2009

Count regression models for recreation demand: an application to Clear Lake

Brian David Pattiz
Iowa State University

Count regression models for recreation demand: an application to Clear Lake

by

Brian Pattiz

A thesis submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Major: Economics

Program of Study Committee:
Catherine Kling, Major Professor
Quinn Weninger
Sarah Nusser

Iowa State University

Ames, Iowa

2009

LIST OF TABLES

ABSTRACT

An important objective for policy-makers is how to allocate resources for the enjoyment of its citizens. Outdoor recreation is a very popular hobby for a lot of people. The sites they travel to for recreational purposes are public sites such as the Clear Lake, located in central Iowa. The users of the lake often care about the quality of the water. It is the goal of the researcher to determine how much they are willing to pay in order to preserve or improve the water quality. The researcher must decide on not only the theoretical methodology, but the appropriate statistical model. The focus of this thesis is using count data models to estimate individuals' willingness to pay. A common count data model is the Poisson model, however it is restrictive and often alternative models must be used. This thesis introduces a new count data model to the literature: The Conway-Maxwell-Poisson regression model. Using the data gathered by individual users at Clear Lake, I contrast this model with a popular alternate to the Poisson model, the negative binomial model.

CHAPTER 1. GENERAL INTRODUCTION

1.1 Introduction

Households enjoy spending their leisure at public sites such as lakes and parks for recreational activities. Due to resource constraints, policy-makers need a reliable estimate of the benefits and costs of actions that affect the quality and quantity. These goods share the “common-property” characteristics of non-excludability and non-depletion; therefore market prices cannot allocate resource usage. Since they are public goods, special methods and techniques are needed to estimate the welfare gained from site usage and quality improvements and the welfare lost from quality detriments and use restrictions.

The act of placing a value on natural resources is sometimes controversial. Some environmentalists reject this practice for moral reasons that are similar to why someone rejects the concept of placing a value on somebody's life. Another argument against this approach is that some believe that natural resources contain intrinsic value. A natural resource contains intrinsic value if '“it is valuable in and for itself-if its value is not derived from its utility, but is independent of any use or function it may have in relation to something or someone else”' (Freeman 2003). While the concept of intrinsic value is sympathetic to those who have certain philosophical inclinations, it is problematic when applying it to natural resource management. The concept of intrinsic value cannot provide answers to many practical issues that policy-makers face. For example, consider the problem of controlling pollution. The regulation of pollution imposes costs to both private industries and taxpayers. Intrinsic value does not provide help on determining the correct level of pollution. Another example is the question of when to allow ecologically sensitive areas to be developed. Land development is important for economic growth but people value open space and habitat, thus tradeoffs must be made. Freeman (2003) advocates the concept of instrumental value instead of intrinsic value. Instrumental value is when the valuation of a natural resource has a goal or purpose. For example, a lake provides services for households by the means of boating and fishing. It also provides a habitat for many creatures. Since people gain utility from those activities and enjoy the existence of a lake to provide a habitat and economic value is based on the welfare of individuals, then the goal is measure the welfare that the lake provides for recreation and its very existence.

1.2 Natural Resource Valuation Methodology

Economists use both revealed preference and stated preference methods in estimating the demand for non-market goods. Revealed preference methods use actual observed behavior of households to reveal their value on the resource. There are two different modeling approaches when using the revealed preference method. The continuous choice approach estimates the quantity of trips an individual takes. Then a demand function can be estimated so that welfare analysis can be performed. The travel cost method (TCM) is the most well-known example. The discrete choice approach models the decision of an individual who chooses among competing sites. The random utility model is commonly used to model that choice.

Stated preference approaches rely on individuals to state their value on hypothetical changes of the resource or to give their expectations on their future use. The contingent valuation method (CVM) is the most widely used stated preference method. It is able to elicit information about the individual's value. The method asks a hypothetical question in which respondents are asked their willingness to pay for the change in a natural resource or willingness to accept compensation to forgo that change. There have been different ways to ask CVM questions. The most common is the dichotomous choice format. This format asks the respondent whether they are willing to pay $X. An answer of "yes" allows the researcher to place a lower bound on the respondent's willingness to pay, while an answer of "no" places an upper bound.

Contingent behavior is another stated preference method that is often used in conjunction with the TCM. A common contingent behavior question is to ask a respondent how many trips they would take if their travel costs changed by $X. Contingent behavior questions can also be used to elicit information on how many trips would be taken if there was a change in one of the environmental attributes of the recreation site.

There are a lot of advantages and drawbacks of using either a revealed preference or stated preference method. Revealed preference methods have the advantage of using data that was generated from actual behavior. Problems of the TCM include what qualifies as a site, dealing with substitute sites, and how to measure the opportunity cost of time.

In contrast, a huge disadvantage for stated preference methods is the hypothetical nature of the questions. A famous critique of CVM was made by Diamond and Hausman (1994). A group of well-known economists, the Blue Ribbon Panel, suggested guidelines for refining the CVM. While guidelines did not quiet every criticism, it has made the practice more generally accepted.

Researchers decide on which method to use based on the problem they are trying to address. The TCM is able to measure the value of using the recreation site, while CVM is able to measure both use value and non-use value. For example, it is possible for an individual to value the existence of a recreation site while never actually visiting it. Krutilla (1967) argues that individuals can have a bequest motivation. They will value the preservation of the site for future generations to use. Researchers used to favor one method over another. However, there has been a trend to use both methods in order to validate the other approach (Haab and McConnell 2002).

1.3 Empirical Background

The goal of non-market valuation studies is to measure the well-being that individuals receive from visiting a recreation site. Researchers are not able to measure the economic value of a site directly from the data generated by observed behavior. That is why they rely on statistical inference. The accuracy of estimation is dependent on both the econometric model chosen and the reliability of data. The respondent is subject to time constraints, the ability to correctly interpret questions, and recall the necessary information. Respondents are often asked to recall how many trips they've taken in a particular year and how much money they spent in a particular month/year. For a nice discussion of potential pitfalls of survey methods see Phaneuf and Smith (2005).

There have been numerous statistical methods that have been used to estimate single site demand models. The trip data that is gathered is in the form of non-negative integers. This had led to the use of count data models, the most common being the Poisson regression model. An assumption of that model is that the conditional mean and the conditional variance are equal. This is known as the equidispersion of counts. This assumption is often violated and alternative count model is needed. A popular alternative is the negative binomial regression model which allows for overdispersion.

1.4 Problem Statement and Thesis Outline

This thesis focuses on the estimation of recreation demand for Clear Lake, located in Clear Lake, Iowa. The lake is used for recreational purposes such as fishing, boating, and beach use. Since the mid-20th century, the quality of the water was detraining. In the summer of 2000, researchers at Iowa State University conducted a survey to users. Using the revealed preference data generated from that survey, I will explore the estimation of recreation demand using count data models including a newer regression model that results from the Conway-Maxwell Poisson distribution.

A brief organization of this thesis follows. Chapter 2 contains brief review the theoretical framework and count models that are used with travel cost models. Chapter 3 develops the Conway-Maxwell Poisson generalized linear model. Chapter 4 discusses the survey that was use for the Clear Lake study. Finally, Chapter 5 will provide the estimation and welfare calculations.

CHAPTER 2. A THEORETICAL AND EMPIRCAL INTRODUCTION TO THE TRAVEL COST MODEL

2.1 Introduction

The travel cost model is a widely used method of estimating the recreational demand of a natural resource. Its underlying premise is the fact that households have to travel to a recreational site to benefit from its amenities. Harold Hotelling in 1947 was the first to suggest using travel costs as implicit prices for recreation sites (Haab and McConnell 2002). Burt and Brewer (1971) were the first to use the travel cost model to estimate demand equations in their application of recreational activity at rivers, lakes and reservoirs in Missouri. There has been a variety of different applications of the travel cost model including sport fishing in the Chesapeake Bay (Bockstael et.al 1990), hunting trips in California (Creel and Loomis 1990), and the effect of wildfires on biking and hiking demand in New Mexico (Hesseln et.al 2003).

2.2 Theoretical Development and Assumptions

2.2.1 The Basic Travel Cost Model

The basic travel cost model is based off standard microeconomic theory. Although the demand for the site could be modeled as an aggregate demand, the model estimates it on the individual level and estimates the economic value of the site by adding up each individual's value. The total economic benefits to users of a site can be obtained by summing up the area under each individual's demand function. Each individual demand function shows the number of visits depends on the cost of visiting the site.

The formal model is set up as a representative consumer maximizing their utility given their income and time constraint,

$$\underset{y}{Max}\ U(y,z,q)$$
$$subject\ to \begin{cases} z + c\cdot y = m + p_w t_w (income\ contraint) \\ t_w + (t' + t'')y = T\ (time\ constraint) \end{cases}, \tag{2.1}$$

where U(·) is a quasi-concave utility function, y is the number of trips to a site, z is a numeraire (private) good, q is the quality of the site, c is the cost of visiting the site, m is a household's income that is exogenous to the model, t_w is the time spent working, t' is the time spent traveling to the site, t'' is the time spent onsite, and T is the total time available. Setting up the Lagrangian and taking the first order condition yields

$$\frac{\partial U/\partial y}{\lambda} = c + \frac{\mu}{\lambda}(t' + t''). \tag{2.2}$$

The Lagrangian multiplier μ represents the marginal utility of time and the ratio μ/λ represents the marginal WTP for time. Therefore, (2.2) can be interpreted as the marginal WTP should be the same as a household's full cost of visit (Freeman 2003). The model yields a demand function dependent on the price of visiting the site, the quality of the site and the income of the individual.

2.2.2 The Role of Time

An important assumption of the model is that opportunity cost of time serves as a proxy for the travel cost. This approach follows the household production framework of Becker (1965). The household production framework states the household's demand for market goods and similarly non-market goods are just vehicle for their consumption process, i.e. they transform the goods into service flows that produce utility (Smith 1991). These service flows are not observable but we can observe a different combination of inputs used in the production process given a change in constraints. Consider the household production activity of watching a movie with the family. Given income and time constraints, we can observe the amount of convenience services such as Netflix that allows families to save time and money instead of driving to the local video store. The household production framework gives a framework for how non-market and market goods relate to each other.

The basic model assumes that the individual can freely substitute labor and leisure at a constant wage rate. This assumption is necessary so that the time spent working is in the same units as the time spent onsite and traveling to the site. Cesario (1976) discovered that empirically, this assumption does not hold. This has led researchers to develop alternative ways of modeling time. A response to Cesario's findings has led to researchers to use a fraction of the wage rate, most commonly one-third (Shaw and Feather 1999). This is still problematic because it assumes that every user has the same tradeoff. Other researchers have addressed the problem in a different way. Bockstael, Strand and Hanemann (1987) argue that an individual's marginal value of time may not be equal to their wage rate because they might not be able to optimally choose the number of hours they can work. Their model allows for the "corner solution", a worker can choose to have a job with fixed hours or choose not to work at all.

Also important is the role of onsite time. The assumption of the basic model is that all visits to the site are of the same duration. This is important because it allows the number of visits to be constant. It also ensures that the price of visiting the site is exogenous to the model. In practice, researchers use people who only make day trips to the site in order for the assumption to hold.

2.2.3 The Role of On-site Quality

The quality of a recreation site influences the amount of pleasure that an individual receives from visiting the site. The basic model yields an individual demand function that is partially dependent on the quality of the site. To be able to know how individual preference of the site quality can be inferred from decisions affected by it, there has to be a way to find the relationship between quality and the utility that an individual receives from visiting the site. The concept of weak complementarity was developed to deal with this issue. The assumption requires that along with visiting the recreation site, an individual consumes a private good. The private good is said to be non-essential, which indicates that there is a price, known as a "choke price", where the consumption of it is zero. When the weak complementary good is not consumed, the user will not value any changes to the quality of the recreation site.

Consider the example of fishing in a lake. The water quality of a lake can affect whether fishing in it is safe. If the private good is the rental of fishing equipment, the individual won't care about the quality of lake if they are not fishing in it. It is generally thought not to be possible to test the plausibility of weak complementarity from data (Haab and McConnell 2002). It is thought to be realistic and intuitive because people only value changes in the quality of the resource if they use it. Measuring quality changes from travel cost data is very difficult. It is usually complimented with contingent behavior data so that the researcher can measure how behavior is changed when the quality changes.

2.2.4 Other Assumptions

The basic model assumes that all travel costs that an individual incurs are for a single site. There also cannot be any substitute sites that are available for the individual to choose from. If there are multiple substitute sites, then the number of trips to the recreation site will depend on the travel costs to substitute sites and can lead to biased estimates (Freeman 2003).

2.3 Count Data Models for Recreation Demand

2.3.1 Why Count Data Models?

Recreational site trip data is generated by a stochastic process; dependent on the sampling method that was used. Researchers assume the dependent variable, the number of trips, is assumed to be distributed continuously or discretely. There have been applications in which the researchers assume that the number of trips is distributed continuously. Early applications used ordinary least squares on aggregate zonal data (Phaneuf and Smith 2005). The Tobit model has also been popular because it is a censored model that allows a large number of observations massed at zero. Count data models are popular because the recreation trips are recorded as non-negative integers. They are also attractive because they assume a semi-log demand functional form, which is very popular in the literature.

The theoretical basis for using count data models is very important for interpretation of estimation results. The problem in using the standard microeconomic approach is that if trips are non-negative integers, differential calculus cannot be used to obtain the optimal consumption bundle. Hellerstein and Mendelsohn (1999) address the problem by adding an additional constraint that the number of trips must be a non-negative integer. Their solution requires that each individual has a set of unobserved factors that given a price, determines the quantity of trips that are taken. This along with observable factors such as price and income will yield a distribution of demand that can be modeled using a count data distribution.

2.3.2 The Poisson Count Regression Model

The Poisson probability density function is

$$Pr(Y = y_i) = \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!}, y_i \in \{0\} \cup \mathbb{Z}^+, \tag{2.3}$$

with i=1,…, n observations. The mean $E[Y] = \lambda_i$ and $Var(Y) = \lambda_i$.To apply (2.3) for the purpose of regression, assume $y_i \overset{i.d.d}{\sim} Poisson(\lambda_i)$, and λ_i is assumed to be a function of a 1 x k vector of covariates $\mathbf{x}_i$ and a k x 1 vector of coefficients $\boldsymbol{\beta}$. The functional form of the parameterization for the conditional mean is

$$E[Y \mid X] = \lambda_i = exp(\mathbf{x}_i'\boldsymbol{\beta}). \tag{2.4}$$

The Poisson model assumes that the conditional mean, λ_i, is equal to the conditional variance. Overdispersion is when the conditional variance exceeds the conditional mean and is considered to be heteroskedastic. The standard approach of estimating the model is using a form of maximum likelihood estimation, either using a Newton-Ralphson algorithm or the iterative reweighted least squares, which is used by the generalized linear model approach. The likelihood equation that is given by

$$\ell(\boldsymbol{\beta}) = \sum_{i=1}^{n} -exp(\mathbf{x}_i'\boldsymbol{\beta}) + y_i\mathbf{x}_i'\boldsymbol{\beta} - ln(y_i!). \tag{2.5}$$

Equation (2.5) is globally concave and guarantees convergence. If the conditional mean function (2.4) is correctly specified, the Poisson model is robust to the presence of overdispersion. This leads to using the pseudo-likelihood estimator. See Cameron and Trevidi (1998) for more details on this estimator.

2.3.3 The Negative Binomial Count Regression Model

In many empirical applications of recreation demand, trip counts are often overdispered. A reason for this is because many users only take a few trips and a few take many trips. Although econometricians have modified the Poisson regression model to deal with overdispersion, a popular alternative has been the use of the negative binomial regression model. The probability mass function for the negative binomial distribution is

$$Pr(Y = y_i) = \frac{\Gamma(y_i + r)}{\Gamma(r)\Gamma(y_i + 1)} p^r (1-p)^{y_i}, y \in \{0\} \cup \mathbb{Z}^+ , \tag{2.6}$$

where $\Gamma(\cdot)$ is the gamma function. The mean and variance of the negative binomial are $E[Y] = \mu = r\frac{1-p}{p}$ and $Var(Y) = r\frac{1-p}{p^2}$. It is common to parameterize r and p in the terms of α and μ. Define $\alpha = 1/r$, then $\mu = (1-p)/(\alpha p)$, solving yields $p = 1/(1+\alpha\mu)$. After the reparametrization, (2.6) becomes

$$Pr(Y = y_i) = \frac{\Gamma(y_i + 1/\alpha)}{\Gamma(1/\alpha)\Gamma(y_i + 1)} \left(\frac{1}{1+\alpha\mu}\right)^{1/\alpha} \left(\frac{\alpha\mu}{1+\alpha\mu}\right)^{y_i}. \tag{2.7}$$

The mean of this parameterization is $E[Y] = \mu$ and $Var(Y) = \mu + \alpha\mu^2$. This is known as the "NB-2" model because it has a quadratic variance function. In this model $\alpha \geq 0$ and if $\alpha = 0$, then it reduces to a Poisson. There are other ways to derive (2.7), including mixing distributions of the Poisson-gamma and Poisson-beta (Johnson, Kemp and Kotz 2005). The derivation from the Poisson-gamma mixture provides some intuition into the usefulness of the NB2. It specifies a Poisson model with an error term in the mean. This error term reflects unobserved heterogeneity and is distributed gamma. The problem of overdispersion could a result of unobserved heterogeneity that is not captured by the Poisson model (Cameron and Trivedi 1998).

The negative binomial model can be estimated using maximum likelihood. The NB2 likelihood function is

$$\ell(\mu_i \mid \alpha, y_i) = \sum_{i=1}^{n} \left[y_i \, ln \left(\frac{\alpha\mu_i}{\alpha\mu_i + 1} \right) - \frac{1}{\alpha} ln(\alpha\mu_i + 1) + ln\Gamma(y_i + 1/\alpha) \right. \\ \left. - ln\Gamma(y_i + 1) - ln\Gamma(1/\alpha). \right. \tag{2.8}$$

The NB2 model is less robust to distributional misspecification than the Poisson model where one could use a pseudo-maximum likelihood estimator. If the conditional mean is correctly specified, βs will be consistent (produce the correct standard errors) but α will not be. See Cameron and Trivedi (1998) for first order conditions, Fisher information matrix, and asymptotic results of the maximum likelihood estimators.

2.3.4 Welfare Analysis for Count Regression Models

The eventual goal of recreation demand studies is to use welfare measures for policy recommendations. For count models there are problems when trying to perform this analysis. One problem is that they do not have to be ability to distinguish users and nonusers from those who report zero trips. The preference structure from those who are nonusers to those who are users that might not take any trips during that particular time period will be different. For these models, the number of trips is guided by a distribution that is not associated with a random error, which suggests that the quantity demanded is itself a random variable (Haab and McConnell 1996). Hellerstein (1999) considers a semi-log demand function with a stochastic error term; this function is continuous and is unlike any semi-log demand function generated from a count distribution. He argues that the demand function without the stochastic error should be considered as a "description of potential behavior" (p.272). Welfare analysis is conducted via deriving the consumer surplus by integrating under the demand function. Since the number of trips is guided by a count distribution, taking the expectation of the distribution will yield the expected number of trips. Thus, consumer surplus can be obtained by integrating under the expected number of trips. Formally, the expected number of trips can be calculated as

$$E[Y] = \lambda_i = \int f(p_y, p_z, m, \varepsilon) g(\varepsilon) d\varepsilon, \tag{2.9}$$

where $f(\cdot)$ is the stepwise demand function resulting from (2.1), whose stepwise structure is influenced by unobserved factors. $g(\varepsilon)$ does not have any distributional assumptions. Consumer surplus can be obtained by taking the expectation of (2.9) with respect to the price. For the semi-log demand function, the choke price is infinite. Let c^* denote the choke price and c_o denote the initial price, and then the consumer surplus can be calculated as

$$\begin{aligned} WTP &= \int_{c_o}^{c*} exp(\mathbf{x}_i{}'\boldsymbol{\beta}) dc \\ &= \left[\frac{exp(\mathbf{x}_i{}'\boldsymbol{\beta})}{\beta_c} \right]_{c=c_o}^{c^* \to \infty} = -\frac{\lambda_i}{\beta_c}, \end{aligned} \tag{2.10}$$

Haab and McConnell (1996) argue that for this to occur, price changes have to be independent of the error structure, which they find unsatisfactory. Despite this debate, these models are used as a way to a good way to fit the data given the discrete, non-negative structure of it.

2.3.5 The Use of On-site Sampling

The use of on-site sampling is popular because it is cheaper in terms of both time and money for researchers to be able to obtain information from users. The problem with an on-site sample is that there are not any households that have not taken any trips. The sample will not reflect the true population and those that visit the site more frequently than others will more likely be sampled. Since the sample is not reflective of the population, the conditional mean is not reflected by $E[Y|\mathbf{X}]$. The problem of truncation deals with the fact that the researcher only deals with households that take a positive number of trips. Grogger and Carson (1991) apply truncated Poisson and negative binomial models to Alaskan fishing trips.

Shaw (1988) was the first to recognize the problem of truncation and endogenous stratification simultaneously. The benefits of correcting for truncation and endogenous stratification is that it allows for the estimation of the benefits of a trip to any individual in a population, the mean number of trips that an individual will take and the theoretical availability of calculating the benefits themselves. Failure to correct for this will result in biases in welfare estimates. Correcting for the biased sample will allow researchers to infer the value by potential users (Englin and Shonkwiler 1995b).

Shaw (1988) and Santos Silva (1997) show that from a population density function $f(y_i \mid \mathbf{x}_i)$, the endogenous stratified probability density function is

$$h(\tilde{y}_i \mid \mathbf{x}_i) = \frac{y_i f(y_i \mid \mathbf{x}_i)}{\int_0^\infty y_i f(y_i \mid \mathbf{x}_i) dy} = \frac{y_i f(y_i \mid \mathbf{x}_i)}{E[y_i \mid \mathbf{x}_i]}. \tag{2.11}$$

Santos Silva (1997) and Patil and Rao (1978) show that the mean for the endogenous stratified sample is

$$E[\tilde{y}_i \mid \mathbf{x}_i] = E[y_i \mid \mathbf{x}_i] + \frac{Var[y_i \mid \mathbf{x}_i]}{E[y_i \mid \mathbf{x}_i]}. \tag{2.12}$$

Herriges and Egan (2006) provide the intuition for (2.12). When researchers use count data models to estimate on-site samples without correcting for truncation and endogenous stratification, the amount of bias to estimate the population average will depend on the relative overdispersion, i.e. the ratio of the variance to the mean. When the probability of sampling a household that visits the site more often than others in the population increases, it will increase the variance.

Shaw shows that the Poisson regression model when adjusted for endogenous stratification has a probability density function of the form

$$f(\tilde{y}_i \mid \mathbf{x}_i) = \frac{exp(-\lambda_i)\lambda_i^{\tilde{y}_i - 1}}{(\tilde{y}_i - 1)!}, \tilde{y} = 1, 2, ... \tag{2.13}$$

where $\hat{\lambda}_i = exp(\mathbf{x}_i'\boldsymbol{\beta})$. The mean and variance for (2.13) are $E[\tilde{y}_i \mid \mathbf{x}_i] = \lambda_i + 1$ and $Var[\tilde{y}_i \mid \mathbf{x}_i] = \lambda_i$. Estimating the on-site Poission regression model is easy because all that is needed to regress the dependent variable $y_i - 1$ on the covariates.

Since the Poisson regression model assumes equidispersion, Englin and Shonkwiler (1995) correct the negative binomial regression model to account for truncation and endogenous stratification. They show the probability density function as

$$g(\tilde{y}_i \mid \mathbf{x}_i) = \frac{\tilde{y}_i \Gamma(\tilde{y}_i + 1/\alpha)}{\Gamma(\tilde{y}_i + 1)\Gamma(1/\alpha)} \left(\frac{1/\alpha}{1/\alpha + \lambda_i}\right)^{1/\alpha} \left(\frac{1}{1/\alpha + \lambda_i}\right)^{\tilde{y}_i} \lambda_i^{\tilde{y}_i - 1}. \tag{2.14}$$

When α_i approaches zero, equation (2.14) will be reduced to (2.13). Thus, testing the statistical significance of α_i can be used to test for overdispersion. The mean and variance for (2.14) are $E[\tilde{y}_i \mid \mathbf{x}_i] = \lambda_i(1 + \alpha_i) + 1$ and $Var[\tilde{y}_i \mid \mathbf{x}_i] = \lambda_i(\alpha_i + \alpha_i\lambda_i + \alpha_i^2\lambda_i + 1)$. There are different ways to parameterize α_i, and the parameterization that is chosen implies an assumption about the functional form of the heteroskedasticity (Cameron and Trivedi 1986) and will result in different estimates in the covariates.

CHAPTER 3. THE CONWAY-MAXWELL-POISSON MODEL

3.1 Introduction

The Poisson distribution is one of the most widely used distributions in statistical applications. The distribution does have its limitations because its mean and variance are equal. In many situations this assumption is not realistic. Thus researchers have used different specifications of the Poisson to deal with this problem. Usually this done by introducing a mixing distribution or estimating the Poisson with extra parameters that account for over-dispersion or under-dispersion. A recent specification of the Poisson is that of the Conway-Maxwell- Poisson distribution. The Conway-Maxwell-Poisson distribution is a two parameter Poisson distribution that was first introduced by Conway and Maxwell (1962). Even though Conway and Maxwell introduced this distribution a long time ago, it was relatively ignored in the literature and its properties were not fully developed. Shmueli et al. (2005) reintroduce this distribution into the literature and develop many properties for it.

The rise in the popularity of the distribution, other than being relatively new, is that the CMP family belongs to the exponential family. Many useful properties of the exponential family have been developed and make the distribution favorable for the use of Bayesian analysis and other statistical inference. Since the reintroduction there has been a recent increase in applications of the distribution including analyzing motor vehicle crashes, electric power reliability, retail predictions of households, and predicting cancer recurrence. See Lord et al. (2008), Guikema and Gofflet (2008), Boatwright, Borle and Kadane (2003) and Rodrigues et al. (2009). This chapter follows the result of Sellers and Schmueli (2008), who extend this distribution in the context of generalized linear models.

3.2 Distributional Properties

The Conway-Maxwell-Poisson (CMP) has a probability mass function of the form

$$Pr(Y = y_i) = \frac{\lambda_i^{y_i}}{(y_i!)^{\phi}} \frac{1}{Z(\lambda_i, \phi)}, y_i \in \{0\} \cup \mathbb{Z}^+, \text{ and} \tag{3.1}$$

$$Z(\lambda_i, \phi) = \sum_{k=0}^{\infty} \frac{\lambda^k}{(k!)^{\phi}}, \tag{3.2}$$

for $\lambda > 0$ and $\phi \geq 0$. The function $Z(\lambda_i, \phi)$ is an infinite series that converges for $\lambda > 0$ and $\phi > 0$. The CMP distribution generalizes other well-known distributions. It is a limiting distribution for the Bernoulli, as $\phi \to \infty$, then $Z(\lambda_i, \phi) = 1 + \lambda_i$. When $\phi = 1$, then $Z(\lambda_i, \phi) = exp(\lambda_i)$ and the distribution nests the Poisson. $Z(\lambda_i, \phi)$ becomes a geometric series when $\phi = 0$ and the distribution nests the geometric distribution. The CMP distribution shows that there is a non-linear relationship between the ratio of successive probabilities as displayed by

$$\frac{Pr(Y = y_i)}{Pr(Y = y_i - 1)} = \frac{\lambda_i}{y_i^{\phi}}. \tag{3.4}$$

Schumeli et al. (2005) shows the moments as being in the form of the recursive formula

$$E[Y^{j+1}] = \begin{cases} \lambda_i E[Y+1]^{\phi-1} & , j = 0 \\ \lambda_i \frac{\partial}{\partial \lambda_i} E[Y^j] + E[Y]E[Y^j] & , j > 0 \end{cases}. \tag{3.5}$$

The mean and variance can be represented as

$$E[Y] = \lambda_i \frac{\partial log(Z(\lambda_i, \phi))}{\partial \lambda_i} \tag{3.6}$$

$$Var(Y) = \lambda_i \frac{\partial E[Y]}{\partial \lambda_i} \quad . \tag{3.7}$$

Sellers and Schumeli (2008) show that the variance can also be specified as

$$Var(Y) = \frac{\partial E[Y]}{\partial log \lambda_i}. \tag{3.8}$$

Since the moments of the distribution do not have a closed form, Schueli et al. show that (3.6) can be approximated by

$$E[Y] \approx \lambda^{1/\phi} - \frac{\phi - 1}{2\phi}, \tag{3.9}$$

and Sellers and Schumeli show (3.7) can be approximated by

$$Var(Y) \approx \frac{E[Y]}{\phi}. \tag{3.10}$$

These approximations are only valid for $\phi \leq 1$ or $\lambda > 10^{\phi}$ (Minka et al. 2003). As shown in (3.10), ϕ measures the dispersion of the mean.

3.3 The Conway-Maxwell-Poisson Generalized Linear Model

Sellers and Schmueli (2008) extend the CMP distribution into the classical GLM framework. A brief summary of the GLM framework can be found in Appendix A. The CMP distribution is a member of the linear exponential family as displayed by:

$$Pr(Y = y_i) = exp\left\{y_i log(\lambda_i) - log(Z(\lambda_i, \phi)) - \phi log(y_i!)\right\} \tag{3.11}$$

Like the Poisson case, the nuisance function is assumed to be normalized and has a logarithm link function. The log-likelihood function is represented as

$$\ell(\boldsymbol{\beta} \mid \lambda_i, \phi) = \sum_{i=1}^{n} \left\{y_i log(\lambda_i) - log Z(\lambda_i, \phi) - \phi log(y_i!)\right\} \tag{3.12}$$

The log-likelihood function can be solved in a few different ways. If the standard Newton-Ralphson algorithm is employed, it must be maximized under the constraint $\phi \geq 0$ (Sellers and Schmueli 2008). However, a far less restrictive approach is the use the IRLS algorithm. Taking the first order conditions of (3.3.2) yields

$$\begin{aligned}\frac{\partial \ell}{\partial \boldsymbol{\beta}} = \mathcal{V}_i &= \sum_{i=1}^{n} \frac{\partial \ell}{\partial \lambda_i} \frac{\partial \lambda_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial g(\mu_i)} \frac{\partial g(\mu_i)}{\partial \boldsymbol{\beta}} \\ &= \sum_{i=1}^{n} \left(y_i - \frac{\partial log Z(\lambda_i, \phi)}{\partial \lambda_i} \right) \frac{1}{Var(Y)} \frac{\partial E[Y]}{\partial log \lambda_i} x_{ij} \\ &= \sum_{i=1}^{n} (y_i - E[Y])\, x_{ij} = 0\end{aligned}$$

$$\frac{\partial \ell}{\partial \phi} = \sum_{i=1}^{n} \left\{ -\frac{\partial log Z(\lambda_i, \phi)}{\partial \phi} - log(y_i!) \right\} \quad = \sum_{i=1}^{n} E[log(Y)!] - log(y_i!) = 0 \tag{3.13}$$

Using the estimates from Poisson regression as initial values, these equations can be solved iteratively with the IRLS algorithm. To estimate the parameters and their corresponding standard errors, the Fisher information matrix is used. The matrix contains the covariance and variance of $\hat{\boldsymbol{\beta}}$, the variance of $\hat{\phi}$ and the covariance between $\hat{\boldsymbol{\beta}}$ and $\hat{\phi}$. The derivation of these estimates can be found in Sellers and Schumeli (2008 Appendix B). Using reweighted least squares, the estimation can be set up similarly to in (A.5) in Appendix A.

3.4 On-site Sampling and Coefficent Interpretation

3.4.1 Correcting for On-site Sampling

The sampling procedure that was used in this thesis was to intercept users on-site. To correct for endogenous stratification, the mean of a distribution has to be specified using (2.11). However, the CMP distribution does not have moments in closed form. Thus, the approximation in (3.9) is used. The probability mass function for the on-site CMP model is

$$h(\tilde{y}_i \mid \mathbf{x}_i) = \frac{\lambda_i^{\tilde{y}} E[y_i]}{Z(\lambda_i, \phi)(\tilde{y}_i - 1!)^{\phi}} \tag{3.14}$$

3.4.2 Interpreting Coefficents

In both the Poisson and the NB2 model the conditional mean is parameterized as $E[Y \mid \mathbf{X}] = \mu_i = exp(\mathbf{x}_i'\boldsymbol{\beta})$. For coefficient interpretations it is useful to look at the marginal effects

$$\frac{\partial E[Y \mid \mathbf{X}]}{\partial x_i} = \beta_i exp(\mathbf{x_i}'\boldsymbol{\beta}).$$

The slope of the demand function varies with the expected demand function. The elasticity, which is the percentage change in the expected number of trips for a unit change in the covariates are

$$\frac{\partial E[Y \mid \mathbf{X}]}{\partial x_i} \frac{1}{E[Y \mid \mathbf{X}]} = \beta_i. \tag{3.15}$$

For interpretation, the CMP model cannot rely on the conditional mean because the relationship between the conditional mean and its covariates are neither additive, like Gaussian regression is, or multiplicative, like Poisson regression is. An alternative approach is to examine the relationship between the fitted values and the changes in the covariates Sellers and Schumeli (2008) show the ratio of the successive probabilities (3.2.3) relate multiplicatively to the covariates by

$$\frac{Pr(Y = y_i \mid x = x_i + 1, \mathbf{X}_{j \neq i} = \mathbf{x})}{Pr(Y = y_i \mid x = x_i, \mathbf{X}_j = \mathbf{x})} = exp(\beta_i). \tag{3.16}$$

The interpretation is that a unit increase in a covariate results in an increase the ratio of successive probabilities that an individual takes a trip to a recreational site by $exp(\beta_i)$. When the CMP distribution reduces to the Poisson, the standard multiplicative interpretation holds. If the distribution reduces to the Bernoulli, the odds-ratio interpretation holds.

CHAPTER 4. THE DATA SET

4.1 Introduction

Clear Lake is the third largest lake in Iowa, located in central region of the state. In the mid 20th century, the lake's water quality was immaculate. By the turn of the century, the water quality had regressed so much that its visibility decreased from 5-8 feet below the surface to 6 inches to one foot (Egan 2004). Water quality deterioration was caused by a variety of sources, including runoff from fertilizer. The poor water quality stems from algae blooms, which can contribute to health concerns, staunch odor and unwelcomed color. The corroded lake also includes the loss of biodiversity.

Regardless of the lake's conditions, it remained a thriving source for recreation. Most users spend their time participating in recreational boating, swimming/beach use, and fishing. As of 2001, Clear Lake generated $30 million a year in tourism revenue for the City of Clear Lake (Egan 2004).

4.2 Survey Design

4.2.1. Survey Structure

The survey was designed to elicit how respondents feel about water quality issues of Clear Lake. To inform the visitors and local residents, a description of the lake's condition was written by limnologists at Iowa State. The survey can be found in Appendix B.

The different types of questions that were asked correspond to different methods of estimating visitors' willingness to pay. The first section consists of travel cost and contingent behavior questions. They are asked to give the number of trips they've taken through different time periods from November 1999 through October 2000. They are also asked how they spend their time at the lake and their visits to substitute sites.

The next section describes various plans that present different scenarios regarding the lakes conditions, including the water clarity, algae blooms and the state biodiversity. The respondent is then asked their willingness to pay for that scenario. These questions correspond to the method of contingent ranking. The survey also contains a contingent valuation method question. They are asked whether or not they would vote yes or no on a hypothetical referendum to maintain the water quality of the lake and avoid degradation as described by one of the contingent ranking questions. The last part of the survey asks respondents their preference regarding water quality issues and whether they would support suggested projects. The survey ends by asking respondents to give their socio-demographic information.

4.2.2. Survey Sampling and Response Rate

The goal of the survey design was to obtain the opinions of visitors and local residents regarding potential water quality improvements. In the summer of 2000, users were intercepted on-site, 1024 users agreed to participate in a mail survey in October. The local residents were randomly sampled from white pages. A total of 900 residents from Clear Lake, IA and Ventura, IA were sampled. All local residents have at least visited the site once. Of the 1024 visitors that were mailed a survey, 26 were returned undeliverable and 662 were returned resulting in a 66% response rate. For the local residents, 443 of the 990 that were mailed were returned with 132 were returned undeliverable.

For determining the final sample that is to be used for empirical analysis, I follow the application of Egan and Herriges (2006). The analysis consists of visitors only. To make sure that those in the sample visited the lake as a day trip, those who drove more than five hours away were removed. This consisted of 19 individuals. Individuals that took more than 52 trips per year, resulting in more than a trip per weekend, were removed. This included 39 individuals. Finally, those that did not answer every question were removed in order to make analysis easier. The final data set has a sample of n=546.

4.3 Model Specification

Models considered in this thesis has the mean number of trips as a function of the travel cost to the site, the individual's income, and socio-demographic characteristics. Each model parameterizes

$$\lambda_i = exp(\beta_o + \beta_C C + \beta_m m + \beta_1 Male + \beta_2 Education + \beta_3 Household), \qquad (4.1)$$

where for i=1,…n individuals, C denotes the roundtrip travel cost, m represents their income, Male is a binary variable representing the gender of the individual, where Male =1 if the respondent is male, and equals 2 otherwise, Education is a coded variable representing the different levels of education, and Household is the total number of people in the household. The total travel cost was computed by taking the total round-trip distance, computed by PC Miller, multiplied by $0.25 plus one-third multiplied by the respondent's wage rate, multiplied by their round-trip travel time. I estimate the CMP model using the approximate conditional mean relationship

$$\hat{y}_i \mid \mathbf{x}_i = \hat{\lambda}_i^{1/\hat{\phi}} - \frac{\hat{\phi}-1}{2\hat{\phi}} \text{ and } \hat{\lambda}_i = exp(\mathbf{x}_i'\boldsymbol{\beta}).$$

4.4 Survey Statistics

The following section will discuss the survey summary statistics for the Clear Lake survey. The following summary statistics can be found in the following table.

Table 1: Summary Statistics

Variable	Mean	Std. Dev.	Minimum	Maximum
Observed Trips (y)	11.80	11.93	1	52
Travel cost (C)	$98.75	$196.52	$10.98	$ 2262.21
Income (m)	$59,750	$37,124	$7500	$200,000
Male	1.37	0.48	1	2
Education	4.77	1.58	1	8
Household	8.85	1.90	1	14

Table 1 provides useful summary statistics. It should be noted that these summary statistics were the result of on-site sampling and should not be inferred to be reflective of the entire population. Between November 1999 and October 2000, the average number of trips was 11.80. With the standard deviation being 11.93, the variance should be quite higher than the mean, indicating evidence that there might be overdispersion. Camerson and Trevedi (1998) recommend that if the sample variance is over twice the sample of the mean, then the counts are likely to be overdispersed. Egan and Herrges (2006) report that since 63% of the users are male and that percentage is much higher than the Iowa population percentage of males, gives more evidence to the bias of on-site sampling. For more information regarding summary statistics including the total spending by users see Azevedo, Herriges and Kling (2001).

CHAPTER 5. ESTIMATION RESULTS

5.1 Testing for Overdispersion

In Poisson regression, it assumed that there is equidispersion – that the conditional mean is equal to the conditional variance. The failure of this assumption is like the presence of heteroskedasticity in linear regression. Consider the independent responses$y_i \sim Poisson(\lambda_i)$, where $\lambda_i = f(\mathbf{x}_i, \boldsymbol{\beta})$. Consider the conditional variance$Var(y_i \mid \mathbf{x}_i) = \lambda_i + \omega\lambda_i^2$. A simple test for extra Poisson variation is:

$$H_o : \omega = 0$$
$$H_a : \omega > 0.$$

Cameron and Trivedi (1986) propose a Lagrangian multiplier test for this hypothesis test, they derive the Lagrange multiplier statistic as

$$L = \left[\frac{\sum_{i=1}^{n} \hat{\tau}_i[(y_i - \hat{\mu}_i)^2 - y_i]}{\sqrt{(2\sum_{i=1}^{n} \hat{\tau}_i \hat{\mu}_i)}} \right]^2$$, where $\hat{\tau}_i$ is a weighted parameter that depends on the alternative distribution. For the negative binomial distribution, $\hat{\tau}_i = 1$. The LM statistic can be reduced to $L = \frac{(\sum_i \mu_i^2 - ny)^2}{2\sum_i \mu_i^2} \sim \chi_1^2$, where n is the sample size (Green 2008). I estimated L = 2.807e+16, with a p-value = 0. Thus, the null hypothesis of no overdispersion is rejected. The Poisson model cannot be a candidate for estimation and the data requires an alternative model. Cameron and Trivedi (1986), Dean and Lawless (1989) and Dean (1992) review the Lagrange Multiplier, likelihood ratio and score tests that have been developed for detecting overdispersion.

5.2 NB2 and CMP Model Comparison

5.2.1. Statistical Inference

Table 2: Negative Binomial Regression Models

Parameter	Corrected for On-Site Sampling	Zero-Truncated	Not Corrected for On-site Sampling
β_o	59.29	2.58**	2.69**
	(698.66)	(0.325)	(0.249)
β_P	-0.236	-0.211**	-0.169**
	(0.038)	(0.035)	(0.025)
β_I	1.26	1.14	1.28
	(1.675)	(1.998)	(1.481)
Male	-3.77	-3.77	-1.93
	(9.438)	(11.257)	(8.074)
School	-3.53	-4.22	-3.73
	(4.829)	(5.779)	(4.464)
Household	-0.291	-0.043	-0.046
	(3.73)	(4.443)	(3.461)
α	8.15×10^{26}	1.54	0.58
	(5.69×10^{29})	(0.191)	(0.055)
χ^2_5	40.34	39.94	46.24
$\Pr > \chi^2_5$	0	0	0
LogLik	-1840.59	-1832.79	-1892.49
AIC	3695.19	3679.58	3798.98

Table 3: Conway-Maxwell-Poisson Regression Models

Parameter	Corrected for On-Site Sampling	Not Corrected for On-site Sampling
β_o	0.105*	0.011
	(0.042)	(0.033)
β_P	-0.059**	-0.069**
	(0.0092)	(0.009)
β_I	0.246	0.265
	(0.018)	(0.014)
Male	-0.187	-0.226
	(0.102)	(0.079)
School	0.036	0.072
	(0.052)	(0.040)
Household	-0.173	-0.126
	(0.042)	(0.032)
ϕ	0.059	0.022
	(0.001)	(0.126)
Pearson χ^2_{540}	6722.25	1185.73
$\Pr > \chi^2_{540}$	0	0
LogLik	-1713.58	-1864.89
AIC	3307	3744

For both tables: ** Significant at the 5% level, * Significant at the 1% level, All parameters except for the constant scaled by 100

The on-site corrected and non-corrected models for the NB2 and the CMP models are shown in Table 2 and Table 3. The CMP model was estimated with code and assistance provided by Kimberly Sellers.

With each model, the price and income coefficients have the expected signs. The amount of trips taken will increase if the average individual's opportunity cost to visit decreases. The amount of trips also increases when the average individual's income increases. The coefficient for price is statistically significant for both models at the 1% level. However, the coefficient for income is not significant. The socio-demographic variables are also not statistically significant for each of the models.

The problem with the corrected negative binomial model is that the value for the intercept coefficient and the α parameter are very high indicating a poor fit. The reason for this is that there is a high frequency of households that only took one trip. This suggests that a zero-truncated negative binomial model would be more appropriate. For the negative binomial model, $Pr(y_i = 0) = (1 + \alpha\mu_i)^{-1/\alpha}$. After rescaling and combing with (2.7) the likelihood becomes,

$$\begin{aligned}\ell(\mu_i \mid \alpha, y_i > 0) = \ & \sum_{i=1}^{n} \Bigg[y_i \, ln \left(\frac{\alpha\mu_i}{\alpha\mu_i + 1} \right) - \frac{1}{\alpha} ln(\alpha\mu_i + 1) + ln\Gamma(y_i + 1/\alpha) \\ & -ln\Gamma(y_i + 1) - ln\Gamma(1/\alpha) - ln(1 - (1 + \mu_i)^{-1/\alpha}), \end{aligned} \tag{5.1}$$

with $\mu_i = exp(\mathbf{x}_i'\boldsymbol{\beta})$.

Table 2 shows the estimation results of this model. The estimation of the dispersion parameter α is 1.54 which is reasonable while the on-site corrected model isn't. It has an AIC of 3680, which is lower than the AIC of the on-site corrected model. Therefore the zero-truncated model fits better than the on-site corrected model.

In terms of model diagnostics, the on-site corrected CMP model has a lower AIC than any other model considered. However, the zero-truncated NB2 model has a lower AIC than the uncorrected CMP model. It should be noted that each model fails a chi-squared goodness to fit test.

5.2.2. Welfare Analysis and Expected Trips

Table 4: Expected Trips and Welfare Analysis

Model	Expected Trips.	Consumer Surplus
NB2	3.18	69.77
Truncated NB2	2.57	55.86
On-site NB2	1.57e28	49.93
CMP	23.51	171.39
On-site CMP	11.32	171.04

Since the on-site corrected versions of the models have been estimated, welfare analysis can now be performed. It should be noted that I estimated the welfare measures for the non-corrected models to illustrate differences. Table 4 illustrates the differences between the models.

Not correcting for truncation or endogenous stratification and truncation leads to a higher consumer surplus estimate for the NB2. This result is intuitive since those that are sampled on-site take more trips than the average household in the population. For the uncorrected model, the consumer surplus is $69.77 and for the corrected model it is $49.93. Since the on-site corrected model fits the data so poorly, it has a standard error of $290.30. For comparison, the consumer surplus for the zero-truncated NB2 model is $55.86.

Unlike the NB2, the uncorrected and corrected CMP differs slightly. The difference between the two models is only $0.35. For the uncorrected CMP and NB2 models, the consumer surplus ranges from $69.77 to $171.39. For the on-site corrected CMP and NB2 models, the consumer surplus ranges from $49.93 to $171.04.

A worthy goal for researchers is to be able to estimate the expected number of trips. By doing so, it allows policy-makers to use resources efficiently in order to accommodate for the use of the site. Intuitively, since this sampling was done on-site, not correcting for endogenous stratification will predict more trips. For the CMP, the results are drastic. For the corrected model, the expected numbers of trips are 11.32, while 23.41 for the uncorrected model. For the NB2 model, the zero truncated model estimates 2.57 trips. In contrast, the uncorrected model predicts 3.18 trips. For the NB2 models, the difference is much smaller.

5.3 Summary and Conclusion

Clear Lake is a popular recreational attraction for Iowans. It offers sporting activities such as boating and fishing and also features a beach for swimming. It used to have pristine water quality in the 1950s, but since then its water quality had degraded. Azevedo, Herriges and Kling (2001) show that people are concerned about the water quality treatment and would favor certain programs. Policy is dependent on knowing how to allocate scarce resources. Recreation demand studies provide a valuable tool in aiding the decisions.

One of the most important decisions that a researcher has is their choice on the econometric model to estimate the demand for the site. The Poisson and negative binomial regression models are the most common in empirical studies of recreation demand. The Conway-Maxwell-Poisson gives an alternative to the negative binomial as a count model that can accommodate overdispersion. In terms of diagnostics, the CMP model fits the best. However, it is important to note the results of Kling (1989). In that paper she finds that consumer surplus measures are sensitive the functional form of the demand function. The Poisson, NB2 and CMP models assume a semi-log demand function. So they are not robust to using other functional forms. Policy-makers must be cautious when choosing one estimate of welfare over the other.

The correction of endogenous stratification and truncation allows for on-site samples to be used for welfare analysis. If consumer surplus is estimated for uncorrected models, it is biased because it will not represent the entire population. For illustration purposes, the consumer surplus was not much difference except for that of the negative binomial model. Herriges and Egan (2006) make a great point that the correction for endogenous stratification assumes that both users and non-users have the same characteristics.

The results from this thesis show that in terms of statistical fit, the on-site corrected CMP model is the best. Unfortunately, the on-site corrected NB2 model is computationally fragile, and provided unreliable estimates. In its place, the zero-truncated NB2 model did the second best in terms of AIC. When estimating the welfare measure consumer surplus, the intuition is that not correcting for on-site sampling would lead to higher estimates; this was the case for the NB2 model but not for the CMP. The bias of on-site sampling shows up more when estimating the expected number of trips. Without correcting for this bias, the model predicts more trips to be taken because it fails to account for non-users or users that might not decide to travel to the site during that time. This is especially true for the CMP model.

APPENDIX A: A BRIEF SUMMARY OF GENERALIZED LINEAR MODELS

Nelder and Wedderburn (1972) were the first to introduce a unified framework for regression models, the generalized linear model (GLM). The main purpose of the GLM is to specify the relationship of the observed dependent variable with the covariates. A particular attractive feature of this class of models is they allow for the violation of the standard Gauss-Markov assumptions. This allows for the "linearization" of a non-linear relationship between the observed dependent variable and the covariates.

GLMs can be characterized into three components; the first component is the stochastic component. Let $y_i \overset{iid}{\sim} f(y_i)$. The assumption is that $f(y_i)$ is a member of a specific parameterization of the linear exponential family. The linear exponential family is expressed as

$$f(y_i;\theta) = exp(a(y_i)b(\theta) + c(\theta) + d(y_i)). \tag{A.1}$$

If $a(y_i) = y_i$, then the distribution is in its canonical or natural form. If there are any parameters included in the distribution, then they are referred to as nuisance parameters. The GLM framework of Nelson and Wedderburn is a canonical parameterization. Their parameterization of the linear exponential family is

$$f(y_i;\theta,\tau) = exp\left\{\frac{\theta y_i - b(\theta)}{a(\tau)} + c(y_i,\tau)\right\}. \tag{A.2}$$

The nuisance parameter τ is used to scale the errors so they follow a particular distribution of the exponential family. For example, for both the Poisson and the binomial case, τ is normalized to 1.

The systematic component is the component that produces the linear predictor. The canonical link function is the function that links the random component to the systematic component. Let the systematic component be $\eta_i = \mathbf{x_i}'\boldsymbol{\beta}$ and the link function be $g(\mu_i) = \mathbf{x}_i'\boldsymbol{\beta}$, where $g(\cdot)$ is a monotonic, differentiable, injective function. Then the inverse link function exists, and relates the mean to the systematic component such that

$$g^{-1}(g(\mu_i)) = g^{-1}(\eta_i) = g^{-1}(\mathbf{x}_i'\boldsymbol{\beta}) = \mu_i = E[y_i].$$

The mean and variance of the GLM parameterization of the linear exponential family are

$$E[Y] = \frac{\partial b(\theta)}{\partial \theta} \text{ and } Var(Y) = c(\tau)\frac{\partial^2 b(\theta)}{\partial \theta^2}.$$

The derivation of these can be found in Dobson and Barnett (2008), GLMs are estimated via the maximum likelihood method, the log-likelihood function is

$$\ell(\boldsymbol{\beta}; \mu, \tau) = \sum_{i=1}^{n} \left\{ \frac{\theta y_i - b(\theta)}{c(\tau)} + d(y_i, \tau) \right\}. \tag{A.3}$$

The first order conditions yield (A full derivation can be found in Hardin and Hilbe (2001)):

$$\frac{\partial \mu_i}{\partial \boldsymbol{\beta}} = \left(\frac{\partial g(\mu_i)}{\partial \mu_i} \right)^{-1} x_{ij}, and$$

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{\beta}} = \mathcal{V}_i = \sum_{i=1}^{n} \left\{ \frac{y_i - \mu_i}{c(\tau)Var(y_i)} x_{ij} \right\}.$$

$\mathcal{V}_i$ is the known as the score function. x_{ij} is the ith observation for the jth covariate. The parameters are obtained using the iterative reweighted least squares algorithm.

The Iterative Reweighted Least Squares algorithm (IRLS) is a type of Newton-Raphson algorithm that uses Fisher Scoring (Expected Hessian matrix) to obtain estimates of the parameters. What is nice about the IRLS is that it does not need starting values for $\hat{\boldsymbol{\beta}}$, rather it uses starting values $\hat{\mu}$, which is easier to implement (Hardin and Hilbe 2001). The

Fisher Information Matrix is $\mathcal{I} = E[\mathcal{V}_i] = \mathbf{X}'\mathbf{W}\mathbf{X}$, where$\mathbf{W}$ is a n x n weighted diagonal matrix. The estimating equation is

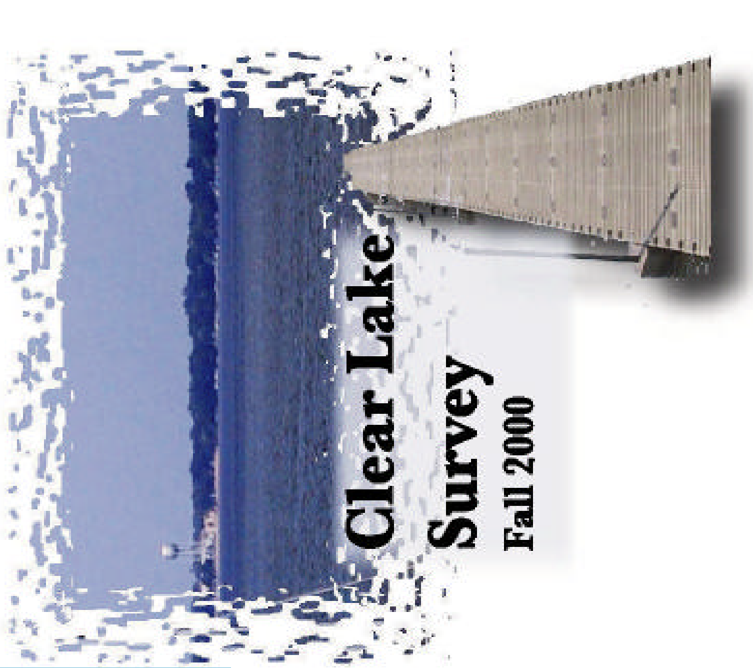
$$\boldsymbol{\beta}^{(m)} = \boldsymbol{\beta}^{(m-1)} + [\mathcal{I}^{(m-1)}]^{-1} + \mathcal{V}^{(m-1)}. \tag{A.4}$$

If $\mathcal{I}^{(m-1)}$ is multiplied to both sides of (3.1.4), the right handed side of the new equation can be expressed as $\mathbf{X}'\mathbf{W}\mathbf{z}$. Thus the estimating equation can be written as

$$\mathbf{X}'\mathbf{W}^{(m-1)}\mathbf{X}\boldsymbol{\beta}^{(m)} = \mathbf{X}'\mathbf{W}^{(m-1)}\mathbf{z}^{(m-1)} \tag{A.5}$$

where $\mathbf{W} = diag\left(\left(\frac{\partial g}{\partial \mu_i}\right)^2 Var(\mu_i)\right)^{-1}$ and $z_i = \partial g/\partial \mu_i(y_i - \mu_i) + g(\mu_i)$

APPENDIX B: THE CLEAR LAKE SURVEY



In order to make intelligent decisions concerning the future of Clear Lake, it is important to understand how the lake itself is used, as well as how this use would be affected by possible changes in the quality of the lake. The answers you give to the questions in this survey are very important in this process. Please try to answer each of the questions below. Finally, please keep in mind that, whenever we refer to Clear Lake, we are referring to the lake itself, not the town.

IN THIS FIRST SECTION, WE WOULD LIKE TO ASK YOU ABOUT how you and other members of your household use the body of water known as Clear Lake.

1. For each of the time periods listed below, please indicate how often you or other members of your household visited the lake.

Time Period	Number of Visits to the Lake
November 1999 through February 2000	
March 2000 through May 2000	
June 2000 through August 2000	
September 2000 through October 2000	
Total	

2. How many of these visits lasted longer than a single day? ______

3. In how many of the past five years did you visit the lake?______

4. Thinking about the past year, while you were visiting Clear Lake, what percentage of your time did you spend:

Activity	Percentage
Fishing	_____%
Sailing	_____%
Recreational boating (water skiing, power boating, jet skiing, etc.)	_____%
Swimming/beach use	_____%
Nature appreciation/viewing	_____%
Snowmobiling and other winter recreation	_____%
Camping	_____%
Picnicking	_____%
Other________________	_____%
	100%

5. During the time period of November 1999 through October 2000, how many visits to the following lakes or reservoirs did you take?

Lake or Reservoir	Number of Visits	Lake or Reservoir	Number of Visits
Lake Okoboji – East and West		Lake Odessa	
Lost Island Lake		Rathbun Reservoir	
Rice Lake		Mississippi River	
Spirit Lake		Minnesota Lakes	
Storm Lake		Wisconsin Lakes	
Tuttle Lake		Other:	
Saylorville Lake			
Lake Red Rock			
Coralville Reservoir			

6. During the course of the next year, how many trips do you expect to make to Clear Lake?

______ trips next year

7. Suppose that the price of visiting Clear Lake increases by $10 per trip (due for example to gas prices, user fees, or equipment costs). How many times would you visit next year?

______trips next year

8. Now suppose that the price of Clear Lake increases by $25 per trip. How many times would you visit next year?

______ trips next year

9. On a typical visit to Clear Lake, how much money do you spend in or near the town of Clear Lake?_________

IN THE FOLLOWING SECTIONS WE WILL ASK YOU SOME QUESTIONS about potential changes to the water quality of Clear Lake during the coming years. First, however, we will give you some information on the current condition of the lake. Please read this information carefully before answering the questions that follow.

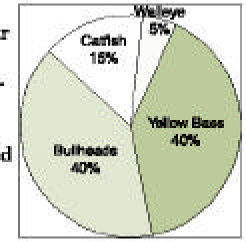
Clear Lake's Current Condition

The quality of a lake can be described in many ways. One measure of water quality is the clarity of the lake water. Water clarity is usually described in terms of how far down into the water an object is visible. The clarity of Clear Lake at the present time is about one-half to one foot. This means that objects are only visible down to about one foot under the surface of the water. The average water clarity of Clear Lake in 1953 was about ten feet.

Another measure of water quality is the amount of nutrients and other substances contained in the water. Water quality degradation can result from a number of sources, including runoff from the surrounding community containing fertilizers used for lawn care and from local agricultural sources. Currently these nutrients contribute to the occurrence of algae blooms in the lake, usually 10 to 12 times per year. Under some circumstances, these blooms can be a health concern, causing skin rashes and allergic reactions. In the past, concerns about bacteria present in Clear Lake have resulted in beach closings.

The overall quality of the water can impact on other conditions of the lake. Poor water quality results in an undesirable color and odor to the lake water. Currently, the color of Clear Lake varies between bright green and brown. The water has a mild odor that many describe as "fishy," with occasional periods of strong odor.

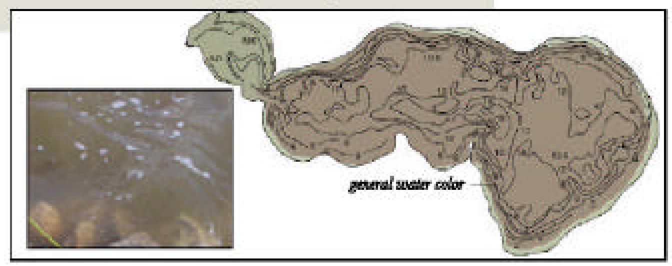
Finally, the quality of the water impacts the variety and quantity of fish in the lake. Currently, Clear Lake has a large quantity of walleye, but the largest percentages of the fish caught in the lake are fish that are considered somewhat less desirable. The chart indicates the type of fish that have been caught in the lake over the past year. While the rate at which fish are caught varies from year to year and from season to season, the typical catch rate has been 1 fish every 2 hours of fishing during the peak fishing months (May and June).



Experts believe that improved water quality would not significantly increase the number of fish in Clear Lake, but would increase the variety of fish species caught, including bass, perch, muskie, and pike.

Overall, the current condition of Clear Lake can be summarized in terms of

Water clarity	objects distinguishable 6 inches to 1 foot under water
Algae blooms	10 to 12 per year
Water color	bright green to brown
Water odor	mild odor, occasionally strong
Bacteria	possible short-term swim advisories
Fish	low diversity, good walleye

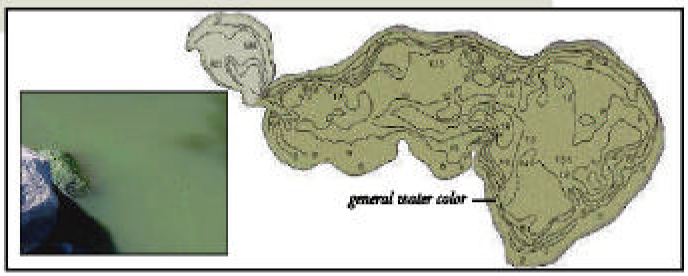


IN THIS SET OF QUESTIONS, WE WOULD LIKE TO ASK YOU about possible changes to the water quality of Clear Lake. Please answer the questions in order and do not go back and revise your earlier answers.

Plan A

If nothing is done to improve the water quality of the lake it is likely to deteriorate over the next decade. Suppose that the conditions at Clear Lake were:

Water clarity	objects distinguishable 1 inch to 5 inches under water
Algae blooms	constant
Water color	fluorescent green
Water odor	always strong
Bacteria	frequent swim advisories and/or beach closings
Fish	low diversity, mostly rough fish



10. Consider all of the recreation trips you made to Clear Lake in the past year. How many trips per year would you have made to Clear Lake if conditions were as described in **Plan A**? _______ trips per year.

IN THE NEXT FEW QUESTIONS, WE WILL BE ASKING YOU HOW you would vote on a special ballot regarding the water quality of Clear Lake. While there is currently no such ballot being considered, we would like you to respond *as if* you were voting on the project and, in each case, as if it were the **only** project available.

When you think about your answer, it is important to keep in mind that people tend to indicate that they would be willing to pay more money when payment is hypothetical than when they're really expected to pay. The idea is that it is very easy for people to say that they support a project when they know they will never have to pay any money based on their response. However, if the proposed payments are real, people may be more inclined to think about other options and what things they would have to give up to make this payment. So in answering the following questions, please keep in mind both the benefits of maintaining Clear Lake's water quality and the impact that passage of the referendum would have on your own pocketbook. In other words, please answer as if this were a **real** referendum and it was the **only** project available.

11. Would you vote "yes" on a referendum to ***maintain*** the current water quality of Clear Lake and ***avoid*** the deteriorated water quality as described under **Plan A**? The proposed project would cost you $50 (payable in five $10 installments over a five year period).

- ❑ NO
- ❑ YES

12. To help us better understand your answers, please indicate the single most important reason for your response to the preceding question:

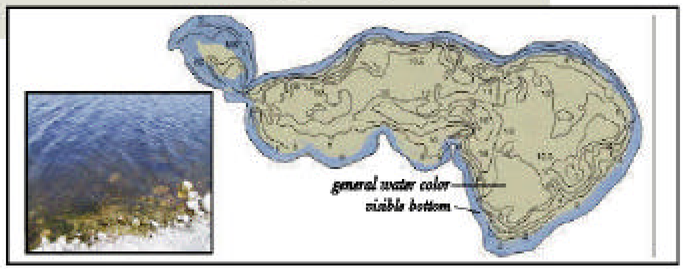
- ❑ In general, avoiding Plan A is *not* a good use of my money.
- ❑ In general, avoiding Plan A is a good use of my money.
- ❑ The plan is not realistic, or unclear
- ❑ The costs of the program should be paid for by those damaging the lake, not by me.
- ❑ I already contribute to environmental causes as much as I can afford.
- ❑ No one should have the right to damage the lake in the first place.
- ❑ Other:________________________________

Plan B

Suppose that investments could be made to actually improve the quality of Clear Lake. These investments might include establishing protection strips along the edge of the lake to reduce runoff from the surrounding area or other structural changes to the lake.

These changes would improve the lake over the next five to ten years to the following conditions:

Water clarity	objects distinguishable 2 to 4 feet under water
Algae blooms	6 to 8 per year
Water color	green to brown
Water odor	occasional mild
Bacteria	occasional swim advisories
Fish	low diversity, good walleye



13. Consider all of the recreation trips you made to Clear Lake in the past year. How many trips per year would you have made to the Lake if conditions were as described in **Plan B**? _______ trips per year

14. Would you vote "yes" on a referendum to *improve* the water quality in Clear Lake to the level described under **Plan B**? The proposed project would cost you $100 (payable in five $20 installments over a five year period).

❑ NO
❑ YES

15. To help us better understand your answers, please indicate the single most important reason for your response to the preceding question:

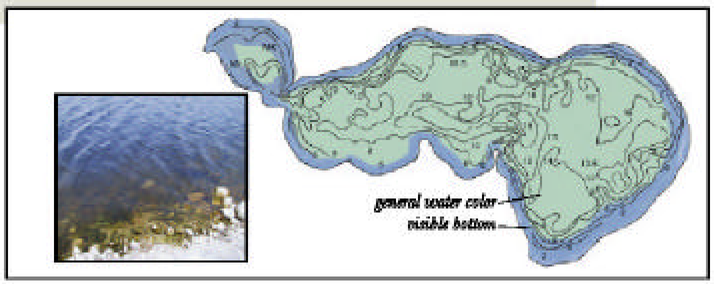
❑ In general, Plan B is *not* a good use of my money.
❑ In general, Plan B is a good use of my money.
❑ The plan is not realistic, or unclear
❑ The costs of the program should be paid for by those damaging the lake, not by me.
❑ I already contribute to environmental causes as much as I can afford.
❑ No one should have the right to damage the lake in the first place.
❑ Other:___________________________

Plan C

Now suppose that additional investments could be made such that conditions at Clear Lake would improve further. These additional changes could include retiring land from agricultural use, and programs to control nutrient runoff from urban and agricultural lands.

Suppose these changes would improve the lake over the next ten to twenty years to the following conditions:

Water clarity	objects distinguishable 6 to 8 feet under water
Algae blooms	3 to 4 per year
Water color	green to blue
Water odor	occasional mild
Bacteria	infrequent swim advisories
Fish	high diversity



16. Consider all of the recreation trips you made to Clear Lake in the past year. How many trips per year would you have made to Clear Lake if conditions were as described in **Plan C**? ________ trips per year.

17. Would you vote "yes" on a referendum to improve the water quality in Clear Lake to the level described under **Plan C**? The proposed project would cost you $200 (payable in five $40 installments over a five year period).

- ❑ NO
- ❑ YES

18. To help us better understand your answers, please indicate the single most important reason for your response to the preceding question:

- ❑ In general, Plan C is *not* a good use of my money.
- ❑ In general, Plan C is a good use of my money.
- ❑ The plan is not realistic, or unclear
- ❑ The costs of the program should be paid for by those damaging the lake, not by me.
- ❑ I already contribute to environmental causes as much as I can afford.
- ❑ No one should have the right to damage the lake in the first place.
- ❑ Other:________________________________

IN THIS SECTION, WE WOULD LIKE TO ASK YOU ABOUT YOUR opinions regarding which lake characteristics are important to you and your views regarding some specific proposals to change Clear Lake.

19. Assume you have a total of 100 importance points to assign to the lake characteristics below. Please indicate the importance of each item by allocating your100 points among the items on this list. To indicate one item is more important to you than another, you should allocate more points to it. You do not need to give points to all of the items, but remember that the total needs to equal 100.

Water clarity	
Hard, clean, sandy lake bottom in swimming areas	
Lack of water odor	
Diversity of wildlife seen at Clear Lake	
Diversity of fish species/habitat	
Quantity of fish caught	
Safety from bacteria contamination/health advisories	
Total	**100**

20. A number of projects have been suggested to accomplish improvements in the lake. How do you feel about the following possibilities?

	Strongly Support	Somewhat Support	Neutral	Somewhat Oppose	Strongly Oppose
Increased park lands and recreational areas	❑	❑	❑	❑	❑
Building of a nature center or environmental park	❑	❑	❑	❑	❑
Purchase of easements for building buffer strips	❑	❑	❑	❑	❑
Increased land idling	❑	❑	❑	❑	❑
Restoration of Ventura Marsh to improve nutrient retention	❑	❑	❑	❑	❑
Non-motor boat days	❑	❑	❑	❑	❑
Increased no-wake zones	❑	❑	❑	❑	❑
Limiting motor horse-power	❑	❑	❑	❑	❑
Lake friendly restrictions on residential development	❑	❑	❑	❑	❑
Repair of storm drains	❑	❑	❑	❑	❑

21. In order to improve water quality in the lake, changes in land use in the watershed may be needed. For example, it's likely that some land will need to be changed to low-impact use. If such changes occur, which of the following land uses do you favor? Please check all that apply.

	Strongly Support	Somewhat Support	Neutral	Somewhat Oppose	Strongly Oppose
Park lands	❑	❑	❑	❑	❑
Additional Conservation Reserve Program acreage	❑	❑	❑	❑	❑
Restored woodlands	❑	❑	❑	❑	❑
Restored prairie	❑	❑	❑	❑	❑
Restored wetlands	❑	❑	❑	❑	❑
Nature conservation area	❑	❑	❑	❑	❑
Constructed ponds	❑	❑	❑	❑	❑
Hunting reserves	❑	❑	❑	❑	❑
Restored riparian zones	❑	❑	❑	❑	❑
Perennial Agriculture	❑	❑	❑	❑	❑
Other	❑	❑	❑	❑	❑

INFORMATION ON YOU AND OTHER MEMBERS OF YOUR household will help us better understand how household characteristics affect an individual's use of Clear Lake and attitudes towards changes in them. It wil also help us to determine how representative our sample is of the state of Iowa. All of your answers are strictly confidential. The information will only be used to report comparisons among groups of people. We will never identify individuals of households with their responses. Please be as complete as possible.

22. Are you

❑ male ❑ female

23. What is your age?

❑ Under 18
❑ 18 – 25
❑ 26 – 34
❑ 35 – 49
❑ 50 - 59
❑ 60 - 75
❑ 76 +

24. What is the highest level of schooling that you have completed? (Please check only one)

❑ Eight years or less
❑ Some high school or less
❑ High school graduate
❑ Some college or trade/vocational school
❑ Two years of college or trade/vocational school
❑ College graduate
❑ Some graduate school
❑ Advanced degree

25. How many adults live in your household (over the age of 18)? ______

26. How many children live in your household (18 or under)?
____

27. If you are currently employed, how many hours a week do you typically work? ______

28. If you are currently employed, do you have the option of working additional hours to increase your total income?
- ❑ No
- ❑ Yes—if so, what would your hourly wage be?
  $______per hour

29. If you answered "no" to question 28, and you could have the option of working more or less hours, which would you prefer?
- ❑ Work more hours
- ❑ Work less hours

30. What was your total household income (before taxes) in 1999?

❑ Under $10,000	❑ $40,000-$49,999
❑ $10,000-$14,999	❑ $50,000-$59,999
❑ $15,000-$19,999	❑ $60,000-$74,999
❑ $20,000-$24,999	❑ $75,000-$99,999
❑ $25,000-$29,999	❑ $100,000-$124,999
❑ $30,000-$34,999	❑ $125,000-$149,999
❑ $35,000-$39,999	❑ Over $150,000

31. Do you own a home in Clear Lake?
- ❑ No
- ❑ Yes, If yes, are you a year-round resident?
  - ❑ No
  - ❑ Yes

Please feel free to make any additional comments about your answers to these questions or about the survey itself. Thank you for your assistance with our Clear Lake Survey.

Comments:

REFERENCES

Azevedo, C.D., J.A. Herriges and C.L. Kling (2001), 'Valuing Preservation and Improvements of Water Quality in Clear Lake', Staff Report 01-SR 94.

Becker, G.S. (1965), 'A Theory of the Allocation of Time', *Economic Journal*. 75, 493-517.

Boatwright, P., S. Borle, and J.B. Kadane (2003), 'A model of the joint distribution of purchase quantity and timing', *Journal of the American Statistical Association*. 98, 564-572.

Bockstael, N.E. and K.E. McConnell (1983), 'Welfare Measurement and Household Production Function Framework', *American Economic Review*, 73, 806-814.

Bockstael, N.E., Strand, I.E., and W.M. Hanemann (1987), 'Time and Recreation Demand Model', *American Journal of Agricultural Economics*, 69, 293-302.

Bockstael, N.E., McConnell, K.E, and I.E. Strand. (1991). Recreation, in J. Braden and C. Kolstad, (eds), *Measuring the Demand for Environmental Quality*, North Holland, Amsterdam.

Bockstael, N.E., Strand,I.E., McConnell,K.E., and Firuzeh Arsanjani (1990), 'Sample Selection Bias in the Estimation of Recreation Demand: An Application to Sportfishing', *Land Economics*, 66, 40-49.

Burt, O. and D. Brewer (1971), 'Estimation of net social benefits from outdoor recreation', *Econometrica*, 39, 813-827.

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

Cameron, A.C and P.K. Trivedi (1998) *Regression Analysis of Count Data*. Cambridge University Press, Cambridge.

Cesario, F.J. (1976), 'Value of Time in Recreation Benefit Studies', *Land Economics*, 52, 32-41.

Conway, R.W. and W.L. Maxwell (1962), 'A queuing model with state dependent service rates', *Journal of Industrial Engineering*, 12, 132-136.

Cox, D.R. (1983), 'Some Remarks on Overdispersion', *Biometrika*. 70, 269-274.

Dean, C. (1992), 'Testing for Overdispersion in Poisson and Binomial Regression Models', *Journal of the American Statistical Association*, 87, 451-457.

Dean, C.B. and J.D. Lawless (1989), 'Tests for Detecting Overdispersion in Poisson Regression Models', *Journal of the American Statistical Association*, 84, 467-472.

Diamond, P. and J. Hausman (1994), 'Contingent Valuation: Is Some Number Better than No Number?', *Journal of Economic Perspectives*, 8, 45-64.

Dobson, A.J and A.G. Barnett (2008), *An Introduction to Generalized Linear Models, 3rd Edition*, CRC Press, London.

Egan, K. (2004). *Recreation demand: On-site sampling and responsiveness of trip behavior to physical water quality measures*. Ph.D dissertation, Iowa State University.

Egan, K. and J. Herriges (2006), 'Multivariate count data regression models with individual panel data from an on-site sample', *Journal of Environmental Economics and Management*, 52, 567-581.

Englin, J. and J.S. Shonkwiler (1995), 'Estimating Social Welfare Using Count Data Models: An Application to Long-Run Recreation Demand Under Conditions of Endogenous Stratification and Truncation', *The Review of Economics and Statistics*, 77, 104-112.

Freeman, A.M. (2003) *The Measurement of Environmental and Natural Resource Values:Theory and Methods*. Resources to the Future: Washington DC.

Green, W.H (2008), *Econometric Analysis, Sixth Edition*, Pearson, Upper Sandle River.

Grogger, J.T. and R.T. Carson (1991), 'Models for Truncated Counts', *Journal of Applied Econometrics*, 6, 225-238.

Guikema, S. D. and J.P. Goffelt (2008), 'A flexible count data regression model for risk analysis', *Risk Analysis*, 28, 213-223.

Haab, T. and K.E. McConnell (1996), 'Count Data Models and the Problem of Zeros in Recreation Demand Analysis', *American Journal of Agricultural Economics*, 78, 89-102.

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

Hardin, J.W. and J. M. Hilbe (2001). *Generalized Linear Models and Extensions*. College Station, TX: Stata Press.

Hellerstein, D. (1999). Can We Count on Count Models?, in *Valuing Recreation and the Environment*, Herriges, J.A. and C.L. Kling, eds. Edward Elgar, Northhampton.

Hellerstein, D. and R. Mendelsohn. (1993), 'A Theoretical Foundation for Count Data', *American Journal of Agricultural Economics*, 75, 604-611.

Hesseln, H., Loomis, J.B., Gonzalez-Caban, A., and S. Alexander. (2003). 'Wildfire effects on hiking and biking demand in New Mexico: a travel cost study.' *Journal of Environmental Management*. 69, 359-368.

Hilbe, J. M. (2005),GNBSTRAT: Stata module to estimate generalized negative binomial with endogenous stratification. Statistical Software Components, Boston College Department of Economics. Available online at http://ideas.repec.org/c/boc/bocode/s456413.html. Last Accessed on 12/3/09.

Hilbe, J.M. (2007), *Negative Binomial Regression*, Cambridge University Press, Cambridge, UK.

Huber, P. (1967), 'The Behavior of Maximum Likelihood Estimates under Nonstandard Condtions', in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*,Volume 1. 221-233.

Johnson, N.L, Kemp, A.W., and S. Kotz (2005), *Univariate Discreet Distributions, 3rd Edition*, John-Wiley, New York.

Kling, C.L. (1989), 'The Importance of Functional Form in the Estimation of Welfare', *Western Journal of Agricultural Economics*, 14, 168-174.

Krutilla, J.V. (1967), Conversation Reconsidered. *American Economic Review*. 57, 777-786.

Lord, D., Guikema, S.D, and S.R. Geedipally (2008), 'Application of the Conway-Maxwell-Poisson generalized linear model for analyzing motor vehicle crashes', *Accident Analysis & Prevention*, 40, 1123-1134.

McConnell, K.E. (1992) 'On-Site Time in the Demand for Recreation', *American Journal of Agricultural Economics*, 74, 467-477.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*. London: Chapman and Hall.

Nelder, J. A. and R.W.M Wedderburn (1972), 'Generalized linear models', *Journal of Royal Statistical Society, Ser. A*, 135, 370-384.

Parsons, G. and A. Wilson (1997), 'Incidental and Joint Consumption in Recreation Demand, *Agricultural and Resource Economics Review*, 26, 1-6.

Patil, G.P. and C.R. Rao (1978), 'Weighted Distributions and Size-Based Sampling with Applications to Wildlife Populations and Human Families, *Biometrics*, 34, 179-189.

Paul, S.R. and R.L. Plackett (1978), 'Inference Sensitivity for Poisson Mixtures', *Biometrika*, 65, 591-602.

Phaneuf, D.J. and V.K. Smith (2005), 'Recreation Demand Models', in K. Mäler and J. Vincent, (eds), *Handbook of Environmental Economics*, Volume 2, North-Holland, Amsterdam.

Rodrigues, J., de Castro, M., Cancho, V.G, and N. Balakrishnan (2007), 'COM-Poisson cure rate survival models and an application to a cutaneous melanoma data', *Journal of Statistical Planning and Inference*, 139, 3605-3611.

Sellers, K.F and G. Schumeli (2008), 'A Flexible Regression Model for Count Data', University of Maryland School of Business Working Paper.

Santos Silvia, J.M.C (1997), 'Unobservables in count data models for on-site samples', *Economic Letters*, 54, 217-220.

Shaw, D. (1988), 'On-site Samples' Regression: Problems of Non-negative Integers, Truncation, and Endogenous Stratification', *Journal of Econometrics*, 37, 211-233.

Shaw,W.D. and P. Feather (1999), 'Possibilities for Including the Opportunity Cost of Time in Recreation Demand Systems', *Land Economics*, 75, 592-602.

Shmueli, G., Minka, T. P., Kadane, J. B., Borle, S., and P Boatwright (2005), 'A useful distribution for fitting discrete data: revival of the Conway-Maxwell-Poisson distribution', *Applied Statistics*, 54, 127-142.

Smith, V.K. (1991), 'Household Production and Environmental Benefit Measurement', in J.B Bradon and C.D Kolstad, (eds), *Measuring the Demand for Environmental Improvement*, North-Holland, Amsterdam.