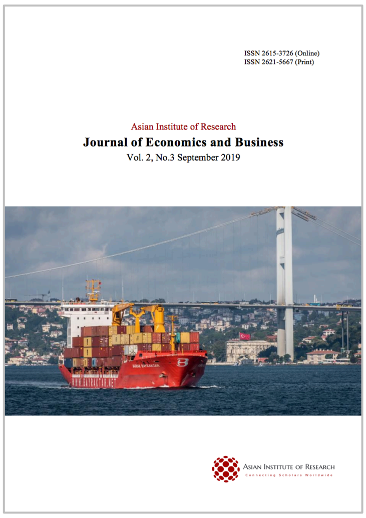


Journal of Economics and Business

Chi, Yeong Nain, and Chang, Guang-Hwa Andy. (2019), Modeling Trip Count Data with Excess Zeros for U.S. Saltwater Recreational Fishing. In: *Journal of Economics and* Business, Vol.2, No.3, 773-785.

ISSN 2615-3726

DOI: 10.31014/aior.1992.02.03.126

Published by:
The Asian Institute of Research

The Asian Institute of Research
Journal of Economics and Business
Vol.2, No.3, 2019: 773-785
ISSN 2615-3726

DOI: 10.31014/aior.1992.02.03.126

Modeling Trip Count Data with Excess Zeros for U.S. Saltwater Recreational Fishing

Yeong Nain Chi[1], Guang-Hwa Andy Chang[2]

[1] Department of Agriculture, Food and Resource Sciences, University of Maryland Eastern Shore, Princess Anne, MD 21853. Tel: 410-651-8186. E-mail: ychi@umes.edu
[2] Department of Mathematics and Statistics, Youngstown State University,Youngstown, Ohio 44555. Email: gchang@ysu.edu

Abstract
Count data, such as recreational fishing trips taken by anglers, is increasingly common in recreational fishing demand analysis. Because of the non-negative integer nature of the recreational fishing trip data, some over-dispersion problems, and truncation of the data at zero trips, count data models are more appropriate for estimating the recreational fishing demand function. This study employed count data models to analyze U.S. saltwater recreational fishing trips with excess zeros, using a cross-sectional data extracted from the 2011 National Survey of Fishing, Hunting, and Wildlife Associated Recreation. Using Akaike Information Criterion and Bayesian Information Criterion, the zero-truncated negative binomial model was selected among other count data models better fitted in this count data for this study. Empirical results of this study provide insight into the determinants of saltwater recreational fishing trips, which can be used in analyzing the social and economic values of saltwater recreational fisheries management.

Keywords: Count Data, Excess Zeros, Hurdle Poisson Model, Negative Binomial Model, Over-Dispersion, Poisson Model, Saltwater Recreational Fishing Trips, Zero-Inflated Models, Zero-Truncated Models

1. Introduction

Saltwater recreational fishing is a popular pastime across the nation that generates significant economic impacts both to local economies and to the nation. According to the report of the 2011 National Survey of Fishing, Hunting and Wildlife-Associate Recreation, saltwater recreational fishing attracted 8.9 million anglers who took 86.2 million trips in 99 million days. A total amount of $10.3 billion was spent on saltwater recreational fishing trips and equipment during that year. Expenditure on trip-related cost totaling $7.3 billion was the highest, Accommodation and food cost $2.4 billion, and transportation cost was $1.5 billion. Other miscellaneous cost such as guide fees, licenses, permits, bait, membership dues and equipment rental were $3.4 billion (U.S. Fish and Wildlife Service, 2014).

Saltwater recreational fishing is usually done with equipment such as rod, reel, bait, hook and line. It was estimated that anglers on equipment for saltwater recreational fishing spent a total of $2.9 billion. A detailed breakdown of this cost comprised of $1.4 billion on main fishing equipment (rod, reel, hook and line), $1.3 billion for special equipment (boats, travel vans etc.) and $217 million for auxiliary equipment (binoculars, camping equipment etc.) (U.S. Fish and Wildlife Service, 2014).

In 2011, saltwater recreational anglers spent an average of 11 days fishing and enjoyed an average of 10 trips. Saltwater recreational anglers spent an average of $824 per angler on trip related costs which was the highest average expenditure cost compared to average expenditure of freshwater recreational anglers and great lake recreational anglers, an average of $74 per day (U.S. Fish and Wildlife Service, 2014).

The most commonly sought fish among saltwater recreational anglers are striped bass, flatfish, redfish, sea trout, bluefish, salmon and mackerel. According to the 2011 National Survey of Fishing, Hunting and Wildlife-Associate Recreation, 2.1 million saltwater recreational anglers fished for striped bass for 18 million days, 2 million anglers fished for flatfish for 22 million days. 1.5 million Anglers fished for redfish for 21 million days and 1.1 million saltwater recreational anglers fished for 15 million days (U.S. Fish and Wildlife Service, 2014).

A comparison of the 2001, 2006 and 2011 National Survey of Fishing, Hunting and Wildlife-Associate Recreation indicated the total number of saltwater recreational anglers decreased significantly from 9.5 million in 2001 to 7.7 million in 2006 and then increased to 8.9 million in 2011. Total expenditures on saltwater recreational fishing trip-related costs and equipment increased slightly from $8.4 billion in 2001 to $8.9 billion in 2006, and also increased to $10.3 billion from 2006 to 2011 (U.S. Fish and Wildlife Service, 2002, 2007, 2014).

Many studies neglect how best to model saltwater recreational fishing trips and get meaningful insights into the behavior of saltwater recreational anglers that affect their saltwater recreational fishing behavior and participation. This engender the reason for this study so as to provide guidelines and create awareness for the proper use of count data models that will lead to more accurate results and get a better understanding of saltwater recreational fishing trips. It may also contribute to a better understanding of current and future angler behavior of saltwater recreational fishing participation and consumption.

2. Count Data Models

In statistics, count data (i.e., saltwater recreational fishing trips) is a statistical data type, a type of data in which the observations can take only the non-negative integer values, and where these integers arise from counting rather than ranking. Particularly in the econometric literature, there has been considerable interest in models for count data that allow for excess zeros in the national survey, such as the 2011 National Survey of Fishing, Hunting and Wildlife-Associate Recreation.

When the statistical requirements are met, the standard ordinary least squares (OLS) technique could be used to estimate the saltwater recreational fishing demand function. Because of the non-negative integer nature of the recreational fishing trip data, some over-dispersion problems, and truncation of the data at zero trips, the standard OLS estimator may be inappropriate, but count data models are more appropriate for estimating the saltwater recreational fishing trips.

Count data modeling techniques have become important tools in empirical studies of economic research and their applicability continues to grow in various areas of economics, e.g., health economics (i.e., the number of doctor visits), labor economics (i.e., labor mobility), financial economics (i.e., the number of reported claims, the number of bank failures), industrial organization (i.e., entry and exit in industries), transportation (i.e., the number of car accidents), and tourism and outdoor recreation (i.e., the number of fishing trips).

The count data models considered in this study including Poisson model, negative binomial model, zero-inflated Poisson (ZIP) model, zero-inflated negative binomial (ZINB) model, zero-truncated Poisson (ZTP) model, zero-truncated negative binomial (ZTNB) model, and hurdle Poisson Model. These count data models were also compared against each other using Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC).

2.1 Poisson Model

The Poisson model is based on the Poisson distribution, was discovered by Simon Denis Poisson and published together with his probability theory in 1838 in his work "Research on the Probability of Judgments in Criminal and Civil Matters" (Good, 1986). In probability theory and statistics, the Poisson distribution is a discrete probability distribution that expresses the probability of a number of events occurring in a fixed period of time.

Poisson distribution is a probability distribution for count data that satisfies the discrete probability distribution (Greene, 2008) represented by,

$$P(Y = y \mid \lambda) = \frac{e^{-\lambda}\lambda^{y}}{y!}$$

where Y is a random variable having Poisson distribution and $\boldsymbol{\lambda}$ is the mean of the distribution. The Poisson regression model is similar to regular multiple regression model except that the dependent variable, y, is an observed count from the Poisson distribution. The most common formulation for λ is the log-linear model,

$$ln(\boldsymbol{\lambda}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k = \mathbf{x}^{\mathbf{T}}\boldsymbol{\beta} \ .$$

The parameter λ is both the mean and variance of the random response variable Y and depends on a set of k explanatory variables, $x_1, x_2, \ldots x_k$, in vector **x**. A sample of observations can be considered as dependent variable vector **Y=** (Y_1, $Y_{2, \ldots,}$ Y_n) that each Y_i is distributed independent Poisson (λ_i) where the expected count of Y_i is E(Y) = λ, i.e.,

$$\lambda = e^{\mathbf{x}^{\mathbf{T}}\boldsymbol{\beta}}, \text{ with } \mathrm{E}(Y \mid \mathbf{x}) = \mathrm{Var}(Y \mid \mathbf{x}) = e^{\mathbf{x}^{\mathbf{T}}\boldsymbol{\beta}}.$$

The Poisson model typically is assumed for count data, but when there are many zeros in the response variable, the mean is not equal to the variance value of the dependent variable, because of over-dispersion, the negative binomial model is suggested instead of the Poisson model.

2.2 Negative Binomial (NB) Model

The negative binomial distribution, like the Poisson distribution, describes the probabilities of the occurrence of whole numbers greater than or equal to zero. Unlike the Poisson distribution, the variance and the mean are not equivalent. This suggests it might serve as a useful approximation for modeling count data with variability different from its mean and it enables the model to have greater flexibility in modeling the relationship between the conditional variance and the conditional mean compared to the Poisson model. The negative binomial model (Greene, 2008) can be expressed as,

$$\mathrm{P}(Y = y \mid \lambda, \alpha) = \frac{\Gamma(y + \alpha^{-1})}{\Gamma(y+1)\Gamma(\alpha^{-1})}\left(\frac{1}{1+\alpha\lambda}\right)^{\alpha^{-1}}\left(\frac{\alpha\lambda}{1+\alpha\lambda}\right)^{y} .$$

where $\lambda = e^{\mathbf{x}^{\mathbf{T}}\boldsymbol{\beta}}$ and α is the dispersion parameter. As the value of dispersion parameter increases, the variance converges to the same value as the mean, and the negative binomial distribution turns into a Poisson distribution. The conditional mean and variance of the negative binomial distribution are,

$$\mathrm{E}(Y \mid \mathbf{x}) = \lambda = e^{\mathbf{x}^{\mathbf{T}}\boldsymbol{\beta}} \text{ and } \mathrm{Var}(Y \mid \mathbf{x}) = \lambda(1 + \alpha\lambda) > \mathrm{E}(Y \mid \mathbf{x}).$$

The Poisson distribution has only one parameter (λ), whereas the negative binomial distribution has two parameters (λ and α). Due to this property, the negative binomial model is more flexible than the Poisson model. Moreover, the Poisson model should have the same mean and variance value and this is not what happens in the real data.

Thus, the negative binomial model can be used instead of the Poisson model when the count data under consideration is over-dispersed.

The other problem with the Poisson and negative binomial models having far more zeros than expected by the distribution assumptions of the Poisson and negative binomial models result in incorrect parameter estimates (Hardin and Hilbe, 2012). Using the zero-inflated models, such as the zero-inflated Poisson model (Lambert, 1992) or the zero-inflated negative binomial model (Hall, 2000), are proposed as a solution for this problem (Loeys et al., 2012). The zero-Inflated models attempt to account for excess zeros, i.e., there is thought to be two kinds of zeros, “true zeros” and “excess zeros”. Therefore, the zero-Inflated models estimate two equations, one for the count data and one for the excess zeros.

2.3 Zero-Inflated Poisson (ZIP) Model

The zero-inflated Poisson model (ZIP) is used to model count data that has an excess of zero counts. Mullahy (1986), Heilbron (1994) and Lambert (1992) pioneered the use of regression model based on the ZIP distribution. Further, theory suggests that the excess zeros are generated by a separate process from the count values and that the excess zeros can be modeled independently. Thus, the ZIP model has two parts, a Poisson count data model and a Logit model for predicting excess zeros (Lambert, 1992). In general, the random variable Y takes on 0 with probability ω and Y takes on a value from Poisson (λ) with probability $(1 - \omega)$, the probability mass function of the random variable Y representing the ZIP model can be expressed as:

$$P(Y = y \mid \lambda) = \begin{cases} \omega + (1-\omega)e^{-\lambda} \ , & \text{if } y = 0, \\ (1-\omega)\left(\dfrac{e^{-\lambda}\lambda^{y}}{y!}\right) \ , & \text{if } y > 0 . \end{cases}$$

The mean and variance of the ZIP model are:

$E(Y \mid \mathbf{x}) = \lambda(1 - \omega)$ and $\mathrm{Var}(Y \mid \mathbf{x}) = \lambda(1 - \omega)(1 + \omega\lambda)$.

If the variance displayed above is greater than the mean, it indicates over-dispersion and the ZIP model is not appropriate in such instance. In such cases the zero-inflated negative binomial model is fitted. Moreover, the non-zero observations may be over-dispersed in relation to the Poisson distribution, biasing parameter estimates and underestimating standard errors. In such a circumstance, the zero-inflated negative binomial model better accounts for these characteristics compared to the ZIP model.

2.4 Zero-Inflated Negative Binomial (ZINB) Model

The zero-inflated negative binomial (ZINB) model has been used for modeling both zero-inflation and over-dispersion in count data. Furthermore, theory suggests that the excess zeros are generated by a separate process from the count values and that the excess zeros can be modeled independently. Greene (1994) gives details of analogous ZINB model. The ZINB distribution is a mixture of a binary distribution that is degenerate at zero and an ordinary negative binomial distribution (Hall, 2000).

With probability ω, the response of the first process is a zero count, and with probability of $(1-\omega)$, the response of the second process is governed by a negative binomial with mean $\boldsymbol{\lambda}$ and can also generate zero counts. The overall probability of zero counts is the combined probability of zeros from the two processes. Thus, the ZINB model for the response y can be written as:

$$P(Y = y \mid \lambda, \alpha) = \begin{cases} \omega + (1-\omega)\left(\dfrac{1}{1+\alpha\lambda}\right)^{\alpha^{-1}} \ , & \text{if } y = 0, \\ (1-\omega)\dfrac{\Gamma(y+\alpha^{-1})}{\Gamma(y+1)\Gamma(\alpha^{-1})}\left(\dfrac{1}{1+\alpha\lambda}\right)^{\alpha^{-1}}\left(\dfrac{\alpha\lambda}{1+\alpha\lambda}\right)^{y} \ , & \text{if } y > 0 . \end{cases}$$

where α (≥ 0) is a dispersion parameter that is assumed not to be depend on covariates. The mean and variance of the ZINB model are:

$$\mathrm{E}(Y \mid \mathbf{x}) = \lambda(1-\omega) \text{ and } \mathrm{Var}(Y \mid \mathbf{x}) = \lambda(1-\omega)(1+\lambda(\omega+\alpha)).$$

In many situations, observed samples may be considered as zero truncated. These kinds of data should be analyzed by a zero-truncated distribution, an alternative distribution to the count data with zero counts cannot be observed. The zero-truncated distribution is a certain distribution having support the set of positive integers. This distribution is applicable for the situations when the data to be modeled originate from a mechanism that generates data excluding zero counts, such as the zero-truncated Poisson distribution (David and Johnson, 1952) and the zero-truncated negative binomial distribution (Sampford, 1955).

2.5 Zero-Truncated Poisson (ZTP) Model

In probability theory, the zero-truncated Poisson (ZTP) distribution is a certain discrete probability distribution whose support is the set of positive integers. This distribution is also known as the conditional Poisson distribution (Cohen, 1960) or the positive Poisson distribution (Singh, 1978). It is the conditional probability distribution of a Poisson distributed random variable, given that the value of the random variable is not zero. Thus, it is impossible for a ZTP random variable to be zero.

Since the ZTP is a truncated distribution with the truncation stipulated as $y > 0$, one can derive the probability mass function from a standard Poisson distribution as follows:

$$\mathrm{P}(Y = y \mid Y > 0) = \frac{e^{-\lambda}\lambda^{y}}{(1-e^{-\lambda})y!}\ .$$

The mean and variance of the ZTP model are:

$$\mathrm{E}(Y \mid Y > 0) = \frac{\lambda}{1-e^{-\lambda}}$$

and

$$\mathrm{Var}(Y \mid Y > 0) = \frac{\lambda(1+\lambda)}{1-e^{-\lambda}} - \frac{\lambda^{2}}{(1-e^{-\lambda})^{2}}\ .$$

2.6 Zero-Truncated Negative Binomial (ZTNB) Model

Given the importance of accounting for over-dispersion in truncated count context, the zero-truncated negative binomial (ZTNB) model is the appropriate model for the analysis of such count data. The ZTNB model is used to model count data for which the value zero cannot occur and for which over-dispersion exists. A detailed discussion of the ZTNB model can be found in Gurmu (1991) and Grogger and Carson (1991). The ZTNB model for the response y can be written as:

$$\mathrm{P}(Y = y \mid Y > 0) = \frac{\Gamma(y+\alpha^{-1})}{\Gamma(\alpha^{-1})y!}\left(\frac{\alpha\lambda}{1+\alpha\lambda}\right)^{y}\frac{(1+\alpha\lambda)^{-\alpha^{-1}}}{\left(1-(1+\alpha\lambda)\right)^{-\alpha^{-1}}}\ .$$

The conditional mean and conditional variance of the ZTNB model are:

$$\mathrm{E}(Y \mid Y > 0) = \frac{\lambda}{1-(1+\alpha\lambda)^{-\alpha^{-1}}}$$

and

$$\mathrm{Var}(Y \mid Y > 0) = \frac{\lambda}{1-(1+\alpha\lambda)^{-\alpha^{-1}}} \frac{1+\alpha\lambda-\lambda(1+\alpha\lambda)^{-\alpha^{-1}}}{1-(1+\alpha\lambda)^{-\alpha^{-1}}}.$$

2.7 Hurdle Poisson Model

A hurdle model (Mullahy, 1986), or two-part model (Heilbron, 1994), is a modified count model in which there are two processes, one generating the zeros and one generating the positive values. The two models are not constrained to be the same. The concept underlying the hurdle model is that a binomial probability model governs the binary outcome of whether a count variable has a zero or a positive value (Shonkwiler and Shaw, 1996). If the value is positive, the hurdle is crossed, and the conditional distribution of the positive values is governed by a zero-truncated count model.

The differences between the hurdle models and the zero-inflated models are that zero and non-zero counts are separately modeling in the hurdle models (Loeys et al., 2012), and also hurdle models assumes that all zero counts are true zeros (Potts and Elith, 2006). The hurdle Poisson model with count variable $\boldsymbol{y}$ has the distribution as:

$$P(Y = y \mid \lambda, \omega) = \begin{cases} \omega & , \quad \text{if } y = 0, \\ \dfrac{(1-\omega)e^{-\lambda}\lambda^{y}}{(1-e^{-\lambda})y!} & , \quad \text{if } y > 0 . \end{cases}$$

If $y > 0$ means the hurdle is crossed then the conditional distribution of the non-zero values is managed by a zero truncated model, $\boldsymbol{\lambda}$ is the mean of the Poisson distribution, and $\boldsymbol{\omega}$ is the probability value of the zero counts. For estimating the parameter values, maximum likelihood method is used. The hurdle Poisson model is nothing but a re-parameterization of the ZIP model. Although for both models the parameters are modeled in the regression framework, hurdle Poisson model is not the same as the ZIP model.

3. Data

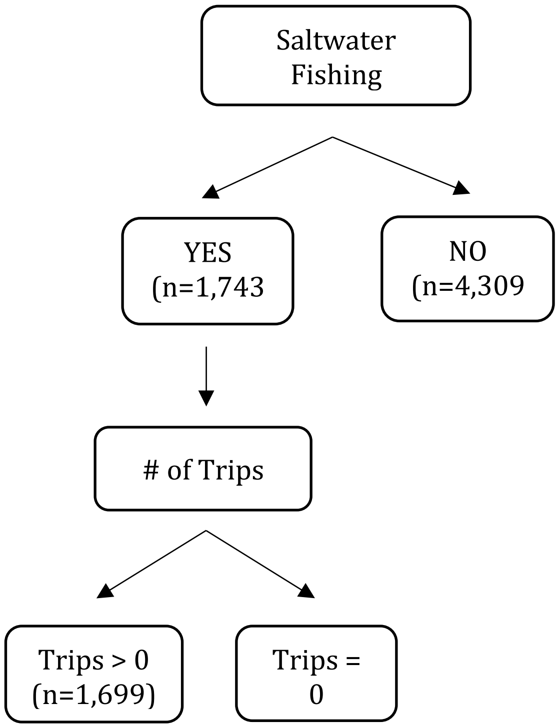
Data used for this study was extracted from the 2011 National Survey of Fishing, Hunting, and Wildlife-Associated Recreation (U.S. Fish and Wildlife Service, 2014), which is developed by the U.S. Fish and Wildlife Service every five years. This type of survey started in 1955 and the 2011 survey is the 12th of its kind. It is one of the comprehensive and most reliable recreation surveys in the United States. Basically, the survey aims to collect information on the frequency of participation and expenditure on fishing activities in the United States as well as the number of anglers, hunters and wildlife watchers.

Data was collected for the survey by the U.S Census Bureau in two phases namely the screening phase and three-detailed wave process. In the screening phase, the U.S. Census Bureau interviewed a sample of 48,600 households in the United States to identify respondents who had participated in wildlife-related activities in the year of 2011 to gather information on fishing, hunting, and wildlife watching participation, expenditures, and socioeconomic characteristics of respondents. Mostly, one adult household member provided information for all members.

The second phase of the survey consisted of three detailed interviews. Interviews were conducted with people who were at least 16 years who were chosen from the screening phase. According to the report from the 2011 National Survey of Fishing, Hunting and Wildlife-Associated Recreation, most of the respondents were interviewed by phone whiles in-person interviews was used for those who couldn't be reached on phones. From this initial phase, 6,052 saltwater recreational anglers were selected for a detailed interview about their participation and expenditures associated with saltwater recreational fishing activities in the United States in 2011.

From this initial phase, 6,052 saltwater recreational anglers were selected for a detailed interview about their participation and expenditures associated with saltwater recreational fishing activities in the United States in 2011, based on the question "Respondent fished in saltwater in the United States in 2011?" and "Total saltwater fishing trips respondent took in the United States in 2011?" (Figure 1).

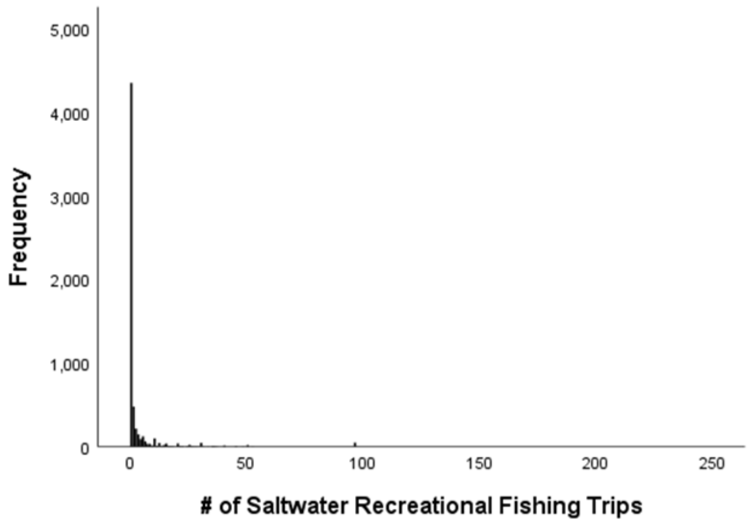
Figure 1. U.S. Saltwater Recreational Fishing Trips



Source: 2011 National Survey of Fishing, Hunting, and Wildlife-Associated Recreation

Figure 2 also shows a sharp decreasing trend of the distribution with most of the count data being zero. Average number of saltwater recreational fishing trips was 3.07 for the total sample (n = 6,052), but average number of saltwater recreational fishing trips was 10.94 for the sample with non-zero cases (n = 1,699). About 72% of respondents reported zero trip (n = 4,353). It could have the possibility of over-dispersion consideration in this count data. In addition, the high percentage of zeros seen from the histogram suggests an issue of excess zeros and/or a zero inflated data.

Figure 2. Distribution of U.S. Saltwater Recreational Fishing Trips



Source: 2011 National Survey of Fishing, Hunting, and Wildlife-Associated Recreation

4. Empirical Results

An analysis of recreational saltwater fishing trips can be beneficial from the use of appropriate econometric analysis and measurement to comprehend the full value of this type of saltwater recreational fishing activities within the framework of saltwater recreational fisheries management and policy. In particular, analyzing saltwater recreational fishing trips in the framework of an angler who allocate the constrained budget to maximize utility improves a better understanding of the tradeoffs made in this process.

According to consumer demand theory, angler attempts to maximize his/her utility from saltwater recreational fishing activities subject to his/her budget constraint. Thus, the maximization of the utility function for saltwater recreational fishing activities can be stated as follows:

$$\text{Max}_{y,\,z} \quad u = u(\boldsymbol{y}, \boldsymbol{z} \mid \boldsymbol{a}, \boldsymbol{s}) \qquad \text{subject to} \qquad \boldsymbol{py} + \boldsymbol{qz} = \boldsymbol{I}$$

where $\boldsymbol{u}(.)$ represents the utility function which is assumed to be continuous, increasing, and quasi-concave, $\boldsymbol{y}$ is quantity demanded of saltwater recreational fishing activities, i.e. the number of saltwater recreational fishing trips to the fishing site, $\boldsymbol{z}$ represents the quantity of all other goods consumed, $\boldsymbol{a}$ is the vector of exogenous attributes of the activity or site, $\boldsymbol{s}$ is the vector of socioeconomic characteristics, $\boldsymbol{p}$ is travel cost of saltwater recreational fishing trips, $\boldsymbol{q}$ is the vector of prices of other goods and services, and $\boldsymbol{I}$ is income.

Then the angler's demand function for saltwater recreational fishing activities can be expressed in terms of saltwater recreational fishing trips (Zawacki et al., 2000) as follows:

$$\boldsymbol{Y} = f(\mathbf{x}, \boldsymbol{\beta}, \varepsilon)$$

where $\boldsymbol{Y}$ is the vector of the dependent variable representing the number of saltwater recreational fishing trips to the fishing site, $\mathbf{x}$ is the vector of independent variables such as travel cost, travel time, socioeconomic factors, and trip characteristics, $\boldsymbol{\beta}$ is a vector of parameters including, but not limited to, the estimated coefficients of the independent variables, and $\boldsymbol{\varepsilon}$ is the vector of the random error term assumed to be independent and identically distributed.

This study employed the count data models to analyze U.S. saltwater recreational fishing trips with excess zeros, using a cross-sectional data extracted from the 2011 National Survey of Fishing, Hunting, and Wildlife Associated Recreation (U.S. Fish and Wildlife Service, 2014). The choice of explanatory variables selected for this empirical analysis was based on the conceptual model of saltwater recreational fishing activities, integrating anglers and fisheries resources and habitats. This conceptual model demonstrates the context of the human-fisheries interaction and provides a framework that identifies utility maximization as the ultimate objective for the anglers in saltwater recreational fishing activities in terms of their participation decisions.

For this study, the dependent variable is the number of saltwater recreational fishing trips, and the explanatory variables include age, household income, male, graduate or professional degree, minority, living in the urban settings, salmon, striped bass, bluefish, flatfish, redfish, sea trout, mackerel, marlin, tuna, mahi-mahi, and shellfish. Description and Descriptive statistics of all parameters used in this empirical analysis are presented in Table 1.

Table 1. Parameter Description and Descriptive Statistics of Count Data Models

Parameter	Description	Mean	Standard Deviation
FISHING TRIPS	The number of saltwater recreational fishing trips	3.070	11.570
AGE	Respondent's age (in year; 16 years old and older)	46.596	16.085
HOUINC	1 if respondent's household income greater than $50,000; 0 otherwise	0.559	0.497
MALE	Respondent's gender; 1 if male; 0 otherwise	0.741	0.438

GRADUATE	Respondent's education level; 1 if graduate or professional degree; 0 otherwise	0.121	0.326
MINOR	Respondent's ethnicity; 1 if minority; 0 otherwise	0.134	0.340
URBAN	1 if respondent lived in the urban settings; 0 otherwise	0.521	0.500
SALMON	1 if Salmon was one type of targeted fish; 0 otherwise	0.023	0.151
SBASS	1 if Striped Bass was one type of targeted fish; 0 otherwise	0.079	0.270
BLUEFISH	1 if Bluefish was one type of targeted fish; 0 otherwise	0.039	0.194
FLATFISH	1 if Flounder, Flatfish, or Halibut was one type of targeted fish; 0 otherwise	0.064	0.245
REDFISH	1 if Red Drum (Redfish) was one type of targeted fish; 0 otherwise	0.031	0.173
SEATROUT	1 if Sea Trout (Weakfish) was one type of targeted fish; 0 otherwise	0.025	0.157
MACKEREL	1 if Mackerel was one type of targeted fish; 0 otherwise	0.018	0.132
MARLIN	1 if Marlin was one type of targeted fish; 0 otherwise	0.007	0.081
TUNA	1 if Tuna was one type of targeted fish; 0 otherwise	0.018	0.131
MAHI-MAHI	1 if Dolphin (Mahi-Mahi) was one type of targeted fish; 0 otherwise	0.016	0.124
SHELLFISH	1 if Shellfish was one type of targeted fish; 0 otherwise	0.027	0.162

Source: Computed by authors using 2011 National Survey of Fishing, Hunting, and Wildlife-Associated Recreation

The software used in this empirical analysis is SAS version 9.4 Window environment. There are various built in procedures including GENMOD, COUNTREG, FMM, and NLMIXED that are commonly used in SAS count data analysis. The NLMIXED Procedure was used in fitting the count data models for this study. Empirical results of the count data models for U.S. saltwater recreational fishing trip are presented in Table 2.

The parameter estimates of count data models with the exception of ZIP, ZINB and Hurdle models have the same interpretation. Each regression coefficient is interpreted as the change in log count in the response variable per unit change in the predictor variable. Another way is to take exponential function of the parameter estimates before interpreting them.

It is often a good practice to compare count data models for the purpose of making the right decision on the one that best fits the data. The most common goodness of fit statistics used in comparison of count data models are Akaike information criterion (AIC), -2 Log-Likelihood + $2k$ where k = number of parameters, and Bayesian Information Criterion (BIC), -2 Log-Likelihood + $k\ ln(n)$ where n = the number of observations. In general, the smaller AIC and/or BIC value refers to the better model.

Table 2. Parameter Estimates (Standard Errors) for Count Data Models

Parameter	Poisson	NB	ZIP	ZINB	ZTP	ZTNB	Hurdle
INTERCEPT	-0.4074 (0.0325)	-1.4900 (0.1398)	1.4798 (0.0342)	1.2562 (0.1300)	-1.2480 (0.0420)	-1.2523 (0.6810)	1.4790 (0.0341)
AGE	0.0055 (0.0005)	0.0106 (0.0023)	0.0026 (0.0005)	0.0059 (0.0020)	0.0083 (0.0006)	0.0097 (0.0036)	0.0026 (0.0005)
HOUINC	0.0621 (0.0156)	-0.0812 (0.0717)	-0.1148 (0.0155)	-0.1259 (0.0621)	0.1163 (0.0186)	-0.1707 (0.1140)	-0.1142 (0.0155)
MALE	0.2091 (0.0198)	0.1277 (0.0801)	0.2428 (0.0198)	0.2399 (0.0689)	0.3041 (0.0254)	0.3580 (0.1209)	0.2432 (0.0198)
GRADUATE	-0.1548 (0.0234)	-0.2395 (0.1109)	-0.3759 (0.0236)	-0.4497 (0.0815)	-0.1544 (0.0271)	-0.6578 (0.1445)	-0.3765 (0.0235)

MINOR	0.8375 (0.0180)	1.4513 (0.0989)	0.4929 (0.0181)	0.5904 (0.0784)	1.0944 (0.0202)	0.8590 (0.1484)	0.4928 (0.0181)
URBAN	0.4068 (0.0159)	0.5890 (0.0701)	0.1985 (0.0159)	0.1974 (0.0603)	0.6148 (0.0197)	0.2700 (0.1084)	0.1981 (0.0159)
SALMON	0.9607 (0.0290)	1.8177 (0.2112)	0.1301 (0.0286)	0.1475 (0.1070)	1.1530 (0.0305)	0.2219 (0.1923)	0.1309 (0.0285)
SBASS	1.2446 (0.0195)	2.1606 (0.1230)	0.4177 (0.0172)	0.5133 (0.0728)	1.4655 (0.0214)	0.7218 (0.1361)	0.4185 (0.0172)
BLUEFISH	0.6205 (0.0227)	1.1062 (0.1717)	0.3982 (0.0199)	0.3609 (0.0893)	0.7001 (0.0239)	0.4919 (0.1695)	0.3973 (0.0199)
FLATFISH	0.7428 (0.0201)	1.5666 (0.1322)	0.3262 (0.0170)	0.3540 (0.0713)	0.8373 (0.0216)	0.4978 (0.1331)	0.3261 (0.0170)
REDFISH	0.8898 (0.0254)	1.5323 (0.1894)	0.2777 (0.0238)	0.2894 (0.1050)	1.0393 (0.0266)	0.4118 (0.1961)	0.2776 (0.0238)
SEATROUT	0.3868 (0.0258)	1.2969 (0.2078)	0.4022 (0.0233)	0.4322 (0.1113)	0.3431 (0.0266)	0.5836 (0.2120)	0.4021 (0.0233)
MACKEREL	0.2853 (0.0279)	1.6216 (0.2426)	0.2186 (0.0255)	0.3605 (0.1182)	0.2145 (0.0292)	0.5261 (0.2267)	0.2183 (0.0255)
MARLIN	0.1959 (0.0465)	1.1953 (0.4229)	0.4611 (0.0411)	0.4962 (0.2123)	0.0667 (0.0481)	0.7532 (0.4451)	0.4601 (0.0411)
TUNA	0.2390 (0.0393)	1.7158 (0.2460)	0.0067 (0.0339)	0.2347 (0.1263)	0.3052 (0.0416)	0.4467 (0.2375)	0.0090 (0.0336)
MAHI-MAHI	0.9468 (0.0375)	1.8329 (0.2902)	0.5060 (0.0332)	0.5424 (0.1484)	1.0595 (0.0394)	0.7055 (0.2933)	0.5047 (0.0332)
SHELLFISH	0.9838 (0.0263)	1.8973 (0.1975)	0.2758 (0.0259)	0.2136 (0.0995)	1.1489 (0.0276)	0.2655 (0.1768)	0.2758 (0.0259)
ALPHA	---	5.8163 (0.1922)	---	1.2638 (0.0422)	---	23.6119 (16.1061)	---
-2 Log-Likelihood	57322	16132	36714	11154	53136	10134	36715
AIC	57358	16170	36786	11228	53172	10172	36787
AICC	57358	16170	36787	11229	53172	10172	36787
BIC	57479	16297	37028	11476	53292	10275	37028

Source: Computed by authors using 2011 National Survey of Fishing, Hunting, and Wildlife-Associated Recreation

The zero-truncated negative binomial model was selected because of lower AIC and BIC values. In the parameter estimates for zero-truncated negative binomial model, the age of respondents appeared to have a positive and significant impact on the saltwater recreational fishing trips. The value of the coefficient for "AGE" (0.0097) suggests that the log count of saltwater recreational fishing trips increases by 0.0097 for each unit increase in age group. It shows that the older a recreational angler's age, the more saltwater recreational fishing trips taken.

The coefficient for "MALE" (0.3580) is statistically significant and indicates that the log count of saltwater recreational fishing trips for male anglers is 0.3580 more than for female anglers. The coefficient for "GRADUATE", -0.6578, is statistically significant and indicates that the log count of saltwater recreational fishing trips for anglers who had graduate or professional degree is 0.6578 less than for non-advanced degree anglers. The coefficient for "MINOR" (0.8590) is statistically significant and indicates that the log count of saltwater recreational fishing trips for minority anglers is 0.8590 more than for non-minority anglers.

The coefficient for "SBASS" (0.7218) is statistically significant and indicates the log count of saltwater recreational fishing trips for striped bass is 0.7218 more than for other fish species. The coefficient for "BLUEFISH" (0.4919) indicates the log count of saltwater recreational fishing trips for bluefish is 0.4919 more than for other fish species significantly. The coefficient for “FLATFISH” (0.4978) is statistically significant and indicates the log count of saltwater recreational fishing trips for flounder, flatfish, or halibut is 0.4978 more than for other fish species. The coefficient for “SEATROUT” (0.5839) indicates the log count of saltwater recreational fishing trips for sea trout is 0.5839 more than for other fish species significantly.

5. Conclusion and Discussion

This study aimed to demonstrate the various range of analysis suitable for count data by modeling saltwater recreational fishing trips taking into consideration the issues of over-dispersion, excess zeros and zero truncation. Count data models are used to model the data in which the dependent variable is count. There are different count data models and is difficult to choose one that best fits the count data. The models considered in this study included Poisson, negative binomial, zero-inflated Poisson, zero-inflated negative binomial, zero-truncated Poisson, zero-truncated negative binomial, and hurdle Poisson models.

The Poisson model is the fundamental model to be fitted for count data where there are no over-dispersion issues and issues with excess zeros. For count data with over-dispersion, a negative binomial model is more appropriate. If excessive zeros are present in the count data, then two part models such as zero-inflated models and hurdle models are more suited to be fitted in such cases. The results of the zero-inflated models and hurdle model are similar but interpretation of hurdle model is much easier. Furthermore, where there is zero truncation of the count data, then zero-truncated models should be rather used.

In selecting a model, there is no fixed yardstick that easily shows the best model. It is not always that simple¸ the best model choice is not easily that obvious as one that performs best in one regard may not be best in another. There are always pros and cons to consider. In selecting the best model, one needs to consider whether that particular model assumptions are met, nature of count data, the relevance of the zeros to the study, over-dispersion, goodness of fit statistics (i.e. AIC, BIC) comparison as well as zero truncation of count data before deciding on the best fit model for the data.

In addition, consistent with the findings of previous studies, males would go fishing more when they participated in saltwater recreational fishing activities. The positive signs on the variable “MINOR” suggested that those recreational anglers who would go fishing more in saltwater recreational fishing activities. In addition, the positive sign on the variable “URBAN” suggested that those who resided in urban settings have a higher demand for saltwater recreational fishing activities. Results also pointed out that respondents who had graduate or professional degree did not have significant effect on U.S. saltwater recreational fishing trips.

Empirical results of this study indicated that the mature minority male living in the urban area would go fishing for striped bass, bluefish, flatfish, redfish, sea trout, mackerel, tuna, and mahi-mahi in U.S. saltwater areas. Therefore, recreational anglers living in urban settings does appear to be a distinguishing factor in saltwater recreational fishing activities. Therefore, saltwater recreational fishery managers should have an opportunity to target this user group in their management plans, expanding a shrinking constituency.

Also, empirical results of this study found that targeting one or more of specific species have positive and significant impacts on U.S. saltwater recreational fishing trips, indicating that demand increases significantly with the presence of fish categories including striped bass, bluefish, flatfish, redfish, sea trout, mackerel, tuna, and mahi-mahi. Thus, the availability of a diversity of species plays an important role in saltwater recreational fishing. Fishery managers should educate the public about the availability or location of diverse habitats to generate continued interest and increased participation in saltwater recreational fishing.

More importantly, healthy fisheries habitat is not only essential for a healthy fishery, but is also an essential part of the fishing experience. Without a healthy fishery based on healthy fisheries habitats the effort will fail. Saltwater recreational fishing adds to mixed activity vacation venues attracting anglers and families with multiple interests. Particularly, saltwater recreational fishing business succeed on the basis of the quality of the fishable resource, the quality of the ancillary experience of nature, comfort and well-directed marketing that matches the venue to the needs of various types of anglers (Cisneros-Montemayor and Sumaila, 2010).

Without adequate fishes and their habitats, there would be far fewer or no participants in saltwater recreational fishing activities. In addition, the purpose of saltwater recreational fishing trips would be expected to have a positive impact on saltwater recreational fishing expenditures. Thus, fisheries habitats and populations can be viewed a critical factor, as with an increase in ecosystem and biodiversity of fisheries, the more saltwater recreational anglers would participate in and consume (Cisneros-Montemayor and Sumaila, 2010).

Many studies neglect how best to model saltwater recreational trips and get meaningful insights into the behavior of saltwater anglers that affect their saltwater fishing trips behavior and participation. This study aims to provide guidelines and create awareness for the proper use of count data models that will lead to results that are more accurate and get a better understanding of saltwater recreational fishing trips that would eventually lead to promotion of saltwater recreational fishing and tourism as a whole.

Results from this study may give better understanding of recreational fishing trips among saltwater anglers and also provide guidelines for saltwater recreational fisheries planning and management. It may also serve as a yardstick in encouraging the proper use of count data models so as to get accurate results and get a better understanding of count data models. Therefore, the empirical results of this study provide insight into the determinants of saltwater recreational fishing trips, which can be used in analyzing the social and economic values of saltwater recreational fisheries planning and management.

References

Cisneros-Montemayor, A. M. and Sumaila, U. R. (2010). A global estimate of benefits from ecosystem-based marine recreation: potential impacts and implications for management. *Journal of Bioeconomics* 12(3), 245-268.

Cohen, A. C., Jr. (1960). Estimating the parameter in a conditional Poisson distribution. *Biometrics* 16(2), 203-211.

David, F. N. and Johnson, N. L. (1952). Extension of a method of investigating the properties of analysis of variance tests to the case of random and mixed Models. *The Annals of mathematical Statistics* 23(4), 594-601.

Good, I. J. (1986). Some statistical applications of Poisson's work. *Statistical Science* 1(2), 157-180.

Greene, W. H. (2008). Econometric analysis (Sixth Ed.). Upper Saddle River, New Jersey: Pearson Education Inc.

Greene, W. H. (1994). Accounting for excess zeros and sample selection in Poisson and negative binomial regression models. NYU Working Paper No. EC-94-10. Available online at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1293115

Grogger, J. and Carson, R. (1991). Models for truncated counts. *Journal of Applied Econometrics* 6(3), 225-238.

Gurmu, S. (1991). Tests for detecting overdispersion in the positive Poisson regression model. *Journal of Business & Economic Statistics* 9(2), 215-222.

Hall, D. B. (2000). Zero-inflated Poisson and binomial regression with random effects: a case study. *Biometrics* 56(4), 1030-1039.

Hardin, J. W. and Hilbe, J. M. (2012). Generalized linear models and extensions (Third Ed.). College Station, Texas: A Stata Press Publication.

Heilbron, D. C. (1994). Zero-altered and other regression models for count data with added zeros. *Biometrical Journal* 36(5), 531–547.

Singh, J. (1978). A characterization of positive Poisson distribution and its application. *SIAM Journal on Applied Mathematics* 34(3), 545-548.

Lambert, D. (1992). Zero-inflated Poisson regression with an application to defects in manufacturing. *Technometrics* 34(1), 1-14.

Loeys, T., Moerkerke, B., Smet, O. D. and Buysse, A. (2012). The analysis of zero-inflated count data: beyond zero-inflated Poisson regression. *British Journal of Mathematical and Statistical Psychology* 65(1), 163-180.

Mullahy, J. (1986). Specification and testing of some modified count data models. *Journal of Econometrics* 33(3), 341-365.

Potts, J. M. and Elith, J. (2006). Comparing species abundance models. *Ecological Modelling* 199(2), 153-163.
Sampford, M. R. (1955). The truncated negative binomial distribution. *Biometrika* 42(1-2), 58-69.
Shonkwiler, J. S. and Shaw, W. D. (1996). Hurdle count-data models in recreation demand analysis. *Journal of Agricultural and Resource Economics* 21(2), 210-219.
U.S. Department of the Interior, U.S. Fish and Wildlife Service, and U.S. Department of Commerce, U.S. Census Bureau. (2002). 2001 National Survey of Fishing, Hunting, and Wildlife-Associated Recreation. Available online at https://www.census.gov/prod/2002pubs/FHW01.pdf [accessed 2 October 2018]
U.S. Department of the Interior, U.S. Fish and Wildlife Service, and U.S. Department of Commerce, U.S. Census Bureau. (2007). 2006 National Survey of Fishing, Hunting, and Wildlife-Associated Recreation. Available online at https://www.census.gov/prod/2008pubs/fhw06-nat.pdf [accessed 2 October 2018]
U.S. Department of the Interior, U.S. Fish and Wildlife Service, and U.S. Department of Commerce, U.S. Census Bureau. (2014). 2011 National Survey of Fishing, Hunting, and Wildlife-Associated Recreation. Available online at https://www.census.gov/prod/2012pubs/fhw11-nat.pdf [accessed 2 October 2018]
Zawacki, W. T., Marsinko, A. and Bowker, J. M. (2000). A travel cost analysis of nonconsumptive wildlife-associated recreation in the United States. *Forest Science* 46(4), 496-506.