Estimating Recreation Demand When Survey Responses are Rounded

Ian B. Page
Office of the Senior Vice President and Provost
University of Maryland, College Park
ibpage@umd.edu

Erik Lichtenberg
Department of Agricultural and Resource Economics
University of Maryland, College Park
elichten@umd.edu

Monica Saavoss
Department of Agricultural and Resource Economics
University of Maryland, College Park
msaavedr@umd.edu

Abstract: Recall data of consumption of cigarettes, alcohol, fresh fruits and vegetables; visits to recreational sites, doctors' offices, and local businesses; household expenditures; and individuals' perceived probabilities of future events often contain reported numbers that appear to be rounded to nearby focal points (e.g., the closest 5 or 10). Failure to address this rounding has been show to produce biased estimates of marginal effects and willingness to pay. We investigate the relative performance of three count data models used with data of the kind typically found in recreation demand studies. We create a dataset based on observed recreational trip counts and associated trip costs that exhibits substantial rounding. We then conduct a Monte Carlo simulation exercise to compare estimated parameters, the average partial effect on an increase in trip cost, and average consumer surplus per trip for three alternative estimators: a standard Poisson model with no adjustment for rounding, a censored Poisson model, and the grouped Poisson model. The standard Poisson model with no adjustment for rounding exhibits significant, persistent bias, especially in estimates of average consumer surplus per trip. The grouped Poisson, in contrast, shows only slight biases and none at all in estimates of average consumer surplus per trip.

Keywords: count data, recreation demand, rounding, heaped data, coarsened data, Poisson

JEL codes: C25, Q26

Selected Paper prepared for presentation at the 2015 Agricultural & Applied Economics Association and Western Agricultural Economics Association Annual Meeting, San Francisco, CA, July 26-28

This material is based upon research supported by the National Institute of Food and Agriculture, Specialty Crop Research Initiative, Award No. 2011-51181-30767.

Estimating Recreation Demand When Survey Responses are Rounded

Introduction

Recall data are ubiquitous in economics. One well known problem with these data is that reported numbers often appear to be rounded to nearby focal points (e.g., the closest 5 or 10), especially for activities that respondents engage in frequently (Blair and Burton, 1987; Burton and Blair, 1991). For example, NOAA's Marine Recreational Information Program survey asks respondents to report the number of days they went fishing during the previous twelve months. Three quarters of the responses of those reporting 10 or more fishing days are divisible by 5; half of the latter are divisible by 10. This rounding or "heaping" phenomenon has been observed in reports of consumption of cigarettes, alcohol, fresh fruits and vegetables, as well as in reports of visits to recreational sites, doctors' offices, and local businesses. Similar problems with rounded or heaped answers have also been observed in reported household expenditures (Browning, Crossley, and Weber, 2003) and individuals' perceived probabilities of future events (Manski and Molinari, 2010).

There is growing evidence suggesting that empirical techniques that do not address heaping will produce biased estimates of marginal effects and willingness to pay. Heitjan and Rubin (1991) identify several empirical strategies researchers can use for heaped data, notably employing either a grouped likelihood function or a likelihood function that accounts for the stochastic rounding in the data. The latter method has generally been more favored. Manski and Molinari (2010) present a general maximum likelihood approach based on reported probabilities being stochastically rounded to the nearest 1, 5, 10, 20, etc. Evans and Herriges (2010) present a latent class count model that divides respondents into "rounders" (who round to the nearest 5) and "non-rounders." They examine the model's performance using Monte Carlo simulations and then estimate trip demand using data from the Iowa Lakes Project as a demonstration. More recently, Drechsler and Kiesl (2014) outlined an imputation strategy to determine actual household income when the degree of rounding might vary between values of 5, 10, 50, 100, 500, and 1000. Another possible approach for count data is the grouped Poisson introduced by Moffatt (1995), which imposes an assumption about what kind of rounding is being used, but can be consistent with the recommendations of Heitjan and Rubin. Yet another possibility is to artificially censor the data at a threshold value beyond which heaping is assumed to occur (Cameron and Trivedi, 1998).

The literature to date has not considered the relative performance of these alternatives. This paper conducts a relative performance test of these alternatives for count data of the kind typically found in recreation demand studies. We create a simulated data set with more pronounced rounding than the data used by Evans and Herriges (2010) but more representative of the degree of rounding found in many recreation demand data sets. Following Manski and Molinari (2010) and Drechsler and Kiesl (2014), we allow for multiple types of rounding. We use that simulated data set to assess biases in estimated parameters, marginal effects and willingness to pay for two relatively easily implemented count data models (censored and grouped Poisson) as well as for a model that ignores the presence of rounding.

We find that ignoring rounding introduces bias into estimates of parameters, average marginal effects, and consumer surplus. Both the grouped Poisson and censored Poisson perform well. The grouped Poisson model performs better than the censored Poisson model, although the censored model is generally easier to implement than the former.

Our findings have important implications for estimating recreation demand. A survey of major datasets such as NOAA's Marine Recreational Fisheries Statistics Survey and Marine Recreational Information Program, and the USDA Forest Service's National Survey on Recreation and the Environment, shows that the incidence of heaping in recreation demand data is widespread and substantial. Yet rounding in reported trip counts is ignored in most recreation demand studies. Our findings suggest that demand estimation can be improved significantly by correcting for heaping in ways that are robust and involve low computational cost.

Background

Rounded, coarsened, or heaped data[1] occur when, for an entire interval of values, the researcher observes a single common value. Rounded data from survey responses are a common example. Another special case is interval censoring, such as when a disease is known to manifest itself at some point between screenings, but the precise timing is unknown. Coarsened data can also occur by the intentional grouping of collected data, for respondent privacy or data management reasons. Data heaping to some extent is inevitable when underlying data are continuous, but the issue becomes magnified when the heaping intervals increase (Heitjan, 1989).

One example of rounded data occurs in the recreational literature. When respondents are asked to recall how many times in a given time period they engaged in a particular activity, they are more likely to report values around focal points such as multiples of 5, 10, or 50. The Alaska Sport Fishing Survey, for example, asks respondents how many recreational fishing trips they took in an entire year, while the National Survey on Recreation and the Environment, asks participants how many times in the past year they engaged in a wide variety of outdoor recreational activities.

Rounded data can also be a problem in health related research using survey data that asks the number of times that respondents engage in an activity such as exercise or smoking. The number of cigarettes smoked per day, for example, is a frequently cited, since respondents

[1] We use all three terms interchangeably.

typically report smoking habits is multiples of ten cigarettes per day.[2] Several papers have explicitly investigated the impact of rounding on estimates of factors associated with smoking (Wang and Heitjan, 2008; Bar and Lillard 2010). Even so, in other health-related contexts, a similar pattern emerges. Moffatt and Peters (2000), for example, observe rounding when respondents are asked about the number of sexual partners they have had over a given period of time. The same phenomenon occurs in examinations of the determinants of exercise, since high frequency exercisers may use rounding to estimate the exact number of times they exercised in a given period of time. Rounding is also observed to occur in recall data on diet, affecting research on in determinants of the number of servings of fruits or vegetables respondents consumed over a given time period.

It is generally possible to observe that rounding is present in a data set by noting a clustering of responses around focal points such as multiples of large even numbers. For a more precise assessment, Roberts and Brewer (2001) present a test for the presence of rounding. The test requires hypotheses of the heaping pattern, such as around multiples of 5, and performs a pairwise comparison of the frequency of responses for the hypothesized category to the frequency of responses one more or one less than the hypothesized category.

Identifying the "rules" respondents follow in rounding their answers or conditions under which rounding tends to occur is more difficult. Both the propensity to round and the rule followed in rounding may be a function of individual characteristics. For example, rounding may be correlated with demographic characteristics such as educational level (Kinkel 2004). Alternatively, rounding may convey uncertainty about an exact number (Manski and

[2] A pattern of responses clustered around focal points is in itself not necessarily indicative of a misreporting problem. In the case of cigarettes smoked in the previous day, it is possible that seemingly rounded responses are in fact accurate because smokers may smoke in increments of whole packs.

Molinari, 2010). In surveys of recreational activities for example, rounding seems to be more pronounced in reports of larger numbers of times an activity is engaged in (e.g., more fishing trips) over a longer period of time (e.g., a year compared to a month). The exact incidence of the rounding can vary in ways that can be difficult identify accurately. For example a response of 100 recreational fishing trips in a year could represent rounding to the nearest 5, 10, 50, or 100; as well as an exact figure.

Estimation methods that ignore rounding can produce parameter estimates that are highly biased, especially when the degree of coarsening is not small (Heitjan 1989, Heitjan and Rubin 1991). Thus, for example, parameter estimates from a Poisson regression that treats reports as precise, i.e., does not correct for rounding, can exhibit substantial bias, especially when the intervals in which rounding occurs are large, unevenly spaced, or of uncertain group widths (Heitjan 1989).

One alternative is the censored Poisson model, which adjusts partially for rounding by treating all observations above a certain cutoff as censored. The rationale for this approach is that individuals tend to round more when they recall higher numbers (Cameron and Trivedi 1998). While this method accounts for the fact that higher numbers are unlikely to be exactly accurate, it has the disadvantage of giving significantly less weight to all observations at or above the censoring point.

A second alternative is the grouped Poisson model, first proposed by Moffatt (1995), which treats rounded responses as part of an interval rather than a single observation. For the grouped Poisson to be effective, the researcher must specify the intervals in which observations are to be rounded, which in practice may be difficult to know. However, this approach has the advantage of better utilizing the available data than the censored Poisson.

A third alternative is a latent class model, which assumes that some respondents rounded their answers while others did not. As with most approaches that impose additional structure, the latent class model can produce more accurate estimates if the structural assumptions about rounding are valid but may introduce additional sources of bias of those assumptions are wide of the mark. Moreover, it can be difficult to implement in practice. Wright and Bray (2002) present an early application of a latent class model for fetal ultrasound measurements, a case in which accurate reporting is not a matter of personal recall ability but of measurement precision. Evans and Herriges (2010) use a latent class approach in recreation demand, and show that the bias imposed by rounding in recreation data can be substantial.

Data and Estimation

Like Evans and Herriges (2010), we investigate the consequences of rounding and potential of alternative estimators in the presence of rounding through a series of Monte Carlo experiments using data based on recreational trip counts. Specifically, we model a discrete variable y when many of the reported values y^* are not accurate because of the incidence of respondent-generated rounding. We then analyze the performance of a standard Poisson regression, a censored Poisson regression, and grouped Poisson regression and determine how well they handle the rounded dataset y^* in terms of bias in estimated coefficient values, average partial effects, and consumer surplus.

Data

We base our simulations on data from the fifth version of the National Survey on Recreation and the Environment (NSRE), a survey conducted by the US Department of Agriculture's Forest Service Southern Research Station. Each version of the NSRE served to collect outdoor recreational information, including participation and frequency of general activities (e.g., bicycling, hiking, and fishing) in the previous twelve months. We use the fifth version because it contained questions about fresh- and saltwater recreational activity. Specifically, we use the data on the number of freshwater recreational trips taken over the preceding 12 months, descriptive statistics of which are shown in Table 1. The distribution of responses followed a pattern indicative of rounding, especially for high trip counts: 55.3% of all responses 5 or greater are divisible by 5 while 84.8% of responses of 20 or more are divisible by 5 and 72.7% are divisible by 10 (Figure 1).

We assume that the true number of freshwater recreation trips taken by household i, y_i, follows a Poisson distribution with a conditional mean $\lambda_i = \exp(\beta_0 + \beta_P P_i)$, where P_i is the cost of a single trip measured in thousands of dollars. We estimate that parameters β_0 and β_P using a Poisson regression and obtain estimates $\beta_0 = 2.8$ and $\beta_P = -3.8$. We randomly draw 10000 observations of trip cost P_i from a uniform distribution ($P_i \sim U[0,1]$) and generate an underlying sample of trip counts y. We then induce rounding by assuming that all trip counts of 3 or more per year are equally likely to be reported precisely or rounded to the nearest number divisible by 5. The distributions of the true and rounded trip counts are presented in Figure 1. The observed distributions show that rounding is much more pronounced in the simulated data than in the actual data. In both cases, rounded responses are much more common for higher trip counts.

Estimation Methods

We investigate the effects of rounding in three econometric specifications: (1) a standard Poisson model that does not adjust for potential rounding; (2) a right-censored Poisson model; and (3) a grouped Poisson model.

Use of the standard Poisson model may be justified if the degree of rounding is sufficiently small that it merely introduces additional noise into the data. At least at first, this assumption does not seem unreasonable, since rounding will cause some people to overstate their trip frequency whereas others will underreport theirs. For the standard Poisson model, elements of the likelihood function assume the form

$$(1)\ \ f(y_i, \boldsymbol{X_i}; \beta) = \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!},$$

where $\boldsymbol{X}_i$ is a vector of explanatory variables and $\lambda_i = \exp(\boldsymbol{X}_i'\beta)$.

As an alternative, Cameron and Trivedi (1998) suggest using a censored Poisson model to adjust for the presence of rounding. In this case, artificially right-censoring the data re-weights the regression to counts that are more likely to be accurately reported. With a censored Poisson model, the observed count is y_i', where

$$y_i' = \begin{cases} y_i & \text{if } \mathrm{y_i} < \mathrm{c} \\ c & \text{if } y_i \geq c \end{cases}$$

and c is the censoring point. The likelihood function thus remains the same for counts below c and becomes

$$(1)\ \ Pr(y_i' = c) = 1 - \textstyle\sum_{a=0}^{c-1} f(a, \boldsymbol{X_i}; \beta)$$

otherwise. In our simulations, rounding can occur when $y_i \geq 3$, so we set the censoring point at $c = 3$.

Another alternative is the grouped Poisson model, first suggested by Moffatt (1995), albeit in a somewhat different context (specifically, in cases where responses are reported categorically in terms of intervals). We repurpose the grouped Poisson model by applying it to data that are rounded to a nearby number, since rounding effectively creates intervals into which reported responses fall. For example, if a researcher knows the true value falls between 13 and 17, either because the survey explicitly restricts answers to this interval or because respondents round to 15 within this range, than the probability is as follows

$$(1)\ \ Pr(13 \leq y_i \leq 17) = \textstyle\sum_{a=13}^{17} f(a, \boldsymbol{X_i}; \beta),$$

where $f(a, \boldsymbol{X_i}; \beta)$ is the probability of count a. In our case, use of the grouped Poisson requires imposing assumptions about interval construction. We estimate the parameters of the grouped Poisson model under the assumption of rounding to the nearest number divisible by 5.

We use the estimated parameters of the model to calculate the partial effect of trip cost on the number of trips and consumer surplus per trip, both of which are averaged over the sample. Confidence intervals for all of these estimates were obtained by 5000 iterations of bootstrapping. Estimates of the parameters of the conditional mean of the underlying Poisson distribution, the average partial effect of trip cost, and average consumer surplus per trip are shown with their bootstrapped standard errors and compared to the true values of each in Table 2.

Figures 2-10 provide additional detail about the performance of each of the three estimators considered here in estimating the trip cost coefficient, the average partial effect of trip cost on the number of trips and consumer surplus per trip. Figures 2, 5, and 8 show how confidence intervals for the estimated trip cost coefficient vary with sample size. Figures 3, 6, and 9 show the distribution of the difference between the estimated and true average partial effects of an increase in trip cost obtained from the bootstrapping procedure. Figures 4, 7, and 10 show the distribution of the difference between the estimated and true average consumer surplus per trip obtained from the bootstrapping procedure. Examination of these distributions indicates whether the model shows a tendency toward under- or overestimating the true values even if the true value lies within an acceptable confidence interval.

Results

As the figures in Table 3 indicate, our simulations show that failure to adjust for rounding results in modest biases in estimates of the trip cost coefficient, marginal effects, and welfare.

Both the censored and grouped Poisson models provide more accurate estimates of all three, while the grouped Poisson in particular exhibits the lowest biases.

The standard Poisson model without an adjustment for rounding gives an unbiased estimate of the intercept and a bias of only 0.7% in the estimated average partial effect of an increase in trip cost. Biases in the trip cost coefficient and average consumer surplus estimates are more substantial, 3.2% and 5.5%, respectively. The estimated trip cost coefficient is significantly different from sample sizes that allow for precise estimates (Figure 2). The estimated average partial effect of an increase in trip cost is biased upward most of the time (Figure 3) while the estimate of consumer surplus per trip is biased upward all the time. Thus, while the bias in these estimates may seem modest, it is also persistent.

In contrast to the standard Poisson with no adjustment for rounding, the censored Poisson gives slightly biased estimate of the intercept as well as of the trip cost coefficient, on the order of 0.4% and 1.1%, respectively (Table 2). These coefficients lie within a 95% confidence interval even as sample sizes get large (Figure 5). The estimated partial effect of an increase in trip cost has a larger bias than the estimate obtained from a Poisson model with no adjustment for rounding, 2.2% compared to 0.7% (Table 2). As in the case of the Poisson model with no adjustment for rounding, the bias occurs in a large majority of the bootstrap replications (Figure 6). The estimate of average consumer surplus per trip obtained from the censored Poisson model is unbiased, however (Table 2). Over- and under-estimated are distributed symmetrically around zero (Figure 7), indicating the absence of a tendency toward either.

The grouped Poisson performs better than either alternative in terms of bias. The estimated intercept is unbiased while the estimated coefficient of trip cost has an average bias of only 0.3%. The true parameter values lie within 95% confidence limits even for very large sample sizes (Figure 8). The bias in the estimated average partial effect of an increase in trip cost is also quite small at 0.4% (Table 2). Its distribution shows a slight asymmetry, with underestimates occurring a little more frequently than overestimates (Figure 9). The estimate of average consumer surplus per trip is unbiased (Table 2) and shows no tendency toward over- or under-estimation (Figure 10).

Conclusion

Recall data often contain reported numbers that appear to be rounded to nearby focal points (e.g., the closest 5 or 10. This phenomenon has been observed in reports of consumption of cigarettes, alcohol, fresh fruits and vegetables as well as in reports of visits to recreational sites, doctors' offices, and local businesses, as well as reported household expenditures and individuals' perceived probabilities of future events. Failure to address this rounding has been show to produce biased estimates of marginal effects and willingness to pay.

This paper conducts a relative performance test of several alternatives for count data of the kind typically found in recreation demand studies. We create a dataset based on observed recreational trip counts and associated trip costs that exhibits substantial rounding. We then conduct a Monte Carlo simulation exercise to compare estimated parameters, the average partial effect on an increase in trip cost, and average consumer surplus per trip for three alternative estimators: a standard Poisson model with no adjustment for rounding, a censored Poisson model, and the grouped Poisson model. The standard Poisson model with no adjustment for rounding exhibits significant, persistent bias, especially in estimates of average consumer surplus per trip. The grouped Poisson, in contrast, shows only slight biases and none at all in estimates of average consumer surplus per trip.

One approach not explored in this paper is the use of a latent class Poisson model that divides respondents into those who round answers and those who do not. Such an approach can be more flexible in terms of the rounding patterns it can accommodate. Absolute performance of latent class models has been investigated by Evans and Herriges (2010), Manski and Molinari (2010), and Drechsler and Kiesl (2014). Performance of these models relative to those that are computationally less burdensome (like the censored and grouped Poisson models investigated here) has not been studied. The sensitivity of that relative performance to misspecification of the rounding "rule" has similarly not been studied. We plan to investigate both extensions in future work.

References

Bar, Haim Y, and Dean R. Lillard. "Accounting for Heaping in Retrospectively Reported Event Data - a Mixture-Model Approach." *Statistics in Medicine*. 31.27 (2012): 3347-3365.

Blair, Edward, and Scot Burton. "Cognitive Processes Used by Survey Respondents to Answer Behavioral Frequency Questions." *Journal of Consumer Research*. 14.2 (1987): 280-288.

Browning, Martin, Thomas F. Crossley, and Guglielmo Weber. "Asking Consumption Questions in General Purpose Surveys." *The Economic Journal*. 113.491 (2003).

Burton, Scot, and Edward Blair. "Task Conditions, Response Formulation Processes, and Response Accuracy for Behavioral Frequency Questions in Surveys." *Public Opinion Quarterly*. 55.1 (1991).

Cameron, Adrian C, and P K. Trivedi. *Regression Analysis of Count Data*. Cambridge, UK: Cambridge University Press, 1998.

Drechsler,Jorg, Hans Kiesl. "Beat the heap - an imputation strategy for valid inferences from rounded income data." *Working Paper.* 2014

Evans, Keith, Joseph A. Herriges "Rounding in Recreation Demand Models: A Latent Class Count Model." *Working Paper.* 2010

Heitjan, Daniel F, and Donald B. Rubin. "Ignorability and Coarse Data." *Ann. Statist*. 19.4 (1991): 2244-2253.

Heitjan, Daniel F. "Inference from Grouped Continuous Data: a Review." *Statist. Sci*. 4.2 (1989): 164-179.

Manski, C.F, and F Molinari. "Rounding Probabilistic Expectations in Surveys." *Journal of Business and Economic Statistics*. 28.2 (2010): 219-231.

Moffatt, Peter G. "Grouped Poisson Regression Models: Theory and an Application to Public House Visit Frequency." *Communications in Statistics - Theory and Methods*. 24.11 (1995): 2779-2796.

Moffatt, Peter G, and Simon A. Peters. "Grouped Zero-Inflated Count Data Models of Coital Frequency." *Journal of Population Economics*. 13.2 (2000): 205-220.

Roberts, John, and Devon Brewer. "Measures and Tests of Heaping in Discrete Quantitative Distributions." *Journal of Applied Statistics*. 28.7 (2001): 887-896.

Wang, H, and D F. Heitjan. "Modeling Heaping in Self-Reported Cigarette Counts."*Statistics in Medicine*. 27.19 (2008): 3789-3804.

Wright, David E, and Isabelle Bray. "A Mixture Model for Rounded Data." *Journal of the Royal Statistical Society: Series D (the Statistician)*. 52.1 (2003): 3-13.

Table 1. Descriptive Statistics, NSRE Freshwater Trips

	Obs.	Mean	Std. Dev.	Min.	Max
During the past 12 months, how many trips did you take to the freshwater body you visited most recently?	725	3.26	17.14	0	365
How much did your most recent trip cost you?	351	124.78	605.88	0	7000

Table 2. Performance of Alternative Estimators in the Presence of Rounding

	True Value	**Standard Poisson**	**Censored Poisson**	**Grouped Poisson**
		Conditional Mean Parameters		
Constant	2.80	2.80**	2.79**	2.80**
Trip Cost	-3.80	-3.68**	-3.76**	-3.81**
		Average Partial and Welfare Effects		
APE	-15.73	-15.62	-15.38	-15.79
CS	1.09	1.15	1.09	1.09

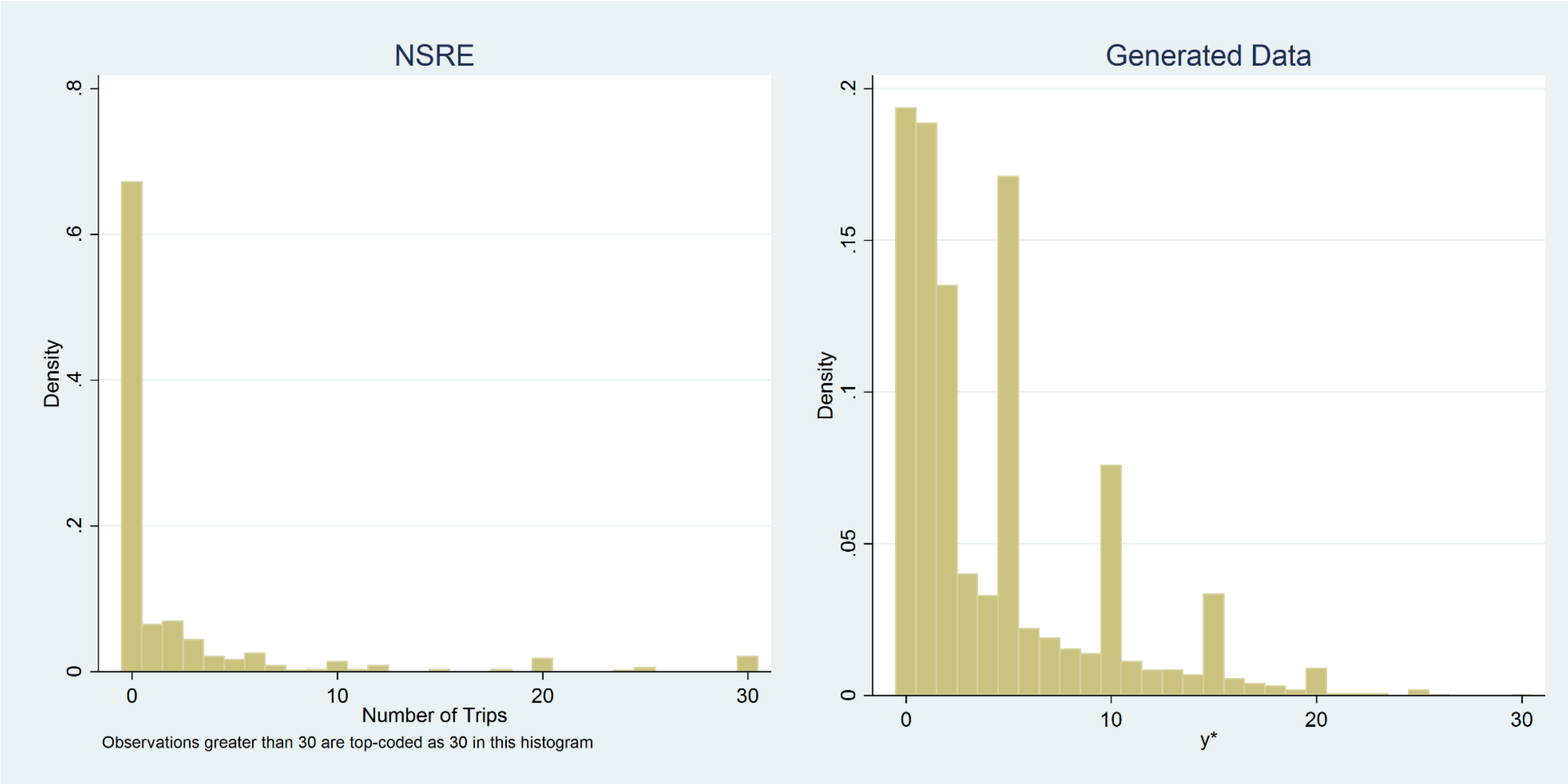

Figure 1. Comparison of Simulated and National Survey of Recreation and the Environment (NSRE) Trip Count Data

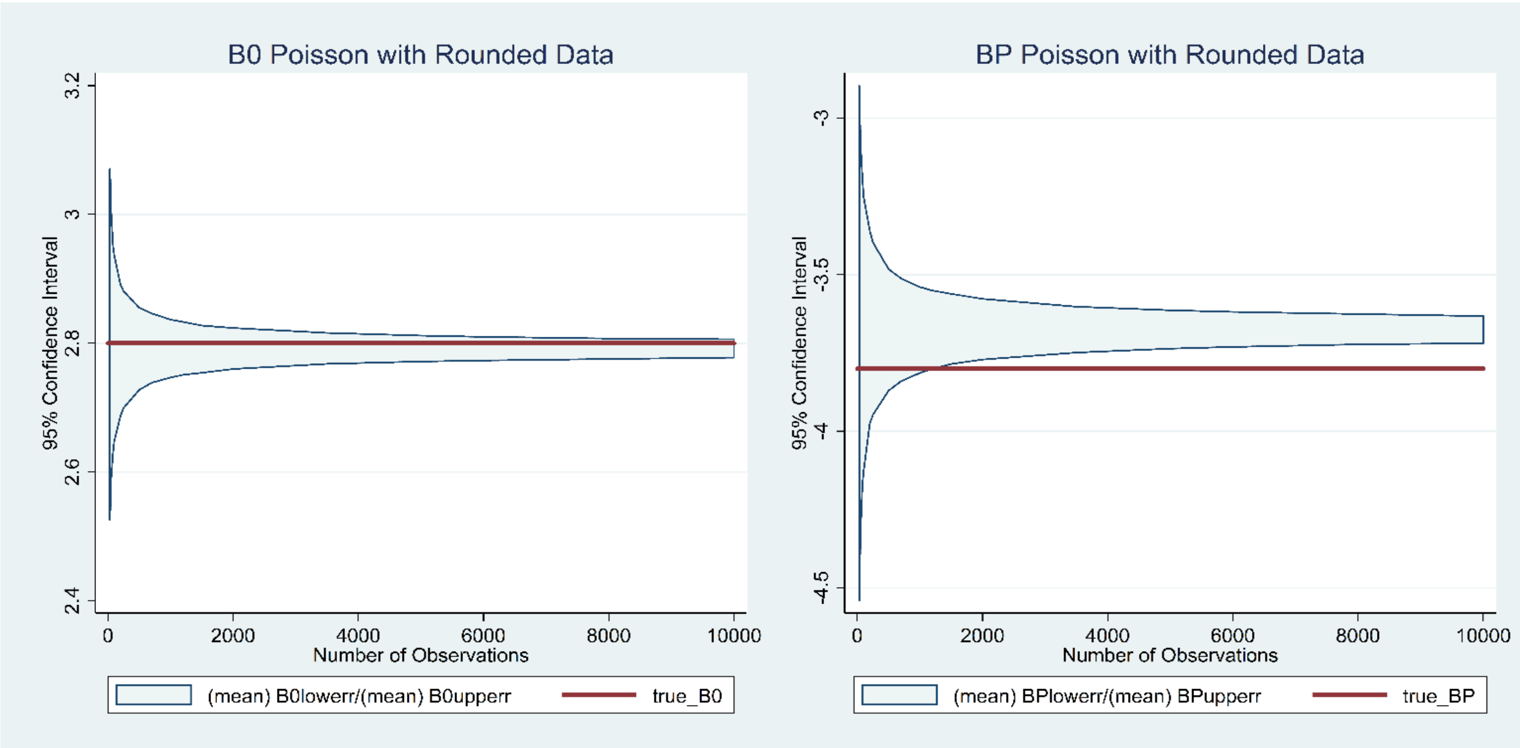

Figure 2. Confidence Intervals for Estimated Coefficients as a Function of Sample Size, Poisson Model

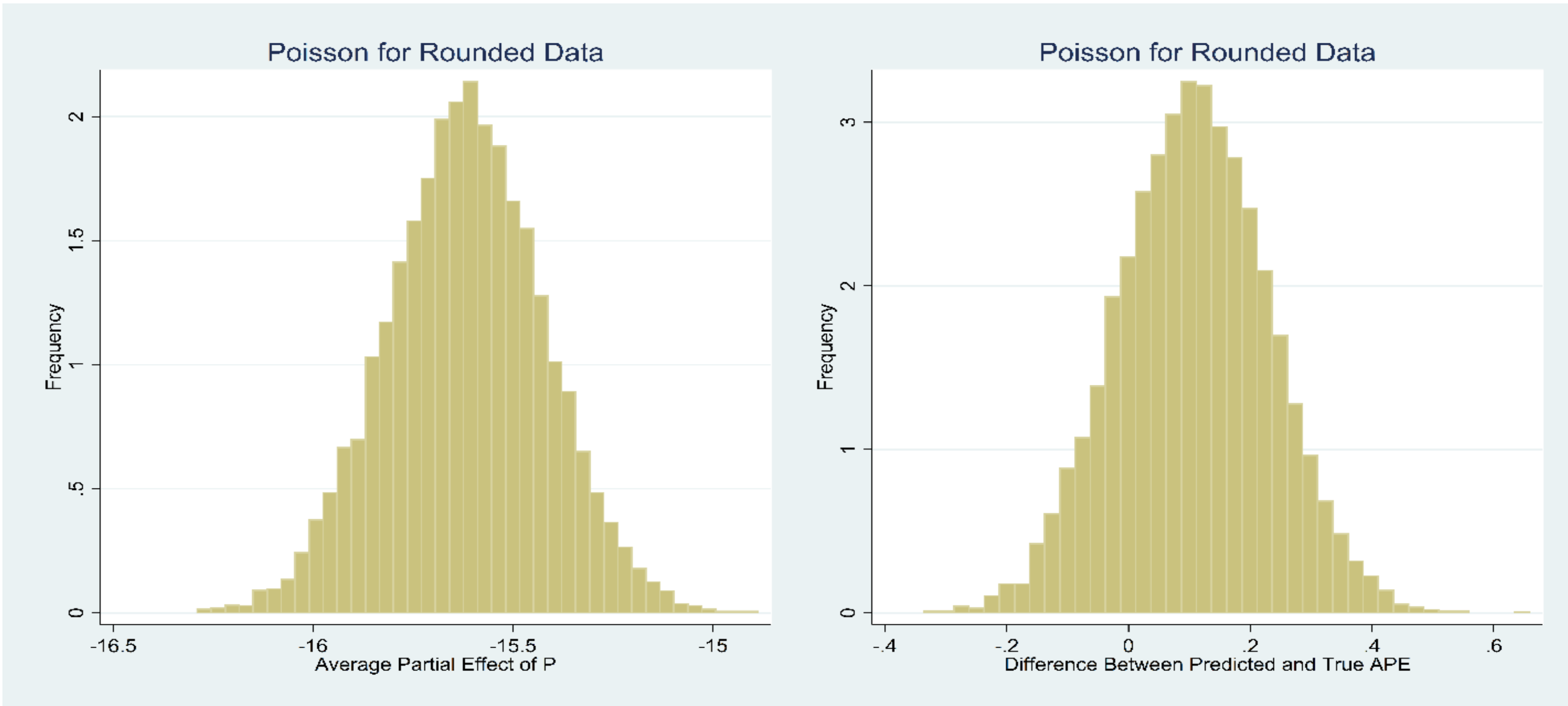

Figure 3. Bias in the Estimated Partial Effect of an Increase in Trip Cost, Poisson Model

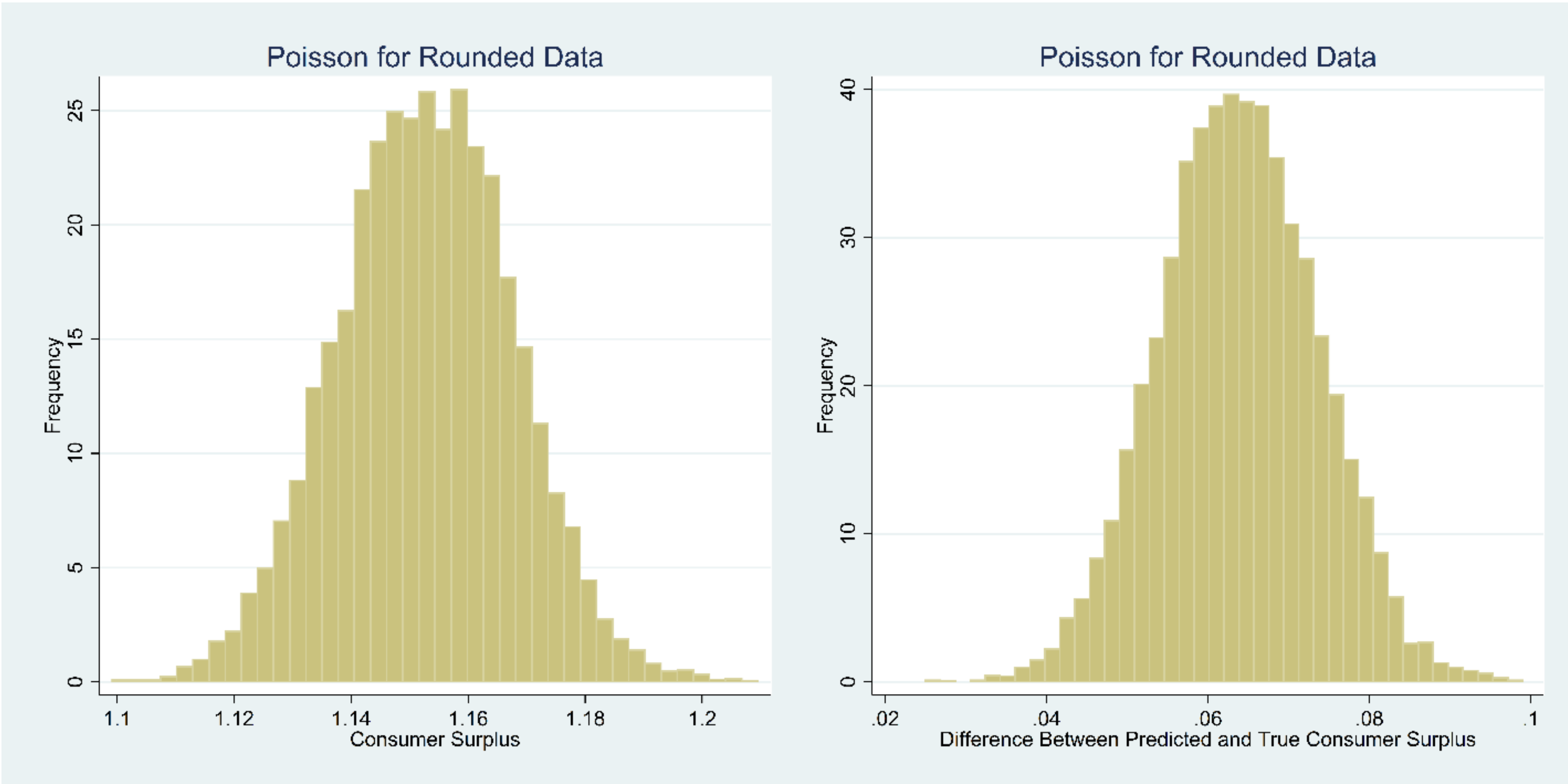

Figure 4. Bias in Estimated Average Consumer Surplus per Trip, Poisson Model

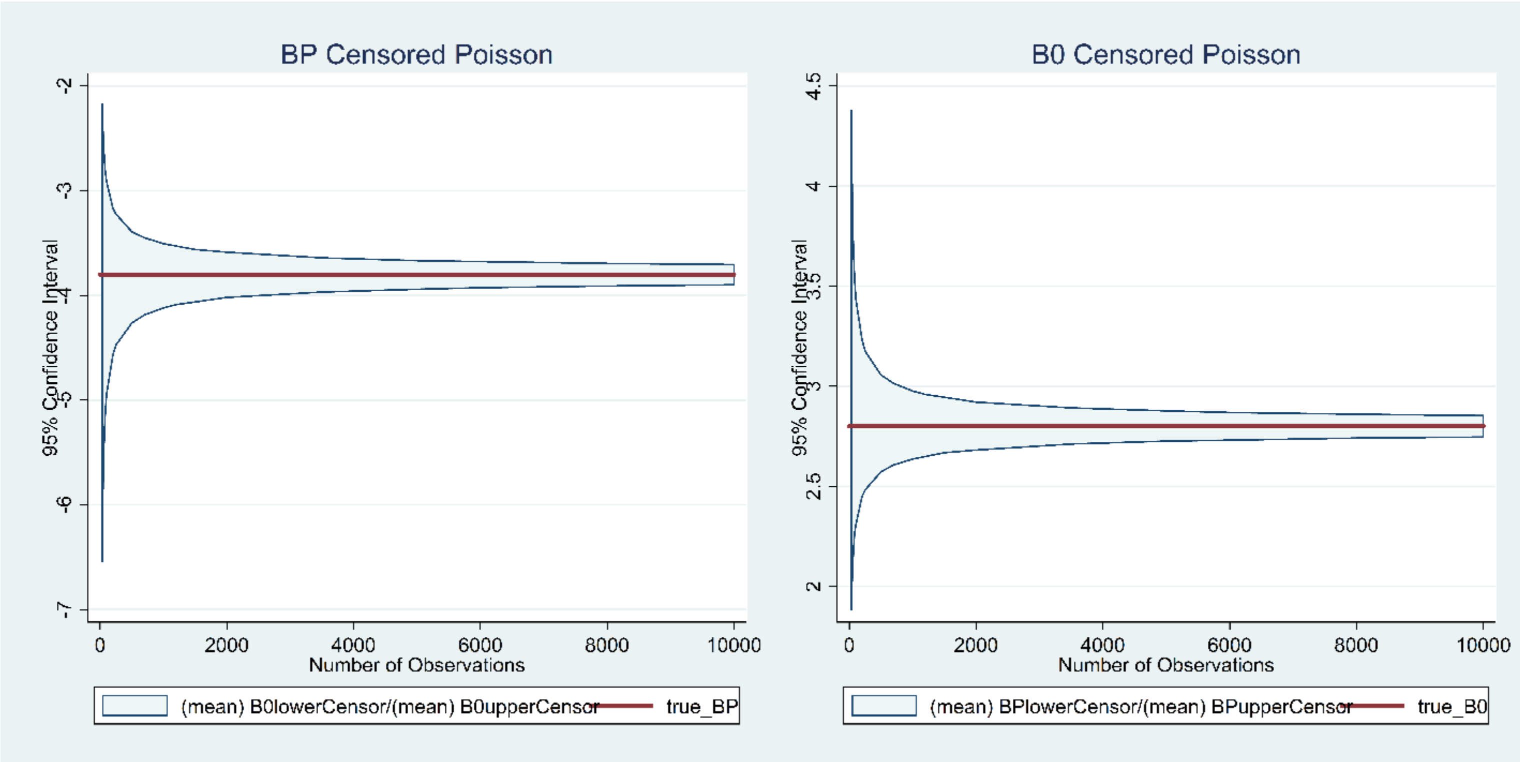

Figure 5. Confidence Intervals for Estimated Coefficients as a Function of Sample Size, Censored Poisson Model

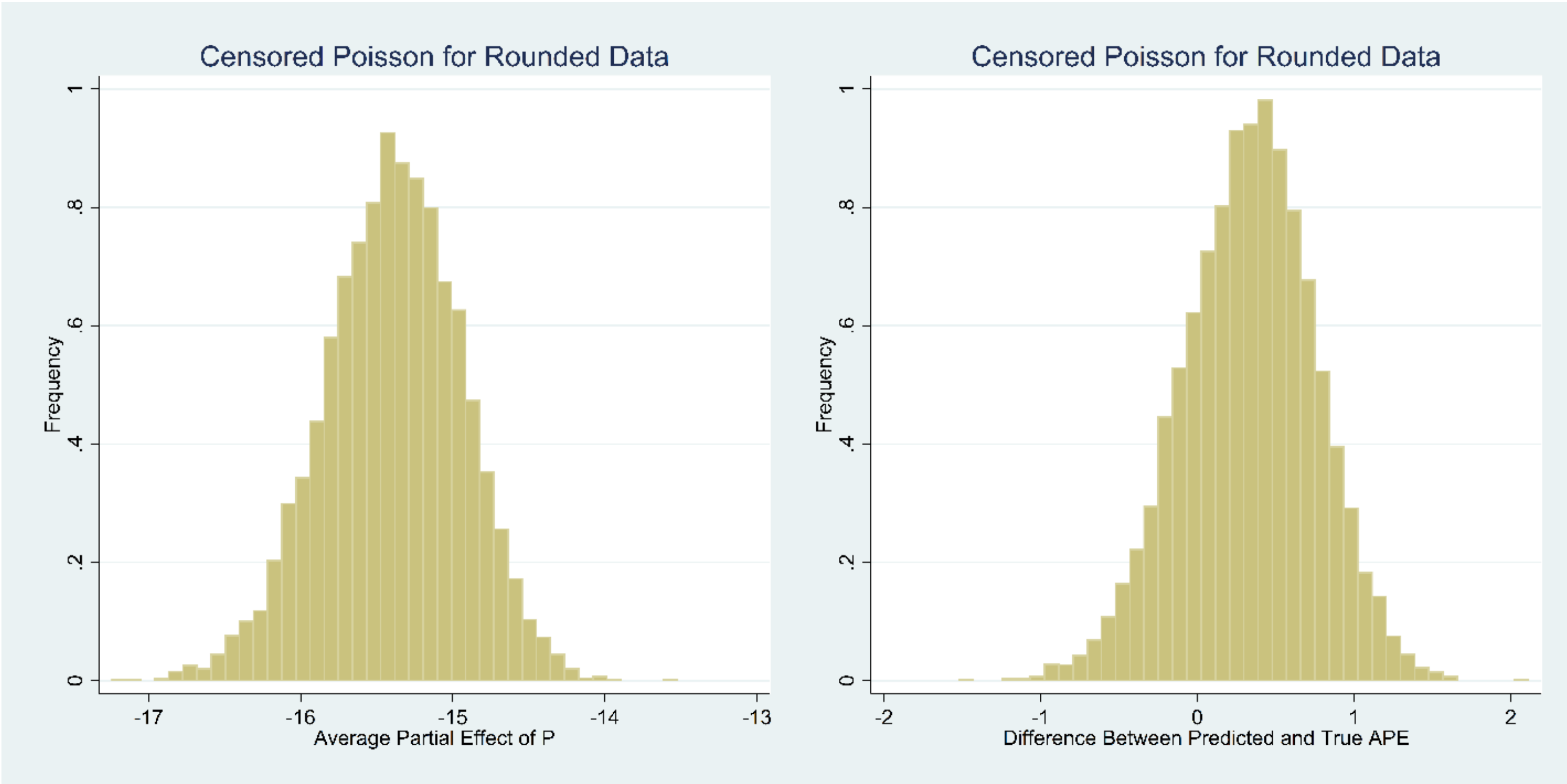

Figure 6. Bias in the Estimated Partial Effect of an Increase in Trip Cost, Censored Poisson Model

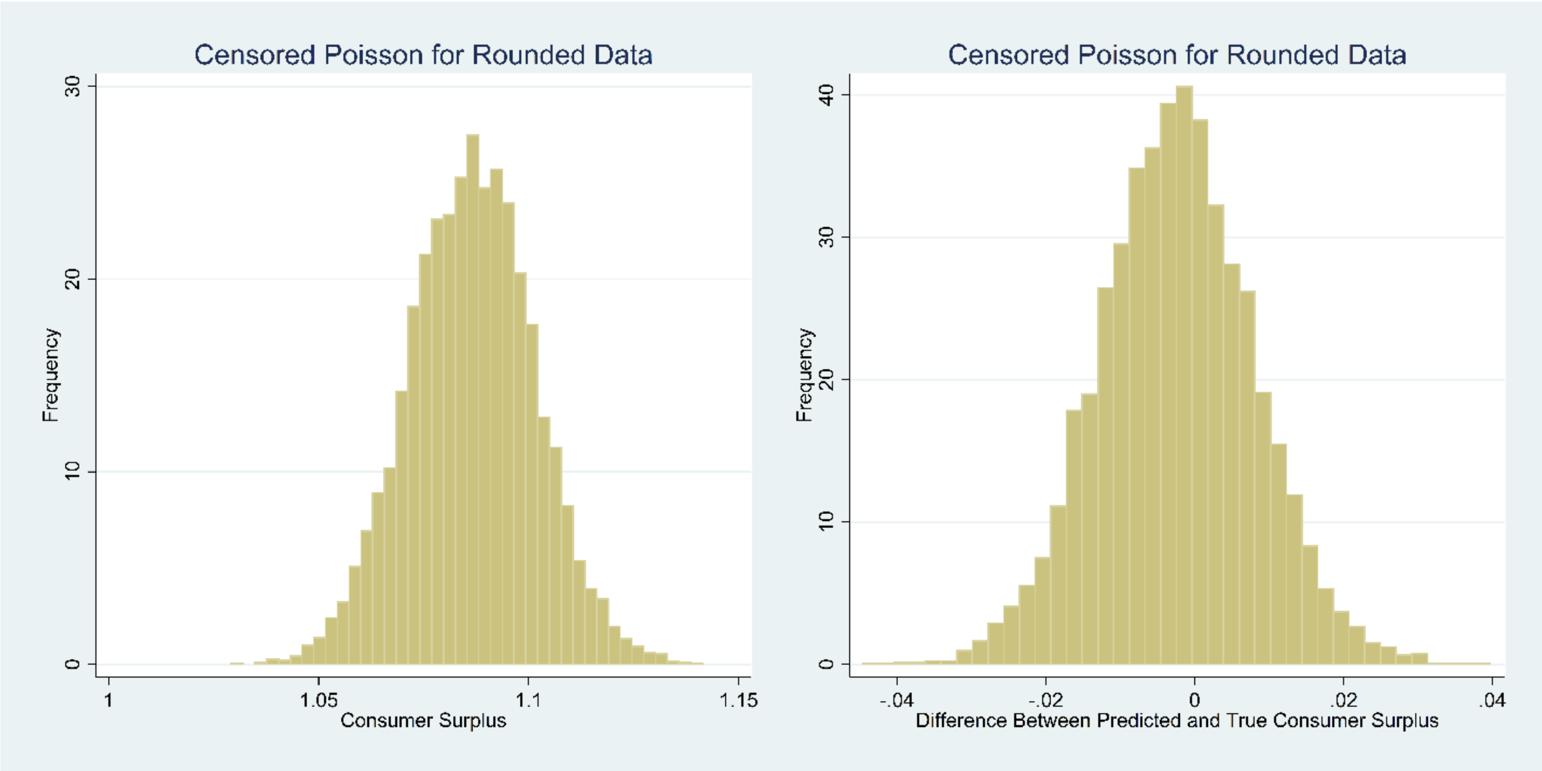

Figure 7. Bias in Estimated Average Consumer Surplus per Trip, Censored Poisson Model

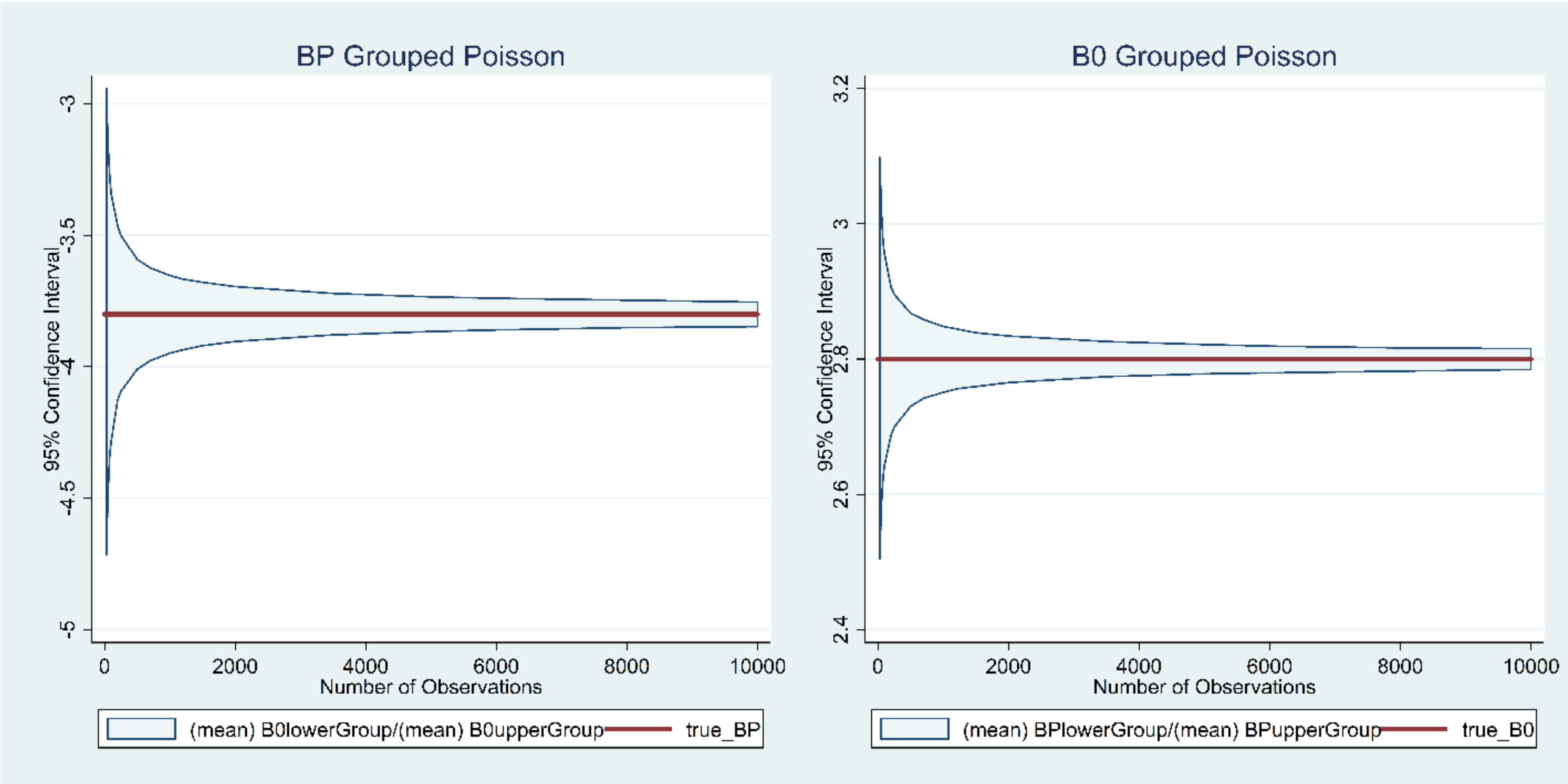

Figure 8. Confidence Intervals for Estimated Coefficients as a Function of Sample Size, Grouped Poisson Model

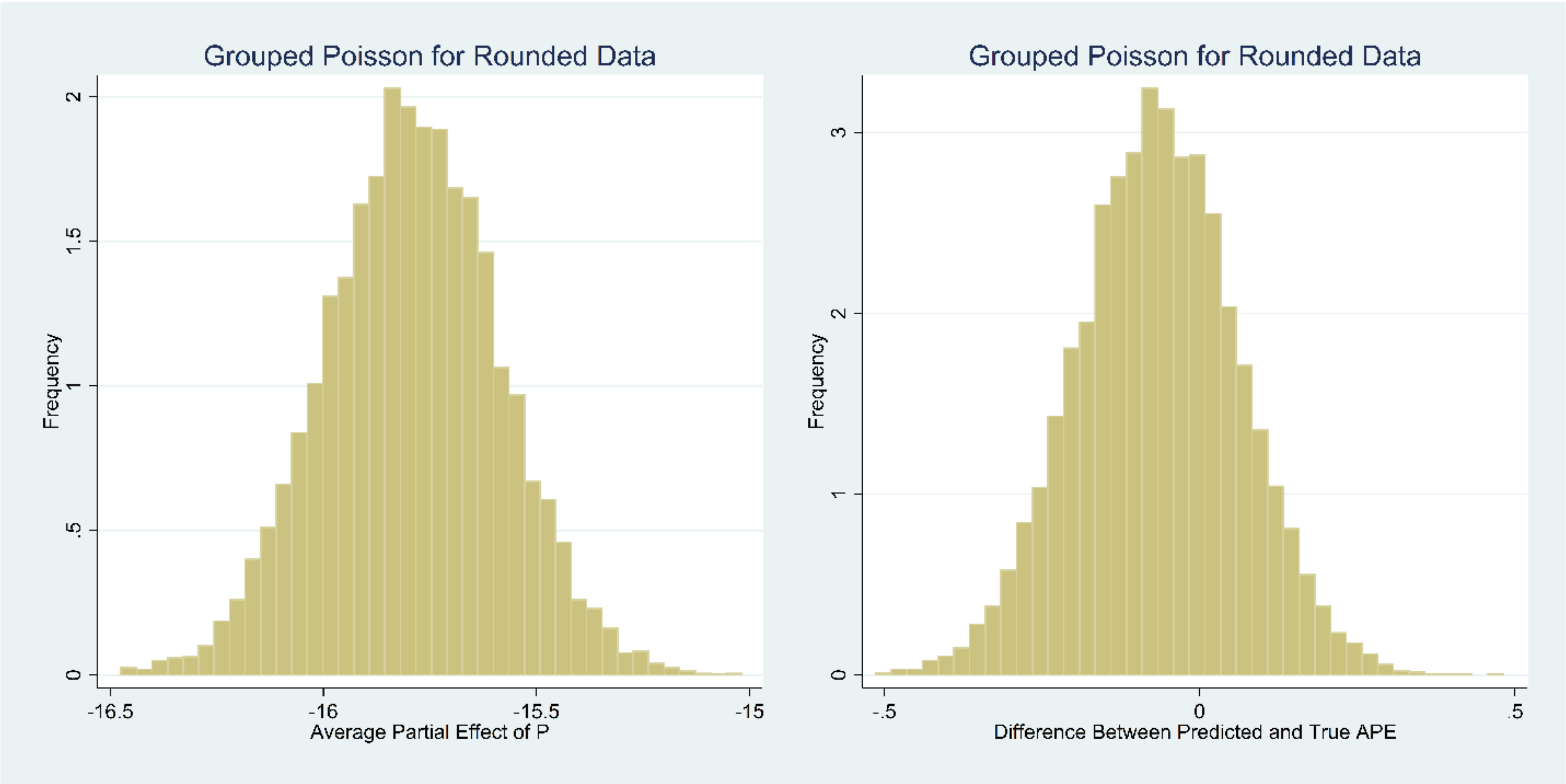

Figure 9. Bias in the Estimated Partial Effect of an Increase in Trip Cost, Censored Poisson Model

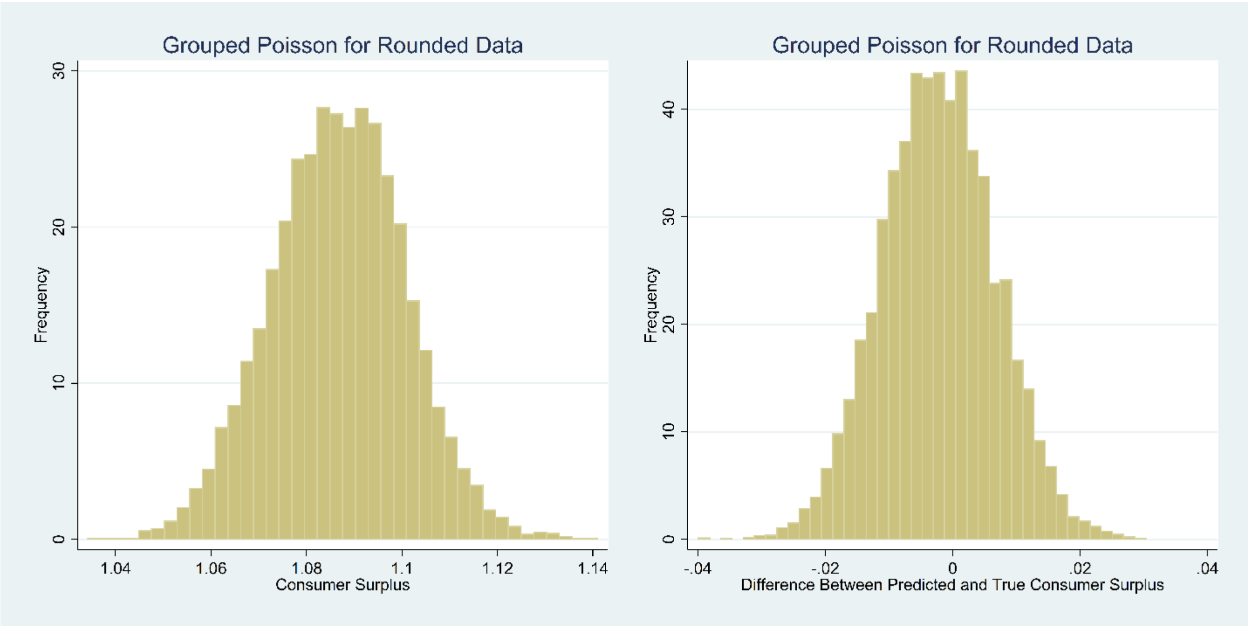

Figure 10. Bias in Estimated Average Consumer Surplus per Trip, Grouped Poisson Model