WESTERN AGRICULTURAL ECONOMICS ASSOCIATION

# PAPERS OF THE
# 1989 ANNUAL MEETING

# WESTERN AGRICULTURAL ECONOMICS ASSOCIATION



COEUR D'ALENE, IDAHO

JULY 9-12, 1989

# THEORETICAL AND EMPIRICAL ADVANTAGES OF TRUNCATED COUNT DATA ESTIMATORS FOR ANALYSIS OF DEER HUNTING IN CALIFORNIA

Michael D. Creel and John B. Loomis

## Introduction

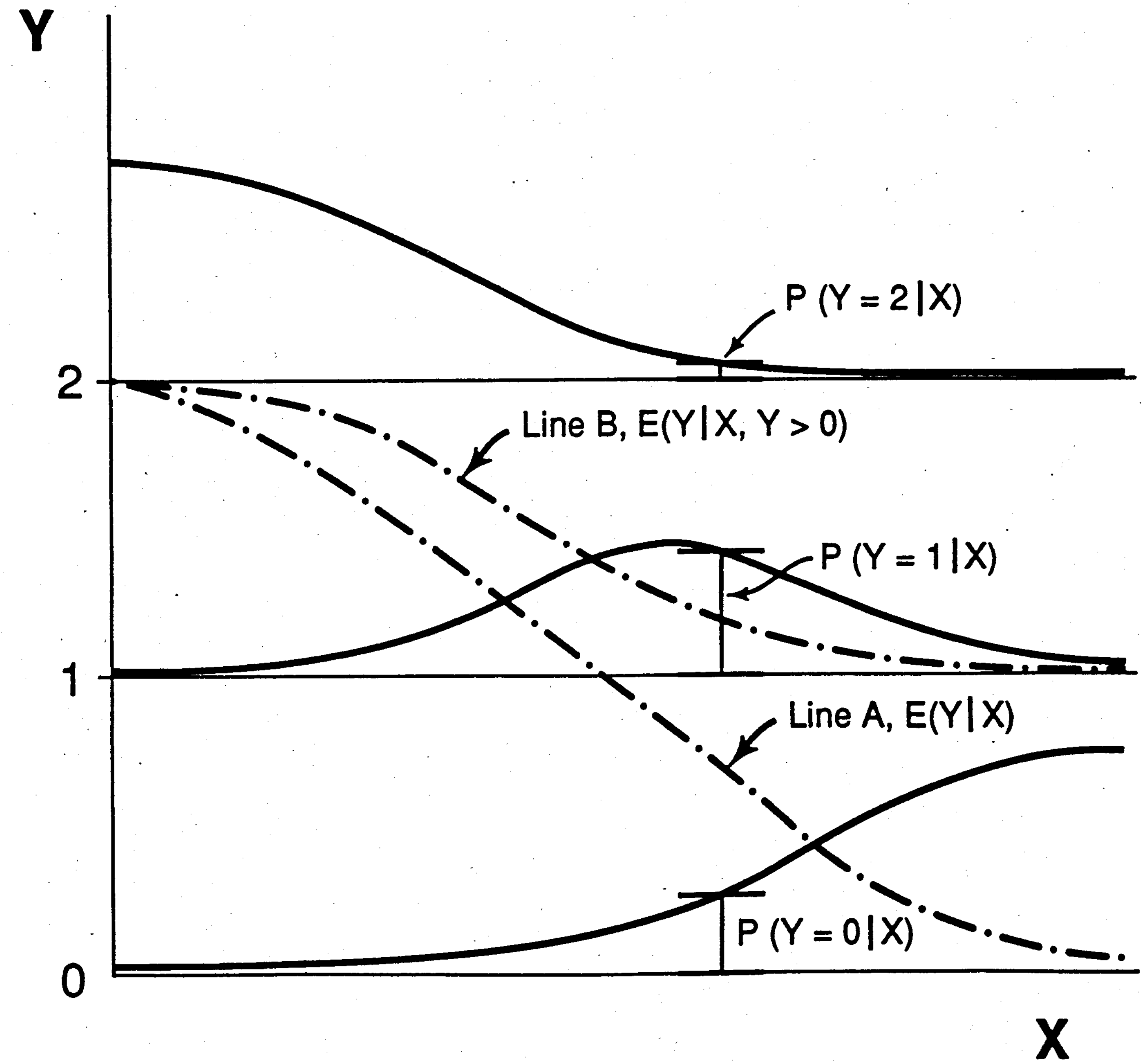
Surveys of users visiting a site are often employed to collect data on recreational demand. Given such a sampling method, no data will be collected for individuals taking zero trips to a site. The sample will therefore be truncated at the zero trip level. A second feature of recreation studies is that the dependent variable is often the count of the number of trips taken over the season or year. As such, it will be a nonnegative integer. The observed dependent variable is therefore the outcome of a data generating process (DGP) based on some unknown probability distribution function defined on the nonnegative integers, which may be termed a count data process. The combination of a truncated sample from a count data process suggests that estimators based on truncated count data distributions may be called for.

Figure 1 serves to illustrate the bias encountered in using an estimator uncorrected for sample truncation as well as to show what a count data process might look like. Let $Y = f(X,\beta,\epsilon)$, and let Y be distributed by some known count data process such that Y can take on the values $i \subset \{0,1,2\}$. Assume X is a matrix of nonstochastic independent variables, $\beta$ is a parameter vector, and $\epsilon$ is a random disturbance. Let the heights the of curved lines above the lines Y = i represent $p(Y=i|X,\beta)$. The heights of these lines sum to one in the vertical dimension. Line A represents $E(Y|X,\beta)$ and line B represents $E(Y|X,\beta,Y>0)$. Let a sample be taken from the population such that Y is observed only if Y > 0. The parameter estimates from a maximum likelihood estimator based on the known distribution of Y that is uncorrected for the sample truncation will be biased and inconsistent since the regression will be approximating line B rather than line A. In the context of modeling demand, if Y is quantity, and X is price, the uncorrected estimator's parameter estimates will most likely overstate Marshallian consumers' surplus since line B everywhere lies outside of line A. If one wishes to estimate potential social benefits after changes in exogenous variables occur rather than estimate the benefits already received by the persons in the sample, line A should be the focus of interest. Shaw (1988b) presents a similar figure for the case of a normally distributed random variable. Truncation bias is further discussed in general by Maddala, and in the recreational demand literature by Smith *et. al.*; Smith; and McConnell and Kling.

Count data estimators may better fit data from a count data process than would a continuous distribution-based estimator since count data estimators restrict positive probability assignment to possible events while continuous distribution estimators give positive probability to fractional and possibly negative values of the dependent variable. The importance of respecting the count data nature of the dependent variable may depend on the problem at hand. As Larson shows, the normal distibution is a good approximation of the Poisson distribution if the Poisson parameter $\lambda$, which is usually parameterized such that $\lambda_i$ is the conditional mean of the dependent variable $Y_i$, is large. Thus, if the DGP is truly a Poisson process, normal MLE may be a suitable procedure if the conditional mean of the dependent variable is "large", but it may not give acceptable results if the conditional mean of $Y_i$ is small. One should be cautious in using normal MLE to model a count data process for which small values of the dependent variable are common. A given count data distribution <u>may</u> be the true distribution underlying the DGP, but any continuous distribution is known *a priori* to be incorrect.

University of California

Figure 1. Truncation bias for a known count data process.



Poisson and negative binomial count data models have been used in numerous recent studies, including Hausman, Hall and Griliches' work on patents issued and Cameron and Trivedi's study of doctor visits. Smith employed a Poisson travel cost model to estimate demand for water-based recreation trips. Grogger and Carson (1987, 1988) introduced truncated Poisson and truncated negative binomial count data estimators. In their papers they present results of using standard untruncated count data estimators as well as their truncated estimators to estimate demand for fishing trips in Alaska. Finally, Shaw (1988a) presented a truncated Poisson estimator appropriate for endogenously stratified samples, and conducted some Monte Carlo simulations.

In this paper we present results of employing Poisson (POIS), truncated Poisson (TPOIS), negative binomial (NB), and truncated negative binomial (TNB) maximum likelihood estimation, as well as ordinary least squares (OLS), nonlinear normal (NLS), and truncated nonlinear normal (TNLS) MLE to estimate a travel cost method demand curve for deer hunting in California. A large sample size (N=2223) allowed us to randomly segment the sample into specification, estimation, and prediction portions. In the following sections we will: i) present versions of Grogger and Carson's estimators for the case of truncation at the zero level and review their properties; ii) describe the specification-estimation-prediction methodology and present the estimation results; iii) compare the out-of-sample predictive performances of the models; and iv) give point estimates of consumers' surplus per trip for the various statistical models.

**The Estimators**

The Poisson probability law is:

$$[1] \qquad f(W=w) = \frac{\exp(-\lambda)\lambda^{w}}{w!}$$

which is a discrete density function defined for w an element of the set of nonnegative integers. The parameter $\lambda>0$ is both the mean and variance of W. If we assume elements of an Nx1 (where N is the sample size) dependent variable vector Y are distributed independent Poisson($\lambda$) and we allow $\lambda$ to vary by observation as a function of an NxK matrix of explanatory variables X and a Kx1 parameter vector $\beta$, as in:

$$[2] \qquad \lambda_i = \exp(X_i\beta)$$

or for the whole sample, in matrix notation:

$$[2.1] \qquad \lambda = \exp(X\beta)$$

we may obtain the standard Poisson likelihood function:

$$[3] \qquad \ln L = -s'\lambda + Y'X\beta - s'\ln[Y!]$$

where s is an Nx1 sum vector and the logarithmic and factorial functions are element-by-element.

If we note that $F(W=0) = f(W=0) = \exp(-\lambda)$, where $F(\bullet)$ indicates the cumulative density function (CDF) and $f(\bullet)$ indicates the probability density function (PDF) we may express the conditional Poisson density function as:

$$[4] \qquad f(W=w|W>0) = \frac{\exp(-\lambda)\lambda^{w}}{w!}\left[\frac{1}{1-\exp(-\lambda)}\right]$$

Maintaining the the above notation and allowing $\lambda$ to vary as in [2], the zero-level truncated Poisson likelihood function may thus be written as:

$$[5] \qquad \ln L = -s'\lambda + Y'X\beta - s'\ln[s-\exp(-\lambda)] - s'\ln[Y!]$$

A characteristic of the standard Poisson model is that the conditional mean of the dependent variable, $\lambda$, is equal to the conditional variance, *i.e.* the variance-mean ratio is unity. If the population exhibits "overdispersion" *i.e.*, the conditional variance is greater than the conditional mean giving a variance-mean ratio greater than unity, and the conditional mean is correctly specified as the true mean of the the DGP, possibly as in [2], the untruncated Poisson model will give consistent estimates of the parameters, but downwardly biased estimates of their standard errors (Gourieroux, Monfort and Trognon, 1984b, pg. 707; Grogger and Carson, 1987, pg. 4). The truncated Poisson estimator is biased and inconsistent in the presence of overdispersion, since it is based on the higher moments of the Poisson distribution through the CDF term in [4], which are incorrect given that the DGP embodies overdispersion (Grogger and Carson, 1988, pg. 7). Analogously, the standard Poisson estimator is biased and inconsistent when applied to a truncated sample since the conditional mean is misspecified as was illustrated in Figure 1. The fact that both estimators are inconsistent if the sample is truncated and overdispersed makes the truncated negative binomial estimator an attractive generalization if these conditions are present.

The negative binomial probability law may be written as

$$[6] \qquad f(Z=y) = \frac{\Gamma(z+1/\alpha)}{\Gamma(z+1)\Gamma(1/\alpha)} (\alpha\lambda)^{z}(1+\alpha\lambda)^{-(z+1/\alpha)}$$

where $\Gamma(\bullet)$ indicates the gamma function. This is a discrete pdf defined for z an element of the set of nonnegative integers, with parameters $\alpha>0$ and $\lambda>0$. The mean of Z is $\lambda$ and its variance is $\lambda+\alpha\lambda^2$, which is everywhere greater than the mean. The variance-mean ratio is $1+\alpha\lambda$, so the degree of overdispersion is an increasing function of both $\alpha$ and $\lambda$. This is the same distribution as found in Lee's equation 4.9 (Lee, pg. 698) where our $\alpha$ is equivalent to Lee's $\delta/(1-\delta)$. It may be derived by letting the Poisson parameter be distributed as a gamma random variable with mean $\lambda$ and variance $\alpha\lambda^2$. As $\alpha\to0$ the gamma distribution becomes degenerate and the negative binomial distribution reduces to a Poisson distribution. If we assume that the elements of the Nx1 dependent variable vector Y are distributed independent negative binomial($\lambda$,$\alpha$) and we let again let $\lambda$ vary by observation as in [2] and estimate $\alpha$ as a constant for the population controlling the level of overdispersion we obtain one of the possible negative binomial statistical models with likelihood function given by:

$$[7] \quad \ln L = s'\ln[\Gamma(Y+s/\alpha)] - s'\ln[\Gamma(Y+s)] - N\ln[\Gamma(1/\alpha)] + \ln(\alpha)s'Y + Y'X\beta - (Y+s/\alpha)'\ln[s+\alpha\lambda]$$

Note that for this density function:

$$[8] \qquad F(Y=0) = f(Y=0) = (1+\alpha\lambda)^{-(y+1/\alpha)}$$

We may use Bayes' theorem to write the conditional negative binomial density function:

$$[9] \quad f(Z=y\,|\,Z>0) = \frac{\Gamma(z+1/\alpha)}{\Gamma(z+1)\Gamma(1/\alpha)} (\alpha\lambda)^{z}(1+\alpha\lambda)^{-(z+1/\alpha)} \left[\frac{1}{1-(1+\alpha\lambda)^{-1/\alpha}}\right]$$

Using the same notation as above and again allowing $\lambda$ to vary as in [2], the zero-level truncated negative binomial likelihood function may be written as:

$$[10] \quad \ln L = s'\ln[\Gamma(Y+s/\alpha)] - s'\ln[\Gamma(Y+s)] - N\ln[\Gamma(1/\alpha)] + \ln(\alpha)s'Y + Y'X\beta - (Y+s/\alpha)'\ln[s+\alpha\lambda] - s'\ln\left[s-(s+\alpha\lambda)^{-1/\alpha}\right]$$

where all operations other than matrix products are element-by-element.

Under a set of regularity conditions (Gourieroux, Monfort, and Trognon, 1984a, appendix), the truncated Poisson and truncated negative binomial estimators will be consistent provided their conditional means are specified as underlies the true DGP (see Grogger and Carson, 1987, appendix). If the dependent variable exhibits overdispersion the truncated Poisson estimator will therefore be inconsistent, since its mean depends on the incorrect higher moments of the Poisson distribution, while the truncated negative binomial estimator will be consistent if the DGP is truly the above truncated negative binomial process and [2] is a correct specification of the population mean. Larson (p. 171, 188) provides some diagrams which aid in forming an intuitive understanding of the differences between the count data models. Next, we turn to an application of these estimators as well as estimators based on the normal distribution to data of deer hunting in California.

**Specification and Estimation**

The data for this study were collected by a mail survey of California deer hunters that purchased a deer hunting license in 1987. It being a mail survey, there is no endogenous stratification (*i.e.*, frequent visitors are no more likely to be sampled than infrequent visitors), so Grogger and Carson's estimators are appropriate rather than Shaw's truncated Poisson estimator. Our focus here is on hunters that took trips to one of 17 X zones, one of the four types of hunting zones in California. X zones are located in northeastern California, for the most part. Due to regulations, there is little scope for hunting in more than one zone during a season, and no such possibility if a hunter wishes to visit an X zone. We view the decision-making process as having two stages; first a potential hunter (all potential hunters defining the population) decides which hunting zone is preferred; then decides how many trips to take to this zone. The unobserved zeros are potential hunters which preferred to hunt in an X zone, but decided not to purchase a license, or having purchased a license, failed to take any trips. In this paper, we are considering only the second stage of the process. Our treating the 17 X zones as one destination may be thought of as imposing untested (but testable) coefficient restrictions across 17 separate models. This was done to facilitate comparing the statistical models, and admittedly may not result in the best estimates for a particular X zone. One may note that a multi-destination model would be inappropriate for this problem, where there is no possibility of visiting other zones.

The general specification of the travel cost model was:

[11] $$Y_i = f(\text{PRICES}_i, \text{SITE QUALILITY}_i, \text{INDIVIDUAL CHARACTERISTICS}_i, \phi, \epsilon_i)$$

or in general matrix notation:

[12] $$Y = f(X, \phi, \epsilon),$$

where X is the matrix of independent variables, $\phi$ is a vector of parameters, and $\epsilon$ is a vector of independent random disturbances.

The dependent variable, Y, the number of trips taken to an X zone, is truncated at the zero level, since individual data records were complete only if at least one trip was taken. A candidate set of twenty explanatory variables was identified based on the above general specification; 2223 observations were complete for this set. The candidate set did not include travel costs to other sites since individuals are restricted to hunt in the zone for which they purchased a license prior to the beginning of the season. These observations were randomly divided into three portions: a specification portion; an estimation portion; and a prediction portion, of 707, 764, and 752 observations respectively. The sample mean of the dependent variable, Y, was 2.76 for the estimation portion of the data. This low mean suggests that the normally distributed specifications may provide a poor approximation to the true DGP.

The statistical models fitted were:

OLS: $Y \sim N(X\beta, \sigma^2 I)$

NLS: $Y \sim N(\mu = \exp(X\beta), \sigma^2 I)$

TNLS: $Y \sim N(\mu = \exp(X\beta), \sigma^2 I)$, Y observed only if $Y > 0.5$

POIS: $Y \sim Pois(\lambda = \exp(X\beta))$

TPOIS: $Y \sim Pois(\lambda = \exp(X\beta))$, Y observed only if $Y > 0$

NB: $Y \sim NB(\lambda = \exp(X\beta), \alpha)$

TNB: $Y \sim NB(\lambda = \exp(X\beta), \alpha)$, Y observed only if $Y > 0$

The nonlinear truncated normal model is a standard lower truncated normal model as discussed in Maddala, with the exception of the nonlinear specification of the mean. The 0.5 lower truncation limit for the TNLS model was chosen as a continuity correction between the limits of 0 or 1, which, as Larson (pgs. 296 and 299) discusses, should allow the normal distribution to better approximate the unknown true count data DGP. An attempt was made to fit a linear ($\mu = X\beta$) truncated normal model, but we were not able to achieve convergence. The count data models were as discussed above. Estimation programs written in the *Gauss* language are available from the authors.

Using the specification data, various combinations of explanatory variables and data transformations were used to fit all the statistical models. A semilog form (for the continuous distribution estimators, OLS, NLS and TNLS) gave a worse fit in all cases. Logarithms of continuous RHS variables gave a slightly better fit in some cases, but were not used due to the arbitrariness of the resulting social benefit estimates. Quadratic forms of some variables (*e.g.*, AGE, and INCOME) did not improve the results.

Ten independent variables including a constant were selected, based on theory and overall performance for each of the seven different statistical models. Some variables which were not significant for all of the statistical models were left in the final variable set to avoid omitted variable bias, as in the case of travel time. It would be possible to improve the fit of any of the models somewhat by selecting a model-specific set of variables, but in the interest of comparability a uniform set was chosen.

These variables were:

-TC, round trip travel cost at $0.22/mile, the sample average.

-TIME, round trip travel time in hours. Experimentation (using the specification data) with wage-weighting of travel time did not fit as well as time alone.

-DAYS, the individual's season average length of trip.

-YEARS, the number of years the individual had previously hunted at the zone.

-BAG, a zero one dummy, equal to 1 if a deer was taken in the 1987 season.

-PASNO, the number of times the individual passed up an opportunity to take a deer.

-DEERSN, the number of deer seen on the last trip of the season.

-INCOME, household income in thousands of dollars.

-SEASON, zone season length in days.

Next, the above set of variables arrived at using the first portion (N=707) of the data was used to estimate all the statistical models, using the estimation portion (N=764) of the data. This was a one-time, first-try estimation for this data, therefore the results are free from any form of pre-test bias. As such, the reported t-statistics and goodness of fit measures are unbiased by the specification search and are a fair means of comparing the statistical models' performance, insofar as the selection of included variables was fair. Estimation results are found in Table 1.

All of the coefficient estimates are of the expected sign for all of the statistical models (a negative sign on the income coefficient is often encountered with travel cost models, see Duffield; Mendelsohn; or Grogger and Carson (1987) for instance). The truncated models are generally more elastic in all variables, which is expected, as in Figure 1. In general, POIS is in closer agreement to NB than it is to TPOIS, and TPOIS is closer to TNB than it is to POIS. This suggests that the effect of truncation is more important than the effect of overdispersion. Given that the t-statistics for alpha, which reflects the level of overdispersion (see above), from NB and TNB are so large, we may reject the hypothesis of no overdispersion, which implies that the t-statistics of both POIS and TPOIS are biased away from zero, which may account for why they are almost always larger in magnitude than the corresponding t-statistics of the negative binomial models. Since the sample exhibits truncation and overdispersion the only potentially unbiased count data model is the TNB. Based on this fact and a good fit and t-statistics, the TNB statistical model seems superior to the alternative count data models. The OLS model is clearly inferior to the alternatives, while NLS and TNLS give results comparable to the count data models.

**Prediction**

The third step of the study was to use the coefficients estimated for the various statistical models using the estimation portion of the data to predict the number of trips taken in the prediction portion of the data (N=752), conditional on the prediction portion's independent variables. This allows us to assess the robustness of the estimators and their relative usefulness in estimating consumers' benefits. Predictive ability is measured by $R^2$, by the difference between total predicted and total actual trips, and by percentage error. The prediction results are found in Table 2.

Based on out-of-sample $R^2$, the count data models predict substantially better than do OLS and TNLS. TNLS performs strikingly poorly out-of-sample, achieving an $\hat{R}$ of only 0.023 and overpredicting trips by 40.9%. NLS predicts about as well as do the count data models. For all the statistical models, the difference between in-sample and out-of-sample $R^2$ is somewhat greater for the truncated models than for the corresponding untruncated models. In terms of predicting total trips taken, the POIS model is clearly superior, and TNLS is clearly inferior.

**Benefit Estimates**

Point estimates of consumers' surplus per predicted trip $(\hat{C}/\hat{Y})$ are found in Table 3. For OLS the formula is $\hat{C}/\hat{Y} = -\hat{Y}/(2\hat{\beta}_{TC})$, where $\hat{Y}$ is the predicted value of Y calculated at the means of the independent variables, and for all the other models the formula is $\hat{C}/\hat{Y} = -1/\hat{\beta}_{TC}$. The truncated models for which $\hat{Y}=E(Y|X)$ rather than $\hat{Y}=E(Y|X,Y>0)$ is appropriate, as was argued in the introduction, give a lower estimate than do the corresponding untruncated models, as is expected from Figure 1. The truncated models attempt to fit a line analogous to Line A of Figure 1, while the untruncated models attempt to fit a line analogous to line B. This effect is most apparent in the NLS-TNLS pair. For the count data models, allowing for overdispersion has little effect on the estimate, as is seen by comparing the pairs POIS-NB and TPOIS-TNB.

**Conclusions**

Several results are indicated by this study. First, accounting for truncation of the dependent variable makes a substantial difference in the coefficient estimates, and subsequently, in benefit estimates, regardless of the choice of statistical model.

TABLE 1
ESTIMATION RESULTS
(t-statistics below estimated coefficient)

| | OLS | NLS | TNLS | POIS | TPOIS | NB | TNB |
|---|---|---|---|---|---|---|---|
| ONE | 4.674 | 1.827 | 2.190 | 1.560 | 1.603 | 1.514 | 1.332 |
| | 6.38 | 7.92 | 5.67 | 9.75 | 8.11 | 7.28 | 3.68 |
| TC | -0.0118 | -0.00578 | -0.0272 | -0.0065 | -0.0134 | -0.0061 | -0.0143 |
| | -3.43 | -2.19 | -3.96 | -6.40 | -8.37 | -5.66 | -8.10 |
| TIME | -0.0517 | -0.0977 | -0.175 | -0.0245 | -0.0326 | -0.0174 | -0.0169 |
| | -1.45 | -3.56 | -2.79 | -2.38 | -2.11 | -1.61 | -1.05 |
| DAYS | -0.154 | -0.0547 | -0.057 | -0.0442 | -0.0481 | -0.0385 | -0.0495 |
| | -4.47 | -3.95 | -3.12 | -6.00 | -5.70 | -4.31 | -3.56 |
| YEARS | 0.0374 | 0.0122 | 0.0186 | 0.00991 | 0.011 | 0.00902 | 0.0106 |
| | 3.58 | 5.46 | 6.32 | 5.19 | 5.26 | 3.50 | 2.56 |
| BAG | -0.252 | -0.104 | -0.189 | -0.0783 | -0.104 | -0.0747 | -0.147 |
| | -1.05 | -1.67 | -2.2 | -1.59 | -1.84 | -1.15 | -1.36 |
| PASNO | 0.226 | 0.0267 | 0.0326 | 0.0335 | 0.03 | 0.0397 | 0.0622 |
| | 3.77 | 4.11 | 3.81 | 4.56 | 3.96 | 2.97 | 2.36 |
| DEERSN | 0.0012 | 0.000962 | 0.001 | 0.000644 | 0.000776 | 0.000511 | 0.000655 |
| | 1.81 | 4.05 | 3.38 | 2.96 | 3.11 | 1.74 | 1.31 |
| INCOME | -0.0187 | -0.00777 | -0.0144 | -0.00626 | -0.00814 | -0.00614 | -0.0108 |
| | -3.79 | -4.43 | -4.67 | -5.37 | -5.54 | -4.17 | -4.11 |
| SEASON | 0.0299 | 0.0237 | 0.0258 | 0.0214 | 0.0332 | 0.0183 | 0.0364 |
| | 0.71 | 1.74 | 1.10 | 2.28 | 2.84 | 1.50 | 1.72 |
| ALPHA | N/A | N/A | N/A | N/A | N/A | 0.216 | 0.754 |
| | | | | | | 8.51 | 5.19 |
| SIGMA | 2.9 | 2.59 | 2.83 | N/A | N/A | N/A | N/A |
| R-SQUARED | 0.242 | 0.389 | 0.341 | 0.354 | 0.403 | 0.332 | 0.326 |
| Log-L | -1892.9 | -1810.4 | -1523.5 | -1539.1 | -1295.8 | -1443.2 | -1145.6 |

TABLE 2 OUT-OF-SAMPLE PREDICTION RESULTS
(total actual trips = 1845)

| | OLS | NLS | TNLS | POIS | TPOIS | NB | TNB |
|---|---|---|---|---|---|---|---|
| R-SQUARED | 0.233 | 0.297 | 0.027 | 0.346 | 0.334 | 0.328 | 0.301 |
| ACT-PRED TRIPS | -121.9 | 50.5 | -775.0 | 16.0 | -132.8 | -99.1 | -74.1 |
| (ACT-PRED)/ACT | -6.6% | 2.7% | -40.9% | 0.9% | -7.2% | -5.4% | -4.0% |

TABLE 3 ESTIMATED CONSUMERS' SURPLUS PER PREDICTED TRIP
(estimation data)

| | OLS | NLS | TNLS | POIS | TPOIS | NB | TNB |
|---|---|---|---|---|---|---|---|
| CS/PRED TRIPS | $117.25 | $172.82 | $36.72 | $153.62 | $74.71 | $163.05 | $70.07 |

Second, the TNLS model is not robust in the sense that it does not perform well out-of-sample, and it would be a poor choice to predict visitation. The conditional mean of this model is

$$[9] \qquad E[Y|X,Y>0] = \exp(X\beta) + \sigma\phi[(X\beta-.5)/\sigma]\Phi[(X\beta-.5)/\sigma]^{-1}$$

where $\phi[\bullet]$ and $\Phi[\bullet]$ are the density and distribution functions of a standard normal random variable, respectively. If the assumption of homoscedasticity is not justified the parameter estimates of this model would be biased through the misspecified distribution function term in the likelihood function. If the DGP is characterized by the conditional variance increasing with the conditional mean, as is embodied in the count data models, the TNLS model will overcompensate for truncation at low levels of the dependent variable, since $\sigma$ will be too large and $\Phi[\bullet]$ will be too small. This overcompensation would serve to make the term $\exp(X\beta)$ in [9] small, which would tend to bias the price coefficient downward. This is reflected in the low consumers' surplus per trip estimate from this model. This bias may be sensitive to the number of outlying (large) observations of the dependent variable, which strongly affect the estimate of $\sigma$, which would explain the lack of robustness exhibited by this model's spectacular overprediction of trips for the out-of-sample data.

A third result is that the data do appear to exhibit overdispersion. We are able to reject the hypothesis of no overdispersion (H0: $\alpha = 0$) with a high degree of confidence using either the NB or TNB models. This leads us to regard the relatively large t-statistics of the POIS and TPOIS models with suspicion, (recalling that these models' standard errors are biased downward if there is overdispersion) as well as to suspect bias in the TPOIS model due to the incorrect higher moments. We suspect significant bias in the coefficient estimates of all the untruncated models due to the truncation. Given these results, and given the poor performance of the TNLS model out-of-sample, we believe the TNB model is the best suited of the models studied to estimating demand and social benefits for this data.

We hope that this paper has left the reader with two impressions. First, that truncated count data models present a useful and perhaps better way to analyze a broad class of problems not limited to recreational demand. Second, we believe that the specification-estimation-prediction methodology should be more widespread. Elimination of specification bias and "clean" goodness-of-fit and t-statistics are necessary to evaluate a model, and are results of the procedure. For smaller samples, the prediction step might be skipped (perhaps at the cost of failing to identify poor models such as our TNLS model). The use of an entire large sample to both specify and estimate a model is an inefficient use of data in that bias is introduced to what are likely satisfactorily efficient estimates.

The above comparison of statistical models has been fairly heuristic. In the future we plan to conduct non-nested testing of the models against each other. Also, there is possibly simultaneity between the Q and some of the regressors. The residuals of the NB and TNB models should be examined to test for the assumed form of overdispersion, and possibly other forms of NB (and implied TNB) models as discussed by Hausman *et. al.* and Lee should be explored. Compensating and equivalent variations should be calculated, as well as should confidence intervals for benefit estimates. While these research areas remain to be addressed, we believe this paper gives sufficient evidence to warrant further work with truncated count data models.

## References


Cameron, A. Colin and Pravin K. Trivedi, "Econometric Models Based on Count Data: Comparison and Application of Some Estimators and Tests," *Journal of Applied Econometrics*, 1(1986): 29-53.

Duffield, John, "Travel Cost and Contingent Valuation: A Comparative Analysis," in *Advances in Applied Micro-Economics*, Vol. 3, V. Kerry Smith and Ann Dryden Witte (eds.), Jai Press, Greenwich, 1984: 67-87.

Gourieroux, C., A. Monfort and A. Trognon, "Pseudo Maximum Likelihood Methods: Theory," *Econometrica*, Vol. 52, No. 3 (May 1984): 681-700.

Gourieroux, C., A. Monfort and A. Trognon, "Pseudo Maximum Likelihood Methods: Applications to Poisson Models," *Econometrica*, Vol. 52, No. 3 (May 1984): 701-720.

Grogger, Jeffrey T. and Richard T. Carson, "Models for Counts From Choice Based Samples," University of California, San Diego, Dec. 1987.

Grogger, Jeffrey T. and Richard T. Carson, "Truncated Counts," University of California, San Diego, July 1988.

Hausman, Jerry, Bronwyn H. Hall and Zvi Griliches, "Econometric Models for Count Data with an Application to the Patents-R&D Relationship," *Econometrica*, Vol. 52, No. 4 (July 1984): 909-938.

Larson, Harold J., *Introduction to Probability Theory and Statistical Inference*, Third Edition, John Wiley & Sons, 1982.

Lee, Lung-Fei, "Specification Test for Poisson Regression Models," *International Economic Review*, Vol. 27, No. 3 (Oct. 1986): 689-706.

McConnell, Kenneth E. and Catherine Kling, "Aggregation Issues: The Choice Among Estimation Approaches," in Benefit Analysis Using Indirect or Imputed Market Methods, Vol. II, Nancy E. Bockstael, W. Michael Hanemann and Ivar E. Strand (eds.), United States Environmental Protection Agency: 35-52.

Maddala, G.S., *Limited Dependent and Qualitative Variables in Econometrics*, Cambridge University Press, New York, 1983.

Mendelsohn, Robert, "An Application of the Hedonic Travel Cost Framework for Recreation Modeling to the Valuation of Deer," in *Advances in Applied Micro-Economics*, Vol. 3, V. Kerry Smith and Ann Dryden Witte (eds.), Jai Press, Greenwich, 1984: 89-101.

Shaw, Daigee, "On-Site Samples' Regression: Problems of Non-Negative Integers, Truncation, and Endogenous Stratification," *Journal of Econometrics*, 37 (1988a): 211-223.

Shaw, Daigee, "Theoretical and Empirical Specifications Issues in Travel Cost Demand Studies: A Comment," (unpublished paper) The Institute of Economics, Academia Sinica, Nankang, Taipei, Taiwan 1988b.

Smith, V. Kerry, "Selection and Recreational Demand", *American Journal of Agricultural Economics*, Vol. 70, No. 1: 29-36.

Smith, V. Kerry, William H. Desvousges and Ann Fisher, "A Comparison of Direct and Indirect Methods for Estimating Environmental Benefits," *American Journal of Agricultural Economics*, Vol. 68, No. 2: 280-290.