Recreation

TP 6470 1982

Ag. Econ file copy

A NOTE ON THE FUNCTIONAL FORM OF TRAVEL COST MODELS WITH ZONES OF UNEQUAL POPULATIONS

Elizabeth J. Strong

May 1982

Abstract

Bowes and Loomis have recently shown that the linear travel cost model contains heteroskedastic errors, and they have presented a correction procedure for the linear model. Since the semilog functional form has been used in other studies of recreation demand, this paper addresses the question of whether the semilog model requires a correction procedure for heteroskedasticity. Statistical tests indicated that heteroskedasticity was virtually nonexistent in the semilog model estimated by OLS, even though serious heteroskedasticity was detected in the linear model. The semilog form was also indicated to be empirically appropriate for this application of the travel cost method.

The author is a Graduate Research Assistent in the Department of Agricultural and Resource Economics, Oregon State University, Corvallis. She wishes to acknowledge the benefits of helpful comments from William G. Brown and an anonymous reviewer. Financial support was provided under Projects 503 and 909 of the Oregon Agricultural Experiment Station. Technical Paper No. 6470.

Presented at AAEA meeting, Logan, Utah, Aug 1-4, 1982.

A NOTE ON THE FUNCTIONAL FORM OF TRAVEL COST MODELS WITH ZONES OF UNEQUAL POPULATIONS

Introduction

As Bowes and Loomis [1980] noted in their contribution, for recreation demand functions of a linear form, the existence of unequal zonal populations in travel cost models leads to the problem of nonconstant error variances when observations are expressed on a per capita basis, i.e., when the zone average method is used. For a model with nonhomogeneous variances, the ordinary least squares (OLS) parameter estimates will be unbiased but inefficient. These properties will hold true of the consumer surplus estimates as well. A weighting of the observations is the well-known solution to the heteroskedasticity problem, where in this case the appropriate weighting factor for each observation is the square root of the associated zone of origin's population. While preserving unbiasedness, the weighted least squares (WLS) approach generates parameter estimates that are equivalent to the generalized least squares (GLS) minimum variance estimators.

In their application of the travel cost method to recreation trips during 1977 down the Westwater Canyon in Utah, Bowes and Loomis illustrated the efficacy of WLS in predicting recreation trips more accurately than OLS for a linear first-stage demand curve. As in previous studies, the OLS linear regression equation grossly mispredicted the total number of trips taken at the actual price per trip of zero, and seemingly overestimated the value of consumer surplus. Further evidence of this latter finding was obtained in a study of warm water fishing in Georgia by Ziemer, Musser, and Hill [1980], who analyzed the appropriateness of three widely-used functional forms for demand equations which were fitted to the 1971 data on fishing patterns. It was demonstrated that the choice of functional form can have a significant effect on the estimated value of consumer surplus. To exemplify, the level of consumer surplus estimated from a linear demand model was nearly 300 percent higher than the value derived from a quadratic form of the model (with a squared travel cost term included) and 200 percent higher than that from a semilog form (involving the natural logarithm of the dependent variable). These findings suggest that the choice of an appropriate functional form for recreation demand equations is an important specification problem to be considered along with the selection of an appropriate set of regressors.

Because Ziemer, Musser, and Hill applied the individual observation method (rather than the zone average method) for specifying the travel cost model, the problem of nonconstant error variances would not be expected to exist in their model. On the other hand, Bowes and Loomis used the zone average approach and removed the heteroskedasticity from the linear demand model by application of the WLS estimation procedure. The purpose of this paper is to illustrate that a transformation on the dependent variable is an alternative to the WLS procedure for removing heteroskedasticity from the linear model of per capita demand. By means of statistical inference, the empirical appropriateness of this curvilinear form of the model will be demonstrated with an application of the travel cost method to steelhead fishing in Oregon.

Discriminating Among Model Specifications

More often than not, in applied econometrics the selection of an appropriate functional form is based on results of statistical tests. This approach is generally reasonable because economic theory provides little guidance on the matter of functional form [Zarembka, 1974]. For example, methods of statistical inference are considered as being necessary in situations where more than one functional form is consistent with a given theoretical model [Gaver and Geisel, 1974]. Several approaches to using statistical inference as a means of model discrimination have been proposed in the literature, but no single method proves to be the "best."

In the simplest case, where competing models differ with respect to the definitions of one or more independent variables, Rao and Miller [1971] contend that the empirically appropriate model is obtained by a comparison of their sums of squared residuals. For the more general case of alternative model specifications with the same dependent variable, conventional methods for choosing among them include first, an examination of the "plausibility" of the regression coefficients and second, selection of the model specification with the lowest residual variance, or, equivalently, the highest multiple correlation coefficient adjusted for changes in the degrees of freedom. Theil [1957] justifies the "maximize R^2" rule by claiming that it will lead, "on the average," to selection of the correct model. Pesaran [1974] argues, however, that conditions under which the adjusted multiple correlation coefficient can adequately serve as a choice criterion are unlikely to be fulfilled in practice. One of these conditions is the existence of a so-called correct model. Pesaran proposes an adaptation of

Cox's [1962] model discrimination method for testing non-nested competing regression models, such as H_o and H_1, where

$$H_o: \; y = Xb_o + u_o; \; u_o \sim N(0,\sigma_o^2 I), \text{ and}$$

$$H_1: \; y = Zb_1 + u_1; \; u_1 \sim N(0,\sigma_1^2 I) \; .$$

The symbols u and σ^2 denote the stochastic error term and its variance, respectively. Cox's procedure for testing non-nested hypotheses is based on a likelihood ratio statistic.

The meaningfulness of these informal decision rules and hypothesis-testing procedures breaks down in cases where a transformation is performed on the dependent variable. For such cases, a procedure was provided by Box and Cox [1964] for estimating a transformation of the dependent variable in the form

$$y^{(\lambda)} = \begin{cases} (y^{\lambda} - 1)/\lambda & \lambda \neq 0 \\ \ln(y) & \lambda = 0. \end{cases}$$

This procedure has been extended to include transformations of the independent variables as well. It requires a search over a reasonable portion of the parameter space (such as $-2 \leq \lambda \leq 2$) for the power transformation which minimizes the calculated error variance by maximizing the value of a log likelihood function.

In the study discussed earlier, Ziemer, Musser, and Hill applied the Box and Cox procedure as a means of distinguishing between the linear and semilog specifications of the travel cost model. The results of this statistical test implied that the semilog model was appropriate for the data while the linear model was inappropriate. The quadratic form was rejected on the basis of a statistically insignificant coefficient to the squared travel cost term.

In an earlier study of wilderness recreation during 1972 in the Desolation Area of California, Smith [1975] evaluated the suitability of three "popular" specifications of the travel cost model -- linear, semilog and log-linear -- for describing the behavior of wilderness recreationists. On the basis of conventional criteria, the linear model of per capita demand was rejected. Although the semilog and log-linear models were judged as being completely acceptable (and the semilog preferable) by conventional standards, results of Pesaran's adapted form of the aforementioned Cox-likelihood ratio test implied that neither model provided a reasonable representation of the recreationists' behavior patterns.

While neither semilog nor log-linear forms of the travel cost model may be suitable for estimating the demand for wilderness services, the former, as Ziemer, Musser, and Hill [1980] illustrated, may be acceptable for estimating the demand for sport fishing. Perhaps these discrepant findings can be explained by a difference in the character of these two types of demand models in addition to their specification differences. Smith [1975] notes that the demand equation for wilderness recreation encompasses all types of wilderness activities. As a result, instead of directly reflecting the demands for specific activities (as the angler demand equation does for sport fishing), it reflects the demand for all of the site's services.

One similarity in the findings of these two separate studies is the implied empirical inappropriateness of linear travel cost models for both wilderness recreation and sport fishing. Repeated rejection of linear demand models raises the question of whether the semilog form of the per capita demand model might not be more appropriate than the linear form. If the semilog form is more appropriate, as previous studies indicate, then the question arises as to whether an application of the WLS estimation procedure is needed for this specification of the demand function.

In this paper it will be shown that the semilog form does not appear to require the use of WLS. That is, a natural logarithmic transformation of the dependent variable seems to move the error variance toward homogeneity. According to Tukey [1957], the conventional purposes of performing transformations on a random variable y are to remove either one or a combination of these undesirable properties: nonadditivity of effect; nonconstancy of the error variance; and asymmetry or nonnormality of the error distribution. In the following section the usefulness of this type of transformation procedure as an alternative to the WLS estimation procedure will be demonstrated with an empirical example.

Application to Steelhead Fishing in Oregon

As indicated above, findings of previous research suggest that the semilog form should be tested as an alternative to the linear specification of travel cost models. Accordingly, in this section the regression results and consumer surplus estimates will be compared for two demand models of the forms

$$q_{ij} = \beta_o + \beta_1 TC_{ij} + \beta_2 I_{ij} + \beta_3 I^2_{ij} + \beta_4 METRO_i + e_{ij}$$

$$\ln q_{ij} = \alpha_o + \alpha_1 TC_{ij} + \alpha_2 I_{ij} + \alpha_3 I^2_{ij} + \alpha_4 METRO_i + \varepsilon_{ij}$$

where:

q_{ij} = annual number of fishing trips per capita from zone i (or subzone i) to river j.

TC_{ij} = average round trip travel costs from zone i to river j.

I_{ij} = average income of respondents from zone i who fished at river j.

$$METRO_i = \begin{cases} 1 & \text{if zone i is a metropolitan area}^{1/} \\ 0 & \text{otherwise.} \end{cases}$$

e_{ij}, ε_{ij} = stochastic disturbance term.

The major zones correspond to single counties or groups of adjacent counties in Oregon. There is a different set of zones for each of the 21 rivers included in the study, but in some instances a particular zone is used for more than one river. The formation of zones in this manner provides a framework for estimating the demand for and value of steelhead fishing on a per river basis. Hence, the character of this model is different than the previously discussed models in that it reflects the demand for a specific recreation service, i.e., steelhead fishing, at each of the 21 sites.

The data on fishing patterns were obtained from four groups of mail questionnaires which were each sent to a different sample of fishing license holders at the end of each three-month period in 1977. (Details of the survey are presented in Sorhus, Brown, and Gibbs, [1981]). Because the number of responses varied among the counties, some of the zones were subdivided into income subzones.

In an earlier study of angler demand in Oregon during 1962, the effects of income on participation rates proved to be curvilinear [Brown, 1976]. That is, participation rates tended to increase with income up to a certain level, and then proceeded to decline as income continued to rise. Accordingly, an income squared term is included in this model. Inclusion of the METRO dummy variable as an intercept shifter is based on the assumption that metropolitan residents exhibit different patterns of recreation behavior than residents of non-metropolitan areas, primarily because a greater variety of alternative recreation opportunities are expected to be available in metropolitan areas. Another factor which may tend to cause higher participation rates for non-metropolitan residents is their closer proximity to uncongested fishing sites.

Before the regression results are presented, the Goldfeld-Quandt test for detecting the existence and magnitude of heteroskedasticity will be performed [Koutsoyiannis, 1977]. This test is applicable if there is *a priori* knowledge of the variable causing the error variance to change. According to Bowes and Loomis [1980], the residual variance for the linear model of per capita cemand is

$$V(e_{ij}) = \sigma^2/N_{ij}$$

where σ^2 denotes the constant variance, and where N_{ij} is the population of zone i for river j. Accordingly, it is hypothesized here that the error variance is inversely related to zonal population size, and that the relevant hypotheses to be tested with the Goldfeld-Quandt test are

$$H_o: V(e_{ij}) = \sigma^2, \text{ and}$$

$$H_1: V(e_{ij}) = \sigma^2/N_{ij}.$$

Preparation for the Goldfeld-Quandt test involves first, an ordering of the sample observations by the magnitude of zonal population size and second, an omission of approximately one-fifth of the middle observations to avoid any effects of overlapping between the two groups of large and small population sizes. The next two steps are to run a separate regression on each group of observations and to calculate the test statistic

$$F^* = SSE_1/SSE_2$$

where SSE_1 and SSE_2 denote the sum of squared residuals in estimating the equations for the observations with smaller and larger population sizes, respectively. The calculated F* is then compared to the critical value of the F distribution with $[(N - d)/2 - k]$ degrees of freedom for both the numerator and denominator, where N is the total number of observations, d is the number of omitted observations, and k is the number of parameters in the model. Both degrees of freedom will not be equivalent if the number of observations in each group are unequal. If F* exceeds the chosen critical value, then the null hypothesis of homoskedasticity can be rejected at the associated level of significance.

In this example, the calculated test statistic for the linear equation is equal to 42.54, which is larger than the 1 percent critical F value with with 29 degrees of freedom. Consequently, the null hypothesis can be rejected at the 1 percent significance level. This result strongly indicates that the linear model of steelhead demand contains heteroskedastic errors. By application of the WLS procedure recommended by Bowes and Loomis, a new error term is defined as

$$\tilde{e}_{ij} = e_{ij} \cdot \sqrt{N_{ij}}$$

of which the variance is

$$V(\tilde{e}_{ij}) = (\sigma^2/N_{ij}) \cdot (\sqrt{N_{ij}})^2 = \sigma^2 .$$

The calculated F* statistic for the WLS regression equation drops to 1.827. The null hypothesis cannot be rejected at the 1 percent level of significance, but it is almost significant at the 5 percent level. On the other hand, the value of F* for the semilog model as defined above is only 1.036; hence, the null hypothesis cannot be rejected at even the 10 percent level. This interesting result implies that it is not necessary to apply the WLS procedure to the semilog model.

On first inspection, the regression results in Table 1 suggest that the semilog model is preferable to the linear equation of both the OLS and WLS regression types. Firstly, in the semilog regression equation the estimated coefficients are all highly significant with the exception of the constant term which was dropped for lack of statistical significance. Secondly, the Durbin-Watson d-statistic indicates non-autocorrelated disturbance terms in the semilog regression, whereas autocorrelation is indicated in both of the linear regressions. The lower and upper Durbin-Watson test bounds at the one percent level are 1.39 and 1.60, respectively.

Because the linear and semilog functional forms are not directly comparable, a statistical test will be employed to distinguish between the two model specifications. The test is based on a transformation of the dependent variable q that allows a direct comparison of the residual sums of squares in estimating both equations of the forms

$$q^*_{ij} = \beta^*_o + \beta^*_1 TC_{ij} + \beta^*_2 I_{ij} + \beta^*_3 I^2_{ij} + \beta^*_4 METRO_i + e^*_{ij}$$

$$\ln q^*_{ij} = \alpha^*_o + \alpha^*_1 TC_{ij} + \alpha^*_2 I_{ij} + \alpha^*_3 I^2_{ij} + \alpha^*_4 METRO_i + \varepsilon^*_{ij}$$

were the transformed variable q^*_{ij} is

$$q^*_{ij} = c \cdot q_{ij}$$

and c is the inverse of the geometric mean of q. To perform this test the linear and semilog equations are estimated with the transformed dependent variables. The resulting sums of squares are then used to compute a test statistic to determine whether or not they are significantly different than each other. The likelihood ratio test statistic is defined as

$$\ell = T/2 \left| \ln(SSE^*_1/SSE^*_2) \right|$$

where T is the number of observations, and SSE_1^* and SSE_2^* represent the residual sums of squares in estimating the transformed linear and semilog equations, respectively.[2/] If the ℓ statistic, which follows a Chi-squared distribution with one degree of freedom, is greater than the chosen critical value, then the null hypothesis of empirical equivalence of the two model specifications can be rejected. Although it may seem to be an *ad hoc* type of procedure, this test statistic actually selects the model with the larger likelihood value. Accordingly, this criterion for choosing one functional form over the other is similar to that of conventional methods for selecting the functional form that maximizes the likelihood value over a reasonable portion of the parameter space [Rao and Miller, 1971]. In this example, the ℓ statistic is equal to 68.60, which exceeds the critical value of 6.635 at the 1 percent significance level. Consequently, the null hypothesis of empirical equivalence can be rejected, and the semilog functional form is chosen because it has the lower (transformed) residual sum of squares.

Another appealing quality of the estimated semilog equation is its advantage of not estimating any negative values of the dependent variable. Although the WLS regression equation most accurately estimated the total number of trips when the negative predictions were included in the calculation, this virtue of WLS was negated when the negative predictions were set equal to zero (see Table 2). In addition to providing a more accurate estimate of total trips, the semilog model also resulted in a slightly lower mean squared error (MSE) for trips than did either of the linear models. For purposes of comparison, the predicted number of trips and the estimated value of consumer surplus which were derived from the linear equation of the OLS regression type are included in Table 2. There are dramatic differences between these results and the estimated levels of demand and consumer surplus derived from the WLS and semilog regressions.

Conclusions

The semilog specification of the travel cost model was tested as an alternative to the linear specification for a model with unequal zonal populations. By means of statistical inference, it was demonstrated that these two functional forms are not empirically equivalent. The statistical test which was employed involved a transformation of the dependent variable in a way that allows a direct comparison of the residual sums of squares in estimating the transformed models. On the basis of the test results, the semilog form was chosen over the linear form as the empirically appropriate specification of the travel cost model. This finding agrees with the results of the Ziemer, Musser, and Hill [1980] study in which individual observations were used to specify an angler demand equation.

In this case, however, the zone average method was used, and hence the question arose as to whether a correction for heteroskedasticity would be needed for the apparently appropriate semilog specification. Previous research has shown that the existence of unequal populations across zones in linear per capita demand models can lead to the problem of heteroskedasticity. While the linear model did contain a significantly heterogeneous error variance, heteroskedasticity was not detected in the semilog model. Consequently, the WLS procedure was applied to the linear model to remove the heteroskedasticity, and the resulting WLS model was then compared to the semilog model on the basis of their mean squared errors in estimating trips. The results suggest that the semilog form of the travel cost model generates somewhat better estimates of total trips. Another advantage of the semilog specification is that the estimated quantity of trips demanded at the actual price per trip is always positive for all zones with participation rates greater than zero.

Table 1. Regression Results for Alternative Functional Forms

Independent Variables	Regression Type		
	OLS Linear	WLS Linear	OLS Semilog
Travel Cost	-.00278 c/	-.00103 c/	-.04358 c/
	(.00097)	(.00038)	(.00846)
Income	-.000014 b/	-.000002	-.00022 c/
	(.000007)	(.000004)	(.00002)
$Income^2$	.324E-09 b/	.537E-10	.481E-08 c/
	(.175E-09)	(.915E-10)	(.758E-09)
METRO	-.09829 c/	-.04919 c/	-1.06349 c/
	(.02753)	(.01345)	(.24292)
Intercept	.31258 c/	.10745 c/	
	(.06749)	(.04014)	
Statistics			
R^2	.3241	.1939	.5317
F(4,76)	9.11 c/		
d	1.1955	1.2139	1.7463
F*(29,29)	42.54 c/	1.827 a/	1.036

Number of observations = 81.

Standard errors are in parentheses below the estimated coefficients.

a/ Denotes significance at the 10 percent level.

b/ Denotes significance at the 5 percent level.

c/ Denotes significance at the 1 percent level.

Table 2. Estimates of Total Fishing Trips and Consumer Surplus for 1977

Regression Type	Total Trips a/	Total Trips b/	Consumer Surplus b/ ($000's)	MSE c/ (000's)
Semilog OLS	521,450	521,450	$11,965	46,775
Linear WLS	506,298	525,879	13,430	50,545
Linear OLS	664,449	822,627	16,800	253,100

Actual number of trips is 506,287.

a/ Predicted negative trips included.

b/ Predicted negative trips excluded.

c/ MSE = $\sum_{i=1}^{81} (\hat{y}_{ij} - y_{ij})^2/81$, where y_{ij} denotes the total number of trips from the ith zone to the jth river.

FOOTNOTES

1/ A zone is metropolitan if it contains at least one of the three metropolitan counties in northwest Oregon. Together, these three adjacent counties account for 40 percent of the total 1977 population of the 36 Oregon counties.

2/ $\alpha_o^* = 0$ in computing SSE_2^*.

REFERENCES

Bowes, Michael D. and John B. Loomis. 1980. "A Note on the Use of Travel Cost Models with Unequal Zonal Populations." *Land Economics* 56 (Nov.): 465-470.

Box, George E.P. and D.R. Cox. 1964. "An Analysis of Transformations." *Journal of the Royal Statistical Society Series* B 26. 211-243.

Brown, William G. 1976. "Economic Implications of Allocation." *Marine Recreational Fisheries*. Sport Fishing Institute: 29-36.

Cox, D.R. 1962. "Further Results on Tests of Separate Families of Hypotheses." *Journal of the Royal Statistical Society Series* B 24. 406-424.

Gaver, Kenneth M. and Martin S. Geisel. 1974. "Discriminating Among Alternative Models: Bayesian and Non-Bayesian Methods." In *Frontiers in Econometrics*. Paul Zarembka ed. New York: Academic Press. pp. 49-77.

Koutsoyiannis, A. 1977. *Theory of Econometrics*. 2nd ed. The MacMillan Press, Ltd.

Pesaran, M.H. 1974. "On the General Problem of Model Selection." *Review of Economic Studies* 41 (April): 153-171.

Rao, Potluri and Roger LeRoy Miller. 1971. *Applied Econometrics*. Belmont: Wadsworth Publishing Co., Inc.

Smith, V. Kerry. 1975. "Travel Cost Demand Models for Wilderness Recreation: A Problem of Non-Nested Hypotheses." *Land Economics* (May): 103-111.

Sorhus, Colin N., William G. Brown and Kenneth C. Gibbs. 1981. *Estimated Expenditures by Salmon and Steelhead Sport Anglers for Specified Fisheries in the Pacific Northwest.* Agricultural Experiment Station. Oregon State University, Corvallis. Special Report 631 (July).

Theil, Henri. 1957. "Specification Errors and the Estimation of Economic Relationships." *International Statistical Institute Review* 25: 41-51.

Tukey, John W. 1957. "On the Comparative Anatomy of Transformations." *Annals of Mathematical Statistics* 28: 602-632.

Zarembka, Paul. 1974. "Transformation of Variables in Econometrics." In *Frontiers in Econometrics.* Paul Zarembka ed. New York: Academic Press. pp. 81-104.

Ziemer, Rod F., Wesley N. Musser, and R. Carter Hill. 1980. "Recreation Demand Equations: Functional Form and Consumer Surplus." *American Journal of Agricultural Economics* (Feb.): 136-141.