AN ABSTRACT OF THE THESIS OF

Kerri L. Weidner for the degree of Master of Science in
Agricultural and Resource Economics presented on
October 10, 1995. Title: A METHOD TO IMPROVE CONSUMER
SURPLUS ESTIMATES FROM TRUNCATED RECREATION DEMAND DATA.

Abstract approved: ___Redacted for Privacy___

William G. Brown

A method is presented in this study for improving
empirical travel cost analyses when truncated data are
available. The analysis expands upon a Monte Carlo study
conducted by Riyadi (1993) which compared the coefficient
and consumer surplus estimates for five travel cost
estimators using both truncated and censored data. Riyadi
showed that if travel costs within origin zones are equal
and the sampling rate is the same for all zones, the
estimated coefficients for the Bowes-Loomis zonal average
travel cost (ZATC) method are identical to estimates from
ordinary least squares applied to the censored sample.

A method related to Riyadi's equality was developed in
this study to expand a truncated data set to include non-
user observations and permit estimation by a censored
estimator, such as the standard negative binomial. In the
Monte Carlo study presented here, the data expansion method
provided substantial efficiency gains for predicted consumer
surplus estimates relative to estimates obtained from

truncated data. The average percent deviation of predicted consumer (CS) from the standard negative binomial fitted to the expanded data were only 31 to 80 percent of the average deviation of the truncated negative binomial CS estimates.

The method was also applied to empirical data. The standard negative binomial fitted to expanded data appeared to give much better empirical estimates of the travel cost coefficient and related consumer surplus estimates than did the truncated negative binomial.

A METHOD TO IMPROVE CONSUMER SURPLUS ESTIMATES
FROM TRUNCATED RECREATION DEMAND DATA

by

Kerri L. Weidner

A THESIS

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Master of Science

Completed October 10, 1995
Commencement June 1996

Master of Science thesis of <u>Kerri L. Weidner</u> presented on <u>October 10, 1995</u>

APPROVED:

Redacted for Privacy

Major Professor, representing Agricultural and Resource Economics

Redacted for Privacy

Chair of Department of Agricultural and Resource Economics

Redacted for Privacy

Dean of Graduate School

I understand that my thesis will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my thesis to any reader upon request.

Redacted for Privacy

Kerri L. Weidner, Author

ACKNOWLEDGMENTS

I would like to express the respect and gratitude I have for Dr. William Brown, my mentor and friend. His unending patience and guidance throughout this research project and his true love of research have been inspirational.

I would also like to thank Dr. Gil Sylvia, Dr. Richard Adams, and Dr. Hal Pritchett for their support and suggestions and the OSU Department of Agricultural and Resource Economics for their partial funding of the research effort. Dr. William Greene, the author of LIMDEP, and Dr. Daniel Hellerstein, an expert in travel cost analysis, both provided me with guidance when I really needed it. Thank you.

My wonderful husband, Gary, deserves special thanks for putting up with the long work days and the weekends we could not spend together and for supporting my desire to make a significant contribution to the field of environmental economics.

TABLE OF CONTENTS

LIST OF TABLES

A METHOD TO IMPROVE CONSUMER SURPLUS ESTIMATES FROM TRUNCATED RECREATION DEMAND DATA

SECTION I: INTRODUCTION

It is widely recognized that omitting non-users from samples of recreation demand data produces biased coefficient and consumer surplus estimates if an uncorrected estimator is applied (see Creel and Loomis (1990) for a good discussion of truncation bias). Truncated count estimators have recently been proposed to account for the missing non-user observations. For example, Shaw (1988) suggests a truncated Poisson estimator which acknowledges endogenous stratification.[1] Grogger and Carson (1991) demonstrate the improved accuracy of coefficient estimates obtained by applying truncated rather than standard count estimators to truncated data. Creel and Loomis (1990) support the use of truncated estimators over standard estimators in an empirical hunting study.[2]

Grogger and Carson (1990) show that while the truncated Poisson is biased and inconsistent in the presence of overdispersion, the truncated negative binomial is both

---

[1] Endogenous stratification occurs when site users who visit the site more frequently are more likely to be included in an on-site survey than those who visit infrequently.

[2] Other studies which discuss the standard and truncated count estimators include Hellerstein and Mendelson, 1993; Gomez and Ozuna, 1993; Hellerstein, 1991; Smith, 1988; Cameron and Trivedi, 1986; Gourieroux, et al., 1984; and Hauseman, et al., 1984.

unbiased and consistent.[3] They also note that as the number
of non-users in the true (untruncated) sample increases (for
a given level of overdispersion) the accuracy of the
estimated coefficients for the truncated Poisson diminishes.
This result suggests that when a large portion of the true
sample consists of non-users and is omitted, there is too
little information to produce accurate estimates. As will
be demonstrated here, the estimation problem carries over to
the truncated negative binomial: When truncation is
substantial, the truncated negative binomial produces highly
variable estimates of the demand coefficients and consumer
surplus relative to estimates from censored data.[4]

Thus, although the use of a truncated estimator is
preferred to ignoring truncation altogether, one can not
expect the truncated estimator to produce results equivalent
to those obtained if the censored sample were available
(given a positive number of non-user observations). A
costly solution is to collect population-wide (censored)
data; however, given the budget constraints of government
agencies this option is often impractical.

In this paper, a method of obtaining a censored sample
without the additional expense of a population-wide survey
is proposed. The procedure originates from a mathematical

---

[3] Overdispersion occurs where the conditional variance of
the dependent variable exceeds the conditional mean.

[4] Here 'censored' refers to recreation demand data which
consists of both site users and non-users (i.e.,
$y_i=0,1,2,..,n$).

identity demonstrated by Riyadi (1993) between the linear demand coefficient estimates of two early estimation methods; the zonal average travel cost method (ZATC) and ordinary least squares (OLS) applied to censored data (OLSALL). Given this identity, the appropriate number of non-user observations in the censored sample can be obtained by iteratively adding non-user observations to the truncated data and applying OLS until the coefficient estimates of OLSALL are identical to the ZATC estimates. Sophisticated estimators, such as the negative binomial, can be applied to the resulting imputed censored data to resolve the problem of too little information.

In the next Section, the ZATC is reviewed and a modification of Riyadi's mathematical proof for the identity of coefficient estimates between ZATC and OLSALL is provided. The procedure for generating non-user observations and a review of count data estimators are also discussed. In Section III the results of a Monte Carlo experiment are presented. The experiment compares the accuracy of predicted consumer surplus estimates obtained from the truncated negative binomial applied to truncated data to estimates from the standard negative binomial applied to imputed censored data. The relative accuracy of the estimates is determined by the root mean squared error criteria (RMSE) and the percent deviation from the true values. The simulation results suggest that the value of the expansion method increases with the proportion of non-

user observations in the true sample. In Section IV the feasibility of applying the method in an empirical setting is demonstrated. A sample of 631 anglers (from a 1989 survey) is expanded to include the imputed non-user observations. Estimates from the truncated data set are compared to estimates from the expanded data set. Finally, the summary and method limitations are given in Section V.

SECTION II: ECONOMETRIC METHODS

Data Expansion Method

Application of the classic zonal average travel cost method (ZATC) involves grouping observations collected from an on-site survey into zones according to the survey respondents' home location (e.g., by zip code or county). Total per capita trips by zone are regressed against the zone averages of explanatory variables such as time and travel costs and socio-economic characteristics. OLS was used in early studies to estimate ZATC models, but later Bowes and Loomis (1980) rightly suggested a least squares correction for heteroskedasticity. Assuming that travel cost is the sole determinant of site participation, the Bowes-Loomis model may be written as

$$\frac{VISITS_t}{\sqrt{POP_t}}*k = \alpha^*\sqrt{POP_t} + \beta^*(TC_t*\sqrt{POP_t}) \quad \textbf{(1)}$$

where $VISITS_t$ = the total sample visits for site users from origin t,
   POP_t = the population of origin t,
   k = the expansion factor (i.e., the inverse of the percent of population sampled)
   TC_t = the average travel cost faced by site users from origin t,
   α^* = the estimated ZATC intercept, and
   β^* = the estimated ZATC cost coefficient.

Riyadi (1993) compared the Monte Carlo performance of five travel cost methods which assumed either a censored normal or truncated normal distribution of site visits and found an identity between the Bowes-Loomis ZATC and OLSALL (OLS applied to censored data) demand coefficient estimates. Riyadi assumed that the total zone population (POP_t) equaled

the zone population of potential recreators. Riyadi's proof is modified here by redefining POP_t as the zone population of site users and non-users (which allows for the usual case where zone site users and non-users are a subset of the zone populations). The equality of the coefficients between the two methods is shown by equating the normal equations for the two methods. The ZATC normal equations are

$$\alpha^* \sum_{t=1}^{T} POP_t + \beta^* \sum_{t=1}^{T} (POP_t * TC_t) = k \sum_{t=1}^{T} VISITS_t \quad \text{(2)}$$

$$\alpha^* \sum_{t=1}^{T} (POP_t * TC_t) + \beta^* \sum_{t=1}^{T} (POP_t * TC_t^2) = k \sum_{t=1}^{T} (VISITS_t * TC_t) \quad \text{(3)}$$

or t=1,2,..,T zones.[5] Given $VISITS_i = \alpha + \beta * TC_i + \epsilon_i$, the corresponding normal equations for OLSALL are

$$n\hat{\alpha} + \hat{\beta} \sum_{i=1}^{n} TC_i = \sum_{i=1}^{n} VISITS_i \quad \text{(4)}$$

$$\hat{\alpha} \sum_{i=1}^{n} TC_i + \hat{\beta} \sum_{i=1}^{n} TC_i^2 = \sum_{i=1}^{n} (VISITS_i * TC_i) \quad \text{(5)}$$

for i=1,2,..,n individual sample user and non-user observations. By assuming equal within-zone costs, the normal equations for OLSALL can be rewritten in terms of zones rather than individual observations:

$$\hat{\alpha} \sum_{t=1}^{T} n_t + \hat{\beta} \sum_{t=1}^{T} n_t * TC_t = \sum_{t=1}^{T} VISITS_t \quad \text{(6)}$$

$$\hat{\alpha} \sum_{t=1}^{T} n_t * TC_t + \hat{\beta} \sum_{t=1}^{T} n_t * TC_t^2 = \sum_{t=1}^{T} (VISITS_t * TC_t).^{[6]} \quad \text{(7)}$$

[5] The proof can be generalized to allow the expansion factor to vary across zones but would increase its complexity without adding much information.

[6] If costs are allowed to vary across individuals the identity will not occur but the coefficient estimates will usually be very close.

Rewriting $n_t = POP_t/k$ for each origin zone and assuming k is constant across zones, equations (6) and (7) are

$$\frac{1}{k}\hat{\alpha}\sum_{t=1}^{T} POP_t + \frac{1}{k}\hat{\beta}\sum_{t=1}^{T} (POP_t * TC_t) = \sum_{t=1}^{T} VISITS_t \qquad \textbf{(8)}$$

$$\frac{1}{k}\hat{\alpha}\sum_{t=1}^{T} (POP_t * TC_t) + \frac{1}{k}\hat{\beta}\sum_{t=1}^{T} (POP_t * TC_t^2) = \sum_{t=1}^{T} (VISITS_t * TC_t) \qquad \textbf{(9)}$$

for t=1,2,..,T zones. Multiplying equations (8) and (9) by k produces the normal equations for ZATC in (2) and (3). This implies that $\alpha^* \equiv \hat{\alpha}$ and $\beta^* \equiv \hat{\beta}$.[7] Note that in this analysis any value of k can be used to produce an identity but only the correct k will generate the correct number of non-users.

Riyadi suggested using the identity to generate a censored data set from a truncated data set. To determine the number of non-user observations by zone, one could add non-user observations to the truncated data (proportional to the zone population of recreators) until the coefficients are identical between the two methods. While this approach works, an easier and identical method is to divide the zone population of recreators by the expansion factor (to determine the total observations which were expected from

---

[7] If the proof is generated without the expansion factor, k, the coefficients for ZATC are exactly 1/kth the coefficients for OLSALL.

each zone) and subtract the observed visitors from each zone
from the total observations. That is,

IMPUTED NON-USERS$_t$ = POP$_t$/k - VISITORS$_t$.[8] **(10)**

Equation (10) can be used with a constant or variable
(across zones) expansion factor. The intuition behind the
method is as follows. If the site users in the sample
originating from zone t accurately represent a portion of
the population of site users from that zone, then the
difference between the total observations which should have
been collected for the zone (POP$_t$/k) and the observations
which were actually collected (VISITORS$_t$) are the sample
non-users. When the true expansion factor(s) is (are) known
and there is no sampling error, the true number of non-users
can be obtained. In practice, calculation of the non-user
observations will vary with sampling error and errors in
calculating the expansion factor.

---

[8] If 1/k is the percent of the origin population of site
users sampled then the number of site users collected is

sample site users$_t$ = 1/k * total site users$_t$.

Had it been possible for non-users to represent themselves
on-site one would expect to collect 1/k of the non-user
population:

sample site non-users$_t$ = 1/k * total site non-users$_t$.

Although the site non-users from each origin are typically
unknown, the total origin users and non-users can often be
computed from license (or other) data and the sample site
non-users can be estimated from Equation 8.

Count Data Estimators

Count estimators have been suggested as an improvement over OLS-based methods (such as the ZATC) because they recognize the discrete, non-negative nature of the dependent variable, visits to the recreation site. Two estimators used to model count data are the Poisson and the negative binomial. The probability law for the Poisson is given by

$$Prob(Y_i = y_i; i = 1,2,..,n) = \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!} \quad (11)$$

where $\lambda_i = \exp(X_i\beta)$ is the Poisson parameter to be estimated for the count variable Y.

Application of the Poisson is limited because $E(\lambda_i)$ is constrained to equal $Var(\lambda_i)$. Violation of this restriction is a form of heteroscedasticity termed overdispersion. Cameron and Trivedi (1986) suggest a method of relaxing the restriction by allowing for unexplained randomness in λ_i such that $\lambda_i = \exp(X_i\beta + \epsilon_i)$. If ϵ_i are distributed as a gamma random variable then $\lambda_i = g(X_i, \epsilon_i)$ has mean λ and variance $\alpha\lambda^2$ (and is also distributed as a gamma random variable) and $f(\lambda_i)$ is distributed negative binomial. The probability law for this parameterization of the negative binomial (the Negbin II model in Cameron and Trivedi (1986)) is

$$F(y) = Pr(Y_i = y) = \int Pr[Y_i = y|\lambda_i] f(\lambda_i) d\lambda_i$$

$$= \frac{\Gamma(y+\frac{1}{\alpha})}{\Gamma(y+1)\Gamma(\frac{1}{\alpha})} (\alpha\lambda_i)^y [1+\alpha\lambda_i]^{-(y+\frac{1}{\alpha})} \quad (12)$$

for Y_i, a member of a set of non-negative integers, where $\Gamma(.)$ represents a gamma function, $\alpha > 0$ is the over-dispersion parameter, $\lambda_i > 0$, $E[Y_i]=\lambda$, and $VAR[Y_i]=\lambda(1+\alpha\lambda)$.

The truncated negative binomial adjusts the estimated probabilities of observed Y_i (given $Y_i > L$, the truncation limit) by the estimate of the probability of being above the truncation limit:

$$Pr(Y_i = y|Y_i>L) = \frac{F(y)}{1-F(L)}. \quad \textbf{(13)}$$

For L=0, the probability distribution for the truncated negative binomial is

$$Pr(Y_i = y|Y_i>0) = \frac{\Gamma(y+\frac{1}{\alpha})}{\Gamma(y+1)\Gamma(\frac{1}{\alpha})}(\alpha\lambda_i)^y[1+\alpha\lambda_i]^{-(y+\frac{1}{\alpha})}[1-F(0)]^{-1}. \quad \textbf{(14)}$$

The truncated Poisson is biased and inconsistent in the presence of overdispersion but the truncated negative binomial is unbiased and consistent if the mean is correctly specified and **(14)** is the true data generating process (Creel and Loomis, 1990). However, as will be shown in the Monte Carlo simulation presented here, the variance of the estimates may be substantial for small $E(\lambda_i)$.

SECTION III: SIMULATION STUDY DESIGN AND RESULTS

This section illustrates the improved efficiency of consumer surplus estimates that can result from use of the data expansion method. A summary of several Monte-Carlo simulations is presented that compares the coefficient and predicted consumer surplus estimates of the truncated negative binomial estimator applied to truncated data (TRNEGB) to the estimates for the standard negative binomial applied to imputed censored data (ICNEGB).

Study Design

The simulated data (site visits) is distributed negative binomial ($VISITS_i=0,1,2..n$). A simple model was adopted of the form $\ln(VISITS_i) = \beta_0 + \beta_{TC}*TC_i + \epsilon_i$ where the travel costs, TC_i, are taken from a 1968 Oregon hunter study, and $\exp(\epsilon_i)$ is gamma distributed with mean one and variance α, the overdispersion parameter (Greene, 1991).[9] A value of -0.015 was selected for β_{TC}. Three values of the constant (β_0) were chosen to achieve $E(\lambda_i) = 0.5$, 1, and 2. A moderate level of overdispersion ($\alpha=0.6$) relative to Grogger and Carson's study was chosen. The gamma distributed error term (ϵ_i) was produced in Limdep using the GTB function. One thousand samples, each consisting of 400

---

[9] These empirical data, taken from Shalloof (1985), generated acceptable results in the 1968 travel cost analysis (i.e., correct coefficient signs and significant t-statistics). The travel costs were updated from $0.06 per mile to $0.22 per mile. The total observations were distributed among nine distance zones in proportion to the zone population of recreators (see Appendix A).

observations, were generated for each $E(\lambda_i)$ by applying $\lambda_i = \exp(\beta_0 -0.015*TC_i + \epsilon_i)$ to Limdep's Poisson random number generator. The true sample consumer surplus for each $E(\lambda_i)$ was computed as $\sum \lambda_i/0.015$ from the censored samples; predicted consumer surplus was calculated as $\sum \hat{\lambda}_i/-\beta_{TC}$. Examples of the computer programming are provided in Appendix B.

Two sets of results were produced for the imputed censored data. For the first set, the expanded data are generated assuming that the true value of k is known and that k is constant across all zones. The expanded data for the second set are generated assuming that k is estimated with error and varies across zones within each sample. The error was incorporated into the data by creating a uniformly distributed expansion factor which varied by ten percent on either side of the true value. The effect of creating error in the expansion factor is to vary the number of non-user observations estimated with the expansion method. The truncated negative binomial estimates and the true values of the coefficients and consumer surplus remain unchanged for the second set of results.

Distribution Check

To check that the simulated data were distributed negative binomial, the average generated number of visits to the site for each $E(\lambda_i)$ across the 1000 samples were compared to the intended averages ($E(\lambda_i)$=0.5, 1, and 2).

The average variance and proportion of non-user observations (% NON-USERS) were also compared.[10] All comparisons are provided in Table 1.

Table 1. Average Intended Values Versus Average Generated Values for Each $E(\lambda_i)$ in the Simulation Study

	$E(\lambda_i)=0.5$	$E(\lambda_i)=1.0$	$E(\lambda_i)=2.0$

$E(\lambda_i)$
  Ave. Intended
  Ave. Generated
	0.50
	0.51
		1.00
		1.02
			2.00
			2.05

$VAR(\lambda_i)$
  Ave. Intended
  Ave. Generated
	0.65
	0.71
		1.60
		1.80
			4.40
			5.13

% NON-USERS
  Ave. Intended
  Ave. Generated
	0.64
	0.65
		0.46
		0.46
			0.27
			0.27

The average of the actual mean number of visits across generated samples, the average proportion of non-users, and the average value of the variance were quite close to the intended values. In addition, the Z-statistic was used to test whether the generated average number of visits differed significantly from the intended average number of visits for each level of $E(\lambda_i)$.[11] At a 5% significance level one would

---

[10] The theoretical proportion of non-user observations is calculated by evaluating the log-likelihood function for the standard negative binomial at y=0 using known α and λ.

[11] The central limit theorem states that regardless of the distribution of X, the distribution of $\bar{X}$ will be approximately normal for large samples (Kmenta, 1986). Thus the 2-tailed test of $H_0:\mu_X=E[\lambda_i]$ against the alternative $H_1:\mu_X \neq E[\lambda_i]$ with 399 degrees of freedom and a critical value of 1.96 is appropriate. The test statistic is $Z=(\bar{X}-\mu_X)/(\sigma_X/\sqrt{n})$.

expect to reject the null hypothesis for about 50 out of 1000 samples. The number of samples out of 1000 which were rejected for $E(\lambda_i)$=0.5, 1, and 2 were 43, 48, and 56, respectively.

Accuracy Tests

The accuracy of the demand coefficient estimates, predicted consumer surplus (CS) estimates, and the estimated proportion of non-user observations were evaluated based on the root mean square error criteria and the percent deviation from the true values. The formulas for root mean square error (Kmenta, 1986) and percent deviation (Hellerstein, 1992) are as follows:

$$RMSE = \sqrt{\sum (y_j - \hat{y}_j)^2 / M}$$

$$AVERAGE \ \% \ DEVIATION = \sum (\frac{ABS(y_j - \hat{y}_j)/y_j}{M}) * 100$$

where M = the number of samples (j=1,2,...M),
   $\hat{y}_j$ = the estimated sample value of y (either sample CS, sample coefficient estimate, or % non-users), and
   Y_j = the true sample value of y.

Smaller values for RMSE and percent deviation obviously indicate better performance.

A problem, not previously discussed in the literature, occurs when comparing the RMSE of the estimates as $E(\lambda_i)$ changes. When $E(\lambda_i)$ increases from 0.5 to 1, for example, the true values of the intercept and consumer surplus also increase. The RMSE formula is based on the absolute difference between the true and estimated values. The absolute difference may be larger for larger $E(\lambda_i)$ than for

small $E(\lambda_i)$ even if the estimate at larger $E(\lambda_i)$ is closer, in percentage terms, to the true value.

To adjust the RMSE to account for the relative difference in true values the following formula is suggested:

$$Relative\ RMSE = \sqrt{\sum (y_j^* - \hat{y}_j^*)^2 / M}$$

where
$$y_j^* = (\bar{y}^*/\bar{y}_{base})^{-1} * y_j$$
$$\hat{y}_j^* = (\bar{y}^*/\bar{y}_{base})^{-1} * \hat{y}_j .$$

$\bar{y}_{base}$ is the average of the sample values of y_j (call these $y_{j_{base}}$) for some base level of $E(\lambda_i)$ and $\bar{y}^*$ is the average of the sample values of y_j for each $E(\lambda_i)$. At the base level $(\bar{y}^*/\bar{y}_{base})^{-1} = 1$ and the Relative RMSE is identical to the RMSE.

As an example, assume that for 2 levels of $E(\lambda_i)$ the true values and the estimated values are as presented in Table 2. Also assume, for simplicity, that M=1 (then $y_{j_{base}} = \bar{Y}_{base}$ and $y_j = \bar{Y}^*$). While the absolute differences between

Table 2. The Incomparability of RMSE Across Levels of $E(\lambda_i)$

	TRUE CS	Ave. $\hat{CS}$	DIFF.	RMSE	% DEVIATION
$E(\lambda_i)=0.5$	15,000	10,000	5,000	5000	33.3%
$E(\lambda_i)=1$	150,000	145,000	5,000	5000	3.3%

the true and average estimated CS are identical and generate identical RMSE estimates for both $E(\lambda_i)$, the percent deviation of the estimated average CS for $E(\lambda_i)=0.5$

is far greater (33.3%) than for $E(\lambda_i)=1$ (3.3%). The RMSE incorrectly suggests that the estimates are equally accurate.

Using $E(\lambda_i)=0.5$ as the base level in the example above, the Relative RMSE estimate does not change for $E(\lambda_i)=0.5$ (RMSE=5000) and the Relative RMSE for $E(\lambda_i)=1$ is

$$Relative\ RMSE = \sqrt{((\frac{150,000}{15,000})^{-1}*150,000 - (\frac{150,000}{15,000})^{-1}*145,000)^2/1}$$

$$= \sqrt{(\frac{150,000}{10} - \frac{145,000}{10})^2/1} = 500.$$

Clearly the Relative RMSE is more suited to situations where true values are changing.

Study Results

The simulation results are presented in Table 3. While the primary emphasis of the study was to evaluate the accuracy of consumer surplus estimates, the performance of the intercept and travel cost coefficient (β_{TC}) may also be of interest and are included in Table 3.

To effect truncation at each level of $E(\lambda_i)$, the non-user observations were dropped from each censored sample and Grogger and Carson's truncated negative binomial estimator was applied. The average estimates and accuracy tests for the coefficients and consumer surplus at each $E(\lambda_i)$ for the truncated estimator are provided under the column headings TRNEGB.

TABLE 3. Accuracy of Estimates from Truncated Data Versus Those from Imputed Censored Data for $\alpha=0.6$ and Three Levels of $E(\lambda_i)$

$E(\lambda_i)=0.5$: 65% NON-USERS	TRNEGB	ICNEGB1	ICNEGB2
E(INTERCEPT) Ave. Estimate Relative RMSE Ave. % Deviation	0.269 0.325 0.593 167	0.269 0.235 0.324 95	0.269 0.445 0.385 116
$E(\beta_{TC})$ Ave. Estimate Relative RMSE Ave. % Deviation	-0.0150 -0.0156 0.0081 44	-0.0150 -0.0145 0.0052 28	-0.0150 -0.0178 0.0061 33
E(TRUE CS) Ave. Estimate Relative RMSE Ave. % Deviation	13,682 2,115 138,658 145	13,682 17,097 11,592 44	13,682 15,473 53,279 116
# Samples for $\beta_{TC}>0$	36	5	5
# Samples for CS>10*TRUE CS	14	3	3
$E(\lambda_i)=1.0$: 45% NON-USERS	TRNEGB	ICNEGB1	ICNEGB2
E(INTERCEPT) Ave. Estimate Relative RMSE Ave. % Deviation	0.962 0.853 0.122 38	0.962 0.947 0.075 22	0.962 1.082 0.084 25
$E(\beta_{TC})$ Ave. Estimate Relative RMSE Ave % Deviation	-0.0150 -0.0140 0.0060 32	-0.0150 -0.0148 0.0042 23	-0.0150 -0.0169 0.0047 25
E(TRUE CS) Ave. Estimate Relative RMSE Ave. % Deviation	27,346 23,033 38,447 70	27,346 31,023 8,044 30	27,346 25,956 9,346 35
# Samples for $\beta_{TC}>0$	12	2	1
# Samples for CS>10*TRUE CS	4	2	1

Table 3--Continued

E(λ_i)=2: 27% NON-USERS	TRNEGB	ICNEGB1	ICNEGB2
E(INTERCEPT) Ave. Estimate Relative RMSE Ave. % Deviation	1.656 1.633 0.043 13	1.656 1.626 0.038 11	1.656 1.702 0.040 12
E(β_{TC}) Ave. Estimate Relative RMSE Ave % Deviation	-0.0150 -0.0142 0.0041 22	-0.0150 -0.0146 0.0036 19	-0.0150 -0.0158 0.0039 20
E(TRUE CS) Ave. Estimate Relative RMSE Ave. % Deviation	54,679 49,894 8,158 31	54,679 60,848 6,333 25	54,679 56,556 6,719 23
# Samples for β_{TC} >0	1	1	0
# Samples for CS>10*TRUE CS	1	1	1

Notes: TRNEGB are estimates from truncated data; ICNEGB1 are estimates from imputed censored data assuming constant, known k; ICNEGB2 are estimates from imputed censored data assuming variable k with error.

The non-user observations for the imputed censored data sets were computed using the data expansion method applied to the truncated portion of the data sets. Travel costs were averaged across individuals within each zone and applied to the non-user observations. The average estimates for the imputed censored samples and accuracy test results at each E(λ_i) are shown under the column headings ICNEBG1 and ICNEGB2 (which are collectively referred to as ICNEGB). The results listed under ICNEGB1 suggest the best possible performance of the expansion method for the simulated data since no error is assumed in either the sampling procedure

or the expansion factor. The results listed under ICNEGB2 represent a more realistic evaluation of the method performance by allowing the expansion factor to vary across zones and introducing error in the data.

For $E(\lambda_i)=0.5$, the truncated negative binomial estimator failed to find a minimum for 63 samples. These samples were deleted for ICNEGB1 and ICNEGB2 to ensure comparability to TRNEGB. Further problems in estimation occurred when calculating the correct sign for the travel cost coefficient and in computing a reasonable consumer surplus (here, defined as less than ten times the absolute value of the true consumer surplus). For the two lower levels of $E(\lambda_i)$, TRNEGB had more difficulty estimating correct signs and a reasonable consumer surplus than did ICNEGB. As $E(\lambda_i)$ increased TRNEGB performance improved due to the decrease in omitted non-user observations.

Grogger and Carson (1991) noted the poor performance of the truncated Poisson estimator given small $E(\lambda_i)$. That result is demonstrated here for the truncated negative binomial in the comparison of TRNEGB and ICNEGB coefficient and consumer surplus estimates. The consumer surplus estimates from the imputed censored data are superior to those of the truncated data at every level of $E(\lambda_i)$, ranging from 30 to 81 percent of the average percent deviation for TRNEGB. As expected, the TRNEGB estimates converge in accuracy to the ICNEGB estimates as the proportion of non-

users declines (as $E(\lambda_i)$ increases). Although inclusion of sampling error and a variable expansion factor increased the variability of the ICNEGB2 estimates relative to ICNEGB1 estimates, the performance of ICNEGB2 remained superior to that of TRNEGB at every level of $E(\lambda_i)$.

SECTION IV: EMPIRICAL ANALYSIS

In this section the expansion method is demonstrated using data from a 1989 angler survey.[12] The data is comprised of 631 Oregon trout anglers who fished in the Willamette Valley, Oregon. Trout fishing was the most popular angling activity in Oregon in 1989 (comprising an estimated 40 percent of the total angling activity) and the Willamette Valley was only one of seven regions in which trout anglers could have fished. Since the survey noted only the region of the fishing experience rather than a specific site and was conducted by mail (which introduces some complications for the expansion method), the estimates obtained here should be considered only for the purpose of illustrating the method.

The observations from the truncated data set were grouped by adjoining zip codes into 65 origin zones from around the state. The populations for the zip codes within each zone were summed and divided by the expansion factor to obtain the total observations (site users and non-users) expected from each zone. The site users were summed by zone and subtracted from the total observations to produce the number of non-user observations for each zone. The total number of non-user observations for all origins summed to

---

[12] The author wishes to thank Chris Carter, Oregon Department of Fish and Wildlife, and Shannon Davis and Hans Radtke, survey implementors, for providing the data. Average travel costs and total visits by zone are presented in Appendix C.

2,357. The average number of visits in the truncated samples is 2.2 visits. For the imputed censored sample the average is 0.5 visits.

The respondents had an opportunity to discuss up to eight fishing trips for a two month period during the year. The reported catch rates and costs (which included travel expenses and purchases made for food and supplies during trips to the Willamette Valley for trout fishing) were averaged across trips for each respondent. The average catch rate and costs were then averaged across respondents within each zone and assigned to the non-user observations (that were calculated using the data expansion method) for each zone. The average income reported by zip code in the 1990 census was averaged across zip codes within each zone and was used for site non-users. Substitute sites were not considered here because the regions were large and the average quality for the Willamette Valley (measured as the average per trip catch rate) did not differ significantly from the average quality for all other regions.

Age and sex were the only other variables from the mail survey relevant to the study. Obviously, there is no way to include these variables in the imputed censored sample. The truncated negative binomial was estimated with age and sex (TRNEGB1) and without age and sex (TRNEGB2) and is compared to the standard estimator (using imputed censored data) in Table 4. The difference in the log-likelihoods between TRNEGB1 and TRNEGB2 and the t-statistics in TRNEGB1 for age

and sex suggest that these variables were fairly poor at explaining visits to the region. The t-statistic for average catch rate for trout is also insignificant for the TRNEGB models. The small travel cost coefficient estimate in TRNEGB2 generated an estimate of $684 for individual per trip consumer surplus.

The coefficient estimates between TRNEGB2 and ICNEGB differed dramatically. The signs on all coefficient estimates for ICNEGB are consistent with theory and the estimates are significant. The larger travel cost coefficient for ICNEGB, relative to TRNEGB models, generated a value of $35 for the individual (per trip) consumer surplus estimate.

Although the true value of consumer surplus is unknown in empirical studies, economists often judge the reasonableness of benefit estimates by comparing these estimates to values obtained from different empirical studies. Based on other studies' empirical estimates of the value of trout fishing, the consumer surplus estimate for TRNEGB is substantially higher than expected; whereas, the consumer surplus estimate for ICNEGB is very similar in magnitude to other study estimates.[13]

The high estimate of consumer surplus and insignificant t-statistics obtained from the truncated data are likely due

---

[13] For example, Loomis, et al. (1986) estimated a range of $34 to $70 and Vaughan and Russell (1982) estimated values between $11 and $24 for individual (per trip) consumer surplus.

to the lack of information available to the truncated estimator when a large proportion of the censored data set is comprised of non-user observations. Use of the data expansion method to reduce the problem of too little information in this empirical analysis produced a more reasonable estimate of consumer surplus.

Table 4. Estimation and Test Results From the 1989 Oregon Angler Survey

	TRNEGB1	TRNEGB2	ICNEGB
INTERCEPT	0.925 (4.067)	0.846 (8.043)	0.393 (2.953)
COST	-0.003 (-5.267)	-0.003 (-5.490)	-0.014 (-44.451)
CATCH	0.001 (0.209)	0.002 (0.259)	0.032 (3.808)
INCOME	-0.831E-05 (-2.794)	-0.820E-05 (-2.794)	-0.228E-05 (-6.889)
AGE	-0.733E-03 (-0.266)	----	----
SEX	-0.030 (-0.364)	----	----
α	0.199 (3.265)	0.199 (3.268)	2.951 (13.099)
Log-Likelihood	-949.55	-949.61	-2550.51
Average CS	$692	$684	$35

Notes: TRNGB1 are estimates from truncated data with sex and age included; TRNEGB2 are estimates from truncated data without sex and age; ICNEGB are estimates from imputed censored data. T-statistics are in parentheses. α is the overdispersion parameter estimate.

SECTION V: SUMMARY

The objective of this study was to provide an alternate
method for obtaining consumer surplus estimates when only
truncated data are available to the researcher. The study
also emphasizes the inaccuracy of truncated estimates which
arises when the proportion of non-users in the untruncated
sample is large.

From the identity found and demonstrated by Riyadi
(1993), the number of non-user observations can be estimated
by equating the coefficient estimates between OLSALL and
ZATC. Alternatively, the non-user observations can be
estimated using the data expansion method that was developed
in this study. The expansion method allocates the non-user
observations across origin zones, offering an attractive
guideline for assigning independent variable values to the
non-user observations.

In the simulation from Section III, the data expansion
method provided substantial efficiency gains for consumer
surplus estimates. The value of the expansion method
increased with the proportion of non-user observations in
the true sample. When the proportion of sample non-users
was large, the truncated negative binomial estimator
(TRNEGB) was six to seven times more likely to produce the
wrong sign on the cost coefficient (and consumer surplus)
than estimates based on the data expansion method (ICNEGB)
and two to three times more likely to produce unreasonable
estimates of consumer surplus. The average percent

deviation of predicted consumer surplus from the true consumer surplus for TRNEGB was 200% to 300% more than ICNEGB for small $E(\lambda_i)$. Even when some error was introduced in estimating the non-user observations, the data expansion method still produced less variable estimates of consumer surplus in the simulation than could be obtained using the truncated negative binomial estimator. Future research to determine the percentage of error necessary to generate more accurate estimates may be beneficial. A second issue to consider for future study is the effects of, and correction for, endogenous stratification which was not treated in this analysis.

The data expansion method and standard negative binomial were also used to fit a travel cost model to data from a 1989 angler survey. These results were compared to results from fitting the truncated negative binomial to truncated data. The average, individual consumer surplus estimate from the expanded data was approximately $35 per trip compared to $684 for the truncated data. Although the true consumer surplus for this empirical study is unknown, the $684 per trip estimate is ten to twenty times larger than most reported empirical estimates (and the estimate calculated from expanded data). The empirical analysis illustrated the difficulties that can arise from relying upon truncated data, especially where a large proportion of non-users is expected.

BIBLIOGRAPHY

Bowes, M. D., and J. B. Loomis. 1980. "A Note on the Use of Travel Cost Models with Unequal Zonal Populations." *Land Economics* 56:465-70.

Cameron, A. C. and P. K. Trivedi. 1986. "Econometric Models Based on Count Data: Comparisons and Applications of Some Estimators and Tests." *Journal of Applied Econometrics* 1:29-53.

Creel, M. and J. Loomis. 1990. "Theoretical and Empirical Advantages of Truncated Count Data Estimators for Analysis of Deer Hunting in California." *American Journal of Agricultural Economics* 72 (May):434-41.

Greene, W. H. 1991. *Limdep Version 6.0: User's Manual and Reference Guide*. Econometric Software, Inc, Bellport, N.Y.

Gomez, I. A. and T. Ozuna, Jr. 1993. "Tests for Overdispersion in Truncated Count Data Recreation Demand Functions." *Journal of Environmental Management* 37:117-25.

Gourieroux, C., A. Montfort, and A. Trognon. 1984. "Pseudo Maximum Likelihood Methods: Application to Poisson Models." *Econometrica* 52,3:701-20.

Grogger, J. T. and R. T. Carson. 1991. "Models for Truncated Counts." *Journal of Applied Econometrics* 6:225-38.

Hauseman, J., Hall, H. H., and Z. Griliches. 1984. "Econometric Models for Count Data with an Application to the Patents-R&D Relationship." *Econometrica* 52 (July):909-38.

Hellerstein, D. 1991. "Using Count Data Models in Travel Cost Analysis with Aggregate Data." *American Journal of Agricultural Economics* (Aug);860-67.

Hellerstein, D. 1992. "Estimating Consumer Surplus in the Censored Linear Model." *Land Economics* 68(1):83-92.

Hellerstein, D. and R. Mendelson. 1993. "A Theoretical Foundation for Count Data Models." *American Journal of Agricultural Economics* 75 (Aug):604-11.

Kmenta, J. *Elements of Econometrics: Second Edition*. Macmillan Publishing Company, New York, N.Y.

Loomis, J. B., C. F. Sorg, and D. M. Donnelly. 1986. "Evaluating Regional Demand Models for Estimating Recreation Use and Economic Benefits: A Case Study." *Water Resources Research* 22(4):431-38.

Riyadi, M. M. 1993. *A Monte Carlo Study of Several Methods for Estimating Linear Demand for Outdoor Recreation from Censored and Truncated Data*. Dissertation (July), Oregon State University, Oregon.

Shalloof, F. M. 1985. *Impact of Various Factors Upon Benefits From Big Game Hunting Estimated By the Travel Cost Method*. Dissertation (May), Oregon State University, Oregon.

Shaw, D. G. 1988. "On-Site Samples' Regression: Problems of Non-negative Integers, Truncation, and Endogenous Stratification." *Journal of Econometrics* 37(Feb):211-23.

Smith V. K. 1988. "Selection and Recreation Demand." *American Journal of Agricultural Economics* (Feb);29-36.

Vaughan, W. J. and C. S. Russell. 1982. "Valuing a Fishing Day: An Application of a Systematic Varying Parameter Model." *Land Economics* 58(4):450-63.

APPENDICES

APPENDIX A: Empirical Cost Data and Non-user
Observations for the Simulation Study

Table 5. Total Zone Observations and Respective Travel Cost
Values for Site Users and Non-users in the Simulation Study

ZONE	SITE USERS	TRAVEL COST	ZONE POPULATION	TOTAL OBSERVATIONS (ROUNDED)	NON-USERS
1	4	67.98	312,347	65	61
2	10	56.76	166,089	34	24
3	4	81.62	100,946	21	17
4	4	45.10	36,727	7	4
5	8	64.90	198,167	41	33
6	24	59.40	556,665	115	92
7	5	36.74	86,283	17	13
8	15	78.76	429,654	89	74
9	5	2.20	40,543	8	3
TOTAL	79	---	1,927,421	400	321

Notes: 400 observations were desired for the total sample
size. This was accomplished by using an expansion factor
(481.848) such that dividing the total angler populations
(assumed to be 10% of the total population for each zone) by
that expansion factor would produce exactly 400
observations. Total zone observations were calculated by
dividing the zone populations by the expansion factor.

APPENDIX B: Programming

The following programs were generated in LIMDEP. Due to
computer and program limitations the data were generated in
sets of 40,000 observations using different seeds for each
set. The same 400,000 error terms were used across $E(\lambda_i)$.
Programs for $E(\lambda_i)=0.5$ are presented here. The time
necessary to generate the data was substantial;
approximately two weeks of continuous iterations for the
error term and several hours apiece for each of the
remaining programs.

Gamma Distributed Error Generation

```
fast
calc; ran(9232246) $
read; nobs=2000; nvar=1; names=gamma; temp=scrtch61.dat;
file=gamma.dat; maxvar=2 $
calc; j=1 $
label; 2 $
sample; 1-20000 $
create; gam3=0 $
calc; i=1 $
label; 20 $
sample; i $
calc; prob=rnu(0,1); gam=gtb(prob, 1.66666666, 1.66666666) $
create; gam3(i)=gam   $
calc; i=i+1 $
goto; 20; i<=2000 $
sample; 1-2000 $
write; gam3; file=gamma2.dat $
delete; gam3 $
append; nobs=2000; nvar=1; names=gamma; file=gamma2.dat $
save; file=gamma61.sav $
??? gamma61.sav is recalled and data saved to gamma61.dat
calc; list;j=j+1 $
go to; 2; j<= 20 $
```


Grouping Errors

```
fast
read; nobs=40000; nvar=1; names=gamma; temp=gamma61.dat;
file=scgamma.dat $
append; nobs=40000; nvar=1; names=gamma; file=gamma62.dat $
append; nobs=40000; nvar=1; names=gamma; file=gamma63.dat $
append; nobs=40000; nvar=1; names=gamma; file=gamma64.dat $
append; nobs=40000; nvar=1; names=gamma; file=gamma65.dat $
append; nobs=40000; nvar=1; names=gamma; file=gamma66.dat $
append; nobs=40000; nvar=1; names=gamma; file=gamma67.dat $
append; nobs=40000; nvar=1; names=gamma; file=gamma68.dat $
append; nobs=40000; nvar=1; names=gamma; file=gamma69.dat $
```

```
append; nobs=40000; nvar=1; names=gamma; file=gamma610.dat $
sample; 1-400000 $
write; gamma; file=gamma.dat $
```

Generating Non-Stochastic Data

```
fast
read; nobs=400; nvar=4 ; names= zone,costi,costz,popz;
temp=scrnon.dat; maxvar=4; file=nonst.dat $
calc; j=1 $
label; 2 $
append; nobs=400; nvar=4; names=zone,costi,costz,popz;
file=zatc.dat $
save; file=bigdat.sav $
??? bigdat.sav is recalled and data saved to bigdat6.dat
??? costi are individual cost values, costz and popz are
??? zone average costs and total zone populations
calc; j=j+1 $
go to; 2; j<=999 $
```

Combining Error and Nonstochastic Terms

```
fast
read; nobs=400000; nvar=1; names=gamma; file=gamma.dat;
temp=scra62.dat;maxvar=10 $
read; nobs=400000; nvar=4; names=zone,costi,costz,popz;
file=bigdat6.dat $
save; file=bigc6.sav $

create;x=(costi/.06)*.22;
cost2=(costz/.06)*.22;lgam=log(gamma) $
calc; mp=xbr(x); int=log(.5)+.015*mp $
??? The .5 in log(.5) above is changed to 1 and 2 for λ=1
and 2
create; xb=int-.015*x; y1=exp(xb+lgam) $
save; file=big56.sav $
sample; 1-40000 $
write; x, y1, cost2, zone, popz;  file=datc561.dat $
sample; 40001-80000 $
write; x, y1, cost2, zone, popz;  file=datc562.dat $
sample; 80001-120000 $
write; x, y1, cost2, zone, popz;  file=datc563.dat $
sample; 120001-160000 $
write; x, y1, cost2, zone, popz;  file=datc564.dat $
sample; 160001-200000 $
write; x, y1, cost2, zone, popz;  file=datc565.dat $
sample; 200001-240000 $
write; x, y1, cost2, zone, popz;  file=datc566.dat $
sample; 240001-280000 $
write; x, y1, cost2, zone, popz;  file=datc567.dat $
sample; 280001-320000 $
write; x, y1, cost2, zone, popz;  file=datc568.dat $
```

```
sample; 320001-360000 $
write; x, y1, cost2, zone, popz;  file=datc569.dat $
sample; 360001-400000 $
write; x, y1, cost2, zone, popz;  file=datc5610.dat $
```

Sample Truncated Negative Binomial Estimation

```
fast
read; nobs=40000; nvar=2; names=x,y1;
      file=datc561.dat $
matrix;   YMEAN=init(100,1,0);     YVAR=init(100,1,0);
          tnbb0=init(100,1,0);    tnbb1=init(100,1,0);
         tnbexp=init(100,1,0);   REALCS=init(100,1,0);
       tsnbalph=init(100,1,0); tnbalph=init(100,1,0);
           pnon=init(100,1,0);     bads=init(100,1,0) $
calc; nolist; ran(5682643); J=1; n1=1; n2=400; chk2=1 $
label; 1 $
sample; n1-n2 $
create;  visits=rnp(y1); rcs=y1/.015 $
create; if (visits=0) pnon1=1; (else) pnon1=0 $
calc; list; YVAR1=VAR(VISITS); YMEAN1=XBR(VISITS);
            NTOTRCS=SUM(RCS); pnon2=(sum(pnon1))/400 $
reject; visits=0 $
poisson; lhs=visits; rhs=one,x;model=n;limit=0;truncation;
par $
MATRIX; NOLIST; A=VECD(VARB); VARB0=PART(A, 1, 1) $
CALC;NOLIST; D1=VARB0; SEB0=SQR(D1) $
create;lambdah=exp(b(1)+b(2)*x); nexpcs=-lambdah/b(2) $
calc; noLIST; ntotexp=sum(nexpcs);
sealph=sqr(varalpha); ttalph=alpha/sealph; b0=b(1);
if [SEB0>100 + b0=0 +  sealph>=100] chk2=0 $
matrix; NOLIST; a=iden(1); alp=a*alpha; ttalph1=a*ttalph;

        rcs2=ntotrcs*a;nbexp1=a*ntotexp; pnon3=a*pnon2;
        YVAR2=A*YVAR1;YMEAN2=YMEAN1*A;
        tnbb01=a*b(1); tnbb11=a*b(2) $
matrix;NOLIST; tnbb0(j,*)=tnbb01;  tnbb1(j,*)=tnbb11;
          tsnbalph(j,*)=ttalph1;tnbexp(j,*)=nbexp1;
            realcs(j,*)=rcs2;     YVAR(j,*)=YVAR2;
            tnbalph(j,*)=alp;      pnon(j,*)=pnon3;
              YMEAN(j,*)=YMEAN2 $
calc; bads1=0; if [chk2=0] bads1=j $
matrix; bads2=bads1*a $
matrix; bads(j,*)=bads2 $
calc; list; j=J+1;  chk2=1; n1=n1+400; n2=n2+400 $
goto; 1; j<=100 $
save; file=mdtt61.sav $
sample; 1-100 $
delete; x, y1, rcs, visits, pnon1, lambdah, nexpcs $
create; nb0=0;   nb1=0; nexp=0; pnonuse=0; real=0 ; MEANY=0;

        vaRY=0; ads3=0; alph=0; talph=0 $
create;nb0=dta(tnbb0) $
```

```
create; nb1=dta(tnbb1) $
create; nexp=dta(tnbexp) $
create; pnonuse=dta(pnon) $
create; real= dta(realcs) $
create; alph=dta(tnbalph) $
create; talph=dta(tsnbalph) $
create; bads3=dta(bads) $
CREATE; MEANY=DTA(YMEAN) $
CREATE; VARY=DTA(YVAR) $
write; nb0, nb1, nexp, pnonuse, real;
file =mdtt611.dat $
write; alph, talph,bads3,VARY,MEANY;
file=mdtt612.dat$
delete;nb0,nb1,nexp,pnonuse,real,alph,talph,bads3,MEANY,VARY
$
```

Average Truncated Estimates Across Samples

```
read;nobs=100;nvar=5;names=nb0,nb1,nexp,pnon,real;
file=mdtt611.dat$
append;nobs=100;nvar=5;names=nb0,nb1,nexp,pnon,real;
file=mdtt621.dat$
append;nobs=100;nvar=5;names=nb0,nb1,nexp,pnon,real;
file=mdtt631.dat$
append;nobs=100;nvar=5;names=nb0,nb1,nexp,pnon,real;
file=mdtt641.dat$
append;nobs=100;nvar=5;names=nb0,nb1,nexp,pnon,real;
file=mdtt651.dat$
append;nobs=100;nvar=5;names=nb0,nb1,nexp,pnon,real;
file=mdtt661.dat$
append;nobs=100;nvar=5;names=nb0,nb1,nexp,pnon,real;
file=mdtt671.dat$
append;nobs=100;nvar=5;names=nb0,nb1,nexp,pnon,real;
file=mdtt681.dat$
append;nobs=100;nvar=5;names=nb0,nb1,nexp,pnon,real;
file=mdtt691.dat$
append;nobs=100;nvar=5;names=nb0,nb1,nexp,pnon,real;
file=mdt6211.dat$
sample; 1-1000 $
WRITE; NB0,NB1,NEXP,PNON,REAL;FILE=T561A.DAT $
delete; nb0,nb1,nexp,pnon,real,alph,talph $

READ;nobs=100;nvar=5;names=alph,talph,bads3,vary,meany;
file=mdtt612.dat $
append;nobs=100;nvar=5;names=alph,talph,bads3,vary,meany;
file=mdtt622.dat$
append; nobs=100;nvar=5; names=alph,talph,bads3,vary,meany;
file=mdtt632.dat$
append;nobs=100;nvar=5;names=alph,talph,bads3,vary,meany;
file=mdtt642.dat$
append;nobs=100;nvar=5;names=alph,talph,bads3,vary,meany;
file=mdtt652.dat$
append;nobs=100;nvar=5;names=alph,talph,bads3,vary,meany;
file=mdtt662.dat$
```

```
append;nobs=100;nvar=5;names=alph,talph,bads3,vary,meany;
file=mdtt672.dat$
append;nobs=100;nvar=5;names=alph,talph,bads3,vary,meany;
file=mdtt682.dat$
append;nobs=100;nvar=5;names=alph,talph,bads3,vary,meany;
file=mdtt692.dat$
append;nobs=100;nvar=5;names=alph,talph,bads3,vary,meany;
file=mdt6212.dat$
sample; 1-1000 $
WRITE; ALPH,TALPH,BADS3,VARY,MEANY; FILE=T562.DAT $

READ; NOBS=1000;NVAR=5;
NAMES=NB0,NB1,NEXP,PNON,REAL;FILE=T561A.DAT $
READ; NOBS=1000;NVAR=5;
NAMES=ALPH,TALPH,BADS3,VARY,MEANY;FILE=T562.DAT $

create; if (bads3>0) ts=1; (else) ts=0 $
calc; list; sts=sum(ts) $
write; ts; file=bad56.dat $
write; alph; file=alph56.dat $
write; vary,meany; vmc56.dat $
READ; NOBS=1000; NVAR=1; NAMES=nb1m0; FILE=nb1562.dat $
SAMPLE; 1-1000 $
create; if (nb1m0>0 + bads3>0) ts4=0; (else) ts4=1 $
calc;list; totsampl=sum(ts4) $
REJECT; bads3>0 $
create; if (nb1>0) KO=1; (else)KO=0 $
calc; list; KB1=SUM(KO) $
sample; 1-1000 $
reject; ts4=0   $
calc; mreal=(xbr(rcs))*10;nmr=-1*mreal $
create; if (nexp>MREAL + nexp<nMR) K1=1; (else) K1=0 $
calc; list; sts=sum(ts4); kcs=sum(k1) $
OPEN; OUTPUT=tnegb56.OUT $
slow
calc; list; sts; KB1; KCS;int=.2692236;al=.6; mnb0=xbr(nb0);

     mnb1=xbr(nb1);mreal=xbr(rcs); mnexp=xbr(nexp);
     malp=xbr(alph); mtalp=xbr(talph); pbb0=mnb0/int;
     pbb1=mnb1/-.015;pbxp=mnexp/mreal $
create; nb0rmse1=(Int-nb0)^2; nb1rmse1=(-0.015-NB1)^2;
        nxrmse1=(Rcs-NEXP)^2 $

calc; list; nb0rmse=sqr((sum(nb0rmse1))/sts);
            nb1rmse=sqr((sum(nb1rmse1))/sts);
            nxrmse=sqr((sum(nxrmse1))/sts) $
create; if (nb0#0) nintall=nb0; (else) nintall=-999 $
create; if (nb1#0) ncostall=nb1; (else) ncostall=-999 $
calc; list;MINnB0=MIN(nINTALL); maxnb0=max(nintall);
            minnb1=min(ncostall); maxnb1=max( ncostall) $
create; pdevb01=((abs(nb0-int))/int)*100;
        pdevb11=((abs(nb1+.015))/.015)*100;
        pdevexp1=((abs(nexp-rcs))/rcs)*100 $
calc; list; pdevb0=XBR(pdevb01);
```

```
                pdevb1=XBR(pdevb11);
                pdevxp=XBR(pdevexp1);
                PFAC=1/(MREAL/13682.02);
                BPFAC=1/(INT/.269223) $
Create; yhstar=pfac*nexp; ystar=pfac*real;
                relexp1=(ystar-yhstar)^2;
                byhstar=bpfac*nb0; bystar=bpfac*int;
                relb01=(bystar-byhstar)^2 $
Calc; list; relexp=sqr((sum(relexp1))/sts);
                Relb0=sqr((sum(relb01))/sts) $
Sample; 1-1000 $
Reject; ts4=0 + nb1>0 $
Calc; list; nomoxp=xbr(nexp) $
```

Estimated Proportion of Non-users, RMSE and % Deviation

```
fast
read; nobs=40000; nvar=2; names=x, y1; file=datC161.dat $
read; nobs=1000; nvar=1; names=aa; file=alph56.dat $
matrix; ka=init(100,1,0); non1=INIT(100, 1, 0) $
calc; nolist; ran(5682643); J=1; n1=1; n2=400; chk2=1 $
label; 1 $
sample; j $
calc; v=sum(aa); vi=1/v; adj=1/(.5/.5) $
??? The first .5, above, is changed to 1 and 2 for E(λi)=1
and 2
sample; n1-n2 $
create;  vis=rnp(y1) $
calc; list; mv=xbr(vis); vv=var(vis); tk=mv^2/(vv-mv) $
reject; vis=0 $
create; unon3=(1+(vis/vi))^-vi;
create;
prob0=(1+((adj*vis)/vi))^-vi;tp1=(1+((adj*vis)/tk))^-tk $
calc; list; nonu=XBR(PROB0); tp2=xbr(tp1) $
matrix; NOLIST; a=iden(1);  ka1=tp2*a; non2=nonu*a;
                unon2=unon3*a $
matrix;NOLIST; non1(j,*)=non2 ; ka(j,*)=ka1;unon1(j,*)=unon2
$
calc; list; j=J+1;  n1=n1+400; n2=n2+400 $
goto; 1; j<=100 $
sample; 1-100 $
create;  adjnon=0; kb=0; unon=0 $
create; kb=dta(ka) $
create; adjnon= dta(non1) $
create; unon=dta(unon1) $
write; adjnon, kb, unon; file=non161.dat $
delete; x, y1 $


read;  nobs=100; nvar=3; names=non,k,unon; file=non161.dat $
append;nobs=100; nvar=3; names=non,k,unon; file=non162.dat $
append;nobs=100; nvar=3; names=non,k,unon; file=non163.dat $
append;nobs=100; nvar=3; names=non,k,unon; file=non164.dat $
append;nobs=100; nvar=3; names=non,k,unon; file=non165.dat $
```

```
append;nobs=100; nvar=3; names=non,k,unon; file=non166.dat $
append;nobs=100; nvar=3; names=non,k,unon; file=non167.dat $
append;nobs=100; nvar=3; names=non,k,unon; file=non168.dat $
append;nobs=100; nvar=3; names=non,k,unon; file=non169.dat $
append;nobs=100; nvar=3; names=non,k,unon;file=non1610.dat $
sample; 1-1000 $
OPEN; OUTPUT=PNONC56.OUT $
create; PNON1=(k-real)^2 $
calc; list; mnon=xbr(unon); NONRmse=sqr((sum(PNON1))/1000) $
CREATE; PDEV1=((ABS(k-real)/k))*100 $
calc; list; pdevNON=XBR(pdev1) $
```

Program to Show the Identity Between OLSALL and ZATC

```
???  Using matrix algebra to generate the estimates
eliminates ???  rounding error which would occur using
canned programs
fast
matrix; nbb0=init(1000, 1, 0);
        nbb1=init(1000, 1, 0) $
read; nobs=40000; nvar=4 ; names= ZONE, costi, costz, popz;
file=datc561.dat $
create; buser=0; new=0; newg=0; totsampl=popz/481.848;
        sqpop=sqr(totsampl) $
calc; nolist; ran(5682643); n1=1; n2=400; n3=1; n4=9; J=1  $
label; 2 $
sample; n1-n2 $
create; visits=rnp(Y1) $
namelist; bold=visits $
matrix; nolist; bmeans=gxbr(bold, zone); bgsize=gsiz(zone) $
create; new=dta(bmeans) $
create; newg=dta(bgsize) $
reject; visits=0 $
matrix; nolist; bgsize2=gsiz(zone) $
sample; n1-n2 $
CREATE; BUSER=DTA(BGSIZE2) $
create; totvis=new*newg;
bnonuser=totsampl-buser; bnucosts=bnonuser*costz;
bnutcsqu=costz^2*bnonuser;
busrcost=buser*costz; butcsqu=costz^2*buser;
bxy=totvis*costz $

calc; nolist;  btotvist=sum(visits);
btotn=sum(bnonuser)+sum(buser);
btottc=sum(bnucosts) + sum(busrcost);
btottcsq=sum(bnutcsqu) + sum(butcsqu); bsumxy=sum(bxy);
bxxinv11=1/(((btotn)*btottcsq) - btottc^2)*btottcsq;
bxxinv12=1/(((btotn)*btottcsq) - btottc^2)*-btottc;
bxxinv22=1/(((btotn)*btottcsq) - btottc^2)*btotn;
bb0=(bxxinv11*btotvist)+(bxxinv12*bsumxy);
bb1=(bxxinv12*btotvist) + (bxxinv22*bsumxy) $

sample; N3-N4 $
```

```
create; bin=(totvis)/sqpop; tc=costz*sqpop $
regress; lhs=bin; rhs=sqpop, tc $

matrix; a=iden(1); bb0a=bb0*a; bb1a=bb1*a; zb0=b(1);
        zb1=b(2) $
matrix; nolist;    olsb0(j, *)=bb0a;
                   olsb1(j, *)=bb1a;
                   zatcb0(j,*)=zb0;
                   zatcb1(j,*)=zb1  $
calc; list; j=j+1; n1=n1+400; n2=n2+400; n3=n3+400;
            n4=n4+400 $
goto; 2; j<=1000 $
sample; 1-1000 $
```

**Program to Generate Estimates From Data with Sampling Error
and a Variable Expansion Factor**

```
fast
read; nobs=40000; nvar=5; names=x, y1,ZONE,COST2,POPT;
file=datcZ561.dat $
DELETE; COST2 $
matrix; tnbb0=init(100, 1, 0); tnbb1=init(100, 1, 0);
tnbexp=init(100, 1, 0); tsnbalph=init(100, 1, 0);
tnbalph=init(100, 1, 0) $
create; nuse=0; tc=0 $
calc; nolist; ran(3580029); J=1;N1=1; N2=400 $
label; 1 $
sample; N1-N2 $
create; u1=0; nu1=0; visits=rnp(Y1) $
reject; visits=0 $
matrix; nolist; bgsize2=gsiz(zone) $
sample; 401-800 $
sample; N1-N2 $
create;  kt=rnu(433.6632, 530.0328) $
CREATE; U1=DTA(BGSIZE2) $
calc; list; su1=sum(u1); eu=su1+1 $
create; nu1=fix ((popt/kt)-u1) $
sample; n1-n2 $
reject; visits=0 $
create; nuse1=0; t1c=0; nuse=0; tc=0; if (visits>0)
nuse1=visits; t1c=x $
matrix; nuse2=copy(nuse1) ; t2c=copy(t1c) $
sample; 1-su1 $
create; nuse=dta(nuse2) $
create; tc= dta(t2c) $
sample; n1 $
calc;  n12=n1+1; nu2=sum(nu1); u2=sum(u1); tobs1=nu2+u2;
            eu2=eu+nu2-1 $
matrix;  nu31=init(nu2, 1, 0); tc1=init(tobs1, 1, 67.98) $
sample; eu-eu2 $
create; nuse=dta(nu31) $
create; tc=dta(tc1) $
sample; n12 $
calc; n12=n1+2; nu2=sum(nu1); u2=sum(u1); tobs1=nu2+u2;
```

```
                eu2=eu2+1; eu3=eu2+nu2-1 $
matrix;   nu32=init(nu2, 1, 0); tc2=init(tobs1, 1, 56.76) $
sample; eu2-eu3 $
create; nuse=dta(nu32) $
create; tc=dta(tc2) $
sample; n12 $
calc; n12=n1+3; nu2=sum(nu1); u2=sum(u1); tobs1=nu2+u2;
                eu2=eu3+1; eu3=eu2+nu2-1 $
matrix;   nu33=init(nu2, 1, 0); tc3=init(tobs1, 1, 81.62) $
sample; eu2-eu3 $
create; nuse=dta(nu33) $
create; tc=dta(tc3) $
sample; n12 $
calc; n12=n1+4; nu2=sum(nu1); u2=sum(u1); tobs1=nu2+u2;
                eu2=eu3+1; eu3=eu2+nu2-1 $
matrix;   nu34=init(nu2, 1, 0); tc4=init(tobs1, 1, 45.1) $
sample; eu2-eu3 $
create; nuse=dta(nu34) $
create; tc=dta(tc4) $
sample; n12 $
calc; n12=n1+5; nu2=sum(nu1); u2=sum(u1); tobs1=nu2+u2;
                eu2=eu3+1; eu3=eu2+nu2-1 $
matrix;   nu35=init(nu2, 1, 0); tc5=init(tobs1, 1, 64.9) $
sample; eu2-eu3 $
create; nuse=dta(nu35) $
create; tc=dta(tc5) $
sample; n12 $
calc; n12=n1+6; nu2=sum(nu1); u2=sum(u1); tobs1=nu2+u2;
                eu2=eu3+1; eu3=eu2+nu2-1 $
matrix;   nu36=init(nu2, 1, 0); tc6=init(tobs1, 1, 59.4) $
sample; eu2-eu3 $
create; nuse=dta(nu36) $
create; tc=dta(tc6) $
sample; n12 $
calc; n12=n1+7; nu2=sum(nu1); u2=sum(u1); tobs1=nu2+u2;
                eu2=eu3+1; eu3=eu2+nu2-1 $
matrix;   nu37=init(nu2, 1, 0); tc7=init(tobs1, 1, 59.4) $
sample; eu2-eu3 $
create; nuse=dta(nu37) $
create; tc=dta(tc7) $
sample; n12 $
 calc; n12=n1+8; nu2=sum(nu1); u2=sum(u1); tobs1=nu2+u2;
                eu2=eu3+1; eu3=eu2+nu2-1 $
matrix;   nu38=init(nu2, 1, 0); tc8=init(tobs1, 1, 78.76) $
sample; eu2-eu3 $
create; nuse=dta(nu38) $
create; tc=dta(tc8) $
sample; n12 $
calc; nu2=sum(nu1); u2=sum(u1); tobs1=nu2+u2;
                eu2=eu3+1; eu3=eu2+nu2-1 $
matrix;   nu39=init(nu2, 1, 0);  tc9=init(tobs1, 1, 2.2) $
sample; eu2-eu3 $
create; nuse=dta(nu39) $
create; tc=dta(tc9) $
```

```
sample; 1-eu3 $
poisson; lhs=nuse; rhs=one, tc; model=n;  par $
create;lambdah=exp(b(1)+b(2)*tc); nexpcs=-lambdah/b(2) $
calc; LIST; ntotexp=sum(nexpcs);
sealph=sqr(varalpha); ttalph=alpha/sealph $
matrix; NOLIST; a=iden(1); alp=a*alpha; ttalph1=a*ttalph;
        nbexp1=a*ntotexp; tnbb01=a*b(1); tnbb11=a*b(2) $
matrix;NOLIST; tnbb0(j,*)=tnbb01; tnbb1(j,*)=tnbb11;
tsnbalph(j,*)=ttalph1;
        tnbexp(j,*)=nbexp1; tnbalph(j,*)=alp $
calc; list; j=J+1;  chk2=1; n1=n1+400; n2=n2+400 $
goto; 1; j<=100 $
sample; 1-100 $
delete; zone, popt, x, y1, visits, lambdah,
nexpcs, nuse1,nuse,kt,u1,nu1,t1c $
create; nb0=0; nb1=0; nexp=0; alph=0; talph=0 $
create;nb0=dta(tnbb0) $
create; nb1=dta(tnbb1) $
create; nexp=dta(tnbexp) $
create; alph=dta(tnbalph) $
create; talph=dta(tsnbalph) $
write; nb0, nb1, nexp, alph, talph; file=ic569.dat$
delete;nb0, nb1, nexp, alph, talph $
```

APPENDIX C: Empirical Cost Data and Non-user
Observations for the Empirical Study

Table 6. Total and Non-user Observations by Zone from the
1989 Oregon Angler Survey Using the Expansion Factor 708.06
(Angler Population Adjusted for in the Expansion Factor)

ZONE	SITE USERS	TRAVEL COST	ORIGIN POPULATION	TOTAL OBSERVATIONS (ROUNDED)	NON-USERS
1	15	44.44	38855	55	40
2	9	18.00	11258	16	7
3	6	37.50	23686	33	27
4	7	24.22	22427	32	25
5	10	87.60	49456	70	60
6	6	53.88	18598	26	20
7	6	36.75	29715	42	36
8	5	28.80	43339	61	56
9	7	32.00	28743	41	34
10	10	64.00	11457	16	6
11	7	27.63	19843	28	21
12	11	56.33	100499	142	131
13	7	72.14	80731	114	107
14	5	62.00	80429	114	109
15	6	28.13	3540	6	0
16	10	41.00	15355	22	12
17	6	69.50	3807	6	0
18	6	37.45	6469	9	3
19	8	35.72	3707	8	0
20	7	14.20	16954	24	17
21	9	28.55	23650	33	24
22	9	44.28	36398	51	42
23	18	68.00	37991	54	36
24	16	37.12	40898	58	42
25	16	46.45	12956	18	2
26	6	14.79	3620	6	0
27	9	42.00	4087	9	0
28	20	37.92	27575	39	19
29	25	89.81	32417	46	21
30	5	183.00	16515	23	18
31	27	40.82	42913	61	34
32	11	21.18	11190	16	5
33	6	18.41	6708	9	3
34	17	37.50	23259	33	16
35	9	50.75	7919	11	2
36	10	21.41	13064	18	8

Table 6--continued

ZONE	SITE USERS	TRAVEL COST	ORIGIN POPULATION	TOTAL OBSERVATIONS (ROUNDED)	NON-USERS
37	6	107.83	10150	14	8
38	5	52.00	5686	8	3
39	6	37.83	7302	10	4
40	8	28.06	28202	40	32
41	12	33.83	49228	70	58
42	21	51.51	35735	50	29
43	14	22.66	48485	68	54
44	13	40.84	11725	17	4
45	10	20.68	17514	25	15
46	9	35.44	87604	124	115
47	9	36.94	93687	132	123
48	7	34.76	51835	73	66
49	9	43.46	72686	103	94
50	7	91.64	82075	116	109
51	15	38.34	130213	184	169
52	8	93.72	83516	118	110
53	7	41.90	4826	7	0
54	7	20.69	14401	20	13
55	5	55.00	14987	21	16
56	6	51.46	18129	26	20
57	17	91.46	69327	98	81
58	5	93.00	52591	74	69
59	8	36.25	35656	50	42
60	9	38.56	46811	66	57
61	6	18.00	3849	6	0
62	7	37.37	24084	34	27
63	6	26.16	18092	26	20
64	10	43.38	21890	31	21
65	12	48.05	18959	27	15
Total	631	----	2,109,273	2,988	2,357