A Distribution Transition Method for Extreme Responses in Recreation Survey Data

Ashley S. Barfield and J. Scott Shonkwiler

Department of Agricultural and Applied Economics

University of Georgia

adro88@uga.edu

Selected Paper prepared for presentation at the 2016 Agricultural & Applied Economics Association Annual Meeting, Boston, Massachusetts, July 31-August 2

Abstract:

Revealed preference methods require survey data on past resource use, and numerous studies have found reported recreation frequency to be overestimated and concentrated on prototype (rounded and calendar-based) values. This paper develops an approach to treat extreme values and rounded responses in survey datasets and thereby improve model fit and resulting welfare estimates. We illustrate how, when modeling single-site trip data, model fit can be improved by transitioning from a discrete to a continuous distribution at a cut-point where response behavior begins to exhibit rounding. We feel this method will be useful for recreation demand research and may have broad applicability to the general analysis of count data.

Introduction

The estimation of economic values for environmental amenities is critical to making informed resource management decisions and accurately assessing environmental damages. Revealed preference methods for measuring economic values require data on observed behavior. Typically such behavior is elicited by surveys of target populations regarding their past use of a natural resource. Since recall periods are generally over a past season or year, the issue of recall accuracy needs to be considered. Systematic biases in reporting past behavior may compromise the methods used to derive values from revealed preference data.

Numerous studies have found that reported recreation frequency has been overestimated. For instance, Connelly and Brown (1995) found that reported angling trips on Lake Ontario were over-estimated by roughly 44% as compared with diary data, with recall bias increasing with user avidity. Hoehn et al. (1996) similarly found recall bias to be associated with respondents' over-statement of Michigan angling trips. Explanations for this bias are concerned mainly with the saliency of the resource, the respondent's strategic behavior (real or imagined), and the respondent's self-delusion (or effort to impress the surveyor) if the activity can be considered glamourous or healthy. Another dimension of recall bias is respondents' tendency to round off responses to end in a zero or five. Tourangeau et al. (2000) report that open-ended questions which require a numerical response may manifest these characteristics: i) the larger the number to be reported the more likely it will be a round value; ii) the distances between successive rounded values increase as the numbers increase.[1]

[1] Tourangeau et al. (2000) further clarify that if respondents round fairly (i.e., if they always round their responses to the nearest round value), due to the uneven spacing of round values, the net effect of rounding will actually be a downward bias in the data. However, if respondents are not rounding fairly but are characteristically rounding up or down, systematic error is introduced into the model in the direction of the rounding. The evidence in the recreational survey response literature generally finds that respondents do not round fairly – they overstate their participation.

To illustrate the patterns observed in recreation survey data, consider the following table of reported trips in three different recreation demand studies:

Table 1. Trips reported in recreation demand studies.

Trips	**Ozuna and Gomez (1995)**	**Moeltner (2006)**	**Parsons et al. (1999)**		**Trips**	**Ozuna and Gomez (1995)**	**Moeltner (2006)**	**Parsons et al. (1999)**
0	417	469	287		**16**	1	0	0
1	68	24	60		**17**	0	0	2
2	38	22	38		**20**	3	1	11
3	34	9	36		**25**	3	2	4
4	17	7	25		**26**	1	0	0
5	13	10	24		**28**	0	0	1
6	11	5	17		**30**	3	0	2
7	2	4	4		**35**	0	0	1
8	8	2	5		**40**	3	0	3
9	1	1	0		**50**	1	0	2
10	13	4	34		**88**	1	0	0
11	2	1	0		**100**	0	1	1[a]
12	5	0	0		**N**	**659**	**563**	**565**
13	0	1	0					[a]Also 150,200,250
14	0	0	1					
15	14	0	4					

The Ozuna and Gomez (1995) study collected data from a random sample of registered boat owners about boating trips to a popular lake in Texas. Moeltner (2006) collected data from a random sample of fishing license holders regarding fishing trips to the trophy section of a local Nevada river. Parsons et al. (1999) collected data from a random sample of Delaware residents regarding their visits to a popular beach.

It is noteworthy how many properties these datasets share. First, there is a large proportion of zeros (indicating a lack of involvement in the recreational activity being studied) which suggests what Sarker and Surry (2004) refer to as a "fast decay" process. Second, there is a disproportionate number of rounded responses and some evidence of rounding to the half-dozen and dozen. Third, there are some very large reported values which are almost all rounded to the nearest 10.

This paper develops an approach to treat both the presence of extreme values and rounded responses that we feel will be of interest to recreation demand modelers and that may have broad applicability to the analysis of other types of count data.

Theoretical Background

A common problem in recreation survey response data is a preponderance of zeros due to non-participation. This excess-zero problem may be addressed by considering a negative binomial (NB) estimator for the recreation demand model or by employing some of the alternative count data estimators suggested by Sarker and Surry (2004).[2] The remaining question is how to treat the rounded responses. Schaeffer and Presser (2003) have claimed that "estimation strategies lead to heaping at common numbers, such as multiples of 5 or 10…these strategies can be considered techniques for 'satisficing'…conserving time and energy and yet producing an answer that seems good enough for the purposes at hand." Similarly, Tourangeau et al. (2000) state that "by reporting their answers as round values, respondents may be consciously attempting to communicate the fact that their answers are at best approximations."

[2] If the zeros are generated by a different process than the non-zero responses (i.e., if some inherent, behavioral difference between users and non-users of a resource is readily identifiable), then hurdle count data models may need to be estimated (Haab and McConnell 2002).

> "There are no established conventions for rounding survey responses. Hence, researchers cannot be sure how much rounding there may be in survey data. Nor can researchers be sure whether respondents round to simplify communication or to convey partial knowledge," (Manski and Molinari 2010).

Vaske and Beaman (2006) provide a summary of their research on the topic of survey response, describing how respondents may answer recall questions about frequencies (such as days of participation) using episode enumeration, formula-based multipliers, and prototypes, which cause their responses to deviate from what occurred in reality. With low participation and episode enumeration (the recall and counting of specific occurrences), episode omission or telescoping may cause response errors. With greater participation and formula-based multipliers (the recall of a frequency rule applied to a time frame), misestimation of the rule or failure to recall exceptions to it may result in response error. With the use of prototypes (a single number used to represent a range of values), response clusters can occur, commonly around 0's and 5's.

All of these approximations can manifest in the data as response "heaping" – where reported numbers occur more often than chance would suggest. As Vaske and Beaman (2006) explain, heaping is likely related to number (or digit) preference: "numbers that a person has a disposition to use or avoid." Indeed, Huttenlocher et al. (1990) found that respondents tend to overuse both multiples of 5 and 10 and numbers associated with calendar events such as weeks or months (7, 14, 21, 30 and 60, for example).[3] These patterns (heaping, rounding and digit

[3] Huttenlocher et al. (1990) also describe the occurrence of forward bias, which results both from response leaping and from response "bounding": the imposition of an upper boundary (self-imposed or otherwise) on reports. This phenomenon (and more generally, response contraction bias) is further explored in Tourangeau et al. (2000).

preference) have all been observed and studied in the demographic, epidemiological and historical literatures (Pudney 2008).

Another manifestation of recall error is response "leaping" – where response heaping increases with reported magnitudes. "For responses under 15, several studies have found limited 0-5 heaping…Above 100, responses may fall largely on 150, 200, 300, and 500 with gaps widening as responses move into the thousands," (Vaske and Beaman 2006).

A number of papers discuss front-end reduction (prevention) of response bias and suggest strategies to improve survey design and implementation (Pudney 2008; Schaeffer and Presser 2003; Miller and Anderson 2002; Tarrant and Manfredo 1993; Chu et al. 1992). With regard to back-end reduction (correction) of response bias, Evans and Herriges (2010) and Barfield and Shonkwiler (2015) provide recent examples. Evans and Herriges (2010) employ a latent class model which assumes respondents are members of either a rounding or non-rounding class. Using generated data experiments, they find that rounding bias can have significant impacts on parameter estimates and resulting welfare measurements, with consumer surplus loss due to site closure being overstated by 5-37 percent. Barfield and Shonkwiler (2015) focused on very large reported counts. In the sample they analyzed, the overall mean number of trips was 2.09 and the mean number of trips by users was 6.31. However, for those 13 respondents who reported 20 or more trips, only one count was not a multiple of five or ten. Through the use of the incomplete beta function, they specified a negative binomial count data model with multiple censored regimes. Their rationale was that for larger counts the regimes should cover larger intervals.

Methodology

Both the Evans and Herriges (2010) latent class model and the Barfield and Shonkwiler (2015) censored regime method have their merits and shortcomings. The Evans and Herriges (2010) approach requires the identification of two regimes: rounders and non-rounders. As the mean number of rounders' trips will likely greatly exceed the mean number of non-rounders' trips, a stochastic specification to identify class membership must be employed. It is unclear how to precisely formulate this, and there is no obvious reason why rounders and non-rounders should have different conditional means. While the Barfield and Shonkwiler (2015) approach allows for the same conditional mean formulation for all respondents and may improve model fit, it suffers from the fact that increasing the size of the intervals will necessarily increase the value of the log likelihood. Since there is no statistical penalty from this approach, the selection and size of the censored regimes can only be based on a reasonableness criterion.[4] The researcher must consequently justify the sizes and positions of multiple intervals.

An alternative approach is to assume that at some cut-point, the distribution of responses changes from a discrete to a continuous distribution. The selection of this single cut-point, where a transition from non-rounding to rounding behavior can be assumed, is again based on a reasonableness criterion. For a discrete distribution, a given integer outcome has a unique probability associated with it, and though the term "count data regression" has become commonplace, it is somewhat misleading. In truth, the count data model is based on a probability mass function with a conditional mean – it is not, in fact, a regression. There is no underlying

[4] Increasing the width of the intervals allows each regime to encompass larger sums of probabilities. Attempting to parameterize the bounds on these regimes will therefore result in the model selecting bounds at the minimum and maximum responses. To avoid this collapsing of regimes, the researcher must determine and impose what they deem to be appropriate interval bounds for their particular dataset.

distribution of outcomes associated with a conditioning variable. By contrast, a regression model with a continuous response variable has a distribution of responses associated with a conditioning variable because the probability of a given response is very small.

For smaller responses where rounding is less likely to have occurred and where an outcome of zero is meaningfully different from outcomes such as one or two, the (less flexible) discrete distribution can reasonably be applied to calculate an exact probability for each response. For larger responses where rounding is more likely to have occurred and there is less certainty that outcomes are exact, the (more flexible) continuous distribution calculates probabilities around each response.[5] Under this formulation, the area of responses defined by the discrete distribution is right truncated, and the area of responses defined by a continuous distribution is left truncated. The form that the likelihood function takes in each of these partitions is therefore determined by the specific mixture of distributions chosen, and will be illustrated with regard to our specific application.

Our statistical approach employs the negative binomial (NB) distribution (see Cameron and Trivedi 2013), which is capable of handling large numbers of zeros and extreme values.[6] Its probability mass function is:

$$\frac{\Gamma(y+\alpha^{-1})\left(\frac{\mu}{\mu+\alpha^{-1}}\right)^{y}\left(\frac{\alpha^{-1}}{\mu+\alpha^{-1}}\right)^{\alpha^{-1}}}{\Gamma(y+1)\Gamma(\alpha^{-1})} \tag{1}$$

[5] If a respondent has engaged in rounding behavior, the researcher may observe that the respondent took, for example, 50 trips to a recreational resource. In fact, there is some distribution of trips around this 50 trip response that better represents the respondent's true pattern of visitation.

[6] For additional discussion of the NB distribution's strengths in this context, please see Sarker and Surry (2004).

where Γ is the gamma function, y is the number of trips to a site, $\mu = E(y)$, and α is a scale parameter capturing overdispersion. Note that if $\alpha = 0$, the NB distribution collapses to a Poisson distribution.

We use the generalized Pearson X^2 statistic (McCullagh and Nelder 1989) to evaluate our model fit. This statistic is chi-square distributed with degrees of freedom equal to the number of observations (i.e. respondents) minus the number of parameters estimated:

(2) $$X^2 = \sum_{i=1}^{n} \{y_i - E(y_i; \hat{\theta})\}^2 / V(y_i; \hat{\theta})$$

where n is the number of respondents, y_i is the number of trips reported by person i, $\hat{\theta}$ is a vector of estimated parameters, and V is variance. This form of the Pearson statistic is preferred to the form based on observed and expected frequencies as it does not require the assignment of data to groups ("bins"). The null hypothesis of this statistic is that the model fits the dataset well; specifically, that the model's predicted values accurately reproduce the dataset's first two moments (the mean and variance). Thus, a low p-value for this statistic indicates that the model fits badly – there is a low probability of error in rejecting this null hypothesis – and vice versa.

Application

Using data from Parsons et al. (1999)[7], we consider day-trip visits to a single site (Cape Henlopen State Park) in our estimation. The numbers of trips reported by the respondents in this survey are shown below in table 2. Again, we see a large number of zeros, possible heaping at rounded numbers and the half-dozen and dozen marks, some extreme values, and increasing

[7] The Parsons et al. (1999) survey was conducted in October, 1997, asking respondents about recreational trips to 62 Mid-Atlantic beaches during 1997 to date.

distances between the larger numbers of reported trips ("leaping"). There is also evidence of overdispersion, with an observed mean of 2 and a variance of over 48.

Table 2. Reported trips to Cape Henlopen State Park, 1997.

Trips Reported	Respondents	Proportion		Trips Reported	Respondents	Proportion
0	378	0.6690		**15**	4	0.0071
1	54	0.0956		**18**	1	0.0018
2	24	0.0425		**20**	4	0.0071
3	26	0.0460		**25**	3	0.0053
4	11	0.0195		**30**	1	0.0018
5	21	0.0372		**35**	1	0.0018
6	10	0.0177		**40**	1	0.0018
8	6	0.0106		**50**	1	0.0018
10	11	0.0195		**72**	1	0.0018
12	5	0.0088		**100**	1	0.0018
14	1	0.0018		**Descriptive Statistics**	**n=565**	**s^2=48.73 mean=2.09**

Barfield and Shonkwiler (2015) fit both standard and generalized variations of the NB distribution to this same dataset and found the (more flexible) generalized NB to be superior. Following their results, we propose the use of the generalized NB for the discrete distribution. For the continuous distribution, we propose the use of the lognormal distribution. The lognormal distribution has the advantage of a long right tail, and the NB and lognormal distributions follow the same conditional mean process. This makes the transition from the discrete to the continuous distribution smoother than in the case of latent class models where there are different conditional means depending on class membership.

For observations at or below the cut-point, *c*, the likelihood is:

(3) $$\frac{\Gamma(y+a)\left(\frac{\mu}{\mu+a}\right)^{y}\left(\frac{a}{\mu+a}\right)^{a}}{\Gamma(y+1)\Gamma(a)\Pr(y\leq c)}\Pr(y\leq c)$$

where $a = \frac{1}{\alpha}\mu^{\Phi}$, μ is the conditional mean, α is the dispersion parameter, and *y* is the number of trips to a site up to the cut-point, c (i.e., $y = 0,1,\ldots,c$). This is the result of multiplying the right truncated generalized NB distribution by the probability of being in that regime—hence, the two probabilities will cancel out.

For observations above the cut-point, the likelihood is:

(4) $$\frac{\exp(-.5(\frac{(\log(y)-\log(\mu))}{\sigma})^{2})(1-\Omega(x,a,c+1))}{(\sigma\sqrt{2\pi})(\Phi\left(\frac{\log(\mu)-\log(c+1)}{\sigma}\right))}$$

where *y* is the number of trips to a site beyond the cut-point (i.e., $y = c+1,\ldots$), $\Omega(x, a, c + 1)$ is the sum of probabilities of the generalized NB distribution from 0 to c, and $\Phi(z)$ is the standard normal cumulative distribution function. In this case, $\Phi(z)$ is the probability that the lognormal distribution is above *c* to account for the left truncation. The term $(1 - \Omega(x, a, c + 1))$ accounts for the probability of being above the cut-point and can be computed using the incomplete beta function (note that *x* is now defined as $(a/(a + \mu))$).

To calculate the Pearson statistic, we must obtain the conditional means and variances of this mixed distribution model.

For the right truncated generalized NB distribution, we refer to the recent work by Shonkwiler (2015), which corrects the second moments as reported by Gurmu and Trivedi (1992) and Cameron and Trivedi (2013). The formulas are as follows for the conditional mean:

(5) $$E[Y|Y \leq c] = \mu - \frac{(c+1)pmf(c+1)}{x\Pr(Y\leq c)} = \mu^{0}$$

and conditional variance:

(6) $$V(Y|Y \leq c) = \mu + \mu^2/\mathrm{a} + (c+1)(\mu^0 - \mu) - (\mu^0 - \mu)^2 - (\mathrm{a}-1)(\mu^0 - \mu)\mu/\mathrm{a}$$

where $pmf(c+1)$ represents the generalized NB probability mass function evaluated at c+1.

For the left truncated lognormal distribution[8], the conditional moments can be written as:

(7) $$E(y|y > c) = \exp(\log(\mu) + .5\sigma^2)\ \frac{\Phi(\sigma + (\log(\mu) - \log(c))/\sigma)}{\Phi((\log(\mu) - \log(c))/\sigma)}$$

(8) $$E(y^2|y > c) = \exp(2\log(\mu) + 2\sigma^2)\ \frac{\Phi(2\sigma + (\log(\mu) - \log(c))/\sigma)}{\Phi((\log(\mu) - \log(c))/\sigma)}$$

This permits straightforward computation of the Pearson statistic.

Results

Our results from the basic generalized NB specification are summarized in table 3 below.

Table 3. Estimates: Generalized negative binomial distribution.

Variable/Parameter	Coefficient	Robust Std. Error	z-Value
Constant	1.4250	1.4686	0.9703
Trip Cost	-0.0354	0.0045	-7.9368
Log Age	0.1283	0.3926	0.3269
Child <10	0.7206	0.2090	3.4486
Retired	-0.5810	0.3190	-1.8216
Student	0.9522	0.3252	2.9278
Alpha	6.4913	0.8720	7.4443
Phi	0.4765	0.1033	4.6141
Pearson Statistic: 621.5, p=0.030	**Log Likelihood = -785.11**		

[8] These formulas are based on the work of Bebu and Mathew (2009) (note what they report as the variance is actually E(y^2)).

In our application of the Distribution Transition Method, the model was fit to the data with a cut-point set at 19, as we believe responses of 20 reported trips or more could exhibit rounding behavior. The results we obtained (table 4) are remarkably similar to those found by the censored regime applications of Barfield and Shonkwiler (2015), and display similar improvements as compared with a simple truncated distribution (described below).

Table 4. Estimates: Mixed generalized negative binomial—lognormal distribution.

Variable/Parameter	Coefficient	Robust Std. Error	z-Value
Constant	1.568	1.4055	1.116
Trip Cost	-0.0340	0.0044	-7.766
Log Age	0.0653	0.3728	0.175
Child <10	0.6897	0.1980	3.483
Retired	-0.5750	0.3091	-1.86
Student	0.9255	0.3198	2.894
Alpha	6.2178	0.7927	7.844
Phi	0.5168	0.1077	4.798
Sigma	1.0608	0.1884	5.631
Pearson Statistic: 477.5, p=0.993	**Log Likelihood = -741.04**		

In the context of over-reporting and extreme values, a common practice in the survey response literature is simply to exclude larger values from the dataset on the basis that they are likely to be unrepresentative of the general (or target) population. To this end, we consider how our results would be affected if we truncated the data at 50 (thereby losing three, extreme-value observations) and estimated a generalized NB model. Our results from this specification are summarized in table 5 below.

Table 5. Estimates: Truncated, generalized negative binomial distribution.

Variable/ Parameter	Coefficient	Robust Std. Error	z-Value
Constant	2.0370	1.3548	1.5036
Trip Cost	-0.0309	0.0041	-7.5161
Log Age	-0.1108	0.3549	-0.3123
Child <10	0.6409	0.1897	3.3783
Retired	-0.5312	0.2958	-1.7961
Student	0.8711	0.3123	2.7891
Alpha	5.9586	0.7362	8.0940
Phi	0.5994	0.1131	5.2982
Pearson Statistic: 488.2, p=0.979	**Log Likelihood = -757.72**		

We have lost information in estimating this model (by eliminating data points), and as a result, the fit is not quite as good as when we incorporate this information under uncertainty using the Distribution Transition Method. Estimated per-trip consumer surplus[9] moves from $29.41 in the mixed generalized negative binomial-lognormal distribution to $32.36 in the truncated distribution, and we observe changes in all of the parameter estimates. While these changes are not statistically different, we have only lost three observations in this particular example. In datasets with large numbers of extreme values to either truncate or model under uncertainty, the disparities in these parameter estimates could become significant with regard to policy implications.

[9] Calculated as the inverse of the estimated coefficient on trip cost, the marginal utility of income.

Discussion

The survey response literature has established that respondents tend to over-report their recreational activities, and correcting for “heaps and leaps” in survey response data is largely an empirical issue. This paper illustrates how, when modeling single-site recreational trip data using a negative binomial distribution, the model’s fit is significantly improved by transitioning from the NB (a discrete distribution) to the lognormal (a continuous distribution) at a cut-point where it is supposed that response behavior begins to exhibit rounding. This approach offers empirical advantages over those methods recently proposed by Evans and Herriges (2010) and Barfield and Shonkwiler (2015).

We find two socio-demographic variables to be significant: Child<10 and Student. We hypothesize this is because Cape Henlopen State Park is a popular vacation destination for those who are, or have children who are, out of school during the summer. Our analysis did not find a statistically significant difference in the parameter or per-trip consumer surplus estimates when extreme values were either truncated or incorporated under uncertainty. However, only three observations were truncated in our particular application, which may not have provided a significant enough loss of information to impact the overall estimation. As we expand this research, we may examine other sites in the rich Parsons et al. (1999) dataset (or different datasets entirely) where the impacts of truncation may be more extensive.

References

Barfield, A., and J. S. Shonkwiler. 2015. “Recreation Survey Response Data: Patterns and Problems.” University of Georgia. Department of Agricultural and Applied Economics. Working paper

Bebu, I. and T. Mathew. 2009. “Confidence intervals for limited moments and truncated moments in normal and lognormal models.” *Statistics & Probability Letters* 79(3):375-380

Cameron, A. C., and P. K. Trivedi. 2013. *Regression analysis of count data, 2nd ed.* Cambridge, NY: Cambridge University Press

Chu, A., D. Eisenhower, M. Hay, D. Morganstein, J. Neter and J. Waksberg. 1992. “Measuring the recall error in self-reported fishing and hunting activities.” *Journal of Official Statistics* 8(1): 19-39

Connelly, N. A., and T. L. Brown. 1995. “Use of angler diaries to examine biases associated with 12-month recall on mail questionnaires.” *Transactions of the American Fisheries Society* 124(3): 413-422

Evans, K., and J. A. Herriges. 2010. “Rounding in recreation demand models: A latent class count model.” Iowa State University. Working paper.

Gurmu, S. and P. K. Trivedi, 1992. “Overdispersion tests for truncated Poisson regression models.” *Journal of Econometrics* 54: 347-370

Haab, T. C., and K. E. McConnell. 2002. *Valuing Environmental and Natural Resources: The Econometrics of Non-Market Valuation*. Cheltenham: Edward Elgar

Hoehn, J. P., T. Tomasi, F. Lupi and H. Z. Chen. 1996. “An economic model for valuing recreational angling resources in Michigan.” Michigan State University. Report to the

Michigan Department of Environmental Quality and to the Michigan Department of Natural Resources

Huttenlocher, J., L. V. Hedges, and N. M. Bradburn. 1990. “Reports of elapsed time: Bounding and rounding processes in estimation.” *Journal of Experimental Psychology: Learning, Memory, and Cognition* 16(2): 196-213

Manski, C. F., and F. Molinari. 2010. “Rounding probabilistic expectations in surveys.” *Journal of Business & Economic Statistics* 28(2): 219-231

McCullagh, P. and J.A. Nelder. 1989. *Generalized Linear Models*. 2nd ed. New York, NY: Chapman & Hall/CRC

Miller, C. A., and W. L. Anderson. 2002. “Digit preference in reported harvest among Illinois waterfowl hunters.” *Human Dimensions of Wildlife* 7(1): 55-65

Moeltner, K. 2006. “Anglers’ Fishing Choices on the Truckee, Carson, and Walker Rivers and the Threat of the New Zealand Mud Snail.” Report prepared for the Pacific States Marine Fisheries Commission and the U.S. Fish and Wildlife Service, June 27, 2006

Ozuna, T. Jr., and I. A. Gomez. 1995. “Specification and testing of count data recreation demand functions.” *Empirical Economics* 20(3): 543-550

Parsons, G. R., D. M. Massey, and T. Tomasi. 1999. “Familiar and favorite sites in a random utility model of beach recreation.” *Marine Resource Economics* 14(4): 299-315

Pudney, S. 2008. “Heaping and leaping: survey response behaviour and the dynamics of self-reported consumption expenditure.” University of Essex. Institute for Social and Economic Research. Working paper

Sarker, R., and Y. Surry. 2004. “The fast decay process in outdoor recreational activities and the use of alternative count data models.” *American Journal of Agricultural Economics* 86(3): 701-715

Schaeffer, N. C., and S. Presser. 2003. “The science of asking questions.” *Annual Review of Sociology* 29: 65-88

Shonkwiler, J. S. 2015. “Moments of the truncated negative binomial distribution.” University of Georgia. Department of Agricultural and Applied Economics. Working paper

Tarrant, M. A., and M. J. Manfredo. 1993. “Digit preference, recall bias, and nonresponse bias in self reports of angling participation.” *Leisure Sciences* 15(3): 231-238

Tourangeau, R., L. J. Rips, and K. A. Rasinski. 2000. *The Psychology of Survey Response.* Cambridge: Cambridge University Press

Vaske, J. J., and J. Beaman. 2006. “Lessons learned in detecting and correcting response heaping: Conceptual, methodological, and empirical observations.” *Human Dimensions of Wildlife* 11(4): 285-296