equals	-1.28	-0.68	0	0.68	1.28
$n = 150$					
T_{1p}	11.44	17.28	21.60	26.36	28.32
T_{3p}	34.28	26.28	23.72	25.48	26.08
AD_p	29.36	23.60	14.40	2.32	0.16
T_{1t}	56.28	85.52	99.16	99.96	99.96
T_{3t}	81.60	94.56	99.64	100.00	99.96
AD_t	57.72	81.76	96.32	99.08	99.24
T_{lat}	100.00	100.00	100.00	100.00	100.00
$n = 300$					
T_{1p}	20.76	29.08	39.40	46.08	49.72
T_{3p}	36.68	35.56	40.52	45.28	47.20
AD_p	34.80	28.50	33.52	12.16	0.52
T_{1t}	88.16	99.04	100.00	100.00	100.00
T_{3t}	94.48	99.76	100.00	100.00	100.00
AD_t	88.64	98.60	99.92	100.00	100.00
T_{lat}	100.00	100.00	100.00	100.00	100.00
$n = 600$					
T_{1p}	36.96	51.76	63.68	73.92	76.40
T_{3p}	47.68	55.20	64.52	73.36	75.28
AD_p	46.92	54.56	61.80	47.84	1.32
T_{1t}	99.84	100.00	100.00	100.00	100.00
T_{3t}	99.92	100.00	100.00	100.00	100.00
AD_t	99.72	100.00	100.00	100.00	100.00
T_{lat}	100.00	100.00	100.00	100.00	100.00

Also note that the tendency in Table 3 for $\mathrm{AD_t}$ to perform better than $T_{1\mathrm{t}}$ when β_0 is negative is no longer present in Table 4.

The above results suggest that $T_{1\mathrm{t}}$ and $T_{1\mathrm{p}}$ are likely to be preferred to $T_{3\mathrm{t}}$ and $T_{3\mathrm{p}}$ unless further work, perhaps involving bootstrapping, is to be undertaken. The statistic $\mathrm{AD_p}$ has little to recommend it.

It is worth briefly discussing the possible responses to heteroskedasticity in the tobit and probit models. One response is to resort to the 'robust' standard errors most associated with [24] in the econometrics literature, as in [22] for example. However, as [9, pp. 673–674] argues, the usefulness of this approach is unclear given that heteroskedasticity implies inconsistent parameter estimation. An alternative response would be to specify f in Equation (2) or Equation (11) and rework the maximum likelihood estimation. A final possibility is the use of an estimation approach capable of producing consistent estimators even in the presence of heteroskedasticity of an unspecified form. The possibilities include the censored least absolute deviations estimator of [20] in the tobit model and the semiparametric estimator of [15] in the probit model.[12] A detailed discussion of [20] and [15], and associated estimators, can be found in, for example [4].

5. Conclusions

The Monte Carlo results of Section 4 suggest that test statistics based on expected second derivatives are to be preferred to test statistics based on first derivatives, unless a large number of observations is available. This conclusion, which is a fairly common one in econometrics, is based on a consideration of behaviour when the null hypothesis is true. Statistics based on first derivatives do have an advantage when computation is considered. This has been stressed by Pagan and Vella [19] and Skeels and Vella [21] as being useful in trying to encourage the diagnostic testing of tobit and probit models. There are other factors which might be of importance as well as ease of computation. One would be the requirements for journal publication of work involving tobit or probit estimation. Further the issue of computational ease does not arise if useful statistics are reported by the software typically available to applied researchers. For example, the open-source econometrics package Gretl automatically reports a normality test after tobit estimation. The nature of the statistic, rather than its ease of calculation, is perhaps the first question to consider.

Notes

1. See, for example [1,12]. Failure of the normality assumption also results in estimator inconsistency. See, for example [2,11].
2. $\hat{e}^{(j)}$, $j = 1, 2, \ldots$, indicates different varieties of generalised residual.
3. The $f'(\hat{\alpha}_0)$ which appears in the implied versions of (5) and (6) can be dropped so that the statistic is invariant to the choice of f.
4. Where 'the difference is in how the rows of $\mathbf{M}$ are constructed' [9, p. 773]. But the $\mathbf{M}$ that Greene subsequently outlines is just $2\sigma^4$ times the previous $\mathbf{M}$, which corresponds to the $b_i \boldsymbol{W}_i$ component of the $\boldsymbol{g}_i$ vector of page 769. It follows that the two statistics presented by Greene [9] are the same.
5. The rejection rate is 1.01% when a nominal significance level of 5% is used.
6. As
$$\frac{(f')^2}{4\sigma^4}\sum_{i=1}^{n}(2d_i + (1-d_i)\lambda_i\delta_i(\delta_i\lambda_i - \delta_i^2 - 1))\boldsymbol{W}_i'.$$
The expectation of this term is the $\underline{\mathbf{Q}}_{\alpha_0,\boldsymbol{\alpha}_1'}$ of Equation (9) after the null is imposed.
7. The expression presented in [19] as a test statistic corresponds exactly to the regression implemented in the Monte Carlo experiments of section 4. Their text in fact refers to an asymptotically equivalent, but numerically different, statistic requiring an artificial multivariate regression, unless $p = 1$. Quite what calculations are actually undertaken by Pagan and Vella is explored in [23].

8. One of these statistics is n times the R^2 from the artificial regression that produces LM_2. In fact this is the statistic outlined in [10, pp. 456–457].
9. The simulations were carried out using a programme written in Gauss [7]. It is available from the author on request.
10. The tables for $n = 300$ and 600 are not included to economise on space. They are available from the author on request. Recall that $E(\chi_1^2) = 1$ and $var(\chi_1^2) = 2$.
11. The exception to this pattern is with $\beta_0 = -1.28$ and $n = 150$.
12. An additional advantage of this approach is that the assumption of normality for the u_i is not required to establish consistency.

References

[1] A. Arabmazar and P. Schmidt, *Further evidence on the robustness of the Tobit estimator to heteroscedasticity*, J. Econometrics 17(2) (1982), pp. 253–258.

[2] A. Arabmazar and P. Schmidt, *An investigation of the robustness of the tobit estimator to nonnormality*, Econometrica 50(4) (1982), pp. 1055–1063.

[3] T.S. Breusch and A.R. Pagan, *A simple test for heteroscedasticity and random coefficient variation*, Econometrica 47(5) (1979), pp. 1287–1294.

[4] A.C. Cameron and P.K. Trivedi, *Microeconometrics: Methods and Applications*, Cambridge University Press, Cambridge, UK, 2005.

[5] A. Chesher and M. Irish, *Residual analysis in the grouped and censored normal linear model*, J. Econometrics 34 (1987), pp. 33–61.

[6] R. Davidson and J.G. MacKinnon, *Convenient specification tests for logit and probit models*, J. Econometrics 25(3) (1984), pp. 241–262.

[7] Gauss, Aptech Systems, Maple Valley, Washington.

[8] L.G. Godfrey, *Misspecification Tests in Econometrics*, Cambridge University Press, Cambridge, UK, 1988.

[9] W.H. Greene, *Econometric Analysis*, 5th ed., Prentice-Hall, Englewood Cliffs, NJ, 2003.

[10] H. Heij, P. DeBoer, P. Franses, T. Kloek, and H.K. VanDijk, *Econometric Methods with Applications in Business and Economics*, Oxford University Press, Oxford, UK, 2004.

[11] D. Holden, *Testing the normality assumption in the tobit model*, J. Appl. Statist. 31(5) (2004), pp. 521–532.

[12] M. Hurd, *Estimation in truncated samples when there is heteroscedasticity*, J. Econometrics 11 (1979), pp. 247–258.

[13] C.M. Jarque, *A test for heteroscedasticity in a limited dependent variable model*, Austral. J. Statist. 23(2) (1981), pp. 159–163.

[14] C.M. Jarque and A.K. Bera, *Efficient specification tests for limited dependent variable models*, Econom. Lett. 9(2) (1982), pp. 153–160.

[15] R.W. Klein and R.H. Spady, *An efficient semiparametric estimator for binary response models*, Econometrica 61(2) (1993), pp. 387–421.

[16] L-F. Lee and G.S. Maddala, *The common structure of tests for selectivity bias, serial correlation, heteroskedasticity and non-normality in the tobit model*, Int. Econom. Rev. 26(part 1) (1985), pp. 1–20.

[17] R. Moffitt, *The estimation of a joint wage-hours labour supply model*, J. Labour Econom. 2(4) (1984), pp. 550–566.

[18] C. Orme, *Efficient score tests for heteroskedasticity in micro-econometrics*, Econometric Rev. 11(part 2) (1992), pp. 235–252.

[19] A. Pagan and F. Vella, *Diagnostic tests for models based on individual data: A survey*, J. Appl. Econometrics 4 (suppl) (1989), pp. s29–s59.

[20] J.L. Powell, *Least absolute deviations estimation for the censored regression model*, J. Econometrics 25(3) (1984), pp. 303–325.

[21] C.L. Skeels and F. Vella, *A Monte Carlo investigation of the sampling behaviour of conditional moment tests in Tobit and Probit models*, J. Econometrics 92(2) (1999), pp. 275–294.

[22] J.H. Stock and M.W. Watson, *Introduction to Econometrics*, 2nd ed., Pearson, Boston, 2007.

[23] C. Wells, *Retesting Fair's (1978) model on infidelity*, J. Appl. Econometrics 18(2) (2003), pp. 237–239.

[24] H. White, *Maximum likelihood estimation of misspecified models*, Econometrica 50 (1982), pp. 1–16.