University of California, Davis
Department of Agricultural Economics

Performance of Optimal MSE Estimators Relative to Alternative Predictors of Recreational Use for National Parks

by

Richard Green and Warren Johnston

Contributed Paper Session at the American Agricultural Economics Association Meetings in Columbus, Ohio, August 10-13, 1975.

Session Number ______________ Session Title ______________________

Performance of Optimal MSE Estimators Relative to Alternative Predictors of Recreational Use for National Parks

Richard Green and Warren Johnston, University of California, Davis

The common use of multiplicative models of recreational use (demand) results in difficulties in projecting future values. Alternative estimators are discussed, and theoretically optimal MSE-estimators are compared with OLS and maximum-likelihood estimators. Empirical results are based on use data for U.S. National Parks from Boyet and Tolley.

Performance of Optimal MSE Estimators Relative to Alternative Predictors of Recreational Use for National Parks

by

Richard Green and Warren Johnston

There has been considerable research activity in the general study area of the economics of outdoor recreation in the decade and a half since Marion Clawson empiricalized Hotelling's seminal idea of the "travel-cost" method for deriving economic demand curves for site-specific recreational activity. We classify much of the activity as being either demand or use oriented, and suggest that many researchers have mistakenly thought the two to be synomous. This error of reasoning leads to considerable difficulty in that efforts focused on demand, in particular, are weak explanators of future use.

Since demand and use may be particularly difficult for non-economists to differentiate we alternatively classify activity into 1) prediction and 2) projection efforts. Prediction involves modelling recreational use to better understand the determinants of demand having the objective either of identifying current or past structure of demand or of developing measures of value for recreational resources. Projection, on the other hand, is confined to forecasting future use--a much more hazardous endeavor since demand structures change over time and supply is not held constant.

Within the literature which applies the "travel-cost" method, there have been numerous concerns over the theoretical bases of such studies and over issues relating more narrowly to variable and model specification. Despite the "felt" progress in this important area of inquiry, baneful theoretical and econometric problems still exist plaguing the progress of applied research efforts. We examine one such consideration, namely, the use of multiplicative models of recreational use (demand).

A wide variety of models are used in applied studies -- linear, logarithmic, transcendental, and mixed. Multiplicative (Cobb-Douglas-like) models, of the form:

$$(1) \quad Y_i = \alpha \prod_{k=2}^{K} X_{ik}^{\beta_k} V_i,$$

which by logarithmic transformation are linearized to facilitate estimation:

$$(2) \quad \ln Y_i = \ln\alpha + \beta_1 \ln X_{i1} + \beta_2 \ln X_{i2} + \ldots + \beta_k \ln X_{ik} + \ln V_i,$$

often are specified for empirical study. Typically, Y_i is the quantity of a specific recreational activity, and the X_i include travel cost as a proxy for price, income, and other explanatory variables thought to affect the quantity of the activity consumed or used by recreationists.

If the objective of the research is to estimate structure, then the above multiplicative specification gives no problems, having the appeal of yielding straight-forward estimates of elasticities. However, if, as often is the case, the objective is to project future use, or to predict the consumer surplus accruing from the activity, OLS estimators of equation (2) may yield misleading, less-than-optimal (in a MSE sense) projections and values.

Although a few applied researchers including Boyet and Tolley were aware of the basic problem and attempted rather ad hoc adjustments to "correct" for it, this problem persists and appears to have been largely ignored even though it may importantly affect empirical estimates of recreational use or value.

The importance of alternative estimators to use and value estimates is obvious. Projections of future use levels by incorporating expected values of independent variables will be affected, as will predictions of value which use mean, or other assigned, values of independent variables. Thus, the problem is of significance to researchers and the clienteles of applied research results -- recreation planners and other decision-makers.

The following sections present OLS estimators and contrast them with estimators presented by Teekens and Koerts which have the desirable theoretical property of being minimum MSE estimators of the conditional mean of the dependent variable in multiplicative models. The theoretical comparison is followed by some empirical results and conclusions are presented for future research in recreational use (demand) studies using multiplicative models.

Alternative Estimators

Several estimation procedures have been developed for multiplicative models of the Cobb-Douglas type including the traditional approach of first estimating the transformed logarithmic equation by least squares and then making a backward transformation to the original specification. Since this method gives rise to many inconsistencies, other methods have been proposed to circumvent these problems. Among them are parametric methods including maximum likelihood (ML) estimates and minimum mean squared error (MSE) predictors (Teekens and Koerts), nonparametric methods (Aigner), and nonlinear estimation techniques which allow for both multiplicative and additive disturbance terms (Goldfeld and Quandt; Kelejian). In this paper consideration will only be given to Teekens and Koerts minimum MSE estimators and their relationship with LS and ML estimators.

A brief theoretical discussion of the various estimation procedures begins by considering the model given in equations (1) and (2) where the following stochastic specification is assumed:

$E(V_i) = 1 \quad , i = 1,\ldots,N;$

$Var(V_i) = w^2;$

$V_i > 0 \quad , i = 1,\ldots,N;$ and

$\ln V_i$ is normally distributed which implies that V_i is lognormally distributed.

(For further problems and implications of this tochastic formulation, see Goldberger, and Teekens and Koerts.)

The model in equation (1) using the property that $x = \exp[\ln x]$, can be rewritten as

$$(3) \quad Y_i = \exp[\ln\alpha + \Sigma\, \beta_k \ln X_{ik} + \ln V_i]$$
$$= \exp[Z_i'\beta + \ln V_i]$$

where $Z_i = (1, \ln X_{i1},\ldots, \ln X_{ik})$, a row vector of the transformed observation matrix of the explanatory variables or a vector whose elements are the projections of the explanatory variables, and $\beta' = (\ln\alpha, \beta_i,\ldots,\beta_k)$. Thus, the conditional mean of Y_i given the vector of explanatory variables can be expressed as $E(Y|Z_i) = \exp\{Z_i'\beta\}$, since $E\{\exp[\ln V_i]\} = EV_i = 1$ by assumption. With this notation the alternative estimators of the conditional mean of the dependent variables can be expressed as[1/]

$$(4) \quad \hat{Y}_p^{LS} = \exp\{Z_p'\hat{\beta}\} \text{; least-squares estimator}$$

$$(5) \quad \hat{Y}_p^{ML} = \exp\{Z_p'\hat{\beta} + \frac{\sigma 2}{2}\} \text{, maximum likelihood estimator}$$

$$(6) \quad \hat{Y}_p^{TK} = \exp\{Z_p'\hat{\beta} + \frac{\sigma 2}{2}(1-3\alpha_p)\} \text{, Teekens and Koerts estimator}$$

where $\hat{\beta}$ = least-squares estimates of the coefficients of the log transformed model, and $\alpha_p = Z'_p (Z'Z) Z_p$ where Z = observation matrix of the transformed model (equation 2).

The above forms assumed that σ^2 is known. But, generally σ^2 is unknown and has to be estimated. Thus, the above expressions need to be modified for this case. The formulas become more complex since estimators of $\exp\{k\sigma^2\}$ are rather complicated from a computational viewpoint. For the situation where σ^2 is unknown, the estimators become:

(7) $\hat{Y}_p^{LS} = \exp\{Z'_p\hat{\beta}\}$, least-squares estimator

(8) $\hat{Y}_p^{ML} = \exp\{Z'_p\hat{\beta} + \frac{N-K}{2N} S^2\}$, maximum likelihood estimator

(9) $\hat{Y}_p^{TK} = \exp\{Z'_p\hat{\beta} + \xi S^2\}$, $\alpha_p \geq \frac{1}{3}$ Teekens and Koerts estimator

$= \exp\{Z'_p\hat{\beta}\}\, g_{N-K}(\frac{N-K+1}{N-K} \xi S^2)$, $\frac{1}{N} \leq \alpha_p < \frac{1}{3}$

where $S^2 = \frac{\hat{e}'\hat{e}}{(N-K)}$ = estimation of the (logarithmic) disturbance variance, σ^2;

$\xi = \frac{1}{2}(1-3\alpha_p)$; and

$g_{N-K}(\frac{N-K+1}{N-K} \xi S^2)$ = Finney's minimum variance unbiased estimator of $\exp\{\xi\sigma^2\}$.

It should be pointed out that when σ^2 is unknown, only "approximate" minimum-MSE predictors can be obtained. However, Monte Carlo experiments performed by Teekens and Koerts and later by Aigner have shown that the TK estimator still dominates the LS and MLE, when the MSE loss function is employed to evaluate alternative estimators.

Some Empirical Comparisons

In order to examine the effects of alternative estimators we have used existent data on recreational use. Elsewhere we report results based on cross-sectional data for 4 California lakes and reservoirs (see Green and Johnston). In this paper we use from 4 to 6 years of cross-sectional data for 12 national parks from Boyet's dissertation. Though dated, these data should provide a stronger application (more valid results) and serve to reveal the effects of our alternative estimators.

The Boyet dissertation, in part, attempts to explain use at selected national parks using a simplified model of the following form:

$$(10) \quad Y_{ij} = \alpha X_{1ij}^{\beta_1} X_{2ij}^{\beta_2} X_{3ij}^{\beta_3} V_{ij}$$

where Y_{ij} = visits to a specific national park from state i in year j

X_{1ij} = distance from state i to the national park which remains constant over time

X_{2ij} = population in state i in year j

X_{3ij} = per capita income in state i in year j.

The results are then used to project aggregate 1980 use by park. Empirical results are also reported in Boyet and Tolley.

We admit our difficulty in trying to reproduce exactly the empirical results reported and therefore will not dwell on a presentation of the estimates of α, β's, and R^2's, etc., since our primary intent here is to compare relative measures of projected use for 1980. One understands the many hazards of trying to project 1980 national park use from models based on data for selected years in the period 1947-1959. Among them are the implicit assumption of invariant elasticities of total use with respect to the three independent variables, drastic changes in supply (e.g., mainland national parks increased from 27 to 35 in number, and other recreational areas also expanded sharply--state parks, recreational water areas, etc.), and changing technology available to consumers (users), including recreational vehicles, the interstate highway system, and better quality highways, in general. These projection assumptions are in addition to the simplified regression model's underlying assumption that the sole purpose of any national park visit was to visit that specific park!

We use the regression results from (10) and projected 1980 levels of population and per capita income for each state to obtain projected use by state for each national park. Aggregate use is then gained by summing projected use over the 48 conterminous states and the District of Columbia.

Results given in Table 1 are expressed in terms of the ratio of Projected Visits in 1980 by the LS and the ML estimators to the TK estimator. Least-squares (LS) estimates of total use are less than the TK estimate by relative magnitudes from 3 percent (Grand Canyon) to 28 percent (Crater Lake) and generally lie about 10 to 15 percent below the TK estimate of total visits for the parks in this selected sample. While the maximum-likelihood (ML) estimate exceeds the TK estimate it is generally closer to it. It exceeds the TK estimate in relative magnitudes ranging from only 2 percent (Bryce Canyon, Grand Canyon, Hot Springs, and Shenandoah) to 15 percent (Crater Lake).

Thus, if one attempts projection of future visits from multiplicative models, very different projections arise by using different estimators. The same caution is also noteworthy in prediction efforts focused on measures of consumer surplus by computing the area under the demand curve. The extreme sensitivity translated into differing measures of consumer surplus should cause concern for empirical researchers who use the multiplicative specification. We suggest that the TK estimator, which is the optimum MSE estimator of the conditional mean, is superior to the LS estimator which, in general, applies to the conditional median (see, Goldberger).

TABLE 1

Relative Projections of 1980 Total Visitation to Selected National Parks,[a/] $\hat{Y}_{LS}/\hat{Y}_{TK}$ and $\hat{Y}_{ML}/\hat{Y}_{TK}$

National Park	Ratio of 1980 Projected Total Visits	
	$\hat{Y}_{LS}/\hat{Y}_{TK}$	$\hat{Y}_{ML}/\hat{Y}_{TK}$
Bryce Canyon	.93	1.02
Crater Lake	.72	1.15
Glacier	.89	1.04
Grand Canyon	.97	1.02
Hot Springs	.81	1.02
Mammoth Cave	.86	1.11
Mesa Verde	.90	1.03
Mount Rainier	.88	1.07
Rocky Mountain	.85	1.12
Sequoia-Kings	.86	1.04
Shenandoah	.92	1.02
Zion	.94	1.03

a/ Based on data contained in Boyet's dissertation.

There is, of course, the additional caution to be appropriately expressed here. It is that multiplicative models are functionally best-suited for, at best, short-run projections, being quite sensitive for projecting over a long period in the future on the basis of hazards expressed previously.

Research in progress by the authors may yield some promising results to overcome some of the problems associated with the common use of multiplicative models in studies of recreational use (demand). Briefly, we hope to combine time series-cross-sectional data and identify systematic changes over time in the structure by applying the Prescott and Cooley varying parameter model. The use of a random coefficient regression model may also hold some promise. In addition, we are using Monte Carlo techniques to better evaluate the alternative predictors discussed in this paper -- least-squares, Teekens and Koerts, and maximun-likelihood--plus the estimator developed by Bradu and Mundlak. Helpful to this effort would be the identification of data bases which might further facilitate this area of inquiry--we invite any suggested sources!

mL
6/11/75

Footnotes

* The authors acknowledge the efforts of Mr. Tim Lowry for programming and computational assistance.

1/ Since the derivations of these estimators are not obvious a brief intuitive explanation is provided. For a rigorous development see Teekens and Koerts. First, the LS estimator is obtained by just replacing β by its LS estimate, $\hat{\beta}$. The ML predictor is obtained by finding the estimator of β that maximizes the lognormal likelihood or equivalently finding β that maximizes the normal likelihood of the transformed model. Thanks are due to Professor Leon Wegge for pointing out this result. It can be shown that on the average the LS estimator underestimates the true conditional mean and the ML estimator overestimates it. The TK estimator is obtained by finding the estimator that minimizes the MSE of an expression that is a weighted average of the LS and ML estimator. An alternative to the TK estimator is the Bradu-Mundlak (BM) estimator which is conceptually very similar to the TK estimator. The BM estimator is

$$\hat{Y}_p^{BM} = \exp\{Z_p'\hat{\beta}\}\ g_{N-K}\ [\tfrac{1}{2}(1-\alpha_p)\frac{N-K+1}{N-K}\ S^2].$$

References

[1] Aigner, Dennis, "Asymptotim Minimum-MSE Prediction in the Cobb-Douglas Model With a Multiplicative Disturbance Term," Econometrica, July 1974, pp. 737-748.

[2] Boyet, Wayne E., "Area Complex and National Park Recreation Demand Projection," unpublished Ph.D. dissertation, Department of Agricultural Economics, North Carolina State University, Raleigh, 1964, 109 pp.

[3] Boyet, Wayne E. and George S. Tolley, "Recreation Projection Based on Demand Analysis," JFE, Vol. 48, No. 4, Part I, November 1966, pp. 984-1001.

[4] Bradu, D. and Y. Mundlak, "Estimation in Lognormal Linear Models," Journal of the American Statistical Association, Vol. 65, 1970, pp. 198-211.

[5] Clawson, Marion, "Methods of Measuring the Demand for and Value of Outdoor Recreation," Resources for the Future Reprint 10, Washington, D.C., 1959.

[6] Goldberger, Arthur S., "The Interpretation and Estimation of Cobb-Douglas Functions," Econometrica, Vol. 33, No. 4, October 1964, pp. 863-868.

[7] Goldfeld, S. and R. E. Quandt, "The Estimation of Cobb-Douglas Type Functions With Multiplicative and Additive Errors," International Economic Review, June 1970, pp. 251-257.

[8] Green, Richard and Warren Johnston, "Alternative Estimation Procedures for Multiplicative Models: Implications for Projecting Recreational Use," paper presented to Western Agricultural Economics Association, Reno, July 1975.

References (Continued)

[9] Kelejian, H. H., "The Estimation of Cobb-Douglas Type Functions With Multiplicative and Additive Errors: A Further Analysis," International Economic Review, February 1972, pp. 179-182.

[10] Meulenberg, M. T. G., "On the Estimation of An Exponential Function," Econometrica, Vol. 33, No. 4, October 1964, pp. 863-868.

[11] Prescott, T. F., and E. C. Cooley, "Systematic (Non-Random) Variation Models Varying Parameter Regression: A Theory and Some Applications," Annals of Economic and Social Measurement, Vol. 2, No. 4, 1973, pp. 463-473.

[12] Teekens, R. and J. Koerts, "Some Statistical Implications of the Log Transformation of Multiplicative Models," Econometrica, Vol. 40, No. 5, September 1972, pp. 793-819.