

STATISTICAL REPORTS

Ecology, 88(11), 2007, pp. 2766–2772
© 2007 by the Ecological Society of America

QUASI-POISSON VS. NEGATIVE BINOMIAL REGRESSION: HOW SHOULD WE MODEL OVERDISPERSED COUNT DATA?

JAY M. VER HOEF¹ AND PETER L. BOVENG

*National Marine Mammal Laboratory, Alaska Fisheries Science Center, National Marine Fisheries Service,
7600 Sand Point Way NE, Building 4, Seattle, Washington 98115-6349 USA*

Abstract. Quasi-Poisson and negative binomial regression models have equal numbers of parameters, and either could be used for overdispersed count data. While they often give similar results, there can be striking differences in estimating the effects of covariates. We explain when and why such differences occur. The variance of a quasi-Poisson model is a linear function of the mean while the variance of a negative binomial model is a quadratic function of the mean. These variance relationships affect the weights in the iteratively weighted least-squares algorithm of fitting models to data. Because the variance is a function of the mean, large and small counts get weighted differently in quasi-Poisson and negative binomial regression. We provide an example using harbor seal counts from aerial surveys. These counts are affected by date, time of day, and time relative to low tide. We present results on a data set that showed a dramatic difference on estimating abundance of harbor seals when using quasi-Poisson vs. negative binomial regression. This difference is described and explained in light of the different weighting used in each regression method. A general understanding of weighting can help ecologists choose between these two methods.

Key words: *covariates; generalized linear models; harbor seals; iteratively weighted least squares; overdispersion; quasi models.*

INTRODUCTION

Ecology is the science of relating organisms to their environment. Often, data on organisms come in the form of counts, and we would like to relate these counts to environmental conditions. Linear regression is commonly used, but may not be the most appropriate for count data, which are nonnegative integers, and hence there is increasing interest in regression models that use Poisson or negative binomial distributions. Count data in ecology are often “overdispersed.” For a Poisson distribution, the variance is equal to the mean. This may be quite restrictive for biological data, which often exhibit more variation than given by the mean. We use the term “overdispersed” for any data set or model where the variance exceeds the mean. A common way to deal with overdispersion for counts is to use a generalized linear model framework (McCullagh and Nelder 1989), where the most common approach is a “quasi-likelihood,” with Poisson-like assumptions (that we call the quasi-Poisson from now on) or a negative binomial model. The

objective of this statistical report is to introduce some concepts that will help an ecologist choose between a quasi-Poisson regression model and a negative binomial regression model for overdispersed count data.

There are many examples of overdispersed count models in ecology, with important applications ranging from species richness to spatial distributions to parasitism. O’Hara (2005) noted the differences between Poisson and negative binomial distributions for species richness, with each being appropriate only when data were simulated from the correct model. Alexander et al. (2000) used a negative binomial distribution with a spatial model of parasitism. White and Bennetts (1996) modeled bird counts with a negative binomial distribution. For trend and abundance estimation for harbor seals, Frost et al. (1999), Small et al. (2003) and Mathews and Pendleton (2006) used Poisson regression, Ver Hoef and Frost (2003) used an overdispersed Poisson regression, and Boveng et al. (2003) used negative binomial regression.

Because overdispersion is so common, several models have been developed for these data, including the negative binomial, quasi-Poisson (Wedderburn 1974), generalized Poisson (Consul 1989), and zero-inflated (Lambert 1992) models. Relationships among some of

Manuscript received 10 January 2007; revised 28 March 2007; accepted 1 May 2007; final version received 22 May 2007.
Corresponding Editor: N. G. Yoccoz.

¹ E-mail: jay.verhoef@noaa.gov

the distributions can be found in Joe and Zhu (2005) and Lord et al. (2005). Despite these developments, the quasi-Poisson and negative binomial models are used most often, largely because they are widely available in software and they generalize easily to the regression case, which we outline in the next section.

The quasi-Poisson model and negative binomial model can account for overdispersion, and both have two parameters. Both are commonly available in software packages such as SAS, S, S-plus, or R. A natural question for the ecologist is: Which should I use? In the example below, we show striking differences between quasi-Poisson regressions and negative binomial regressions for a particular harbor seal data set. There is surprisingly little guidance in the statistical literature, especially for the regression case. Gardner et al. (1995) found little practical difference, but preferred a negative binomial model when a distributional form is required. Terceiro (2003) compared models using a Kolmogorov-Smirnov goodness-of-fit measure, and found cases where each model fit better. Potts and Elith (2006) found that a zero-inflated model was better than either quasi-Poisson or negative binomial for modeling abundance of a rare plant species, but they point out that zero-inflation is a special type of overdispersion that may be most appropriate when occurrence is rare (a specific mechanism creating excessive zeros).

For any given data set, information theoretic approaches such as Akaike information criteria (AIC; Akaike 1973) or Bayesian information criteria (BIC; Schwarz 1978) might be considered to choose between a quasi-Poisson model and a negative binomial. These approaches depend on a distributional form and a likelihood; however, quasi models are only characterized by their mean and variance, and do not necessarily have a distributional form. For this reason, Burnham and Anderson (2002:67) developed quasi-AIC (QAIC), but they only used it to compare within the quasi class of models (e.g., for subset selection of covariates), and not between quasi models and models with distributional forms. Nevertheless, Sileshi (2006) compared QAIC for quasi-Poisson to AIC for negative binomial, though the validity of this approach has not been demonstrated. In theory, any model selection method that depends on full distributional likelihoods, such as Bayes factors (Raftery 1995) or minimum description length (Rissanen 1978), including the information theoretic approaches, would not help choose between a quasi-Poisson and negative binomial model.

With the lack of a demonstrated information theoretic approach, one could adopt predictive or goodness-of-fit criteria as used by Gardner et al. (1995), Terceiro (2003), and Potts and Elith (2006), to choose between a quasi-Poisson and negative binomial model. However, a good understanding of the theoretical differences between them can form the basis for an a priori decision based on scientific purposes, which we explore in this article.

QUASI-POISSON AND NEGATIVE BINOMIAL REGRESSION

Both quasi and negative binomial models can be framed as generalized linear models. Let Y be a random variable such that

$$\begin{aligned} E(Y) &= \mu \\ \text{var}(Y) &= v_{\text{Poi}}(\mu) = \theta\mu \end{aligned} \quad (1)$$

where $E(Y)$ is the expectation of Y , $\text{var}(Y)$ is the variance of Y , $\mu > 0$ and $\theta > 1$. $E(Y)$ is also known as the “mean” of the distribution. Although $\mu > 0$, the data themselves can be any nonnegative integer. In Eq. (1), θ is an overdispersion parameter. The close relationship between Eq. 1 and the expectation and variance of a Poisson distribution, along with the use of a log link function, justify calling this a “quasi-Poisson” model, denoted as $Y \sim \text{Poi}(\mu, \theta)$. The quasi-Poisson model is characterized by the first two moments (mean and variance [Wedderburn 1974]), but Efron (1986) and Gelfand and Dalal (1990) showed how to create a distribution for this model; however, it requires reparameterization. Estimation often proceeds from the first two moments and estimating equations (Lee and Nelder 2000). The quasi model formulation has the advantage of leaving parameters in a natural, interpretable state and allows standard model diagnostics without a loss of efficient fitting algorithms.

We will denote the random variable Y having a negative binomial distribution as $Y \sim \text{NB}(\mu, \kappa)$, with a parameterization such that

$$\begin{aligned} E(Y) &= \mu \\ \text{var}(Y) &= v_{\text{NB}}(\mu) = \mu + \kappa\mu^2 \end{aligned} \quad (2)$$

where $\mu > 0$ and $\kappa > 0$. Here, the overdispersion (the amount in excess of μ) is the multiplicative factor $1 + \kappa\mu$, which depends on μ (in contrast to the quasi-Poisson). From Eqs. 1 and 2, an important difference is that for $\text{Poi}(\mu, \theta)$ the variance $v_{\text{Poi}}(\mu)$ is linearly related to the mean, whereas for $\text{NB}(\mu, \kappa)$ (where NB stands for negative binomial) the variance $v_{\text{NB}}(\mu)$ is quadratic in the mean. Hence, an important diagnostic is to plot $(Y_i - \mu_i)^2$ against μ_i . Often, this plot is messy, so we recommend binning μ_i into categories and averaging $(Y_i - \mu_i)^2$ within categories, as we do in Fig. 1A. To understand the implications further, we turn to the regression formulation.

One of the reasons these are two popular models is that the mean for both models is a single parameter that can vary as a function of covariates. For quasi-Poisson regression, we assume $Y_i \sim \text{Poi}(\mu_i, \theta)$ where we let the mean μ_i for the i th observation vary as a function of the covariates for that observation. Because the mean $\mu_i > 0$, it is natural to model

$$\mu_i = \exp(\beta_0 + \beta_1 x_{1,i} + \cdots + \beta_p x_{p,i}).$$

Generalizing, we can write this as the vector of mean parameters $\boldsymbol{\mu} = g^{-1}(\mathbf{X}\boldsymbol{\beta})$, where g^{-1} is the exponential function, \mathbf{X} is a design matrix of both continuous and

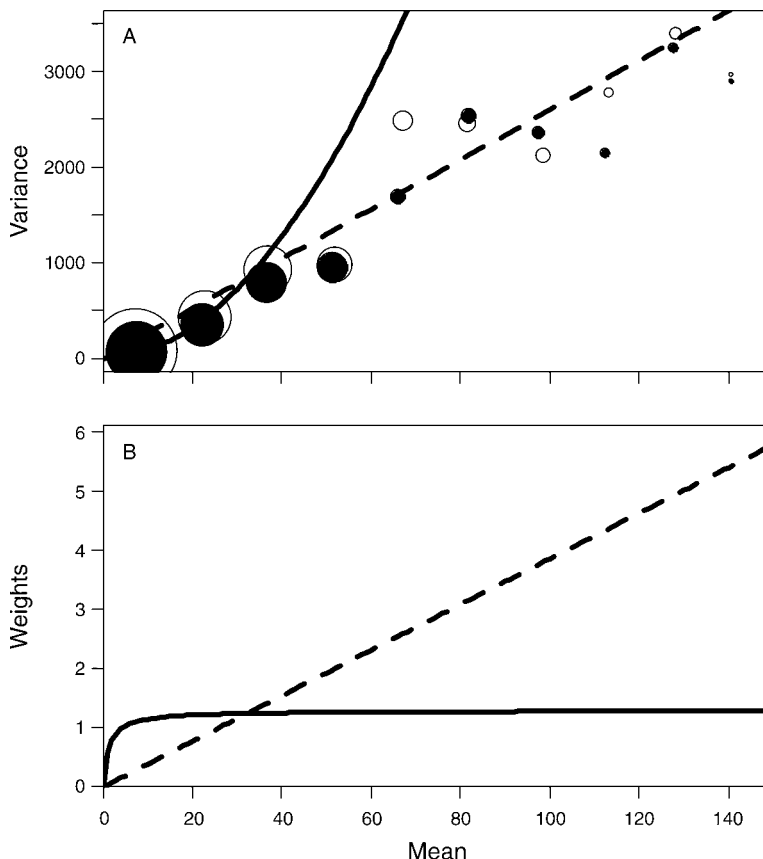


FIG. 1. (A) Estimated variance-to-mean relationship for the example data set. The circles are averaged squared residuals in 10 mean categories, where the solid circles are for quasi-Poisson and the open circles are for the negative binomial. The diameters of the circles are proportional to the number of values that were averaged. (B) Estimated weights as a function of the mean for the example data set. For both panels, the dashed line is the quasi-Poisson regression model, and the solid line is the negative binomial regression model.

categorical covariates, and β is a vector of parameters (regression coefficients). The i th row \mathbf{x}'_i of \mathbf{X} contains the covariates for the i th observation. Alternatively, we could write $g(\mu) = \mathbf{X}\beta$ where g is the log function, and it is called the link function. This is a fairly general specification, and g can take on various forms, but here we only consider the log link. For negative binomial regression, we assume $Y_i \sim \text{NB}(\mu_i, \kappa)$, where we let the mean μ_i vary as a function of covariates. Because $\mu_i > 0$, we again let $g(\mu) = \mathbf{X}\beta$ where g is the log link function.

One of our main questions in the choice of these models is: How much does the use of negative binomial vs. quasi-Poisson affect the fitting of the regression coefficients β ? The different mean/variance relations suggest that regression coefficients might be fit differently between negative binomial and quasi-Poisson because fitting these models uses weighted least squares, and these weights are inversely proportion to the variance. Thus, negative binomial and quasi-Poisson will weight observations differently.

To see how, we use the fact that generalized linear models and quasi models can be estimated using iteratively weighted least squares (IWLS; see, e.g.,

McCullagh and Nelder 1989:40). One iteration $j + 1$ in IWLS is given by

$$\hat{\beta}^{[j+1]} = (\mathbf{X}'\mathbf{W}^{[j]}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}^{[j]}\tilde{\mathbf{y}}^{[j]} \quad (3)$$

where the i th element of $\tilde{\mathbf{y}}^{[j]}$ is

$$\tilde{y}_i^{[j]} = \eta_i^{[j]} + \frac{(y_i - \mu_i^{[j]})}{\frac{\partial g^{-1}(\eta_i^{[j]})}{\partial \eta_i^{[j]}}}$$

$$\eta_i^{[j]} = \mathbf{x}'_i \hat{\beta}^{[j]} \quad \mu_i^{[j]} = g^{-1}(\mathbf{x}'_i \hat{\beta}^{[j]})$$

and $\mathbf{W}^{[j]}$ is a diagonal matrix with elements

$$\frac{\left[\frac{\partial g^{-1}(\eta_1^{[j]})}{\partial \eta_1^{[j]}} \right]^2}{v(\mu_1^{[j]})} \dots \frac{\left[\frac{\partial g^{-1}(\eta_n^{[j]})}{\partial \eta_n^{[j]}} \right]^2}{v(\mu_n^{[j]})}$$

Suppressing the iteration superscripts, for both negative binomial and quasi-Poisson, $\partial g^{-1}(\eta_i)/\partial \eta_i = \exp(\eta_i) = \mu_i$. Thus, for quasi-Poisson we obtain the following:

$$\mathbf{W} = \text{diag} \left(\frac{\mu_1^2}{\theta \mu_1} \dots \frac{\mu_n^2}{\theta \mu_n} \right) = \text{diag} \left(\frac{\mu_1}{\theta} \dots \frac{\mu_n}{\theta} \right) \quad (4)$$

where *diag* indicates diagonal elements only, all other elements being zero, and for negative binomial we obtain

$$\begin{aligned} \mathbf{W} &= \text{diag} \left(\frac{\mu_1^2}{\mu_1 + \kappa \mu_1^2} \dots \frac{\mu_n^2}{\mu_n + \kappa \mu_n^2} \right) \\ &= \text{diag} \left(\frac{\mu_1}{1 + \kappa \mu_1} \dots \frac{\mu_n}{1 + \kappa \mu_n} \right). \end{aligned} \quad (5)$$

Eqs. 4 and 5 provide a very useful comparison between negative binomial and quasi-Poisson. For quasi-Poisson, weights are directly proportional to the mean, and for negative binomial, weights have a concave relationship to the mean; that is, very small mean values get very little weight, but as the mean increases, weights level off to $1/\kappa$. We find this to be the most useful property in comparing the negative binomial to quasi-Poisson for regression applications. We now give an example to illustrate these differences and their effect on a real application.

EXAMPLE

Since the early 1990s, the National Marine Mammal Laboratory has conducted annual surveys of harbor seals in Alaska, rotating annually among five regions: (1) southern southeast Alaska; (2) northern southeast Alaska; (3) Gulf of Alaska; (4) Bristol Bay; and (5) Aleutian Islands. Boveng et al. (2003) described the methodology and reported results from a survey in the Gulf of Alaska region. In examining data from each region over the years, we modeled counts as a function of date, time of day, and tide, and compared the differences between a quasi-Poisson regression and a negative binomial regression. In general, there were not great differences. However, a data set from southern Southeast Alaska in 1998 showed striking differences that depended on the regression method.

For the southern Southeast Alaska data set, harbor seals were counted from aircraft from 18 August to 27 August 1998 (see Plate 1). Harbor seals use traditional haul-out sites, and 423 such sites were identified in the survey region. Let us denote as a random variable the count, Y_{ij} , from the i th site and the j th count. For both quasi-Poisson and negative binomial regression, we assumed

$$E(Y_{ij}) = \mu_{ij}$$

as in Eqs. 1 and 2. We allow the mean to be a positive function of covariates,

$$\mu_{ij} = \exp(\mathbf{x}'_{ij} \boldsymbol{\beta})$$

where \mathbf{x}_{ij} is a vector of measured covariates for the j th count of the i th site, and $\boldsymbol{\beta}$ is a vector of parameters. We fit the following model for each regression method:

$$\begin{aligned} \mathbf{x}'_{ij} \boldsymbol{\beta} &= \beta_{0,i} + \beta_{1,i} x_{1ij} + \beta_{2,i} x_{1ij}^2 + \beta_{3,i} x_{1ij}^3 + \beta_{4,i} x_{2ij} + \beta_{5,i} x_{2ij}^2 \\ &\quad + \beta_{6,i} x_{2ij}^3 + \beta_{7,i} x_{3ij} + \beta_{8,i} x_{3ij}^2 + \beta_{9,i} x_{3ij}^3 \end{aligned}$$

where $\beta_{0,i}$ is an effect for each site, x_{1ij} is day from 15 August for the j th count at the i th site, x_{2ij} is the time of day, in fractional hours since midnight, for the ij th count, and x_{3ij} is the relative tide height for the ij th count, defined as the height of the tide (in meters, always positive) relative to the low tide from the nearest tide station at the time of the count. To keep the model flexible (similar to the use of generalized additive models by Boveng et al. 2003), we considered a cubic polynomial for each environmental factor: date, time of day, and relative tide height. Obviously, time itself is not part of the “environment.” The effect of date is related to seal molt, with the peak in early August (e.g., Ver Hoef and Frost 2003). Time of day is related to solar gain during haul out, with peaks usually around midday (e.g., Boveng et al. 2003). Low tide often exposes isolated, rocky reefs that keep seals safe from terrestrial predators, so peak haul out is often near low tide (e.g., Boveng et al. 2003).

All models were fit using SAS PROC GLIMMIX (SAS Institute 2003). The estimated overdispersion parameter for quasi-Poisson regression was $\hat{\theta} = 25.91$, and the estimated variance parameter for negative binomial regression was $\hat{\kappa} = 0.7717$. For negative binomial regression, from Eq. 2, the amount of overdispersion changes with μ ; if $\mu = 10$ then $1 + \hat{\kappa}\mu = 8.717$. For these data, overdispersion for quasi-Poisson and negative binomial regression was equal at $\mu \approx 32$. Notice that the amount of overdispersion is quite high. Such overdispersion can be caused by several factors (see Eberhardt [1978] for a discussion), including animals acting as a group, seals having individual responses to covariates (e.g., an individual response to date, tide, and so on), or “lurking covariates” (factors that affect all animals but were not measured). All of these are likely to be operating here, but the extent of each is unknown. It is also unknown whether these factors act as a constant overdispersion (e.g., θ for the quasi-Poisson) or act as an increasing overdispersion in the mean (e.g., $1 + \kappa\mu$ for the negative binomial); however, see Fig. 1A and *Discussion*.

Using $\hat{\theta} = 25.91$ and $\hat{\kappa} = 0.7717$, we plotted $v_{\text{Poi}}(\mu)$ (Eq. 1) and $v_{\text{NB}}(\mu)$ (Eq. 2) against μ (Fig. 1A). We also plotted averaged squared residuals $(Y_i - \hat{\mu}_i)^2$ for 10 mean categories $0 < \hat{\mu}_i \leq 15$, $15 < \hat{\mu}_i \leq 30$ and so on, which should help diagnose a linear or quadratic relationship between the mean and variance. The diameters of the circles are proportional to the number averaged for the i th category, and all categories have at least 30 samples. It appears that for small values of $\hat{\mu}_i$, the negative binomial error structure fits slightly better, but for larger values of $\hat{\mu}_i$, the quasi-Poisson fits much better. We can also see from Fig. 1A that for means of less than 32 seals, the quasi-Poisson will have a higher variance, and for means above 32, the negative binomial will have a

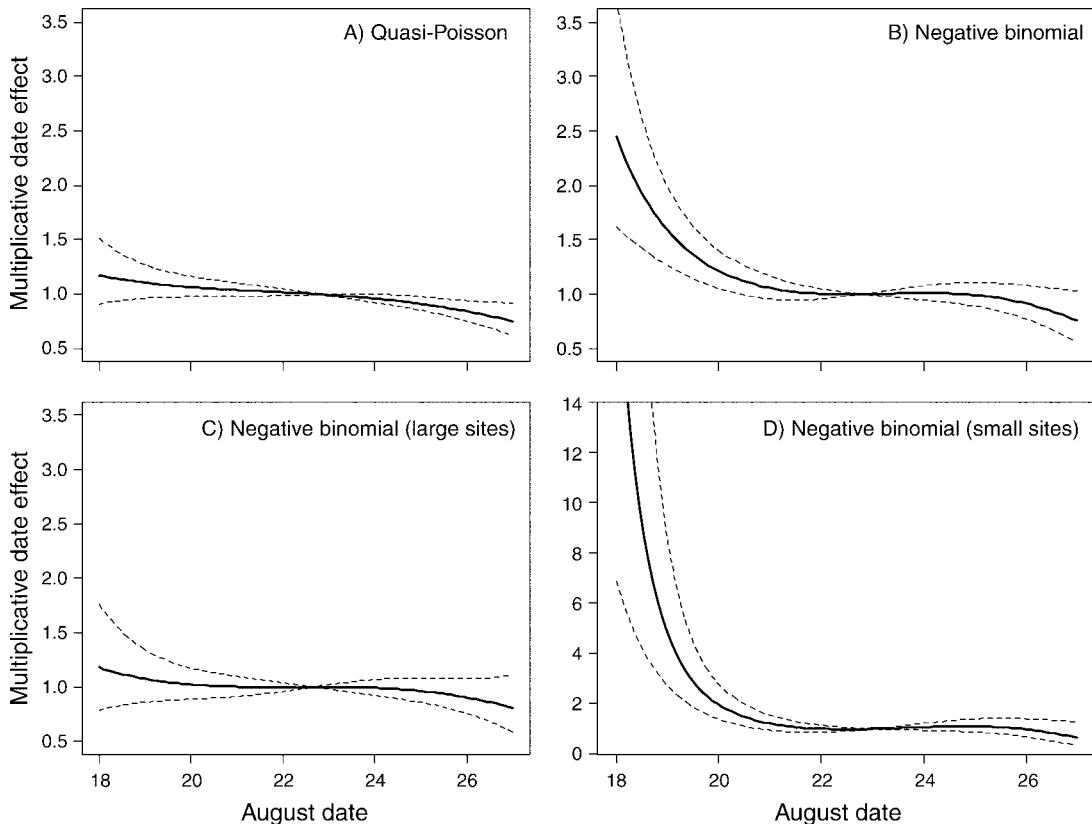


FIG. 2. Fitted effect of date on harbor seal counts. The fitted effect is centered so that the multiplicative effect is 1 for the mean date. (A) Fitted effect for quasi-Poisson model. (B) Fitted effect for negative binomial model for all sites. (C) Fitted effect for negative binomial model for 226 large sites. (D) Fitted effect for negative binomial model for 197 small sites.

higher variance. What does this mean for fitting the covariates? The same parameter estimates define the weights, as a function of the mean, in Eqs. 4 and 5. These weights, as a function of the mean, are given in Fig. 1B. Clearly, a quasi-Poisson regression gives greater overall weight to larger sites (i.e., sites with more seals) than does a negative binomial regression. From Fig. 1B, the negative binomial will give sites with means less than 32 more weight relative to the quasi-Poisson, while the quasi-Poisson will give sites with means greater than 32 more weight than the negative binomial. This is due to their different assumptions about how the variance is related to the mean, shown in Fig. 1A. For the sequel, we focus on the estimated effect of date, which is given for the quasi-Poisson regression in Fig. 2A, and for the negative binomial regression in Fig. 2B.

We created Fig. 2 over the range of the observed dates: 18–27 August. Let $x_{1,[1]}$ be the earliest recorded observational date, and let $x_{1,[101]}$ be the latest recorded observational date, with 99 evenly spaced (fractional) dates in between. We created the following matrix:

$$\mathbf{M}_1 = \begin{pmatrix} 0 & \cdots & 0 & x_{1,[1]} - \bar{x}_1 & x_{1,[1]}^2 - \bar{x}_1^2 & x_{1,[1]}^3 - \bar{x}_1^3 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & x_{1,[101]} - \bar{x}_1 & x_{1,[101]}^2 - \bar{x}_1^2 & x_{1,[101]}^3 - \bar{x}_1^3 & 0 & \cdots & 0 \end{pmatrix}$$

where $\bar{x}_1 = 22.8$ is the mean August date, and the non-zero columns correspond to the date parameter estimates in $\hat{\boldsymbol{\beta}}$. The fitted effect of date for 101 values evenly spaced over the range of observed dates, holding all other effects constant, and centered around the mean, is given by

$$\mathbf{f}_1 = \mathbf{M}_1 \hat{\boldsymbol{\beta}}. \tag{6}$$

The variance for \mathbf{f}_1 in Eq. 6 is given by the diagonal elements of

$$\text{cov}(\mathbf{f}_1) = \mathbf{M}_1 \hat{\mathbf{C}} \mathbf{M}_1' \tag{7}$$

where $\hat{\mathbf{C}} = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}$ is the estimated covariance matrix among the parameter estimates $\hat{\boldsymbol{\beta}}$; that is, the final (converged) values from Eq. (3). The 95% confidence interval for each fitted value was formed by taking 1.96 times the standard error (square root of the estimated variance). Then, the fitted values and confidence intervals were put back on the original scale by exponentiation.

From Fig. 2, the multiplicative effect of date on 18 August was nearly 2.45 using negative binomial regression, while it was only around 1.17 using quasi-Poisson regression. This has a dramatic effect on estimates of harbor seal abundance. As described in Boveng et al. (2003), observed harbor seal counts are



PLATE 1. Hauled-out harbor seals photographed during aerial surveys. Photo credit: NOAA National Marine Mammal Laboratory.

adjusted to conditions of the covariates that yield the maximum counts (we call these the optimum conditions). Observed counts are adjusted to optimum conditions because it is not possible to be at every site at low tide at noon every day of every annual survey period due to tidal fluctuations, weather, and logistics (Ver Hoef and Frost 2003). The optimum date for both quasi-Poisson and negative binomial regression was the earliest date, 18 August. However, the adjusted estimate of harbor seal abundance using a quasi-Poisson regression was 38 884 on 18 August, while using a negative binomial regression it was 80 609. Clearly, the choice of the regression method has a large impact on our abundance estimate.

The difference in shapes between Fig. 2A and Fig. 2B can be explained. The adjustment for the large sites is strongly influenced by the smaller sites for the negative binomial regression; from Fig. 1B, sites with means of 10 to 32 get essentially the same weight as sites with means 32 to 100. In fact, we can check this by looking at the fits after dividing the data into the 197 sites that have raw means of less than or equal to 32 and the 226 sites that have raw means of greater than or equal to 32. If we make separate adjustments for large sites using negative binomial regression, we obtain a date effect as shown in Fig. 2C, which is almost identical to Fig. 2A. For the 226 large sites, we get an adjusted estimate of 34 239 for 18 August. If we make a separate adjustment for the small

sites using negative binomial regression, we obtain a date effect as shown in Fig. 2D; note the change in the scaling of the y -axis. For the 197 small sites, we get an adjusted estimate of 84 343 for 18 August, which is not reasonable in comparison to the large sites or on biological grounds. If we make the adjustment to 21 August for the small sites, the estimate is 5091, more than a 15-fold drop in three days, which does not fit our understanding of harbor seal movements or dynamics.

The preceding analysis, of splitting sites into large and small, might suggest that combining all sites into one analysis was wrong in the first place, and hence it was inappropriate to use a negative binomial regression on all of the data. However, we point out that we have rarely witnessed such a dichotomy between large and small sites in many other analyses of similar data. An alternative explanation is that the difference between large and small sites was purely random, which seems more likely to us.

DISCUSSION

So, which is better: quasi-Poisson regression or negative binomial regression? There is no general answer. However, for the example that we gave, we think that the quasi-Poisson regression is better. A diagnostic plot of the empirical fit of the variance (using averaged squared residuals) to mean relationship, shown in Fig. 1A, suggests the quasi-Poisson is a better fit to the overall

variance-mean relationship. More importantly to us, though, is the fact that the negative binomial gives smaller sites more weight relative to quasi-Poisson and allows smaller sites to have a greater effect on adjustments for negative binomial regression. Our goal is to estimate overall abundance, which is dominated by the larger sites, and we prefer to have adjustments dominated by the effects at those larger sites, which is what happens for quasi-Poisson regression. Fig. 1B is crucial to our understanding of how each regression works, and it frames our decision. We believe this will be useful to other ecologists when it comes time to decide between quasi-Poisson and negative binomial regression methods.

Ultimately, choosing among quasi-Poisson regression, negative binomial, and other models is a model selection problem. In the introduction, we pointed out problems with methods based on likelihoods. However, other approaches that do not depend on distributions, such as cross-validation (Vehtari and Lampinen 2003) or the methods of Gardner et al. (1995), Terceiro (2003), and Potts and Elith (2006) could be used, in addition to diagnostic plots as given in Fig. 1A. In this article, we also point out that an important way to choose an appropriate model is based on sound scientific reasoning rather than a data-driven method. We believe that understanding the difference in weighting between quasi-Poisson and negative binomial regression provides such an example.

ACKNOWLEDGMENTS

We thank Ward Testa and Jeff Breiwick for helpful reviews of this manuscript. This project received financial support from the National Marine Fisheries Service, NOAA.

LITERATURE CITED

- Akaike, H. 1973. Information theory and the extension of the maximum likelihood principle. Pages 267–281 in B. N. Petrov and F. Czaki, editors. Proceedings of the International Symposium on Information Theory. Akademia Kiadoo, Budapest, Hungary.
- Alexander, N., R. Moyeed, and J. Stander. 2000. Spatial modeling of individual-level parasite counts using the negative binomial distribution. *Biostatistics* 1:453–463.
- Boveng, P. L., J. L. Bengtson, D. E. Withrow, J. C. Cesarone, M. A. Simpkins, K. J. Frost, and J. J. Burns. 2003. The abundance of harbor seals in the Gulf of Alaska. *Marine Mammal Science* 19:111–127.
- Burnham, K. P., and D. R. Anderson. 2002. Model selection and multimodel inference: a practical information theoretic approach. Second edition. Springer-Verlag, Berlin, Germany.
- Consul, P. C. 1989. Generalized Poisson distribution: properties and applications. Marcel Dekker, New York, New York, USA.
- Eberhardt, L. L. 1978. Appraising variability in population studies. *Journal of Wildlife Management* 42:207–238.
- Efron, B. 1986. Double exponential families and their use in generalized linear regression. *Journal of the American Statistical Association* 81:709–721.
- Frost, K. J., L. F. Lowry, and J. M. Ver Hoef. 1999. Monitoring the trend of harbor seals in Prince William Sound, Alaska, after the *Exxon Valdez* oil spill. *Marine Mammal Science* 15:494–506.
- Gardner, W., E. P. Mulvey, and E. C. Shaw. 1995. Regression analyses of counts and rates: Poisson, overdispersed Poisson, and negative binomial models. *Psychological Bulletin* 118:392–404.
- Gelfand, A. E., and S. R. Dalal. 1990. A note on overdispersed exponential families. *Biometrika* 77:55–64.
- Joe, H., and R. Zhu. 2005. Generalized Poisson distribution: the property of mixture of Poisson and comparison with negative binomial distribution. *Biometrical Journal* 47:219–229.
- Lambert, D. 1992. Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics* 34:1–14.
- Lee, Y., and J. A. Nelder. 2000. The relationship between double-exponential families and extended quasi-likelihood families, with application to modeling Geissler's human sex ratio data. *Applied Statistics* 49:413–419.
- Lord, D., S. Washington, and J. Ivan. 2005. Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle crashes: balancing statistical fit and theory. *Accident Analysis and Prevention* 37:35–46.
- Mathews, E. A., and G. W. Pendleton. 2006. Declines in harbor seal (*Phoca vitulina*) numbers in Glacier Bay National Park, Alaska, 1992–2002. *Marine Mammal Science* 22:167–189.
- McCullagh, P., and J. A. Nelder. 1989. Generalized linear models. Second edition. Chapman and Hall, New York, New York, USA.
- O'Hara, R. B. 2005. Species richness estimators: How many species can dance on the head of a pin? *Journal of Animal Ecology* 74:375–386.
- Potts, J., and J. Elith. 2006. Comparing species abundance models. *Ecological Modelling* 199:153–163.
- Raftery, A. E. 1995. Bayesian model selection in social research (with discussion by Andrew Gelman, Donald B. Rubin, and Robert M. Hauser). Pages 111–195 in P. V. Marsden, editor. *Sociological methodology* 1995. Blackwell, Cambridge, Massachusetts, USA.
- Rissanen, J. 1978. Modeling by the shortest data description. *Automatica* 14:465–471.
- SAS Institute. 2003. SAS version 9.1. SAS Institute, Cary, North Carolina, USA.
- Schwarz, G. 1978. Estimating the dimension of a model. *Annals of Statistics* 6:461–464.
- Sileshi, G. 2006. Selecting the right statistical model for analysis of insect count data by using information theoretic measures. *Bulletin of Entomological Research* 96:479–488.
- Small, R. J., G. W. Pendleton, and K. W. Pitcher. 2003. Trends in abundance of Alaska harbor seals, 1983–2002. *Marine Mammal Science* 19:344–362.
- Terceiro, M. 2003. The statistical properties of recreational catch rate data for some fish stocks off the northeast US coast. *Fishery Bulletin* 101:653–672.
- Vehtari, A., and J. Lampinen. 2003. Expected utility estimation via cross-validation. Pages 701–710 in J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith, and M. West, editors. *Bayesian statistics 7*. Oxford University Press, Oxford, UK.
- Ver Hoef, J. M., and K. J. Frost. 2003. A Bayesian hierarchical model for monitoring harbor seal changes in Prince William Sound, Alaska. *Environmental and Ecological Statistics* 10:201–219.
- Wedderburn, R. W. M. 1974. Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika* 61:439–447.
- White, G. C., and R. E. Bennetts. 1996. Analysis of frequency count data using the negative binomial distribution. *Ecology* 77:2549–2557.

SUPPLEMENT

SAS code and data set for obtaining the results described in this paper (*Ecological Archives* E088-171-S1).