INTER-AMERICAN TROPICAL TUNA COMMISSION
COMISIÓN INTERAMERICANA DEL ATÚN TROPICAL

WORKING GROUP TO REVIEW STOCK ASSESSMENTS

7TH MEETING

LA JOLLA, CALIFORNIA (USA)
15-19 MAY 2006

DOCUMENT SAR-7-07e

MODELING SHARK BYCATCH: THE ZERO-INFLATED NEGATIVE BINOMIAL REGRESSION MODEL WITH SMOOTHING

Modeling shark bycatch : the zero-inflated negative binomial regression model with smoothing

Mihoko Minami*, Cleridy E. Lennert-Cody†, Wei Gao‡, and Marlon Román-Verdesoto†

April 25, 2006

Abstract

The zero-inflated negative binomial (ZINB) regression model with smoothing is introduced for modeling count data with many zero-valued observations. Methods for estimation of confidence interval bands for model coefficients, and guidance on model selection and on estimation of the negative binomial size parameter are also presented. Use of the ZINB regression model is illustrated with shark bycatch data from the eastern Pacific Ocean tuna purse-seine fishery for 1994-2004. These data are characterized by a large percentage of zero-valued observations and also large non-zero counts. To demonstrate the utility of the ZINB regression model for the standardization of catch data, standardized temporal trends in bycatch rates estimated with the ZINB regression model are compared to those obtained from fitting Poisson, negative binomial and zero-inflated Poisson regression models to the same data. Comparison of trends among models suggests that the negative binomial regression model could be more likely to overestimate model coefficients. We investigate the reasons that fitting the negative binomial regression model to catch and bycatch data with many zero-valued observations could result in poor estimation and conclude that modelling the mechanism of extra zeros explicitly is important for standardization of count data with many zero-valued observations.

1 Introduction

Catch data on non-target species, and some target species, may be characterized by many zero-valued observations, but also include large values when aggregations of animals are caught. Modeing these data is essential to the estimation of trends in catch rates and for understanding processes that lead to increased, or decreased levels of catch. Count data have been modeled with Poisson and negative binomial distributions (*e.g.*, Walsh and Kleiber, 2001; Ward and Myers, 2005) or aggregated by fishing effort and modeled with a lognormal distribution (*e.g.*, Simpfendorfer *et al.*, 2002). However, depending on the skewness of the data and the proportion of zero-valued observations, neither the Poisson nor the negative binomial distributions may adequately describe the data, and even with an added constant, data involving large proportions of zero-valued observations will not be well-approximated by a lognormal distribution. Although the true stochastic processes that generated the data are usually not known, for species such as sharks, which may be encountered relatively infrequently, yet sometimes in large aggregations, delta-F models or zero-inflated models may provide a better fit to the data.

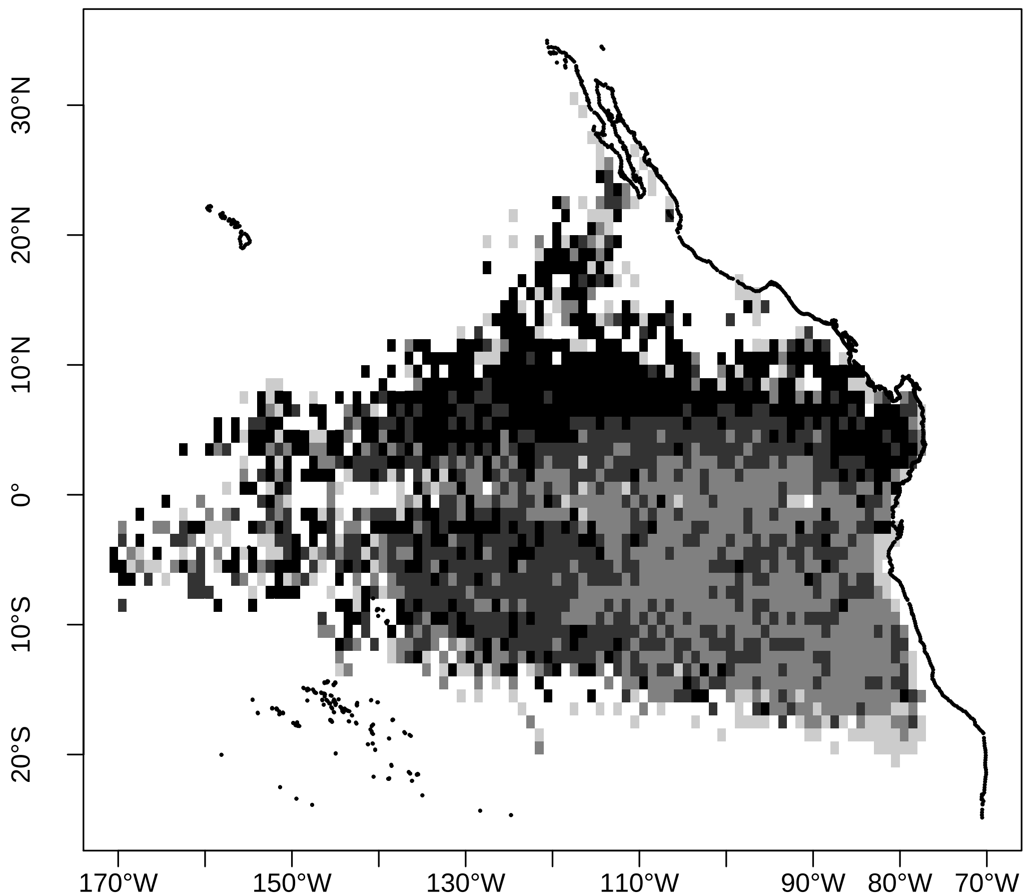
*The Institute of Statistical Mathematics and the Graduate University for Advanced Studies, Tokyo, Japan.
†Inter-American Tropical Tuna Commission, La Jolla, California, U.S.A.
‡Northeast Normal University of China, Changchun, JL, People's Republic of China

Delta-F and zero-inflated models differ somewhat in both their formulation and interpretation. Delta-F models are two-part models that describe the probability of no catch separately from the probability of positive catch. The probability of no catch is typically assumed to follow a logistic model. Positive catches are typically assumed to follow a log-linear model based on either the Poisson or the negative binomial distribution (*e.g.*, Barry and Welsh, 2002; Hoey *et al.* 2002; Punt *et al.*, 2000) for count data, or the lognormal or gamma distribution for real-valued catch data (*e.g.*, Lo *et al.*, 1992; Stefansson, 1996; Ortiz and Arocha, 2004). Mathematically, the latter is particularly simple because the lognormal and gamma distributions have no probability mass exactly at zero. However, both the Poisson and the negative binomial distributions have probability mass at zero, and thus either a zero-truncated Poisson distribution or a zero-truncated negative binomial distribution must be used to model the positive values (*e.g.*, Grogger and Carson, 1991). In terms of their interpretation, delta-F models make a distinction between covariates associated with no catch and those associated with non-zero catch.

Zero-inflated models are also expressed in two parts: the probability of being in a 'zero-state' (*e.g.*, no catch), and the probability of being in an 'imperfect-state' where positive events (*e.g.*, catch) may occur, but are not certain. That is, the imperfect-state includes *both* zero and nonzero values. The zero-state is typically modeled with a logistic, while depending on the context, the imperfect-state may be modeled with a binomial distribution (*e.g.*, for aggregated binary outcomes; Hall, 2000) or a *complete* Poisson or negative binomial distribution for un-aggregated count data (Agarwal *et al.*, 2002; Greene, 1994). These models are referred to as zero-inflated Poisson (ZIP) and zero-inflated negative binomial (ZINB) models, respectively. Because the Poisson is a special case of the negative binomial distribution, the ZINB can be viewed as a more flexible extension of the ZIP. In terms of interpretation, zero-inflated models make a distinction between covariates associated with the perfect state (no catch) and covariates associated with the imperfect state in which catch can occur, but is not certain. Conceptually, zero-inflated models may be more appropriate for catch data of infrequently-encountered species because processes leading to catch of these species are sometimes poorly understood and, therefore, difficult to model; it may be better known when catch will not occur or when it might occur than when it will occur. ZIP models may be appropriate for species that are caught infrequently, but when present occur in small groups, whereas ZINB models may better describe the data of species that when present, can occur in large aggregations.

In this manuscript we introduce the ZINB model with smoothing. The ZINB model with smoothing is an extension of the classical generalized additive model (GAM; Hastie and Tibshirani, 1991). GAMs are one of several tools frequently used to standardize catch per unit effort (CPUE) data, (*e.g.*, Maunder and Punt, 2004). To fit the ZINB model, we employ thin plate regression splines (Wood, 2003), a variant of smoothing splines that avoids complications associated with the treatment of 'knots.' We also present methods for estimating confidence interval bands for model coefficients, and we provide guidance on selecting smoothing parameters and estimating the negative binomial size parameter. To illustrate the use of the ZINB as a tool for CPUE standardization, we estimate temporal trends in the bycatch per set of silky sharks in the eastern Pacific Ocean (EPO) purse-seine fishery for tunas associated with floating objects. We compare characteristics of ZINB regression models fitted to these data, with and without smoothing, to characteristics of ZIP, negative binomial and Poisson regression models fitted to the same data. Partial dependence plots (Hastie *et al.*, 2001) are used to summarize temporal trends in bycatch per set for each of the models, taking into consideration the average effects of other predictors. Comparison of temporal trends among models illustrates important differences in the way in which the negative binomial and the ZINB fit highly skewed count data.

Figure 1: Silky shark bycatch per purse-seine set by 1^o areas of the EPO.



2 Data

Data on the incidental mortality of silky sharks (nominally *Carcharhinus falciformis*; see Appendix 1) collected by IATTC observers onboard large tuna vessels of the international purse-seine fleet between 1994 and 2004 were used to demonstrate the ZINB model. Observers go to sea aboard the largest size category of fishing vessels (> 363 metric tons fish-carrying capacity) in order to collect data on the incidental mortality of dolphins and details of fishing operations. Additionally, these observers collect data on the local environment, the amounts and species of tuna caught, and, since 1993, the bycatches of non-mammal species. The term bycatch will be used herein in place of 'catch' to refer to the incidental mortalities of any non-target species. Target species for this fishery are yellowfin tuna (*Thunnus albacares*), skipjack tuna (*Katsuwonus pelamis*), and bigeye tuna (*Thunnus obesus*).

Purse-seine sets are categorized into three types according to the intent of the fishermen. Fishermen may target tunas associated with marine mammals, tunas associated with floating objects, or unassociated schools of tunas. Floating objects include both fish-aggregating devices (FADs) and flotsam, although since 1996, more than 80% of the objects used have been estimated to be FADs (IATTC, 2005a). FADs are

Table 1: Predictors used in the analysis of silky shark bycatch per set.

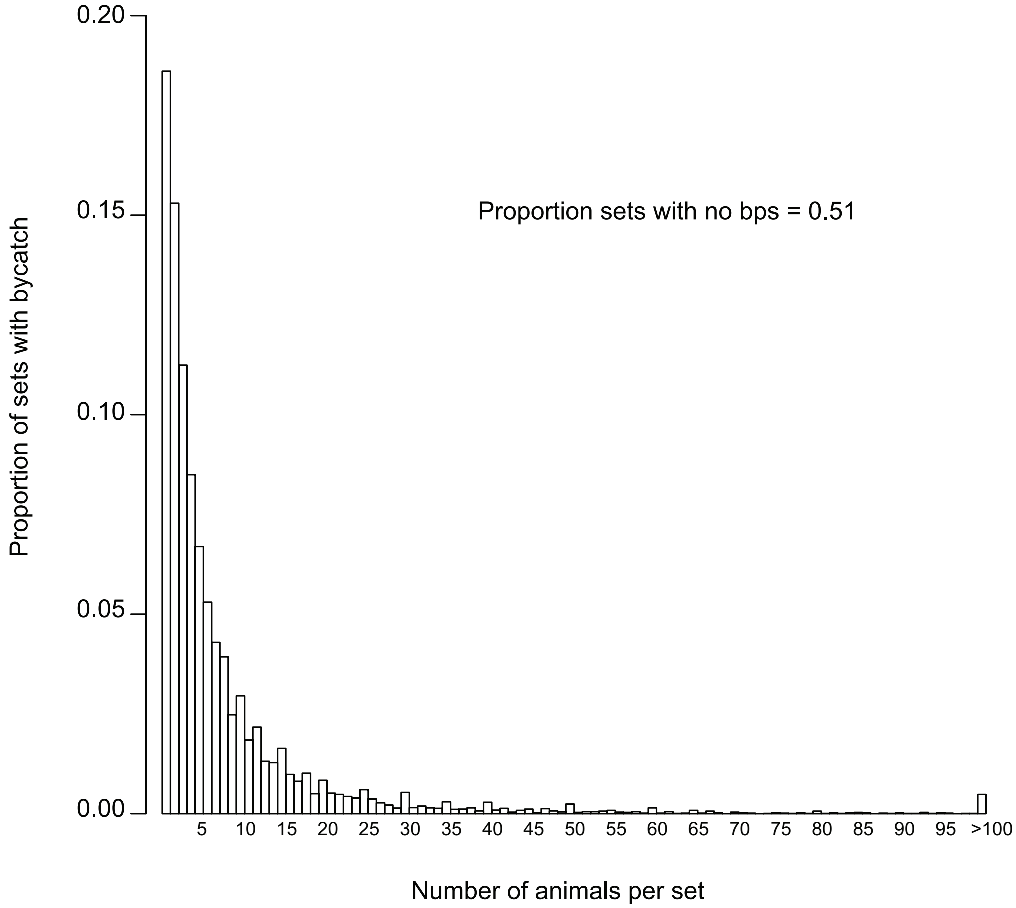
Predictor	Abbreviation	Type	Description
Year	year	Categorical	1994-2004.
Season	season	Categorical	Trimesters: January-April; May-August; September-December.
Calendar day	date	Continuous	Day of the set (1 - 365).
Latitude	lat	Continuous	Latitude in decimal degrees.
Longitude	lon	Continuous	Longitude in decimal degrees.
Time	time	Continuous	Start time of the set (local time, 24 hour clock).
Sea surface temperature	sst	Continuous	Degrees Centigrade.
Net depth	netdpth	Continuous	Approximate depth of the bottom of the net below the water's surface (fathoms).
Floating object depth	objdpth	Continuous	Approximate depth of the bottom of the floating object below the water's surface (meters).
Amount of tunas	logtuna	Continuous	Metric tons of tuna species caught (target and non-target species).
Amount of non-silky shark bycatch	lognonsilky	Continuous	Numbers of non-silky shark bycatch (excluding tuna species).
Floating object density – proxy 1	unqobjnum	Continuous	Number of unique object numbers within a 5º area around the set location and one month prior to the set date (Appendix 1).
Floating object density - proxy 2	meddisttravel	Continuous	Median distance traveled by vessels between objects within a 5º area around the set location and one month prior to the set date (Appendix 1).

typically equipped with some form of relocation equipment, such as a radio beacon or a satellite transmitter. We demonstrate the use of the ZINB model with data from purse-seine sets on tunas associated with floating objects (hereafter referred to as 'floating object' sets). In the last decade, floating object sets were largely made within two longitudinal bands north and south of the equator, extending from the coast to as far offshore as approximately $160^o - 170^o$W (Watters, 1999; Figure 1). Sampling coverage for data on non-mammal bycatch in floating object sets by IATTC observers over this 11-year period was generally greater than 64% annually (IATTC, 2006). After processing, data on 32,148 floating object sets made between 1994 and 2004 were available for analysis. Further details of the species identification and data processing can be found in Appendix 1.

The silky shark bycatch data are characterized by many zero-valued observations and a long right tail (Figure 2). Although a range of species may be found in the bycatch of all three types of purse-seine sets (IATTC, 2004), the bycatch rates of some species groups, including sharks, are estimated to be much greater in floating object sets than in the other two set types. Floating objects are thought to serve as 'attractants' (*e.g.*, Rountree, 1989; Freon and Dagorn, 2000; Marsac *et al.*, 2000) leading to the formation of large aggregations of some species. Annually, the percentage of sets with no reported silky shark bycatch has increased from approximately 40% between 1994 and 1998 to over 60% since 2001. Overall, 51% of sets had no bycatch of silky sharks (Figure 2). When silky shark bycatch did occur, sets involving up to about 20 animals were relatively common (Figure 2). The large percentage of zero-bycatch sets, combined with the fact that occasional sets had bycatches of 10s to 100s of animals, do not lend the analysis of these data to simple models that are sometimes used for count data (*e.g.*, Poisson).

A total of 12 predictors were used in this analysis. These predictors are discussed briefly below; a more detailed description of each predictor is given in Table 1. Proxies used to describe the local environment included latitude, longitude, time of day, calendar date, and sea surface temperature. The natural logarithm of the amount of total tuna catch (target and non-target tuna species) and the amount of non-silky shark bycatch were included as proxies for the size of the object-associated community. In addition, year of the set, gear characteristics (depth of the purse-seine net, depth of the floating object) and two predictors approximating the local density of floating objects were also included.

Figure 2: Frequency distribution of silky shark bycatch per set (bps), 1994-2004.



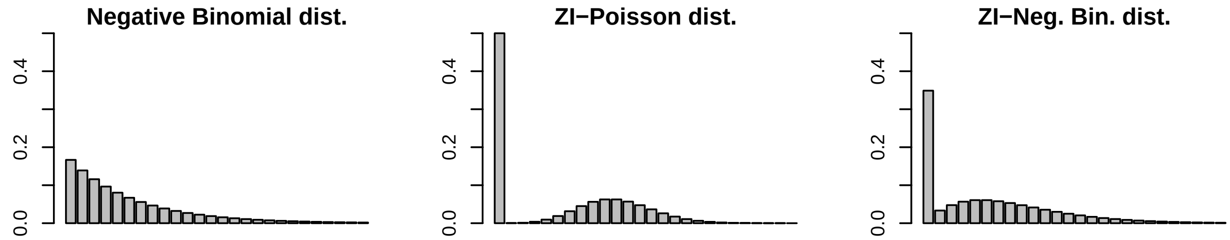
3 Methods

3.1 Zero-inflated negative binomial distribution

A zero-inflated distribution for count data is a mixture of two distributions, the delta distribution on zero (the distribution that takes only the value zero) and a distribution on the non-negative integers (*i.e.*, including the value zero). Its probability function is expressed as:

$$f(y) = p\ \delta_0(y) + (1-p)\ q(y) \qquad (1)$$

Figure 3: Probability functions of the negative binomial, ZIP and ZINB with the same mean and variance.



where $\delta_0(y)$ is the probability function of the delta distribution on zero, that is,

$$\delta_0(y) = \begin{cases} 1 & \text{for} \quad y = 0 \\ 0 & \text{for} \quad y = 1, 2, \cdots, \end{cases}$$

$q(y)$ is the probability function of the distribution on non-negative integers and p and $1-p$ are the mixing probabilities of $\delta_0(y)$ and $q(y)$, respectively. It can be seen that there are two states: a perfect state (or zero state) and an imperfect state. A sample can be in the perfect state with probability p and the imperfect state with $1-p$. If a sample is in the perfect state, it takes only the value zero; if it is in the imperfect state, it follows $q(y)$, where $y \geq 0$.

When the distribution for the imperfect state is the Poisson, $f(y)$ is the ZIP distribution, and when the distribution for the imperfect state is the negative binomial (NB):

$$q(y|\mu, \theta) \quad = \quad \frac{\Gamma(\theta + y)}{\Gamma(a)\Gamma(y+1)} \left(\frac{\theta}{\theta + \mu}\right)^{\theta} \left(\frac{\mu}{\theta + \mu}\right)^{y} \qquad \text{for} \quad y = 0, 1, 2, \cdots. \tag{2}$$

where μ and θ are the mean and the size parameters, respectively, $f(y)$ is the ZINB distribution. As the size parameter, θ, goes to $+\infty$, or equivalent, $1/\theta$ approaches 0, the NB distribution is reduced to the Poison distribution, and thus, the ZIP and the ZINB are nested.

Re-writing (1), we have

$$f(y) = \begin{cases} p + (1-p)q(0) & \text{for} \quad y = 0 \\ \\ (1-p)q(y) & \text{for} \quad y = 1, 2, \cdots. \end{cases} \tag{3}$$

The mean and the variance of the zero-inflated distribution $f(y)$ are $\mathrm{E}[Y] = (1-p)\mu$ and $\mathrm{Var}[Y] = (1-p)\tau + p(1-p)\mu^2$, where μ and τ are the mean and the variance, respectively, of $q(y)$.

Since the mean and the variance of the NB distribution are μ and $\mu + \frac{1}{\theta}\mu^2$, respectively, the variance of the ZINB distribution can be re-expressed as:

$$\mathrm{Var}[Y] = \quad (1-p)\mu + (1-p)\left(p + \frac{1}{\theta}\right)\mu^2 \quad = \mu^* + \frac{p + \frac{1}{\theta}}{1-p}\,\mu^{*2}$$

where

$$\mathrm{E}[Y] = \quad (1-p)\mu \quad \equiv \ \mu^*.$$

Thus, the variance of the ZINB distribution is a quadratic function of the ZINB mean and is in the same form as that of the NB distribution. Figure 3 depicts probability functions of the NB distribution

($\mu = 5, \theta = 1$), the ZIP distribution ($\mu = 10, p = 0.5$) and the ZINB distribution ($\mu = 7.5, p = 1/3, \theta = 3$). These three distributions have the same mean (= 5) and the same variance (= 30), however, the probability functions have quite different shapes. Note that the variance function for the ZIP, $\mu^* + \dfrac{p}{1-p}\mu^{*2}$, has the same form as that of the NB and the ZINB. This suggests that simply specifying the variance function in the framework of generalized linear models is not sufficient to distingush these distributions.

3.2 Zero-inflated negative binomial regression model

Lambert (1992) proposed the ZIP regression model in which p is related to covariates using a logistic regression model, and a loglinear regression model is used to relate the Poisson mean to covariates in the imperfect state. For count data that are highly skewed with a heavy right tail (*e.g.*, Figure 2), the NB distribution instead of the Poisson distribution, may be considered for the imperfect state. The probability function for a ZINB regression model is expressed as:

$$f(y_i|B_i, G_i, \boldsymbol{\beta}, \boldsymbol{\gamma}, \theta) = \begin{cases} p_i + (1-p_i)q(0|\mu_i, \theta) & \text{for} \quad y_i = 0 \\ \\ (1-p_i)q(y_i|\mu_i, \theta) & \text{for} \quad y_i = 1, 2, \cdots \end{cases} \tag{4}$$

where

$$\begin{aligned} q(y_i|\mu_i, \theta) &= \frac{\Gamma(\theta + y_i)}{\Gamma(\theta)\Gamma(y_i+1)} \left(\frac{\theta}{\theta+\mu_i}\right)^{\theta} \left(\frac{\mu_i}{\theta+\mu_i}\right)^{y_i} \\ \log(\mu_i) &= B_{i0} + B_{i1}\beta_1 + \cdots + B_{ik_\beta}\beta_{k_\beta} = B_i\boldsymbol{\beta} \qquad \text{and} \\ \text{logit}(p_i) = \log\frac{p_i}{1-p_i} &= G_{i0} + G_{i1}\gamma_1 + \cdots + G_{ik_\gamma}\gamma_{k_\gamma} = G_i\boldsymbol{\gamma}. \end{aligned}$$

B_i and G_i are row vectors containing covariate values for the i^{th} observation, for NB and logistic regression models, respectively. The maximum likelihood estimates for $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ and θ are obtained by maximizing the log-likelihood function $L(\boldsymbol{\beta}, \boldsymbol{\gamma}, \theta|\boldsymbol{y}, \boldsymbol{B}, \boldsymbol{G}) = \sum_{i=1}^{N} \log f(y_i|B_i, G_i, \boldsymbol{\beta}, \boldsymbol{\gamma}, \theta)$ with respect to $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ and θ. Appendix 2 gives the gradient vector and the Hessian matrix of the log-likelihood function, which are used in gradient-based optimization methods.

The EM algorithm (Dempster *et al.*, 1977) is an alternative to gradient-based optimization methods such as quasi-Newton methods for finding maximum likelihood estimates. The convergence of EM algorithm can be slow, but it is easy to implement and its update formula provides information on the statistical properties of the estimates. Here we introduce a random variable Z that takes the value 1 if the observation is in the perfect state and 0 otherwise. The random variable Z is only partially observable; its value is known if y is strictly greater than zero, but it is not known otherwise. If Z were fully observable, the log-likelihood based on y and Z would be given by

$$\begin{aligned} \log L_c(\boldsymbol{\beta}, \boldsymbol{\gamma}, \theta|\boldsymbol{y}, \boldsymbol{Z}, \boldsymbol{B}, \boldsymbol{G}) &= \sum_{i=1}^{N} \{Z_i \log p_i + (1-Z_i)\log(1-p_i)\} + \sum_{i=1}^{N}(1-Z_i)\log q(y_i|\mu_i, \theta) \\ &= L_l(\boldsymbol{\gamma}|\boldsymbol{Z}, \boldsymbol{G}) + L_b(\boldsymbol{\beta}, \ \theta|\boldsymbol{y}, \boldsymbol{Z}, \boldsymbol{B}). \end{aligned}$$

Thus, the estimation problem would be separated into two parts: the estimation of $\boldsymbol{\gamma}$ by the logistic regression model with Z as the reponse variable (*i.e.*, $L_l(\boldsymbol{\gamma}|\boldsymbol{Z}, \boldsymbol{G})$), and the estimation of $\boldsymbol{\beta}$ and θ by the negative binomial regression model with y as the response variable and $1 - Z$ as weight (*i.e.*, $L_b(\boldsymbol{\beta}, \ \theta|\boldsymbol{y}, \boldsymbol{Z}, \boldsymbol{B})$).

The EM algorithm maximizes the log-likelihood $L(\boldsymbol{\beta}, \boldsymbol{\gamma}, \theta|\boldsymbol{y}, \boldsymbol{B}, \boldsymbol{G})$ based on observed data $\boldsymbol{y}$ by iteratively maximizing the conditional expectation $\mathrm{E}\left[\log L_c(\boldsymbol{\beta}, \boldsymbol{\gamma}, \theta|\boldsymbol{y}, \boldsymbol{Z}, \boldsymbol{B}, \boldsymbol{G})|\, \boldsymbol{y}, \boldsymbol{\beta}^{(k-1)}, \boldsymbol{\gamma}^{(k-1)}, \theta^{(k-1)}\right]$, where the expectation is taken with respect $\boldsymbol{Z}$ given $\boldsymbol{y}$, and $\boldsymbol{\beta}^{(k-1)}$, $\boldsymbol{\gamma}^{(k-1)}$ and $\theta^{(k-1)}$ (parameter values after $k-1$ iterations). The maximization of this conditional expectation can be achieved in two separate steps once the conditional expectation of $\boldsymbol{Z}$ is computed.

The EM algorithm is easily implemented in any software that has functions for the NB and the logistic regression models. At the kth iteration, the EM algorithm involves the following steps:

E-step: Compute $z_i^{(k)} = \mathrm{E}\left[Z_i|y_i, \boldsymbol{\beta}^{(k-1)}, \boldsymbol{\gamma}^{(k-1)}, \theta^{(k-1)}\right]$

$$= \begin{cases} \dfrac{p_i^{(k-1)}}{p_i^{(k-1)} + \left(1 - p_i^{(k-1)}\right) \; q\left(0|\mu_i^{(k-1)}, \theta^{(k-1)}\right)} & \text{for} \quad y_i = 0 \\ 0 & \text{for} \quad y_i = 1, 2, \cdots \end{cases}$$

M-step for $\boldsymbol{\beta}$: Obtain estimates $\boldsymbol{\beta}^{(k)}$ by fitting the NB regression model, using weights $1 - z_i^{(k)}$ and response variable y_i. Update the estimate of θ, $\theta^{(k)}$, if θ is unknown (see below).

M-step for $\boldsymbol{\gamma}$: Obtain estimates $\boldsymbol{\gamma}^{(k)}$ by fitting the logistic regression model, using response variable $z_i^{(k)}$.

Several details regarding implementation of the ZINB are worth noting. Some standard generalized linear model packages may not include the NB distribution because, unlike the Poisson distribution and the binomial distribution, the NB distribution does not belong to the exponential family when θ is unknown. In the freeware R (http://www.r-project.org/), the `glm.nb` function of the MASS library computes the maximum likelihood estimate for both the coefficients and θ. Because the EM algorithm can be slow to converge, for the analysis of the shark bycatch data, we used the EM algorithm with `gam` function of the `mgcv` library for several iterations to obtain good initial values, and then switched to a quasi-Newton method using the `optim` function of the MASS library to get faster convergence. The `gam` function of `mgcv` library estimates θ by the method of moments, and it is more likely to give reasonable estimates of the model parameters even when `glm.nb` function of MASS library fails to do so.

3.3 Zero-inflated negative binomial regression model with smoothing

To incorporate more flexibility into the ZINB regression model, we employ a smoothing method to allow for smoothed functions of some variables. Smoothing splines (Wahba, 1990) are estimated by finding the maximizer of a penalized log-likelihood function that is the sum of a measure of fitness (the log-likelihood) and a penalty for wiggliness. Smoothing splines provide an excellent means for estimation and inference with models such as the ZINB, however, there are obstacles to the adoption of smoothing splines in practical work. Smoothing splines to n data points use n basis functions, thus requiring the estimation of n parameters. This involves prohibitively high computational cost, especially in the case of multivariate smoothing, and often causes computational instability.

Thin plate regression splines (t.p.r.s., Wood, 2003) are optimal low rank approximations to smoothing splines that are constructed by a simple transformation and truncation of the basis that arises from the solution of the smoothing spline problem. Thus, t.p.r.s. approximate smoothing splines with a much smaller number of basis functions and are computationally efficient and stable. t.p.r.s. also avoid the cumbersome problems associated knot placement. With t.p.r.s. the value of the smoothed function of the i^{th} observation for the j^{th} variable, say $x_i^{(j)}$, is expressed as

$$s_j(x_i^{(j)}) = B_i^{(j)} \boldsymbol{\beta}^{(j)} \tag{5}$$

where $s(\cdot)$ denotes a smoothed function, $B_i^{(j)}$ is the row vector consisting of values of basis functions corresponding to $x_i^{(j)}$, and $\boldsymbol{\beta}^{(j)}$ is the vector of parameters for the variable j^{th}. If we combine row vectors

$B_i^{(j)}$ $(j = 1, ..., k_{\boldsymbol{\beta}})$ and covariates vectors for the non-smoothed terms into a new covariate vector B_i, and combine parameter vectors $\boldsymbol{\beta}^{(j)}$ $(j = 1, ..., k_{\boldsymbol{\beta}})$ and parameters vectors for non-smoothed terms into a new parameter vector $\boldsymbol{\beta}$, the NB part of the ZINB regression model can be expressed in the same way as that without smoothing: $log(\mu_i) = B_i\boldsymbol{\beta}$. Following similar reasoning, the logistic part of the ZINB regression model with smoothing can be expressed as: $\mathrm{logit(p_i) = G_i}\boldsymbol{\gamma}$.

In order to avoid overfitting, we use the following penalized log-likelihood for the ZINB regression model with t.p.r.s., which takes into account the wiggliness penalty:

$$L_p(\boldsymbol{\beta}, \boldsymbol{\gamma}; \boldsymbol{y}) = \sum_{i=1}^{n} \log f(y_i; \boldsymbol{B}_i, \boldsymbol{G}_i, \boldsymbol{\beta}, \boldsymbol{\gamma}, \theta) - \frac{1}{2}\sum_{j=1}^{k_\beta} \lambda^{(j)} \boldsymbol{\beta}^{(j)T} S_\beta^{(j)} \boldsymbol{\beta}^{(j)} - \frac{1}{2}\sum_{j=1}^{k_\gamma} \nu^{(j)} \boldsymbol{\gamma}^{(j)T} S_\gamma^{(j)} \boldsymbol{\gamma}^{(j)}, \tag{6}$$

where k_β and k_γ are the numbers of smoothed functions, $S_\beta^{(j)}$ and $S_\gamma^{(j)}$ are smoothing matrices, bmB_i and bmG_i are matrices consisting of basis functions and non-smoothed covariates, $\lambda^{(j)}$ and $\nu^{(j)}$ are smoothing parameters, for the logistic and the negative binomial regression parts, respectively. Given values of $\lambda^{(j)}$ and $\nu^{(j)}$, we estimate $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ by maximizing (6).

For t.p.r.s., as with smoothing splines, smoothing parameters play a key role in controling the trade-off between flexibility of the model and the ability of the model to generalize to new data. Wood (2004) proposed a stable and efficient multiple smoothing parameter estimation method for GAMs which is implemented in the `gam` function. For smoothing parameter selection the `gam` function employs the Unbiased Risk Estimator (UBRE, Wood, 2004; Hurvich *et al.* 1998). We used the EM algroithm with this `gam` function. Because the EM algorithm is an iterative fitting procedure, iterative smoothing parameter selection using the UBRE is equivalent to find the smoothing parameter values that minimize

$$V_u = -\frac{2}{n}\sum_{i=1}^{n} \log f(y_i; \boldsymbol{B}_i, \boldsymbol{G}_i, \boldsymbol{\beta}, \boldsymbol{\gamma}, \theta) - \frac{2}{n}\mathrm{tr}\{I - A_\beta\} - \frac{2}{n}\mathrm{tr}\{I - A_\gamma\} \tag{7}$$

where A_β and A_γ are the influence matrices or hat matrices for the negative binomial regression and logistic regression parts, respectively. The estimates for coefficients are obtained as the maximizer of the penalized log-likelihood (6) given the smoothing parameter values.

In practice problems can arise because the `gam` uses method of moments estimation for θ, if θ is not specified. Convergence of the EM algorithm is not guaranteed if the method of moments estimate is used for θ. To overcome this problem, we estimate $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ using the EM algorithm for fixed values of θ, and then compare the Generalized Information Criteria (GIC) described in the next subsection among models to find the value of θ that has the best generalization ability.

3.4 Model selection

To compare generalization abilities of the different models, we use the GIC of Konishi and Kitagawa (1996). Akaike's information criterion (AIC, Akaike, 1974) is a widely-used model selection tool. GIC generalizes AIC to estimation methods other than maximum likelihood. The smaller the GIC value, the better the fit of the model to the underlying distribution of the data. The GIC for the ZINB with t.p.r.s. is given by

$$\mathrm{GIC} = -2\sum_{i=1}^{n} \log f(y_i; \boldsymbol{B}_i, \boldsymbol{G}_i, \boldsymbol{\beta}, \boldsymbol{\gamma}, \theta) + 2\mathrm{tr}\left\{\mathrm{M}^{-1}\mathrm{R}\right\}$$

where

$$M = -\left.\frac{\partial^2 L_p(\boldsymbol{\beta}, \boldsymbol{\gamma}; \boldsymbol{y})}{\partial(\boldsymbol{\beta}, \boldsymbol{\gamma})\partial(\boldsymbol{\beta}, \boldsymbol{\gamma})^T}\right|_{(\boldsymbol{\beta}, \boldsymbol{\gamma}) = (\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}})} \tag{8}$$

$$R = \frac{\partial L_p(\boldsymbol{\beta}, \boldsymbol{\gamma}; \boldsymbol{y})}{\partial(\boldsymbol{\beta}, \boldsymbol{\gamma})} \frac{\partial \sum_{i=1}^{n} \log f(y_i; \boldsymbol{B}_i, \boldsymbol{G}_i, \boldsymbol{\beta}, \boldsymbol{\gamma}, \theta)}{\partial(\boldsymbol{\beta}, \boldsymbol{\gamma})^T}\Bigg|_{(\boldsymbol{\beta}, \boldsymbol{\gamma})=(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}})}$$

$$\frac{\partial^2 L_p(\boldsymbol{\beta}, \boldsymbol{\gamma}; \boldsymbol{y})}{\partial(\boldsymbol{\beta}, \boldsymbol{\gamma})\ \partial(\boldsymbol{\beta}, \boldsymbol{\gamma})^T} = \frac{\partial^2 \log f(\boldsymbol{y}; \boldsymbol{\beta}, \boldsymbol{\gamma})}{\partial(\boldsymbol{\beta}, \boldsymbol{\gamma})\partial(\boldsymbol{\beta}, \boldsymbol{\gamma})^T} - \text{diag}\Big(\lambda^{(1)}\mathrm{S}_\beta^{(1)}, \cdots, \lambda^{(\mathrm{k}_\beta)}\mathrm{S}_\beta^{(\mathrm{k}_\beta)}, 0, \cdots, 0, \nu_{(1)}S_\gamma^{(1)}, \cdots, \nu_{(k_\gamma)}S_\gamma^{(k_\gamma)}, 0, \cdots, 0\Big) \quad \text{and}$$

$$\frac{\partial L_p(\boldsymbol{\beta}, \boldsymbol{\gamma}; \boldsymbol{y})}{\partial(\boldsymbol{\beta}, \boldsymbol{\gamma})} = \frac{\partial \log f(\boldsymbol{y}; \boldsymbol{\beta}, \boldsymbol{\gamma})}{\partial(\boldsymbol{\beta}, \boldsymbol{\gamma})} - \text{diag}\Big(\lambda^{(1)}\mathrm{S}_\beta^{(1)}\boldsymbol{\beta}^{(1)}, \cdots, \lambda^{(\mathrm{k}_\beta)}\mathrm{S}_\beta^{(\mathrm{k}_\beta)}\boldsymbol{\beta}^{(\mathrm{k}_\beta)}, 0, \cdots, 0, \nu_{(1)}S_\gamma^{(1)}\boldsymbol{\gamma}^{(1)}, \cdots, \nu_{(k_\gamma)}S_\gamma^{(k_\gamma)}\boldsymbol{\gamma}^{(k_\gamma)}, 0, \cdots, 0\Big).$$

3.5 Statistical testing

For the models without smoothing, we can consider the following testing options:

- Testing a ZIP regression model against ZINB alternatives;
- Testing a NB regression model against ZINB alternatives.

For testing a ZIP regression model against ZINB alternatives, we denote $\alpha = 1/\theta$ and conduct a statistical test of the null hypothesis $H_o : \alpha = 0$ (ZIP regression model) against $H_a : \alpha \neq 0$. The log-likelihood ratio test can be performed by fitting ZINB models with $\alpha = 0$ and without this restriction. Ridout *et al.* (2001) introduced a score test procedure that can be performed without having to fit a ZINB regression model.

The ZINB model and the NB regression model are not nested, and thus, the log-likelihood ratio test and score test can not be applied. Instead, Vuong's test (Vuong, 1989; Long, 1997) may be used. Let $\widehat{P}_{ZINB}(y_i|\boldsymbol{x}_i)$ and $\widehat{P}_{NB}(y_i|\boldsymbol{x}_i)$ be the predicted probability of y_i given $\boldsymbol{x}_i$ by a ZINB regression model and a NB regression model, respectively. Let

$$m_i = \log\left(\frac{\hat{P}_{ZINB}(y_i|\boldsymbol{x}_i)}{\hat{P}_{NB}(y_i|\boldsymbol{x}_i)}\right), \quad \bar{m} = \frac{1}{N}\sum_{i=1}^{N} m_i \quad \text{and} \quad s_m = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(m_i - \bar{m}))^2}.$$

Then, the Vuong statistic

$$V = \frac{\sqrt{N}\bar{m}}{s_m}$$

can be used to test the null hypothesis $H_o : \mathrm{E}(m) = 0$. The statistic V asymptotically follows a standard normal distribution if $\mathrm{E}(m) = 0$. Thus, we would chose the ZINB regression model if $z_\alpha \leq V$ where z_α is the critical value for the standard normal distribution at level α.

3.6 Asymptotic covariance matrix

The asymptotic covariance matrix for $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ can be estimated by the following so-called 'sandwich' formula: (Huber, 1981; Hampel *et al.* 1986):

$$V = M^{-1}QM^{-1}$$

where M is given by (8) and

$$Q = \frac{\partial L_p(\boldsymbol{\beta},\boldsymbol{\gamma};\boldsymbol{y})}{\partial(\boldsymbol{\beta},\boldsymbol{\gamma})}\frac{\partial L_p(\boldsymbol{\beta},\boldsymbol{\gamma};\boldsymbol{y})}{\partial(\boldsymbol{\beta},\boldsymbol{\gamma})^T}\bigg|_{(\boldsymbol{\beta},\boldsymbol{\gamma})=(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\gamma}})}.$$

Since smoothed functions are linear functions of the coefficients (5), their variances are given by:

$$\mathrm{Var}\left(\hat{s_j}(x_i^{(j)})\right) = B_i^{(j)}\mathrm{Var}\left(\boldsymbol{\beta}^{(j)}\right)B_i^{(j)T},$$

where $\mathrm{Var}\left(\boldsymbol{\beta}^{(j)}\right)$ is the corresponding block in V.

3.7 Partial dependence plots

To study the marginal effect of a predictor on the response variable (*e.g.*, CPUE), we can use partial dependence plots (Hastie *et al.*, 2001). We denote by X_c covariates whose effects on the response variable we would like to summarize and other covariates in the model by X_s. We denote by $f(X_s, X_c)$ the expected value of the response variable obtained from a model with covariates X_s and X_c . The partial dependence of $f(X)$ on X_s is defined as

$$f_s(X_s) = \mathrm{E}_{X_c} f(X_s, X_c).$$

This can be estimated by

$$f_s(X_s) = \frac{1}{n}\sum_{i=1}^{n} f(X_s, x_{iC}),$$

where $\{x_{1C}, x_{2C}, \cdots, x_{nC}\}$ are the values of X_C occurring in the data set used to fit the model. Thus, a partial dependence plot summarizes the effect of X_s on the response variable, having accounted for the average effects of the other X_c predictors.

4 Results

To compare the fit of the various models to the shark bycatch data, we divided the data into two groups: a training data set and a test data set. Models were fitted using the training data, and predicted distributions were computed and investigated using the test data. The models we fitted were a Poisson regression model with smoothing (Poisson with t.p.r.s.), NB regression models with and without smoothing (NB with t.p.r.s., NB), zero-inflated Poisson regression model with smoothing (ZIP with t.p.r.s.), and ZINB regression models with and without smoothing (with t.p.r.s., ZINB, ZINB). The response variable was silky shark bycatch per set (number of animals per set). The predictors 'latitude,' 'longitude,' 'time' and 'date' were included as smoothed terms into the models that used t.p.r.s.. The number of basis functions for each smoothed term with t.p.r.s. was set to 10 (default of the `gam` function in `mgcv` package for univariate smoothing). 'year' was treated as a factor with 11 levels (1994 to 2004). For the zero-inflated models, the same set of covariates was used for both parts of the models. For models without smoothing, we used maximum likelihood to estimate θ. For models with smoothing, we first computed the method of moments estimate (mme) for θ, and then fitted models with various values of θ around the mme, and chose the model with the smallest GIC value.

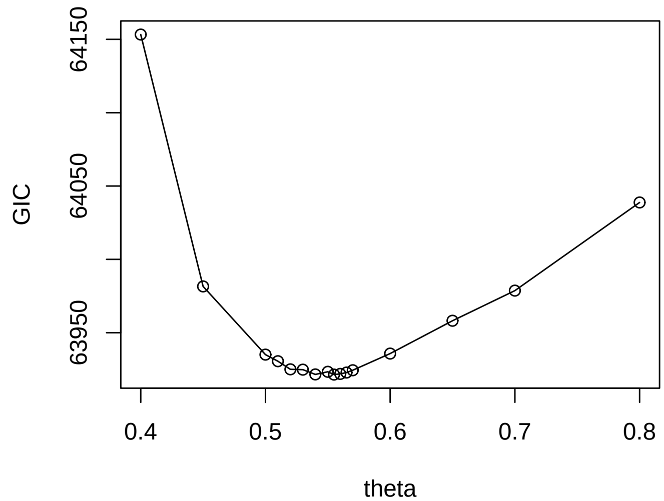
4.1 Model comparison

The magnitude of the GIC values obtained from the test data were in the following order (Table 2):

Table 2: Summary statistics of model fit. Shown for each model are the estimate of θ (where applicable), the log-likelihood for the training data, the GIC for the test data, the difference in GIC from the best (least) GIC value of the eight models. For NB and ZINB models, methods of estimation are also listed.

	Model	theta	log–likelihood	GIC	Difference in GIC	Estimation method for theta
1	Poisson with t.p.r.s.	–––	−81848.7	100000.00<	40000<	
2	NB	0.313	−32867.5	65818.42	1897.0	mle
3	NB with t.p.r.s.	0.330	−32571.9	65280.37	1358.9	best GIC
4	ZIP with t.p.r.s.	–––	−56388.9	100000.00<	40000<	
5	ZINB	0.580	−32345.7	64826.90	905.4	mle
6	ZINB with t.p.r.s.	0.400	−31947.1	64153.29	231.8	
7	ZINB with t.p.r.s.	0.800	−31915.8	64038.78	117.3	
8	ZINB with t.p.r.s.	0.555	−31862.4	63921.46	0.0	best GIC

Figure 4: GIC values for different θ for the ZINB with t.p.r.s.

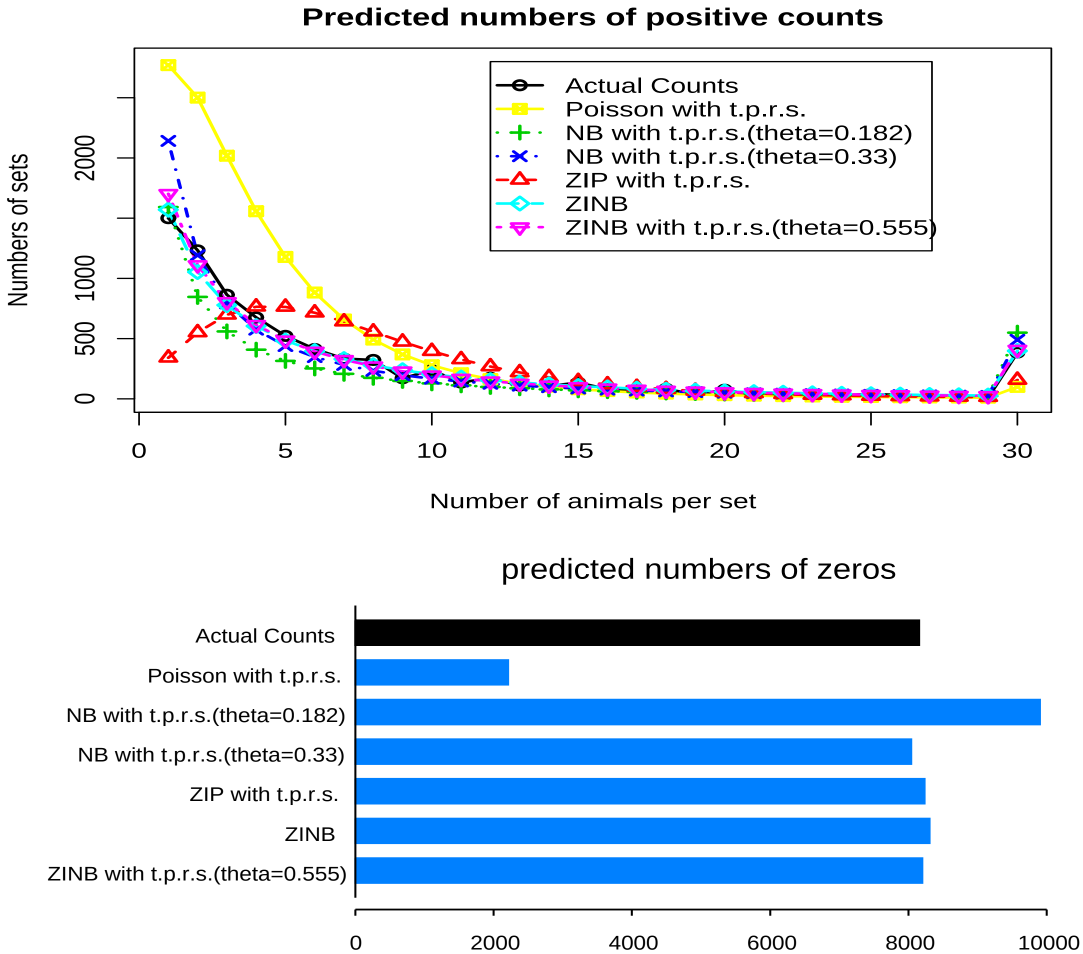


$$\text{ZINB} < \text{NB} << \text{ZIP} << \text{Poisson}$$
$$\text{Model with t.p.r.s.} < \text{Model without t.p.r.s.}$$

suggesting that the ZINB regression model with smoothing provides the best fit to the shark bycatch data. The ZINB with t.p.r.s. ($\theta = 0.555$) has the lowest GIC value, that is, it is the best model with respect to GIC among the eight models; the Poisson with t.p.r.s. and ZIP with t.p.r.s. have the greatest GIC values. Although it is zero-inflated, the ZIP with t.p.r.s. has a much larger GIC value than the NB, even without t.p.r.s. Models with t.p.r.s. had smaller GIC values than corresponding models without t.p.r.s. Because the differences in the log-likelihood are large compared to the number of parameters for these data, the order of the GIC values among the eight models is the same as that of -2 times log-likelihood values.

Figure 4 depicts the values of the GIC for different values of θ for the ZINB with t.p.r.s. The GIC takes the smallest value at $\theta = 0.555$.Note that differences in GIC in this plot are much less than differences in GIC among various models shown in Table 2.

Figure 5: Actual and predicted frequencies of silky shark bycatch per set by the Poisson with t.p.r.s., the ZIP with t.p.r.s., the NB with t.p.r.s. ($\theta = 0.182$), NB with t.p.r.s ($\theta = 0.33$), the ZINB and the ZINB with t.p.r.s.($\theta = 0.555$). The upper plot shows the predicted frequencies of positive counts and the lower plot the predicted number of zero counts.



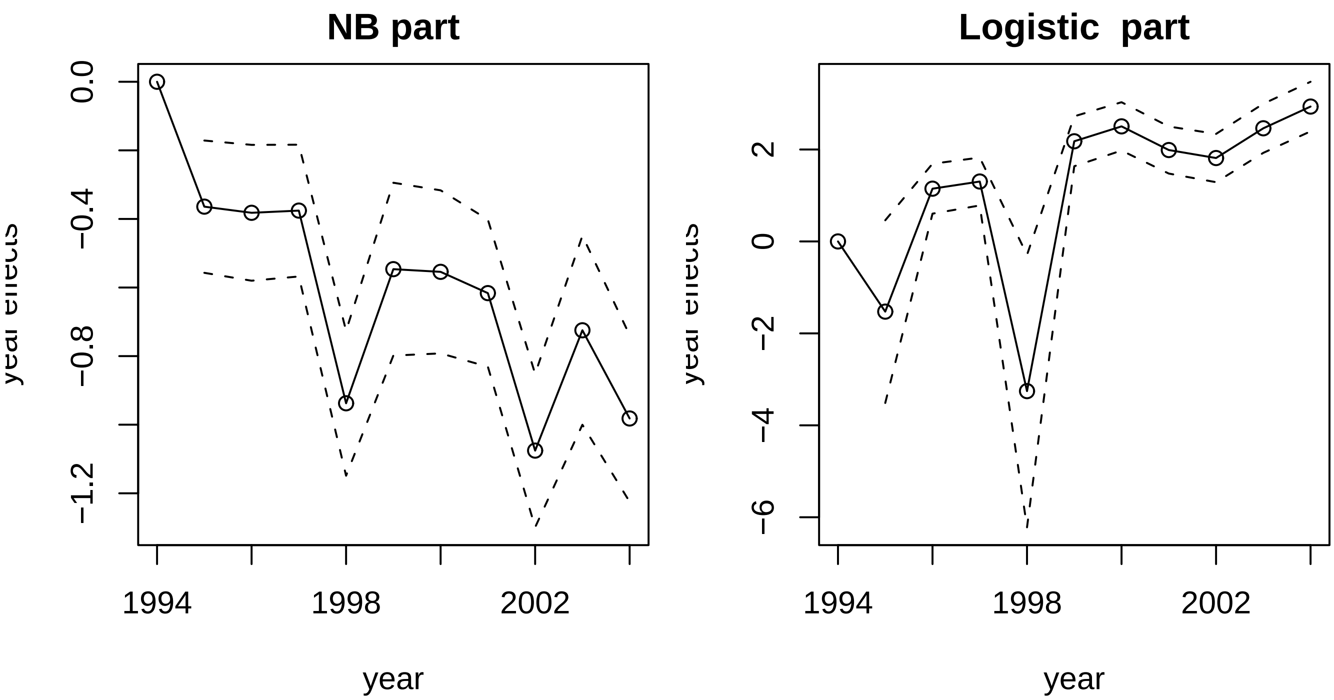
In order to see how well the models predict bycatch per set, we computed the predicted frequencies for the test data:

$$\sum_{i=1}^{N_{test}} f(y|B_i^{test},\ G_i^{test},\ \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}, \theta),$$

where $f(y|B_i, G_i, \boldsymbol{\beta}, \boldsymbol{\gamma}, \theta)$ is the probability function (4), N_{test} is the sample size of the test data, B_i^{test} and G_i^{test} are the vectors of values for basis functions of t.p.r.s. and covariates, as described in section 3.2, and $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\gamma}}$ are the estimated coefficients using the training data. The ZINB with and without t.p.r.s. capture the actual distribution well, compared to the other models (Figure 5). The predicted distribution by the ZIP model with t.p.r.s. exhibits the worst fit to the positive counts. The NB with t.p.r.s. ($\theta = 0.33$) fits relatively well for counts larger than or equal to two, but over-predicts the number of ones, and under-predicts the number of zeros, while the NB with t.p.r.s. ($\theta = 0.182$) under-predicts for positive bycatch and over-predicts the number of zeros. The zero-inflated models perform the best with respect to predicting sets with no bycatch. The NB with t.p.r.s. over-predicted the number of zeros when θ was small (0.182) and under-predicted the number of zeros when θ was the minimizer of GIC ($\theta = 0.33$). The Poisson with t.p.r.s. over-predicted positive counts up to 12 and severely under-predicted the number of zeros.

4.2 Estimated coefficients for the ZINB model with t.p.r.s.

Figure 6: Estimates of the year effects from the ZINB regression model with smoothing ($\theta = 0.555$) and their approximate confidence bands (estimate +/- twice standard error).



With the exception of net depth ('netdpth'), median distance travelled ('meddisttravel') and object depth ('objdpth'), most covariates were significant in both parts of the model (Tables 3-4). Warmer sea surface temperatures ('sst') contributed to increased bycatch by contributing to a decrease in the probability of being in the perfect state and an increase in the bycatch per set in the imperfect state. The greater the amount of tuna catch ('logtuna') and bycatch of non-target species ('lognonsilky'), the greater the shark bycatch, since the probability of being in the perfect state was found to decrease with increases in both tuna catch and non-silky shark bycatch, and the amount of silky shark bycatch in the imperfect state increased with increases in tuna catch and non-silky shark bycatch. The proxies for local floating object density, in particular, the number of unique object numbers ('unqobjnum'), showed greater shark bycatch at lesser values since the probability of being in the perfect state increased and shark bycatch in the imperfect state decreased as the number of unique object numbers increased.

An overall decreasing trend in shark bycatch per set was estimated (Figure 6). Estimates of the year effects in negative binomial regression part show a decreasing trend. This implies that bycatch counts have decreased over the 1994 to 2004 period in sets in which bycatch could have occurred (*i.e.*, sets in the imperfect state). Estimates of the year effects in the logistic regression part show an increasing trend. This implies that the probability that sets would be made in a state in which no bycatch could occur (*i.e.*, the zero or perfect state) increased over the 1994 to 2004 period. In both parts, the year effect coefficients for 1998 depart from the general tendency; however, these departures balance each other. We believe that because of the El Niño event in 1998, effects of environmental covariates on bycatch per set might have been different in this year (see Discussion). The overall importance of 'year,' as measured by change in GIC, is shown in Table 4.

To see the importance of variables expressed with smooth functions, we computed the difference in GIC for models with and without each smoothed term. The difference of GIC values from the full model can be considered as an indicator of relative importance of the variable in the model. The predictor 'location' had the greatest differences in GIC for this data set (Table 4). Excluding 'time' had little effect on the GIC, which is not surprising since most floating object sets are made early in the morning so few samples exist at other time periods.

Table 3: Estimated coefficients of linear and categorical predictors for the ZINB with t.p.r.s., their standard errors (std. error), and approximate z-values.

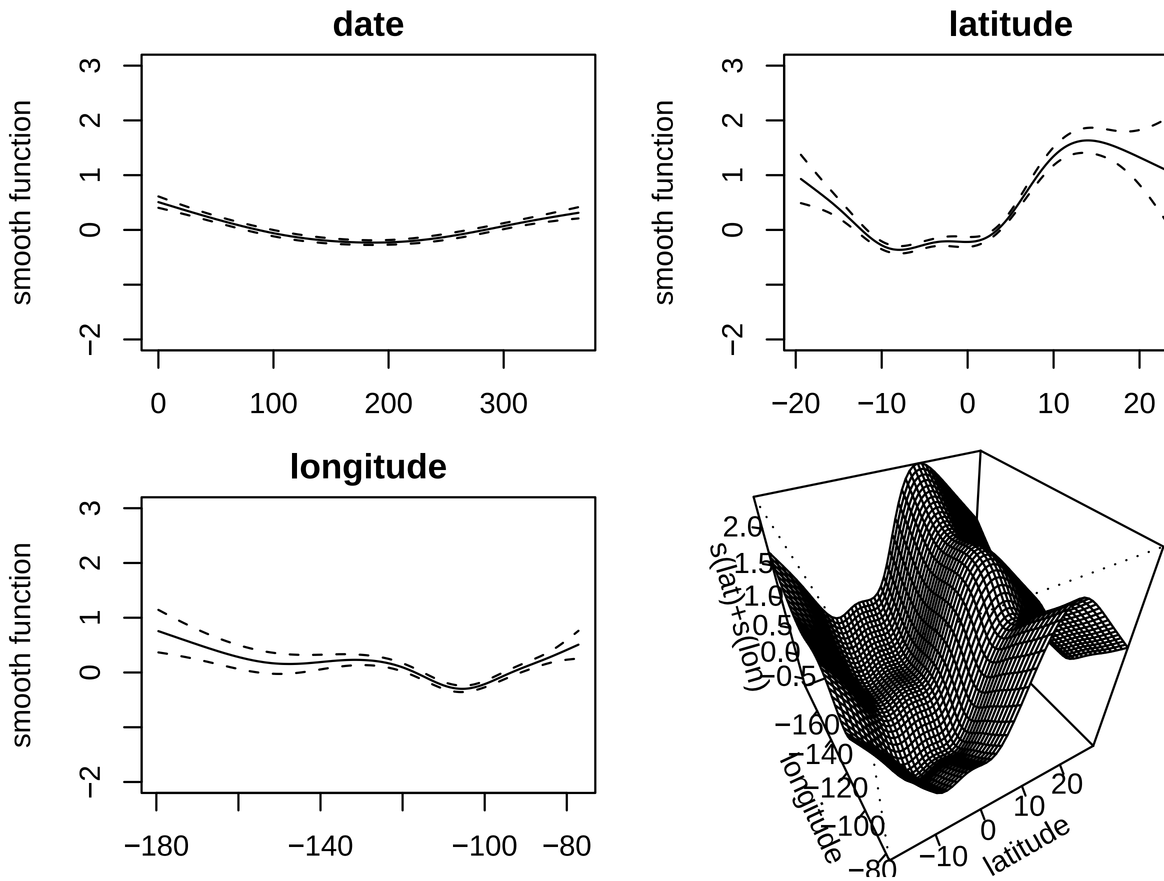
negative binomial regression part				logistic regression part			
covariate	coefficient	std. error	z−value	covariate	coefficient	std. error	z−value
Intercept	−3.67	0.482	−7.60	Intercept	6.63	1.071	6.19
year 1995	−0.36	0.096	−3.78	year 1995	−1.53	0.994	−1.54
year 1996	−0.38	0.099	−3.86	year 1996	1.15	0.271	4.23
year 1997	−0.38	0.096	−3.91	year 1997	1.30	0.260	5.01
year 1998	−0.94	0.106	−8.88	year 1998	−3.26	1.481	−2.20
year 1999	−0.55	0.126	−4.34	year 1999	2.18	0.272	8.02
year 2000	−0.55	0.119	−4.66	year 2000	2.50	0.263	9.52
year 2001	−0.62	0.107	−5.75	year 2001	1.99	0.256	7.77
year 2002	−1.08	0.112	−9.62	year 2002	1.81	0.263	6.91
year 2003	−0.72	0.138	−5.26	year 2003	2.46	0.267	9.22
year 2004	−0.98	0.121	−8.08	year 2004	2.93	0.270	10.88
netdpth	0.00	0.001	−0.24	netdpth	0.00	0.002	0.16
sst	0.17	0.017	10.21	sst	−0.30	0.039	−7.77
objdpth	−0.01	0.002	−3.35	objdpth	−0.01	0.005	−1.64
logtuna	0.28	0.023	12.41	logtuna	−0.14	0.033	−4.21
lognonsilky	0.11	0.013	8.76	lognonsilky	−0.23	0.030	−7.72
unqobjnum	−0.00304	0.00042	−7.20	unqobjnum	0.00271	0.00064	4.22
meddisttravel	0.00006	0.00038	0.15	meddisttrave	−0.00422	0.00145	−2.90

Table 4: Significance of year (categorical) and smoothed covariates (difference in GIC) computed for the test data set.

negative binomial regression part				logistic regression part			
subtracted covariate	GIC	from fulll modele in deviance	effective degree of freedom	subtracted covariate	GIC	difference in deviance	effective degree of freedom
year	64146.26	224.80	10.00	year	64384.79	463.33	10.00
date	64057.00	125.54	8.24	date	63950.93	29.47	7.46
location	64746.90	825.44	16.28	location	64235.06	313.60	15.31
time	63929.61	8.15	7.78	time	63933.04	11.58	1.01

Within the regions where confidence bands are fairly narrow, the results suggest that the probability of being is the perfect state was greatest in the first half of the year, greatest around the equator with respect

Figure 7: Estimated smoothed curves from the negative binomial regression part of the ZINB regression model with smoothing. Solid lines represent the estimated smoothed functions and dotted lines the approximate confidence bands (estimate +/- twice the standard error).



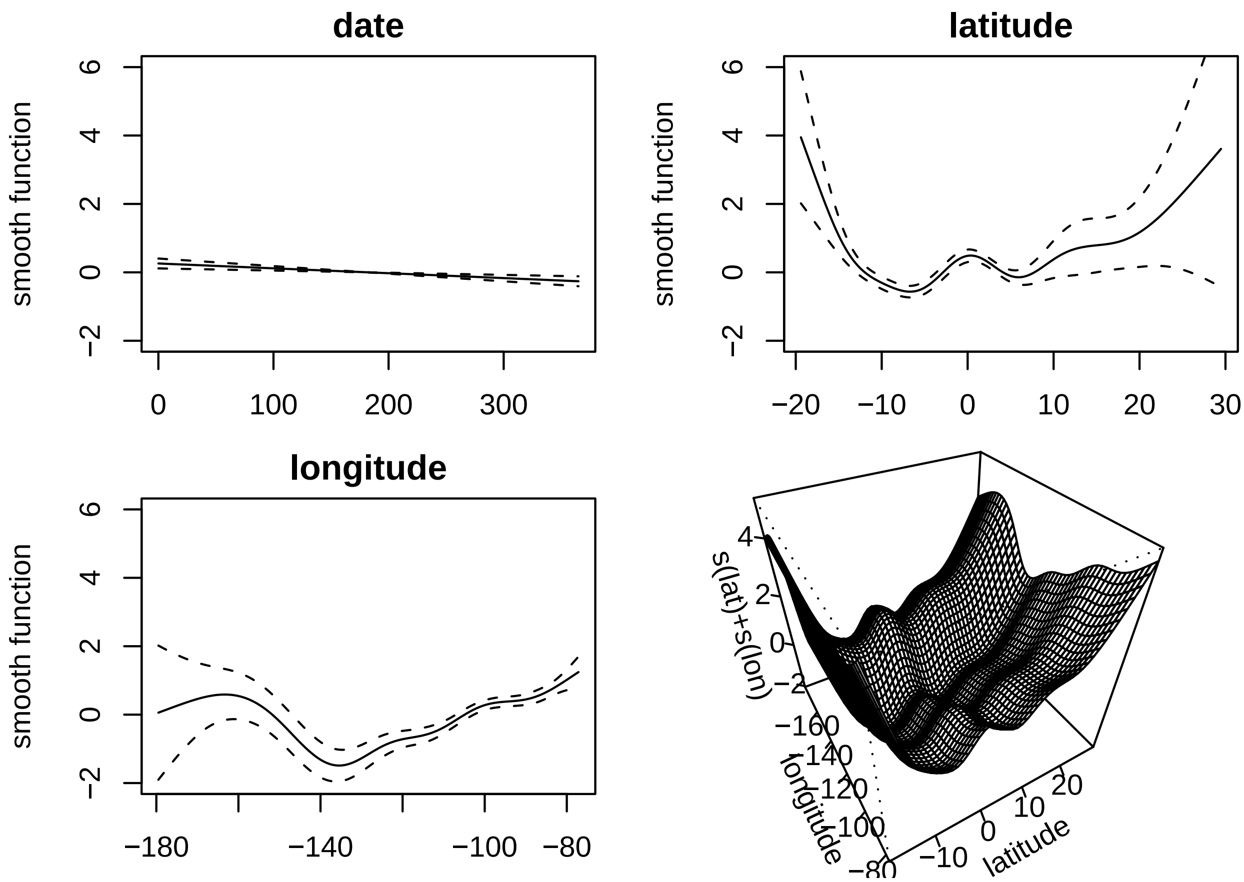
to latitude, and greatest coastally with respect to longitude (Figures 7-8). The amount of shark bycatch in the imperfect state was least in the middle part of the year, increased from south to north across the equator (with respect to latitude), and was least around $100^o - 110^o$W (with respect to longitude). Confidence bands become quite large in the regions of sparse data.

4.3 An example of partial dependence plots: temporal trend

By constructing a partial dependence plot for the year effects we can compare differences in the marginal effect of the predictor 'year' between models. Thus, the partial dependence plot makes possible a comparison of trends in standardized average bycatch per set between models. From Figure 9 it is apparent that all models indicate a decreasing temporal trend in bycatch per set, although the rate of decrease varies by model over the first part of the time series. With the exception of the NB models, the trends are more similar than are the expected counts (Figure 5) and the GIC (Table 2), showing a decrease from a bycatch per set of approximately eight animals in 1994 to a bycatch per set of approximately three to four animals from 2002 to 2004.

The difference between the trends from the NB and the ZINB regression models (Figure 9) appears to be due to the way in which these two types of models accommodate variability caused by many zero-valued observations, and the specification of the mean structure. In this example it is likely that the NB regression model overestimates the trend. This phenomena will be investigated in the following section. Regardless, Figure 9 illustrates the point that fitting multiple models is informative and good practice,

Figure 8: Estimated smoothed curves from the logistic regression part of the ZINB regression model with smoothing. Solid lines represent the estimated smoothed functions and dotted lines the approximate confidence bands (estimate +/- twice the standard error).



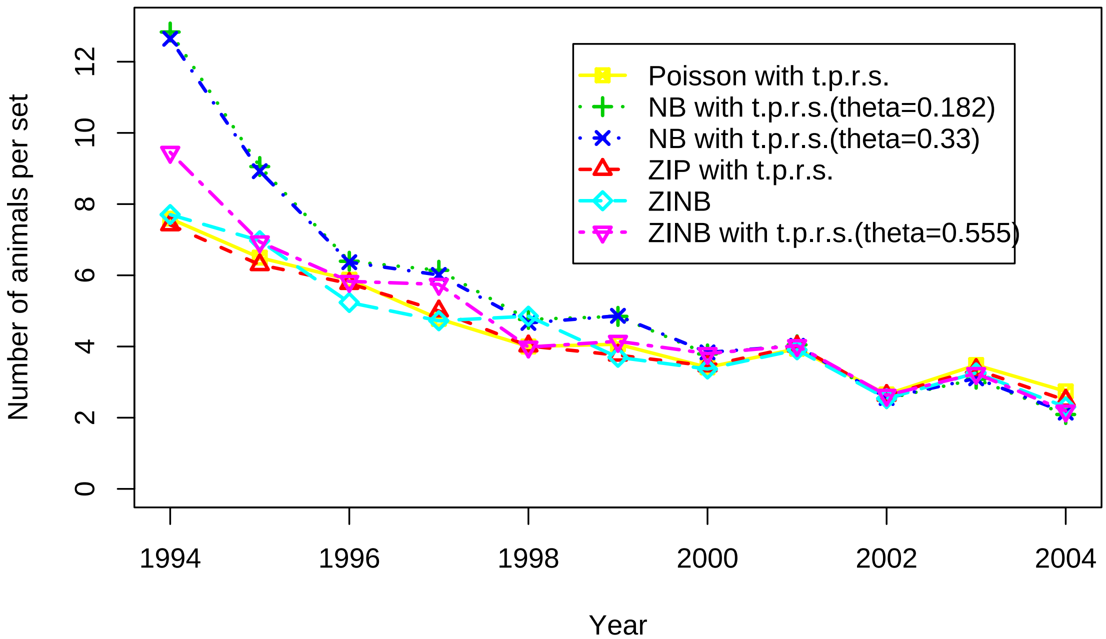
particularly since the true stochastic processes that generated the data are not likely to be known.

5 Overestimation by NB regression models

As shown in Figure 9, the NB regression model may overestimate the trend while the Poisson regression model, equivalent to the NB regression model with the infinite value of size parameter, may not. To investigate this phenomena, we fitted the NB regression model with various fixed values of size parameter to the data. For simplicity, smoothing functions were not used here, and latitude and longitude were included into models as categorical variables defined based on the estimated smooth functions in the previous section.

Table 5 shows that the estimates vary with the value of size parameter. The smaller the value of size parameter, the larger the standardized average is for 1994. For NB regression model, it is often said that the value of size parameter does not affect estimation of parameters in the mean structure because size parameter and parameters in the mean structure are orthogonal (Lawless, 1987). However, this is true only if the negative binomial model is appropriate for the data and the mean structure is properly specified. As regards the shark bycatch data, the proportion of zero-valued observations has changed over time (see Data section), thus, the mean structure cannot be properly expressed with the NB regression model. Moreover, the NB regression model accomodates the presence of extra zero-valued observations in the data by increasing its variance through the size parameter θ.

Figure 9: Estimates of the trend in the standardized average silky shark bycatch per set for the various models.



For the data with many zero-valued observations such as the shark bycatch data, overestimation (or unreliable estimation) of trends by fitting the NB regression model can be explained as follows:

1. When a NB regression model is fitted to data with many zero-valued observations, the estimate for size parameter θ could be very small.

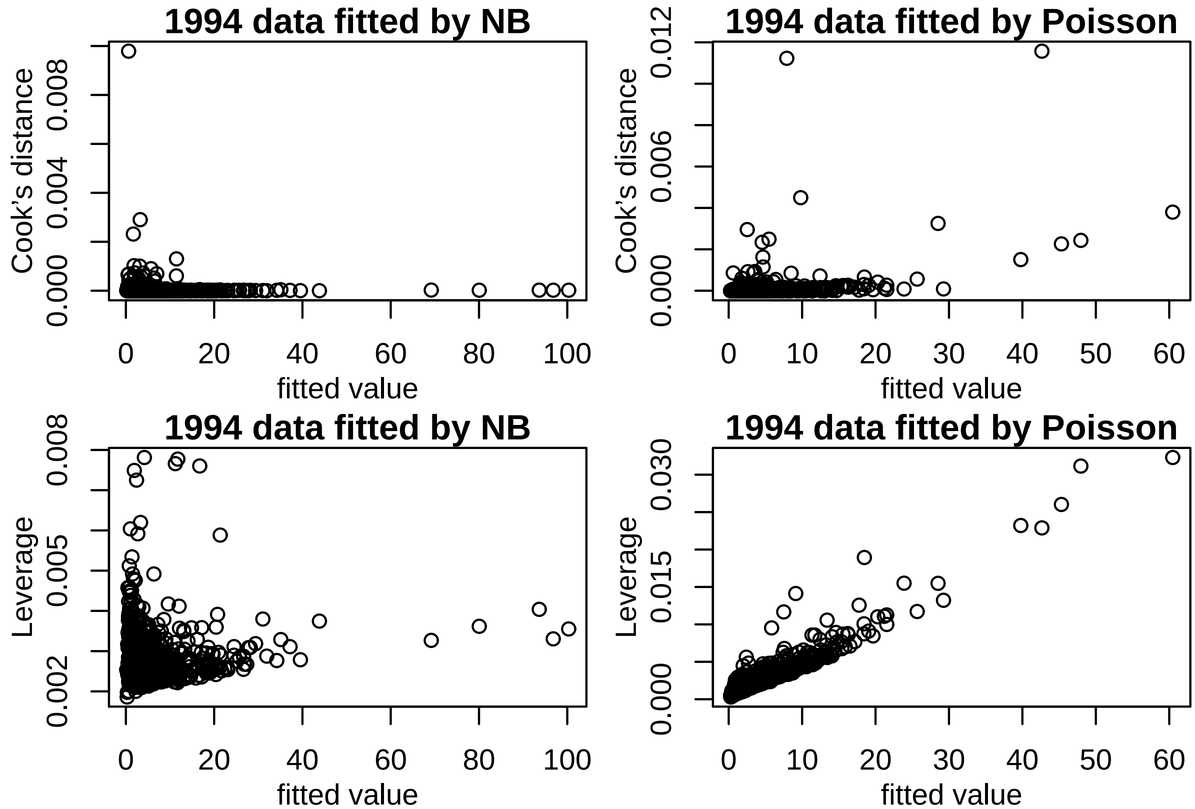
2. When the size parameter value is small, observations with small fitted values have the greatest influence on the estimation, and the estimates reflect only local structure corresponding to those observations.

3. When the proportions of the zero states depend on covariates, the mean structure assumed for the NB regression model does not properly express that of the data no matther which link function is

Table 5: The standardized average silky shark bycatch per set for NB regression models with various values of size parameter. The fourth column shows the ratio (percentage) of the standardized average of 2004 to that of 1994 and the fifth column its 95% confidence intervals by normal approximation of the estimates under the models.

values of size parameter	standardized average		the ratio (2004/1994)	95% CI for the ratio
	1994	2004		
0.01	11.46	2.08	18.2%	(13.2, 25.1)
0.157 (MME)	11.21	2.11	18.8%	(14.5, 24.5)
0.318 (MLE)	10.70	2.17	20.3%	(16.8, 24.6)
1.00	9.72	2.32	23.8%	(21.2, 26.9)
$+\infty$ (Poisson)	7.01	2.77	39.5%	(37.6, 41.5)

Figure 10: Cook's distances and leverages for 1994 data. The left column is for the NB regression model with MLE and the right column is for the Poisson regression model.



used. Thus, the fitted coefficients are inconsistent.

Theoretical description and explanation of the above are given in appendix 3.

For the shilky shark bycatch data, the estimate of the size parameter for ZINB model $\hat{\theta}_0^{\text{ZINB}}$ is 0.56, while the method of moments and the maximum likelihood estimates for the NB regression model are $\hat{\theta}_{\text{MM}}^{\text{NB}} = 0.18$ and $\hat{\theta}_{\text{MLE}}^{\text{NB}} = 0.32$, respectively. Unfortunately, we have only weak predictors of shark bycatch per set, a problem perhaps encountered in other studies as well. Thus, the estimate of size parameter in the ZINB model is also small. However, the both estimates for NB regression model are much smaller.

The influence of particular observations on the estimation can be explored graphically using Cook's distance and leverage. The larger the values of Cook's distance and leverage, the more influential the observations. Figure 10 shows that for the NB model with the MLE for θ, observations with small fitted values are more influential, while for the Poisson model, observations with large fitted values are more influential. Thus, it follows from Appendix 3 that the estimates for the NB model are more likely to reflect only local structure as a result of small fitted values.

6 Discussion

In this paper we have introduced the ZINB regression model with smoothing, and demonstrated its use on shark bycatch data from the eastern Pacific Ocean tuna purse-seine fishery for 1994-2004. For these data we found that the ZINB provided a better fit, as measured by GIC, than Poisson, zero-inflated Poisson, and negative binomial models. Estimated temporal trends in silky shark bycatch per set, after accounting for the average effects of other predictors were, however, more similar across models. The exception to this was for the NB which appeared to overestimate the decreasing trend in catch rate.

In spite of its superior performance in fitting these data, we have found that the ZINB is not without its limitations. Shark bycatch occurs in other types of purse-seine sets, but in lesser amounts (IATTC, 2004). For example, approximately 90% or more of unassociated sets over the 1994-2004 period had no silky shark bycatch. The fit of a ZINB regression model to the silky shark bycatch data in unassociated sets, however, was poor (as measured by GIC). This suggests that for data dominated by zero-valued observations, the ZINB may not be appropriate or may need to be modified. The possibility of modelling such data with a right-truncated ZINB regression model is being explored.

The estimated year effect coefficients for 1998 from the ZINB regression model with smoothing suggest that the likelihood that bycatch would occur was greater in 1998 than in other years, but that when bycatch did occur, the amounts were somewhat smaller, on average (Figure 6). This may reflect a change in the group size of sharks around floating objects in response to El Niño conditions. Although we do not know the residence time of sharks at floating objects nor whether sharks are attracted to the objects to forage, it is possible that during El Niño conditions, the environment and prey characteristics are such that somewhat smaller aggregations of sharks were formed at floating objects, but that more objects were possibly 'attractive' to sharks. Thus, an El Niño effect might be much greater (albeit short-term) on bycatch rates than that of an overall decreasing trend. Another possible explanation is that El Niño conditions may have brought about an influx of silky sharks into the EPO from elsewhere.

The implications of the decreasing temporal trend in bycatch per set for the silky shark in the EPO are unclear. It is well known that changes in bycatch rates may be due to change in the density of sharks, the catchability of sharks or both (*e.g.*, Campbell, 2004). Although we have attempted to control for factors that may affect catchability, bycatch per set may not index abundance for several reasons. First, our proxies of object density do not account for stealing and/or sharing of objects between vessels (Appendix 1). Any temporal trend in the stealing or sharing of objects may be confounded with trends in shark density. Second, there may be a temporal trend in the proportion of sharks released alive. It is believed that the majority of sharks encircled with the purse-seine are killed as a result of being crushed by the weight of the catch within the pursed net and the brailer, being left on the deck for extended periods of time before further handling and/or being finned. In this fishery, releasing sharks 'unharmed' has been encouraged since 2000 (IATTC, 2000), but finning has only been restricted since 2005 (IATTC, 2005b), and so, any effect should be limited to the latter part of the time series. Finally, although we believe that these data represent bycatch per set of *C. falciformis*, only limited data are available from before 2005 to confirm observers' at-sea species identifications (Appendix 1). We note that the decreasing trend in the silky shark bycatch per set from floating object set data is consistent with previous trend estimates for longline fisheries from both the western and eastern Pacific (Matsunaga and Nakano, 1999; Ward and Myers, 2005), but conflicts with recent trend estimates from the western Pacific (FRCC, 2004), further complicating interpretation.

Population dynamics modeling of the silky shark population in the Pacific Ocean would help to determine the current population status, or the research most likely to help in its determination. With respect to population dynamics modeling, one limitation immediately apparent is that the spatial structure of the silky shark population in the Pacific Ocean is not well known. In particular, it has been suggested that the silky shark is much more abundant near land than in the open ocean (FAO, 1984). However, a widespread distribution across the Pacific is suggested by longline CPUE data (FRCC 2004 and references therein) and from purse-seine data (Figure 1). In addition, it has been proposed that there exists spatial structure in the distribution of nursery areas, with juveniles found in nursery areas offshore, but adults found further offshore still (Oshitani *et al.*, 2003). Thus, the definition of 'stocks' for the silky shark in the Pacific remains problematic. Assuming bycatch per set and longline CPUE reflect population trends, one possible explanation for the divergent trends across the Pacific Ocean is that the decreasing trends in the EPO purse-seine data may reflect the combined effects of coastal(WildAid, 2005), distant water longline, and purse-seine fisheries, yet movement rates of the silky shark population are too low for these effects to propagate into the western Pacific Ocean, where overall catch rates may be lower.

Acknowledgements

We thank William Bayliff and Simon Hoyle for helpful comments that improved this manuscript, and Nickolas Vogel and Mauricio Orozco-Zöller for data base assistance. This work was in part supported by ISM Project Reseach 2005.

References

Akaike, H. 1974. A new look at the statistical model identification. IEEE Trans. Auto. Control AC-19, 716-723.

Agarwal, D.K., Gelfand, A.E., and Citron-Pousty, S. 2002. Zero-inflated models with application to spatial count data. Environmental and Ecological Statistics 9:341-355.

Barry, S.C., Welsh, A.H. 2002. Generalized additive modeling and zero inflated count data. Ecological Modelling 157:179-188.

Campbell, R.A. 2004. CPUE standardisation and the construction of indices of stock abundance in a spatially varying fishery using general linear models. Fisheries Research 70:209-227.

Compagno, L.J.V. 1984. FAO species catalogue. Vol 4. Sharks of the world. An annotated and illustrated catalogue of shark species known to date (Part 2. Carcharhiniformes):251-655.

Dempster, A.P., Laird, N.M. and Rubin, D.B. 1977. Maximum likelihood from incomplete data via the EM algorithm (with discussion). Journal of the Royal Statistical Society B 39:1-38.

FAO 1984. Sharks of the World. An Annotated and Illustrated Catalogue of Shark Species Known to Date. Part 2 - Carcharhiniformes. Compagno, L.J.V., editor. FAO Fisheries Synopsis No. 125, Vol. 4.

F.R.C.C. 2004. The Present Condition of International Fisheries Resources, Fisheries Agency, Fisheries Research Coordination Center, Japan.

Freon and Dagorn, 2000. Associative behaviour of pelagic fish: facts and hypotheses. *In* Le Gall, J-Y, Cayre, P., and Taquet, M. (editor), Peche Thoniere et Dispositifs de Concentration de Poissons, Colloque Caraibe-Martinique, Trois-Illets, 15-19 October, 1999. Int. Fran. Recherche Exploitation Mer.

Grogger, J.T. and Carson, R.T. 1991. Models for truncated counts. Journal of Applied Econometrics 6:225-238.

Greene, W.H. 1994 Accounting for excess zeros and sample selection in Poisson and negative binomial regression models. Working paper NO. 94-10. New York; Stern School of Business, New York University, Department of Econometrics.

Hall, D.B. 2000. Zero-inflated Poisson and binomial regression with random effects: A case study. Biometrics 56:1030-1039.

Hampel, F.R., Ronchetti, E.M. Rousseuw, P.J. and Stahel, W.A. 1986. Robust Statistics — The approach based on influence functions. Wiley, New York.

Hastie, T.J. and Tibshirani, R.J. 1991. *Generalized Additive Models*. Monographs on Statistics and Applied Probability 43. Chapman Hall, London. 335pp.

Hastie, T, Tibshirani, R, Friedman, J. 2001. *The Elements of Statistical Learning; Data Mining, Inference and Prediction.* Springer. 533pp.

Hoey, J.J., Pritchard, E., Brown, C. and Showell, M. 2002. Collected Volume of Scientific Papers, ICCAT 54 (4):1199-1211.

Huber, P.J. 1981. *Robust Statistics*. Wiley, New York.

Hurvich, C.M., Simonoff, J.S. and Tsai, C.L. 1998. Smoothing parameter selection in nonparametric using an improved Akaike information criterion. Journal of Royal Statistical Sociey, Series B, 60, 271-293.

IATTC 2000. Resolution on Bycatch, June, 2000. Inter-American Tropical Tuna Commission. [Available at: http://iattc.org/PDFFiles/C-00-08

IATTC 2004. Annual Report of the Inter-American Tropical Tuna Commission, 2002. Inter-American Tropical Tuna Commission, La Jolla, California, U.S.A. 92037. [Available from: http://www.iattc.org].

IATTC 2005a. Fishery Status Report No. 3. Inter-American Tropical Tuna Commission, La Jolla, California, U.S.A. 92037. [Available from: http://iattc.org]

IATTC 2005b. Resolutions on the Conservation of Sharks Caught in Association with Fisheries in the Eastern Pacific Ocean, June 2005. Inter-American Tropical Tuna Commission. [Available at: http://iattc.org/PDFFiles 05-03-Sharks.pdf]

IATTC 2006. Annual Report of the Inter-American Tropical Tuna Commission, 2004. Inter-American Tropical Tuna Commission, La Jolla, California, U.S.A. 92037. In prep.

Konishi, S. and Kitagawa, G. 1996. Generalized information criteria in model selection, Biometrika, 83(4), 875-890.

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

Lawless, JF. 1987. Negative binomial and mixed Poisson regression. The Canadian Journal of Statistics 15:209-225.

Lo, N.C.H., Jacobson, L.D., and Squire, J.L. 1992. Indices of relative abundance from fish spotter data based on delta-lognormal models. Canadian Journal of Fisheries and Aquatic Sciences 49:2515-2526.

Long, S.J. 1997. Regression Models for Categorical and Limited Dependent Variable, Sage Publications.

Maunder, M.N. and Punt, A.E. 2004. Standardizing catch and effort data: a review of recent approaches. Fisheries Research 70:141-159.

Marsac, F., Fonteneau, A. and Menard, F. 2000. Drifting FADs used in tuna fisheries: an ecologial trap? *In* Le Gall, J-Y, Cayre, P., and Taquet, M. (editor), Peche Thoniere et Dispositifs de Concentration de Poissons, Colloque Caraibe-Martinique, Trois-Illets, 15-19 October, 1999. Int. Fran. Recherche Exploitation Mer.

Matsunaga, H. and Nakano, H. 1999. Species composition and CPUE of pelagic sharks caught by Japanese longline research and training vessels in the Pacific Ocean. Fisheries Science 65: 16-22.

Minami, M. and Lennert-Cody, C. E. 2006. Properties for a zero-inflated negative binomial regression model. In prep.

Ortiz, M. and Arocha, F. 2004. Alternative error distribution models for standardization of catch rates of non-target species form a pelagic longline fishery: billfish species in the Venezuelan tuna longline fishery. Fisheries Research 70:275-297.

Oshitani, S., Nakano, H., and Tanaka, S. 2003. Age and growth of the silky shark Carcharhinus falciformis from the Pacific Ocean. Fisheries Science 69:456-464.

Punt, A.E., Walker, T.I., Taylor, B.L., Pribac, F. 2000. Standardization of catch and effort data in a spatially-structured shark fishery. Fisheries Research 45:129-145.

Ridout, M., Hinde, J. and Demétrio, C. G. B. 2001. A Score Test for Testing a Zero-Inflated Poisson Regression Model Against Zero-Inflated Negative Binomial Alternatives. Biometrics 57: 219-223.

Román-Verdesoto, M.H. 2003. Spatial distribution of silky shark (/it Carcharinus falciformis) bycatch in the tuna purse-seine fishery of the eastern Pacific Ocean. Proceedings of the 54th Annual Tuna Conference, Lake Arrowhead, California, May 13-16, 2003. [Available from: http://swfsc.nmfs.noaa.gov/tunaconf.html].

Román, M.H., Vogel, N.W., Olson, R.J. and Lennert-Cody, C.E. 2005. A novel approach for improving shark bycatch species identifications by observers at sea. PFRP 10(3):4-5.

Román-Verdesoto, M. and Orozco-Zöller, M., Submitted. Bycatches of sharks in the tuna purse-seine fishery of the eastern Pacific Ocean reported by observers of the Inter-American Tropical Tuna Commission, 1993-2004. Data Report II. Inter-American Tropical Tuna Commission.

Roundtree, R.A. 1989. Associations of fishes and fish aggregation devices: effects of structure size on fish abundance. Bulletin of Marine Science 44:960-972.

Schindler, D.E., Essington, T.E., Kitchell, J.F., Boggs, C. and Hilborn, R. 2002. Sharks and tunas: Fisheries impacts on predators with contrasting life histories. Ecological Applications 12(3):735-748.

Scott, M.D. and Chivers, S.J. 1990. Distribution and herd structure of the bottlenose dolphins in the eastern tropical Pacific Ocean. In: The Bottlenose Dolphin: Biology and Conservation. Reynolds, J.E. III, Wells, R.S. and Eide, S.D. (eds.). University Press of Florida, pp. 387-402.

Simpfendorfer, C.A., Hueter, R.E., Bergman, U., Connett, S.M.H. 2002. Results of a fishery-independent survey of pelagic sharks in the western North Atlantic, 1977-1994. Fisheries Research 55:175-192.

Stefansson, G. 1996. Abundance indices from groundfish survey data: combining the GLM and delta approaches. ICES Journal of Marine Science 53:577-588.

Terceiro, M. 2003. The statistical properties of recreational catch rate data for some fish stocks off the northeast U.S. coast. Fishery Bulletin, 101, 653-672.

Vuong, Q.H. 1989 Likelihood ratio tests for model selction and non-nested hypotheses. Econometrica, 57, 307-333.

Wahba, G. 1990 Spline models for observational data. CBMS-NSF Regl Conf. Ser. Appl. Math., 59.

Walsh, W.A. and Kleiber, P. 2001. Generalized additive model and regression tree analyses of blue shark (*Prionace glauca*) catch rates by the Hawaii-based commercial longline fishery. Fisheries Research 53:115-131.

Ward, P. and Myers, R.A. 2005. Shifts in open-ocean fish communities coinciding with the commencement of commercial fishing. Ecology 86:835-847.

Watters, G. M. 1999. Geographical distributions of effort and catches of tunas by purse-seine vessels in the eastern Pacific Ocean during 1965-1998. Inter-American Tropical Tuna Commission. Data Report 10.

WildAid, 2005. At rock bottom: The declining sharks of the eastern tropical Pacific. WildAid, 450 Pacific Ave., Suite 201, San Francisco, California, USA 94133

Wood, S.N., 2000. Modelling and smoothing parameter estimation with multiple quadratic penalties. Journal of the Royal Statistical Society, Series B 62(2):413-428.

Wood., S.N. 2003. Thin-plate regression splines. Journal of the Royal Statistical Society, Series B 65(1):95-114.

Wood, S.N. 2004. Stable and efficient multiple smoothing parameter estimation for generalized additive models. J. Amer. Statist. Ass. 99:637-686.

Ye, Y. Pitcher, R. Dennis D. and Skewes, T. 2005. Constructing abundance indices from scientific surveys of different designs for the Torres Strait ornate rock lobster(*Panulirus ornatus*) fishery, Australia, Fisheries Research, 73:187-200.

Appendix 1

In this Appendix we describe details related to observers at-sea species identifications and to data processing.

Silky sharks were identified both directly and indirectly from “species” codes reported by observers. Detailed species information for sharks, including diagnostic characteristics for silky sharks (Román, 2003; Román *et al.*, 2005), available between March, 2000 and March, 2001, and from January, 2005, indicate that silky sharks (*Carcharhinus falciformis*) dominated the shark bycatch during those periods (Román-Verdesoto and Orozco-Zöller, submitted; IATTC unpublished data). These data indicated that almost all sharks identified by observers at sea as silky shakrs were properly identified. However, these data also indicate that observers commonly misidentify silky sharks at sea, frequently reporting silky sharks as “blacktip” sharks (*Carcharhinus limbatus*). Misidentification occurs primarily because the common name used by fishermen for silky sharks is ‘punta negra’ (blacktip), and, although undesirable, observers appear to take species identification cues from fishermen. Given the generally oceanic distribution of this fishery (Watters, 1999) and the known distribution of *C. limbatus* which is primarily continental and insular shelves in temperate and tropical waters (Compagno, 1984), bycatches of *C. limbatus* would be expected to occur only very infrequently in floating object sets. Although we are not certain that this species identification error extends back to 1994, all available information (coastal distribution of *C. limbatus*, consistent observer training, long tradition of fishermen’s common name) indicate that this is very likely the case. Thus, we assume that all sharks reported as “blacktip” sharks by observers were in fact silky sharks. The presence of any true *C. limbatus* in the bycatch in very coastal waters has little effect on shark bycatch in floating object sets because of the almost exclusively oceanic distribution of floating object sets. Although irrelevant prior to 2005, the IATTC has recently expanded its data collection procedures to include documentation of species-specific characteristics for confirm observers' at-sea indentifications (Roman *et al.*, 2005), and additional information on the fate of sharks brailed onto the vessel's deck.

For non-tuna species, bycatch of sharks is typically reported by observers in numbers of animals, in three length categories (total length): small ($<$ 90 cm), medium (90-150 cm) and large ($>$ 150 cm). Infrequently, bycatch amounts may be reported in metric tons. These values are later converted to numbers of animals, using crude length-weight relationships; the error on these conversions is undoubtedly large and likely unestimable. The primary reason observers report bycatches in tons instead of numbers may be because bycatches were so large that they were unable to count the number of animals. However, not all conversions of tons to numbers yielded large bycatches, and the reasons for reporting small bycatches in tons instead of numbers are unclear. Because of the assumed large error on the conversions and uncertainty in all cases as to why reports were made in tons (*e.g.*, numerical errors such as an observer recording 0.01 tons when he intended to record 0.1 can not be excluded), we exclude from this analysis sets for which silky shark bycatch was initially reported in tons of animals. The percentage of such sets varies by purse-seine set type, but was on average annually 0.44% for floating object sets and was always less than 1% annually for the 1994 to 2004 period.

Some data were also excluded for other reasons. Sets for which data were not available on all the predictor variables of interest (Table 1) were excluded prior to analysis. In addition, sets for which the sum of target tunas was less than 0.01 metric tons were excluded because it is assumed that more fish were present when the set was initiated but the fish, and possibly the non-tuna species, escaped encirclement with the net. Repeat sets on the same floating object were also excluded. Further, sampling coverage was lowest in 1993, the year the non-mammal bycatch sampling program was initiated; data for that year are not included in this analysis. For those floating object sets between 1994 and 2004 that had bycatch information, on average, annually 58% of the silky shark bycatch and 69% of the floating object sets were retained for analysis.

Two proxies of local floating object density were computed. The first was the number of unique object numbers within a 5^o area centered on the set location and one month prior to the set date. Ideally, the number of unique objects in a given area and time window would be computed. However, this was not possible because the data do not allow objects to be tracked across vessel trips, nor do the data identify objects shared with or stolen from other vessels. Thus, objects placed in the water by one vessel may be set upon by another vessel resulting in two object numbers in the database for only one object. The degree to which vessels have stolen and/or shared objects over the last decade is unknown. If stealing/sharing of objects is minimal, the number of unique object numbers should approximate the local density of objects. The second proxy was the median distance traveled by vessels between objects within a 5^o area around the set location and one month prior to the set date. Ideally, it would be possible to compute the distance between objects. However, objects are identified in the data only when they are visited by a vessel. Thus, the distance the vessel traveled between objects is known, but not the instantaneous location of all objects at any given time. If stealing and/or sharing of objects occurs relatively frequently, the median distance traveled between objects may be a better proxy of local object density.

Appendix 2

Here we give the first and second derivatives of $l(\boldsymbol{\beta}, \boldsymbol{\gamma}, \theta) \equiv \log f(y|B_i, G_i, \boldsymbol{\beta}, \boldsymbol{\gamma}, \theta)$, where f is the probability function for the ZINB regression model given by (4), and the information matrix.

The first derivatives:

$$\begin{aligned}
\frac{\partial l(\boldsymbol{\beta},\boldsymbol{\gamma},\theta)}{\partial \boldsymbol{\beta}} &= (1-r)\frac{y-\mu}{\theta+\mu}\cdot \theta B_i^T \\
\frac{\partial l(\boldsymbol{\beta},\boldsymbol{\gamma},\theta)}{\partial \boldsymbol{\gamma}} &= (r-p)G_i^T \\
\frac{\partial l(\boldsymbol{\beta},\boldsymbol{\gamma},\theta)}{\partial \theta} &= (1-r)\left(\log\left(\frac{\theta}{\theta+\mu}\right)-\frac{y-\mu}{\theta+\mu}+\Psi(\theta+y)-\Psi(\theta)\right)
\end{aligned}$$

where $\Psi(\theta)$ is the digamma function, $\Psi(\theta)=d\log\Gamma(\theta)/d\theta$, and r is the posterior mean of Z given y:

$$r=\begin{cases} \dfrac{p_i}{p_i+(1-p_i)\ \ q\left(0|\mu_i,\theta\right)} & \text{for} \quad y_i=0 \\ 0 & \text{for} \quad y_i=1,2,\cdots. \end{cases}$$

The second derivatives:

$$\begin{aligned}
\frac{\partial^2 l(\boldsymbol{\beta},\boldsymbol{\gamma},\theta)}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^T} &= (1-r)\frac{(a+y)(r\mu-1)\theta\mu}{(\theta+\mu)^2}B_i^T B_i \\
\frac{\partial^2 l(\boldsymbol{\beta},\boldsymbol{\gamma},\theta)}{\partial \boldsymbol{\gamma}\partial \boldsymbol{\gamma}^T} &= r(1-r)G_i^T G_i - p(1-p)G_i^T G_i \\
\frac{\partial^2 l(\boldsymbol{\beta},\boldsymbol{\gamma},\theta)}{\partial \boldsymbol{\beta}\partial \boldsymbol{\gamma}^T} &= r(1-r)\frac{\theta\mu}{\theta+\mu}B_i^T G_i \\
\frac{\partial^2 l(\boldsymbol{\beta},\boldsymbol{\gamma},\theta)}{(\partial \theta)^2} &= r(1-r)\left(\log\frac{\theta}{\theta+\mu}+\frac{\mu}{\theta+\mu}\right)^2+(1-r)\left(\Psi'(\theta+y)-\Psi'(\theta)+\frac{\mu}{\theta(\theta+\mu)}+\frac{y-\mu}{(\theta+\mu)^2}\right) \\
\frac{\partial^2 l(\boldsymbol{\beta},\boldsymbol{\gamma},\theta)}{\partial \theta\partial \boldsymbol{\beta}^T} &= -r(1-r)\left(\log\frac{\theta}{\theta+\mu}+\frac{\mu}{\theta+\mu}\right)\frac{\theta\mu}{\theta+\mu}B_i+(1-r)\frac{y-\mu}{(\theta+\mu)^2}\mu B_i \\
\frac{\partial^2 l(\boldsymbol{\beta},\boldsymbol{\gamma},\theta)}{\partial \theta\partial \boldsymbol{\gamma}^T} &= -r(1-r)\left(\log\frac{\theta}{\theta+\mu}+\frac{\mu}{\theta+\mu}\right)G_i
\end{aligned}$$

The information matrix:

$$I_{NB}=\begin{pmatrix} I_{\gamma,\gamma} & I_{\gamma,\beta} & I_{\gamma,\theta} \\ I_{\beta,\gamma} & I_{\beta,\beta} & I_{\beta,\theta} \\ I_{\theta,\gamma} & I_{\theta,\beta} & I_{\theta,\theta} \end{pmatrix}$$

where

$$\begin{aligned}
I_{\gamma,\gamma} = -\mathrm{E}_f\left[\frac{\partial^2 \log f}{\partial \boldsymbol{\gamma}\partial \boldsymbol{\gamma}^T}\right] &= p(1-p)G_i^T G_i - dG_i^T G_i \\
I_{\beta,\beta} = -\mathrm{E}_f\left[\frac{\partial^2 \log f}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^T}\right] &= (1-p)\frac{\theta\mu}{\theta+\mu}B_i^T B_i - d\left(\frac{\theta\mu}{\theta+\mu}\right)^2 B_i^T B_i \\
I_{\beta,\gamma} = -\mathrm{E}_f\left[\frac{\partial^2 \log f}{\partial \boldsymbol{\beta}\partial \boldsymbol{\gamma}^T}\right] &= -d\frac{\theta\mu}{\theta+\mu}G_i^T B_i \\
I_{\theta,\theta} = -\mathrm{E}_f\left[\frac{\partial^2 \log f}{(\partial \theta)^2}\right] &= -d\left(\log\frac{\theta}{\theta+\mu}+\frac{\mu}{\theta+\mu}\right)^2
\end{aligned}$$

$$-(1-p)\sum_{y=0}^{+\infty} q(y)\left(\Psi'(\theta+y)-\Psi'(\theta)+\frac{\mu}{\theta(\theta+\mu)}+\frac{y-\mu}{(\theta+\mu)^2}\right)$$

$$I_{\theta,\beta} = -\mathrm{E}_f\left[\frac{\partial^2 \log f}{\partial\theta\partial\boldsymbol{\beta}}\right] = d\left(\log\frac{\theta}{\theta+\mu}+\frac{\mu}{\theta+\mu}\right)\frac{\theta\mu}{\theta+\mu}B_i$$

$$I_{\theta,\gamma} = -\mathrm{E}_f\left[\frac{\partial^2 \log f}{\partial\theta\partial\boldsymbol{\gamma}}\right] = d\left(\log\frac{\theta}{\theta+\mu}+\frac{\mu}{\theta+\mu}\right)G_i$$

and

$$d = \frac{p(1-p)q_0}{p+(1-p)q_0}.$$

Appendix 3

The estimate for size parameter when the NB model is fitted to ZINB data

Property 1 *Suppose that samples are drawn from ZINB(p, μ_0, θ_0) distribution. If NB(μ, θ) distribution is fitted to the samples, then as the sample sizes increase to $+\infty$:*

1. *The method of moment estimator $\hat{\theta}_{MM}$ for the size parameter converge to*

$$\theta_0\left(\frac{1-p_0}{1+\theta_0 p_0}\right)(\equiv \theta^*).$$

   We note that $\theta^* < \theta_0$.

2. *The maximum likelihood estimator $\hat{\theta}_{MLE}$ converges to $\theta^\dagger$ which satisfies*

$$\theta^\dagger < \theta_0.$$

The above property implies that when the NB model is fitted to ZINB data, both estimates for size parameter tend to be smaller than that of tje ZINB data. The proof is given in Minami and Lennert-Cody (2006).

Why the estimation is unstable when the size parameter is small

The partial derivatives of the log-likelihood function (i.e. score functions) for the NB regression model are given by

$$\frac{\partial l(\boldsymbol{\beta},\theta|y_i)}{\partial\boldsymbol{\beta}} = \frac{y_i-\mu_i}{\theta+\mu_i}\cdot\theta B_i^T \qquad \mu_i = \exp(B_i\boldsymbol{\beta}).$$

The maximum likelihood estimate is the solution of the equations obtained by equating the sum of the above partial derivatives for all samples to zero. Thus, the above partial derivatives suggest the characteristics of the maximum likelihood estimator for the NB regression model. At the MLE, residuals $y_i - \mu_i$ are balanced in a sense that the weighted sum of the residuals is zero. If size parameter θ is substantially large, it is unlikely that weights will be very large for some observations, and thus, the estimation is relatively robust. On the other hand, if the size parameter θ is small, weights could be very large for observations with small μ_i and those observations could be very influential for estimation.