ECONOMIC THEORY, APPLICATIONS AND ISSUES

Working Paper No. 15

OLS and Tobit Estimates: When is
Substitution Defensible Operationally?

by

Clevo Wilson and Clem Tisdell

May 2002

THE UNIVERSITY OF QUEENSLAND

Working Paper No.15

OLS and Tobit Estimates: When is Substitution
Defensible Operationally?

by

Clevo Wilson[†] and Clem Tisdell[†]

May 2002

† School of Economics, The University of Queensland, Brisbane 4072 Australia.
  Email: c.tisdell@economics.uq.edu.au

OLS AND TOBIT ESTIMATES: WHEN IS SUBSTITUTION DEFENSIBLE OPERATIONALLY?

Abstract

Field data are used to illustrate that, other things constant, regression results using Ordinary Least Squares (OLS) converge to Tobit estimates as the number of zeros in the regressand decrease. Tobit estimates are theoretically superior to OLS estimates when using censored data. However, if little difference exists between OLS and Tobit estimates, OLS may be operationally acceptable. OLS may even be optimal in a bounded rationality sense because the extra cost of using Tobit analysis may be less than the extra benefits from a very slight increase in accuracy.

OLS AND TOBIT ESTIMATES: WHEN IS SUBSTITUTION DEFENSIBLE OPERATIONALLY?

1. Introduction

The use of a Tobit model is recommended on theoretical grounds in preference to OLS models for data sets with censored samples (Gujarati, 1995). A sample in which information on the regressand is available only for some observations is known as a censored sample. Since the first use of the contingent valuation approach in the mid-1960s for environmental and health valuation studies, use of Tobit models has grown rapidly, especially since the early 1970s[1]. The use of OLS models in the case of censored sample data sets make OLS estimates biased and inefficient, thus violating the basic tenets of *Best Linear Unbiased Estimator* (BLUE) conditions. However, OLS estimates become biased and inefficient depending on the number of zeros in relation to the number of observations in the data set. The greater is the number of zeros in relation to the total number of observations, the greater is the instability of the OLS estimates and vice versa. In cases where the number of zeros is low, then the difference between OLS and Tobit estimates is usually marginal. This is particularly relevant for certain willingness to pay (WTP) data sets with no or few zeros in the dependent variable. Furthermore, the removal of protest bids from the sample by certain researchers further reduces the number of zeros in the dependent variable. In such cases (depending on the number of zeros in the regressand), OLS may also be used, although a Tobit analysis is the theoretically preferred method.

In this article we use two field survey cross sectional data sets to show how OLS and Tobit estimates vary depending on the number of zeros in the regressand. There are many practical implications stemming from this observation. One issue is that researchers may have to use more than one software package if a Tobit model has to be used since some commonly used software such as Microfit and SPSS do not run Tobit models. In such cases acquiring and mastering additional software is not only time consuming, but expensive, especially to researchers and institutions in developing countries. This is especially so if software has to be purchased specifically to run Tobit models[2]. With software prices increasing and licenses having to be renewed yearly, the use of specific software to run Tobit models can be prohibitive, especially in developing countries. Furthermore, some of the software that run Tobit models is still user-unfriendly. All this could prevent or delay research work being conducted in this area. Therefore,

if the OLS and Tobit estimates are almost identical (or only show marginal differences) when the number of zeros in the regressand is small, then it may be rational to substitute OLS for Tobit estimates.

2. A Brief Discussion of the Tobit Model: Issues Involved

The Tobit model (extension of the Probit model) was developed by James Tobin in 1958 (Tobin, 1958)[3]. The model was developed in order to handle cross sectional data sets where some observations in the sample lacked data or had zero values for the dependent variable.

For example, in determining the costs incurred by farmers due to ill health resulting from exposure to pesticides during handling and spraying or how much money an individual farmer is willing to pay to avoid exposure to pesticides poses a problem. This occurs when using OLS because not all farmers would have suffered from ill health and, therefore, some would not have incurred any costs. Similarly, not all farmers are willing to pay to avoid exposure to pesticides and the resulting illnesses because some do not incur any costs. Hence, in such cases zero bids are not uncommon. Therefore, there are zeros involved in the dependent variable either because farmers have zero costs or they are unwilling to pay anything (a zero value) to avoid exposure to pesticides. As a result, we have two groups of farmers, one consisting of, say, n_1 who have incurred costs due to ill health or are willing to pay to avoid exposure to pesticides (in other words farmers with non zero bids) and those farmers consisting of, say, n_2 who have not incurred any costs or who have given zero bids. In such a case, if these data were to be used as the dependent variable to estimate, for example, farmers' willingness to pay to avoid exposure to pesticides, the regress and is a censored sample (Gujarati, 1995).

Amemiya (1984, p.5) points out that the presence of zeros in the dependent variable "destroys the linearity assumption so that the least squares method is clearly inappropriate". As a result, a Tobit analysis is used in preference to OLS. The standard Tobit model can be defined as follows (Amemiya, 1984):

$$y_i^* = x_i'\beta + u_i, \qquad i = 1,2, \ldots, n, \qquad (1)$$

$$y_i = y_t^*, \quad \text{if RHS} > 0$$
$$y_i = 0, \quad \text{otherwise}$$

where y_i^* is a non-observable random variable. It is theoretically incorrect to estimate regression (1) using only n_1 observations and not using the remaining n_2 observations because the OLS estimates of the parameters obtained from the subset of n_1 observations will be biased as well as inconsistent. As Gujarati (1985) points out "the bias arises from the fact that if we only consider the n_1 observations and omit the others, there is no guarantee that $E(u_i)$ will be necessarily zero. And without $E(u_i) = 0$ we cannot guarantee that the OLS estimators will be unbiased". This is shown in Gujarati (1995). Therefore, if a regression is conducted considering only the observed variables (i.e. n_1) then the resulting intercept and the coefficients are bound to be different than if all observations (i.e. $n_1 + n_2$) were included. Tobit regression models have been designed to handle such data sets and involve the method of maximum likelihood.

3. Field Survey WTP Data Showing Similarities and Differences in OLS and Tobit Estimates: Pesticide Exposure, and Sea Turtle Conservation as Case Studies

In this section, two field survey cross sectional data sets are used to show the differences in OLS and Tobit estimates and to illustrate that the differences in the estimates depend on the number of zeros in the dependent variable. The data were transformed into square roots.

The first data set was collected in 1996 in Sri Lanka for a study on farmers' WTP to avoid exposure to pesticides. The number of observations is 203. In the dependent variable there were no zero bids. The regression analysis is conducted with six independent variables that are believed to influence the WTP for pollution control (Wilson, 2002). Since, the purpose of the paper is to show how zeros affect the coefficients, the independent variables are not named, but numbered from A1 to A6. Zeros were introduced to show how the estimates changed as the

number of zeros increased. In the first instance, eight (4%) zeros were introduced at 25 observation intervals. Thereafter, the numbers of zeros were increased to 16 (8%), 24 (12%), 32 (16%), 40 (20%), 48 (24%), 56 (28) and 101 (50%). The results are shown in Table 1. Another experiment was also conducted by placing zeros in the beginning and at the end of the data set. The placing of zeros in the data set did not result in noticeable differences in estimates until they were more than 25% of the total number of observations.

The second data set was obtained from a field survey in Mon Repos, Queensland in 2000 to determine tourists' WTP for sea turtle conservation in Australia (Tisdell and Wilson, 2002). The number of observations was 333. There were 69 (21%) zeros[4]. In order to conduct the experiment, zeros were substituted with bids given by the respondents. This was done randomly. This made it possible to introduce zero bids randomly to the data set. Initially, eight (2.3%) zeros were introduced, thereafter, eight variables were added at regular intervals up to 25% of the total number of observations as shown in Table 2. One experiment was also conducted with 50% of zeros in the data set (Table 2).

4. Conclusions

As the results show, the number of zeros in the dependent variable in these cases have to be significantly large for differences in estimates between OLS and Tobit analysis to emerge. The estimates are not as sensitive as might be expected to censored data. It seems that OLS estimates could be substituted for Tobit when the number of zeros in the dependent variable are not large. However, establishing an acceptable 'cut off' point remains analytically elusive. This is because the data differ (in both the dependent and the independent variables) from sample to sample. Furthermore, benefits and costs of extra accuracy vary from situation to situation. Nonetheless, the article establishes the fact that the net benefit of using OLS can exceed that of using Tobit when the number of censored observations is relatively low. In such situations, OLS can be optimal from a bounded rationality viewpoint (cf. Tisdell, 1996, Ch. 1). This is similar to a sampling problem mentioned by Moroney (1956, p.173) who states that attention has to be paid to the "problem of achieving the maximum economy of inspection effort compatible to the degree of risk willingly faced". Similarly, Anderson et al. (1996, p.10) point out that "the cost of data acquisition and the subsequent statistical analysis should not exceed the savings generated by using the information to make a better decision". In general, statistical analysis and

evaluation techniques involving greater accuracy should only be applied if the extra benefits exceed the extra costs (cf. Baumol and Quandt, 1964; Tisdell, 1996, Ch. 1).

References

Amemiya, T., 1984, Tobit Models: A Survey, Journal of Econometrics, 24, 3-61.

Anderson, D. R., Sweeney, D. J. and Williams, T. A., 1996, Statistics for Business and Economics, West Publishing Company, New York.

Baumol, W. and Quandt, R., 1964, Rules of Thumb and Optimally Imperfect Decisions, American Economic Review, 54, 23-46.

Gujarati, D. N., 1995, Basic Econometrics, International Third Edition, McGraw-Hill, Inc. New York.

Moroney, M. J., 1956, Facts from Figures, Penguin, Third and Revised Edition, Harmondsworth, UK.

Tisdell, C., 1996, Bounded Rationality and Economic Evolution – A Contribution to Decision Making, Economics and Management, pp. 3-18, Edward Elgar, Cheltenham, UK.

Tisdell, C. and Wilson, C., 2002, Economic, Educational and Conservation Benefits of Sea Turtle-Based Ecotourism: A Study Focused on Mon Repos, Coopertaive Research Centre for Sustainable Tourism, Griffith University, Gold Coast, Queensland (In press).

Tobin, J., 1958, Estimation of Relationships for Limited Dependent Variables, Econometrica, 26, 24-36.

Wilson, C., 2002, Pesticide Avoidance: Results from a Sri Lankan Study with Health Policy Implications, In: Hall, D. C and Moffitt, L. J (editors) Economics of Pesticides, Sustainable Food Production, and Organic Food Markets, Advances in the Economics of Environmental Resources, Volume IV, Elsevier Science, New York.

Acknowledgements

Notes

1. Amemiya (1984, p.7) states that "especially since the 1970s, numerous applications of the standard Tobit model have appeared in economic journals, encompassing a wide range of fields of economics".

2. It is interesting to note that Amemiya (1984, p. 4) claims that "a recent advance in computer technology" has "made estimation of large-scale Tobit models feasible".

However, it must be pointed out that in most developing countries such computer technology remains prohibitively expensive even to this day.

3. Tobin applied his model to data on 735 non-farm households obtained from surveys of consumer finances. His dependent variable was the ratio of total durable goods expenditure to disposable income and the independent variables were the age of the head of the household and the ratio of liquid assets to disposable income.

4. OLS and Tobit regression analyses showed that only the estimates in the 'constant' showed a significant change. They were: OLS = 0.18 (0.60) and Tobit = 0.57 (1.48). The T-ratios are reported in parentheses.

TABLE 1: WTP TO AVOID EXPOSURE TO PESTICIDES: ALTERNATIVE SCENARIOS FOR VARIATIONS IN NUMBER OF CENSORED OBSERVATIONS – TOBIT AND OLS COMPARED

No Zeros

	Co-efficient		T-Ratio	
	OLS	Tobit	OLS	Tobit
Constant	2.38	2.38	13.83	14.08
A1	0.072	0.072	1.97	2.01
A2	0.017	0.017	1.08	1.10
A3	0.011	0.011	0.66	0.67
A4	0.178	0.178	2.90	2.97
A5	0.000	0.000	2.19	2.23
A6	0.088	0.088	2.42	2.46

Eight Zeros (4%)

	Co-efficient		T-Ratio	
	OLS	Tobit	OLS	Tobit
Constant	2.91	2.93	5.09	5.01
A1	0.03	0.03	0.28	0.26
A2	0.11	0.11	2.04	2.07
A3	0.03	0.03	0.54	0.56
A4	0.44	0.45	2.16	2.17
A5	0.00	0.00	1.00	1.00
A6	0.03	0.03	0.30	0.28

Sixteen Zeros (8%)

	Co-efficient		T-Ratio	
	OLS	Tobit	OLS	Tobit
Constant	2.28	2.26	2.94	2.74
A1	0.09	0.10	0.58	0.62
A2	0.06	0.07	0.91	0.90
A3	0.09	0.10	1.22	1.25
A4	0.28	0.29	1.01	0.98
A5	0.00	1.00	1.72	1.72
A6	0.10	0.10	0.64	0.61

Twenty Four Zeros (12%)

	Co-efficient		T-Ratio	
	OLS	Tobit	OLS	Tobit
Constant	2.01	1.94	2.20	1.92
A1	0.10	0.10	0.52	0.50
A2	0.06	0.06	0.69	0.68
A3	0.08	0.10	0.95	0.98
A4	0.16	0.16	0.51	0.46
A5	0.00	0.00	1.62	1.61
A6	0.04	0.03	0.21	0.16

Thirty Two Zeros (16%)

	Co-efficient		T-Ratio	
	OLS	Tobit	OLS	Tobit
Constant	2.26	2.20	2.14	1.79
A1	0.10	0.13	0.48	0.52
A2	0.00	0.00	0.02	0.01
A3	0.00	0.00	0.00	0.01
A4	0.03	0.01	0.01	0.02
A5	0.00	0.00	1.00	0.98
A6	0.08	0.08	0.37	0.33

Forty Zeros (20%)

	Co-efficient		T-Ratio	
	OLS	Tobit	OLS	Tobit
Constant	2.54	2.55	2.22	1.81
A1	0.25	0.32	1.03	1.08
A2	0.81	0.00	0.07	0.03
A3	0.32	0.04	0.28	0.31
A4	0.65	0.12	0.16	0.24
A5	0.63	0.00	0.67	0.63
A6	0.08	0.09	0.36	0.31

Forty Eight Zeros (24%)

	Co-efficient		T-Ratio	
	OLS	Tobit	OLS	Tobit
Constant	1.92	1.73	1.57	1.09
A1	0.19	0.26	0.75	0.79
A2	0.04	0.05	0.34	0.39
A3	0.08	0.11	0.66	0.70
A4	0.09	0.07	0.21	0.14
A5	0.00	0.00	0.97	0.92
A6	0.00	0.02	0.02	0.08

Fifty Six Zeros (28%)

	Co-efficient		T-Ratio	
	OLS	Tobit	OLS	Tobit
Constant	1.55	1.18	1.21	0.67
A1	0.05	0.09	0.21	0.26
A2	0.02	0.03	0.23	0.18
A3	0.02	0.03	0.17	0.19
A4	0.22	0.25	0.49	0.41
A5	0.00	0.00	0.56	0.52
A6	0.16	0.20	0.59	0.55

101 (50%) Zeros

	Co-efficient		T-Ratio	
	OLS	Tobit	OLS	Tobit
Constant	2.00	1.89	1.41	0.66
A1	0.25	0.54	0.84	0.90
A2	0.04	0.08	0.32	0.30
A3	0.00	0.00	0.01	0.01
A4	0.06	0.22	0.11	0.23
A5	0.00	0.00	1.72	1.65
A6	0.04	0.14	0.13	0.24

Note: Negative signs are not reported since it has no bearing on the results of this study.

TABLE 2: WTP FOR SEA TURTLE CONSERVATION: ALTERNATIVE SCENARIOS FOR VARIATIONS IN NUMBER OF CENSORED OBSERVATIONS – TOBIT AND OLS COMPARED

	No Zeros				Eight (2.5%) Zeros				Sixteen (5%) Zeros			
	Co-efficient		T-Ratio		Co-efficient		T-Ratio		Co-efficient		T-Ratio	
	OLS	Tobit	OLS	Tobit	OLS	Tobit	OLS	Tobit	OLS	Tobit	OLS	Tobit
Constant	1.25	1.25	4.19	4.23	1.27	1.25	3.69	3.53	1.27	1.26	3.94	3.77
B1	0.03	0.03	0.30	0.30	0.03	0.03	0.30	0.30	0.06	0.06	0.58	0.59
B2	0.16	0.16	2.19	2.21	0.16	0.16	2.47	2.47	0.19	0.19	2.39	2.36
B3	0.12	0.12	0.72	0.73	0.12	0.12	0.85	0.90	0.11	0.11	0.59	0.59
B4	0.06	0.06	0.69	0.70	0.06	0.06	1.07	1.15	0.11	0.12	1.18	1.28
B5	0.06	0.06	0.80	0.81	0.06	0.06	0.22	0.22	0.01	0.00	0.13	0.05

	Thirty Three (10%) Zeros				Forty One (12.5%) Zeros				Forty Nine (15%) Zeros			
	Co-efficient		T-Ratio		Co-efficient		T-Ratio		Co-efficient		T-Ratio	
	OLS	Tobit	OLS	Tobit	OLS	Tobit	OLS	Tobit	OLS	Tobit	OLS	Tobit
Constant	1.27	1.25	3.78	3.39	1.22	1.17	3.61	3.08	1.36	1.35	3.93	3.39
B1	0.16	0.18	1.40	1.41	0.18	0.21	1.59	1.62	0.21	0.25	1.80	1.85
B2	0.21	0.24	2.47	2.35	0.24	0.27	2.95	2.85	0.20	0.21	2.37	2.17
B3	0.26	0.29	1.35	1.39	0.23	0.28	1.22	1.30	0.18	0.22	0.92	0.96
B4	0.14	0.17	1.43	1.56	0.16	0.16	1.34	1.46	0.16	0.19	1.58	1.64
B5	0.02	0.03	0.25	0.35	0.00	0.00	0.03	0.04	0.03	0.02	0.32	0.21

	Fifty Eight (17.5%) Zeros				Eighty Three (25%) Zeros				166 (50%) Zeros			
	Co-efficient		T-Ratio		Co-efficient		T-Ratio		Co-efficient		T-Ratio	
	OLS	Tobit	OLS	Tobit	OLS	Tobit	OLS	Tobit	OLS	Tobit	OLS	Tobit
Constant	1.33	1.31	3.81	3.19	1.32	1.31	3.61	2.77	0.86	0.48	2.35	0.69
B1	0.22	0.26	1.81	1.86	0.26	0.34	2.04	2.12	0.16	0.27	1.25	1.14
B2	0.23	0.25	2.69	2.50	0.25	0.30	2.82	2.59	0.12	0.17	1.36	0.99
B3	0.14	0.16	0.69	0.69	0.09	0.11	0.46	0.41	0.06	0.05	0.33	0.14
B4	0.12	0.15	1.24	1.25	0.09	0.13	0.90	0.95	0.04	0.06	0.46	0.29
B5	0.00	0.02	0.06	0.20	0.02	0.04	0.25	0.37	0.00	0.00	0.02	0.00

Note: Negative signs are not reported since it has no bearing on the results of this study.

PREVIOUS WORKING PAPERS IN THE SERIES

ECONOMIC THEORY, APPLICATIONS AND ISSUES

1. Externalities, Thresholds and the Marketing of New Aquacultural Products: Theory and Examples by Clem Tisdell, January 2001.

2. Concepts of Competition in Theory and Practice by Serge Svizzero and Clem Tisdell, February 2001.

3. Diversity, Globalisation and Market Stability by Laurence Laselle, Serge Svizzero and Clem Tisdell, February 2001.

4. Globalisation, the Environment and Sustainability: EKC, Neo-Malthusian Concerns and the WTO by Clem Tisdell, March 2001.

5. Globalization, Social Welfare, Labor Markets and Fiscal Competition by Clem Tisdell and Serge Svizzero, May 2001.

6. Competition and Evolution in Economics and Ecology Compared by Clem Tisdell, May 2001.

7. The Political Economy of Globalisation: Processes involving the Role of Markets, Institutions and Governance by Clem Tisdell, May 2001.

8. Niches and Economic Competition: Implications for Economic Efficiency, Growth and Diversity by Clem Tisdell and Irmi Seidl, August 2001.

9. Socioeconomic Determinants of the Intra-Family Status of Wives in Rural India: An Extension of Earlier Analysis by Clem Tisdell, Kartik Roy and Gopal Regmi, August 2001.

10. Reconciling Globalisation and Technological Change: Growing Income Inequalities and Remedial Policies by Serge Svizzero and Clem Tisdell, October 2001.

11. Sustainability: Can it be Achieved? Is Economics the Bottom Line? by Clem Tisdell, October 2001.

12. Tourism as a Contributor to the Economic Diversification and Development of Small States: Its Strengths, Weaknesses and Potential for Brunei by Clem Tisdell, March 2002.

13. Unequal Gains of Nations from Globalisation by Clem Tisdell, Serge Svizzero and Laurence Laselle, May 2002.

14. The WTO and Labour Standards: Globalisation with Reference to India by Clem Tisdell, May 2002