International Statistical Review

International Statistical Review (2022), **90**, 2, 328–345 doi: 10.1111/insr.12485

Constructing Flexible, Identifiable and Interpretable Statistical Models for Binary Data

Henry R. Scharf[1], Xinyi Lu[2], Perry J. Williams[3] and Mevin B. Hooten[2,4]

[1]*Department of Mathematics and Statistics, San Diego State University, San Diego, CA, USA*
[2]*Department of Statistics, Colorado State University, Fort Collins, CO, USA*
[3]*Department of Natural Resources and Environmental Science, University of Nevada, Reno, NV, USA*
[4]*U.S. Geological Survey, Colorado Cooperative Fish and Wildlife Research Unit, Department of Fish, Wildlife, and Conservation Biology, Colorado State University, Fort Collins, CO, USA*
Correspondence to: Henry R. Scharf, Department of Mathematics and Statistics, San Diego State University, San Diego, CA. Email: henryrscharf@gmail.com

Summary

Binary regression models are ubiquitous in virtually every scientific field. Frequently, traditional generalised linear models fail to capture the variability in the probability surface that gives rise to the binary observations, and remedial methods are required. This has generated a substantial literature composed of binary regression models motivated by various applications. We describe an organisation of generalisations to traditional binary regression methods based on the familiar three-part structure of generalised linear models (random component, systematic component and link function). This perspective facilitates both the comparison of existing approaches and the development of flexible models with interpretable parameters that capture application-specific data-generating mechanisms. We use our proposed organisational structure to discuss concerns with certain existing models for binary data based on quantile regression. We then use the framework to develop and compare several binary regression models tailored to occupancy data for European red squirrels (*Sciurus vulgaris*).

Key words: Binary response; generalised linear model; hierarchical models; quantile regression.

1 Introduction

In almost every area of scientific application, binary regression models are used to understand how the probability of a 'success' depends on covariates of interest. In most cases, generalised linear models (GLMs Nelder & Wedderburn, 1972) are employed to learn about this probability function. The traditional three-part structure of random component, systematic component and link function provides researchers flexibility in the parametric form of the probability function. In practice, these components are often specified based on commonly used defaults and without consideration of the real natural phenomena that generated the observed binary data. Moreover, the most widely used GLMs often require rigid assumptions that are unrealistic. The need

for flexible binary regression models that relax assumptions about the data-generating process has resulted in a substantial literature motivated by various applications. We organise generalisations of traditional binary regression methods based on the familiar three-part structure of GLMs. We use our proposed organisational structure to discuss concerns with certain existing models for binary data based on quantile regression. Then, we use the framework to develop and compare several binary regression models tailored to occupancy data for European red squirrels (*Sciurus vulgaris*), and score each model based on its predictive ability. The conclusions from this exploration highlight the need for researchers to consider flexibility, interpretability and predictive power when selecting among several candidate models.

1.1 Background and Notation

Denote by $\Pr(y_i = 1)$ the probability that a binary response random variable y_i is 1 (i.e. that we observe a 'success' for the i-th datum), where $i = 1, \ldots, n$ indexes the observations. The goal of a binary regression analysis is to infer the relationship between these probabilities and a vector of relevant covariates, $\mathbf{x}_i$. Thus, interest lies in learning about the characteristics of a map $p: \mathcal{X} \to [0, 1]$, parameterised by a (possibly infinitely long) vector, $\boldsymbol{\theta}$. The data are generally assumed to be conditionally independent given the covariates, leading to a log-likelihood of the form $l(\boldsymbol{\theta}|\mathbf{y}) = \sum_{i=1}^{n} y_i \log(p_i) + (1 - y_i)\log(1 - p_i)$, where $\mathbf{y} = (y_1 \ldots y_n)'$ and $p_i = p(\mathbf{x}_i, \boldsymbol{\theta})$ with unknown parameters, $\boldsymbol{\theta}$. Likelihood-based approaches are often used to fit the model to data, and software packages for both frequentist and Bayesian paradigms are available to facilitate model fitting.

The most common framework employed to specify a functional form for the probability surface is a GLM given by

$$y_i \sim f_Y(\eta(\mathbf{x}_i)), \quad \eta(\mathbf{x}) = \mathbf{x}'\boldsymbol{\beta}, \quad g(\mathbb{E}y_i) = \eta(\mathbf{x}_i). \tag{1}$$

Generalised linear models are often described as consisting of these three parts. The first part, called the 'random component', specifies the probability density of the observed data. The second part, called the 'systematic component', defines a function $\eta(\mathbf{x})$ related to the expected value of the data. The third part, called the 'link function', is represented by the invertible function $g(\cdot)$, and its purpose is to align the systematic component with the support of the expected value of the response distribution. GLMs are used for a wide range of response types; however, the remainder of the manuscript focuses on binary data, and thus, f_Y is always taken to be Bernoulli. For a more in-depth treatment of GLMs, see, for example, McCullagh & Nelder (1983) and Agresti (2002).

In many applications, the traditional linear systematic component, paired with a symmetric link function such as the logit, is a too rigid a form to adequately explain the variation in probability of an observed binary response. In these situations, traditional approaches must be generalised to allow for more flexible probability functions. Many new methods seek to account for variation in binary data showing evidence of extra heterogeneity beyond what is explainable with simple linear effects and a logit or probit (i.e. Gaussian inverse CDF) link (e.g. Bazán *et al*., 2010). These methods have come from a variety of different areas, with applications in psychology (e.g. Fahrmeir & Raach, 2007), ecology (e.g. Augustin *et al*., 1996; Komori *et al*., 2016), economics (e.g. Khan, 2013) and many other disciplines. We show how generalisations of traditional GLMs can be usefully grouped based on which of the three GLM components are modified to allow for additional flexibility in the probability function.

First, increased flexibility in the functional relationship between the probability of success and the covariates may be achieved through polynomial relationships (e.g. quadratic) and

non-parametric approaches that relax the linear form of $\eta(\mathbf{x})$ (Hastie & Tibshirani, 1986; Hefley *et al*., 2017; Wood, 2017). Second, increased flexibility can arise through alternative specifications of the link function, $g(\mathbb{E}y_i)$, to include additional unknown parameters to be estimated from the data. While several approaches have been proposed for binary regression models that accommodate asymmetric (Chen *et al*., 1999; Maalouf & Trafalis, 2011; Komori *et al*., 2016) or heavy-tailed link functions (Wang & Dey, 2010), a clear motivation for specifying a model with a non-standard link function comes when researchers have knowledge about the data generation process (e.g. a complimentary log–log link for a latent count process; Section 1.2).

Finally, flexibility in the probability function may be introduced by modifying the response distribution, or random component. For binary data, the marginal distribution of the responses must be Bernoulli. However, relaxing the assumption of conditional independence allows researchers to account for residual dependence in the observations while controlling for covariate effects (e.g. Augustin *et al*., 1996). One important reason residual dependence may be present in the data is the existence of important but unaccounted for covariates. Models for dependent binary data can be achieved through the introduction of random effects in the systematic component. Frequently, random effects are indexed by known or unobserved class structures in the data, or spatio-temporal information (e.g. Diggle *et al*., 1998; Goldenberg *et al*., 2010).

Crucially, sources of flexibility in the probability function introduced through multiple components do not act independently. For example, two different link functions, g_a and g_b, paired with two potentially non-linear systematic components $\eta_a(\mathbf{x})$ and $\eta_b(\mathbf{x})$ produce exactly the same probability function if

$$g_b^{-1}(\eta_b(\mathbf{x})) = g_a^{-1}(\eta_a(\mathbf{x})), \tag{2}$$

because the probability of success $p_i = \mathbb{E}y_i = g^{-1}(\eta(\mathbf{x}_i))$ is the same under each pair of links and systematic functions. A corollary is that a non-linear systematic component paired with one type of link function has an equivalent representation using a linear systematic component with another link function. Thus, the specification of the systematic component and the link function must be made holistically so that parameters are identifiable from binary response data.

To facilitate comparison of the broad range of approaches and to aid practitioners interested in developing custom models for specific applications, we describe a hierarchical formulation for binary regression in Section 1.2 that makes use of auxiliary variables. Representing models through auxiliary variables is illuminating when selecting an appropriate link function. Auxiliary variables may correspond to interpretable features of a data generation process and include more interpretable parameters than traditional formulations.

1.2 Auxiliary Variable Construction

Statistical models for binary data are sometimes specified using an auxiliary variable construction such that

$$y_i|z_i = 1_{z_i > 0}, \quad z_i|\boldsymbol{\beta} \sim f_Z(z_i|\mathbf{x}_i, \boldsymbol{\beta}), \tag{3}$$

where $1_{z_i > 0}$ is an indicator function that is equal to 1 when z_i is positive and 0 otherwise, and the conditional distribution f_Z is a member of a family of probability density functions with parameters that depend on $\mathbf{x}_i$ and $\boldsymbol{\beta}$. The probability distribution for $y_i|\boldsymbol{\beta}$ is Bernoulli with probability of a success equal to the probability that the auxiliary variable, z_i, exceeds 0, or $1 - F_Z(0|\boldsymbol{\beta})$, where F_Z is the cumulative distribution function for f_Z. An early well-known example of such a hierarchical construction came from Albert & Chib (1993) who noted that probit regression arises when f_Z is chosen to be a Gaussian density with mean

$\mathbb{E}z_i|\boldsymbol{\beta} = \mathbf{x}_i'\boldsymbol{\beta}$ and variance 1. This is due to the symmetry of the Gaussian distribution, which yields $1 - \Phi(-\mathbf{x}'\boldsymbol{\beta}) = \Phi(\mathbf{x}'\boldsymbol{\beta})$. The benefit of the hierarchical specification in this case is that a conjugate prior exists for $\boldsymbol{\beta}$ (multivariate Gaussian), and model fitting can proceed from a Bayesian perspective using a Markov chain Monte Carlo (MCMC) algorithm comprised entirely of Gibbs updates, obviating the need to adjust tuning parameters required by other methods such as Metropolis–Hastings random walks. The approach has been used in a wide variety of applications including species occupancy models in ecology (e.g. Hooten *et al*., 2003; Dorazio & Rodríguez, 2012; Johnson *et al*., 2013).

Hierarchical representations of binary regression models share a close connection with GLMs. In general, a hierarchically specified binary regression model is equivalent to a GLM if there exists a link function and systematic component such that $g^{-1}(\eta(\mathbf{x}_i)) = 1 - F_Z(z_i = 0|\boldsymbol{\beta})$. When, as is commonly the case, the expected value of the auxiliary variable depends linearly on the covariates (i.e. $\mathbb{E}z_i|\boldsymbol{\beta} = \mathbf{x}_i'\boldsymbol{\beta}$) and f_Z is a member of a location family of probability distributions, the model marginalised over the latent variables is guaranteed to be a GLM with systematic component, $\eta(\mathbf{x}_i) = \mathbf{x}_i'\boldsymbol{\beta}$, link function $g(\mathbb{E}y_i) = -F_Z^{-1}(1 - \mathbb{E}y_i)$ (or equivalently, inverse link function $g^{-1}(\eta(\mathbf{x}_i)) = 1 - F_Z(-\eta(\mathbf{x}_i))$).

The flexibility of a hierarchical specification allows for the development of binary regression models motivated by the unobserved natural phenomena that give rise to observed binary data. For example, consider an application in ecology where the data are site-specific observations of the presence ($y_i = 1$) or absence ($y_i = 0$) of a species of interest, and the covariates are features of the landscape measured at each site (e.g. vegetation cover and elevation). If we assume that observations are made with perfect detection (i.e. if a species is present in any abundance at site i, $p_i = 1$, and similarly, $p_i = 0$ when the species is absent), then we could define presence as an indicator that the unobserved abundance of the species, z_i, is greater than 0. A natural choice for f_Z would often be a family of discrete, non-negative probability mass functions such as the Poisson distribution with rate parameter $\lambda = e^{\mathbf{x}'\boldsymbol{\beta}}$ (as in Royle & Nichols, 2003).

For the case of Poisson-distributed auxiliary variables, the probability of success is $p_i = \Pr(z_i > 0|\boldsymbol{\beta}) = 1 - \exp\{-e^{\mathbf{x}_i'\boldsymbol{\beta}}\}$. This corresponds to the special case of a GLM with linear systematic component paired with a complimentary log–log (cloglog) link function. Alternatively, one might specify a negative binomial distribution for z_i, in which case the model construction presents a generalisation of the typical GLM through the inclusion of a link function with an additional unknown parameter.

There are multiple equivalent auxiliary variable specifications of hierarchical models for binary data. The relationship given by equation (2) has an analogous representation in the auxiliary variable framework. Two different auxiliary variables, $z_i^{(a)}$ and $z_i^{(b)}$ with respective CDFs, $F_Z^{(a)}$ and $F_Z^{(b)}$, describe the same probability function for y_i i

$$F_Z^{(a)}(z_i^{(a)} = 0) = F_Z^{(b)}(z_i^{(b)} = 0), \ \forall\, \mathbf{x}_i \in \mathcal{X}. \tag{4}$$

1.3 Scientific Interpretation

In almost every binary regression analysis, there will be interest in estimating the true probabilities of success for each observed combination of covariates, as well as predicting probabilities for new combinations. In addition, researchers are often interested in the change in probability of success for small perturbations in the covariates. For these questions, the particular parameterisation of p is immaterial. That is, provided a sufficiently flexible model for the probability map has been specified, it does not matter, asymptotically, what particular

distribution is chosen for the auxiliary variables. Equivalently, given a sufficiently flexible systematic component, it does not matter what the link function is, as long as it is invertible and maps the support of $\mathbb{E}y_i$ to the real line. Thus, if the questions of scientific interest concern only the raw probability function, the primary modelling considerations will be that the algorithm for fitting the model to data is efficient and numerically stable and that there are useful tools available for assessing goodness of fit (e.g. Conn *et al*., 2018; Wright *et al*., 2019).

The probability surface provides predictions for the probabilities of successes for both observed and unobserved combinations of covariates and is often of interest. An underutilised function in the analyses of binary data is the gradient of the probability function, $\nabla_{\mathbf{x}} p(\mathbf{x}, \boldsymbol{\theta})$. The gradient of the probability function is analogous to the coefficients in traditional linear regression in that it describes how the probability of success changes for small perturbations in the levels of the covariates. However, unlike in traditional linear regression, the probability function is a non-linear function of the covariates, and thus, the value of the gradient is not constant but rather depends on the value of $\mathbf{x}$. One useful region of the gradient surface to consider is the vicinity of the mean of the observed covariates, although features of the gradient surface near the extremes of the observed covariate values may also be scientifically relevant. For example, in actuarial applications, extreme regions of the covariate space might correspond to particularly high or low risk individuals.

In addition, the parameters $\boldsymbol{\theta}$ themselves may offer additional opportunities for scientific learning. For example, the regression coefficients in a traditional logistic regression model represent the change in log-odds of the response for an increase of 1 unit in the associated covariate. As the example involving Poisson random variables in the previous section showed, the auxiliary variables may have a useful parametric interpretation corresponding to some unobserved natural process. The most useful models are those that are parameterised in a way that yields useful interpretation and/or admits efficient algorithms for model fitting. For instance, in the occupancy example mentioned in Section 1.2 that assumes Poisson-distributed auxiliary variables with mean $e^{\mathbf{x}'\boldsymbol{\beta}}$, β_j can be interpreted as the linear effect on log-abundance of increasing x_j by 1 unit, holding all other covariates constant.

2 Modifying the Link Function

One common concern about traditional link functions is the potentially restrictive assumption of symmetry. Symmetric link functions make the implicit assumption that there exists a sub-space of $\mathcal{X}$ defined by $\mathbf{x}'\boldsymbol{\beta} = 0$ around which changes in covariate values, $\Delta\mathbf{x}$, result in changes to the probability function of a magnitude that only depends on the length of the vector $\Delta\mathbf{x}$. In particular, changes in probability for shifts $\Delta\mathbf{x}$ and $-\Delta\mathbf{x}$ are the same in size but in opposite directions (i.e. $p(\Delta\mathbf{x}) = 1 - p(-\Delta\mathbf{x})$). Inverse link functions defined through location family auxiliary variables for which $1 - F_Z(-z) = F_Z(z)$ are examples of symmetric link functions.

There are several useful ways to relax the assumption of symmetry through the link function. Komori *et al*. (2016) defined a new inverse link function of the form

$$g^{-1}(\eta(\mathbf{x})) = \frac{\exp\{\eta(\mathbf{x})\}+\kappa}{1+\exp\{\eta(\mathbf{x})\}+\kappa} \tag{5}$$

that modifies traditional logistic regression. As $\kappa > 0$ grows, an increasing amount of 'right' skewness enters the link function, in that $1 - p(-\Delta\mathbf{x}) < p(\Delta\mathbf{x})$. The cloglog link function is an example of a 'left'-skewed link function (i.e. $1 - p(-\Delta\mathbf{x}) > p(\Delta\mathbf{x})$), where the degree of skewness is determined by the systematic component, rather than a free parameter.

Prentice (1976) proposed a two parameter model for skew link functions that allowed for skewness to occur in either direction and was among the first such generalisations of traditional symmetric approaches. Figure 1(c) and 1(g) shows examples of left-skewed and right-skewed probability curves, respectively. Active development of new methodology for skew link functions is ongoing (e.g. Lemonte & Bazán, 2018).

Introducing additional flexibility through parameters in the link function can indeed result in better-fitting models, yielding probability functions and gradients thereof that more closely align with the true data-generating mechanism. However, the skewness parameter used in the inverse link function of Komori *et al*. (2016) can be difficult to interpret directly, and its introduction complicates the log-odds interpretation of the regression coefficients. In addition, the skewness parameter can be highly confounded with the intercept term, resulting in potentially unstable estimation algorithms. An alternative method to account for skewness is based on an auxiliary variable hierarchical model in which the distribution of $z_i|\boldsymbol{\beta}$ is asymmetric. As noted already, specifying Poisson-distributed auxiliary variables with conditional mean $e^{\mathbf{x}_i\boldsymbol{\beta}}$ is equivalent to a GLM with a cloglog link, but in this case, the skewness in the probability function has a parametric interpretation. It arises as a direct result of a data-generating process, which consists of observing whether a count process is positive or zero. In addition, the regression coefficients can be interpreted as the linear effect of each covariate on log-abundance in ecological applications. When the application does not present a natural choice for the auxiliary variable distribution, phenomenological models that incorporate asymmetric probability functions can still be constructed using the hierarchical representation (e.g. Chen *et al*., 1999; Xing & Qian, 2017).

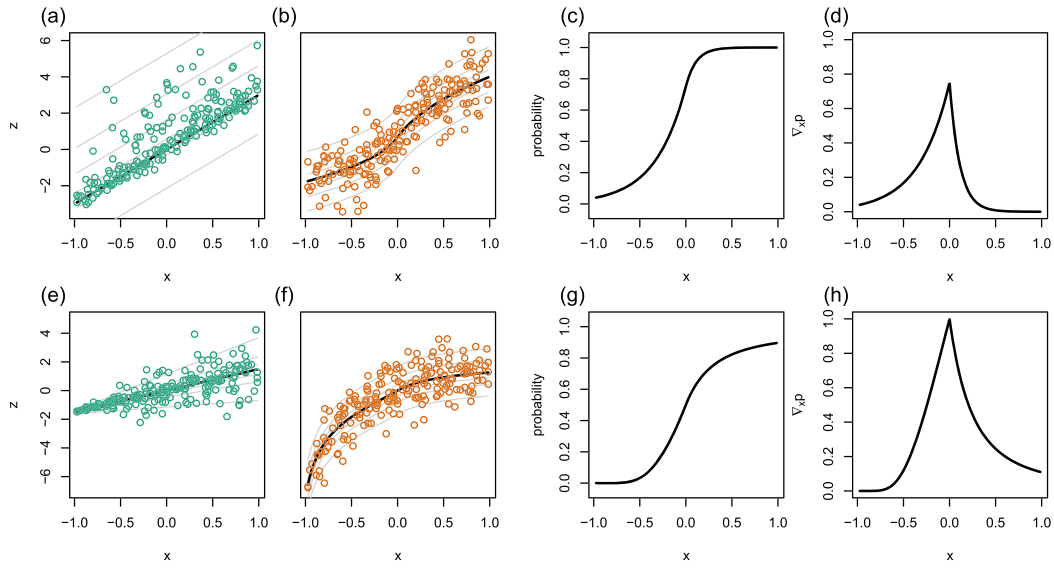

Figure 1. *Plots (a) and (b) show two different auxiliary variable distributions that result in the same left-skewed probability function (c) and probability gradient (d). Plot (a) shows auxiliary variables that vary linearly with the covariate, x, and arise from an asymmetric Laplace distribution. Points represent realisations of the auxiliary variables, and lines represent the mode and several quantiles of the auxiliary variables as a function of the covariate. Plot (b) shows normally distributed auxiliary variables that have a non-linear relationship with the covariate. Plot (e) shows an example of heteroskedastic, normally distributed auxiliary variables with means that vary linearly with the covariate. Plot (f) shows an equivalent representation of the right-skewed probability function (g) and gradient (h) using normally distributed, homoskedastic auxiliary variables with means that are a non-linear function of the covariate*

It is possible, in the pursuit of model flexibility, to take a non-parametric view of the conditional auxiliary variable distribution and allow $z_i|\boldsymbol{\beta}$ to arise from any arbitrary location family of probability distributions with location $\mathbf{x}_i'\boldsymbol{\beta}$. This permits the same level of flexibility as the non-parametric mean approach of Choudhuri *et al*. (2007). Care must be taken when introducing flexibility through the auxiliary variables so that the parameters $\boldsymbol{\theta}$ remain identifiable. For example, as we discuss in Section 2.2, quantile regression approaches for binary data, which attempt to make minimal assumptions about the auxiliary variable conditional distribution (e.g. Manski, 1985; Benoit & Van den Poel, 2012; Padellini & Rue, 2019), have model parameters that are only identifiable up to a multiplicative constant.

2.1 Issues with Parameter Identifiability

Non-identifiability can sometimes be difficult to anticipate in hierarchically specified models. Indeed, non-identifiability may have sometimes gone undetected in the literature, potentially leading to unsubstantiated scientific conclusions.

Specifying a model for binary data in a hierarchical form can provide intuition about the model and aid in comparing closely related models. However, it is important to note that many seemingly distinct statistical models based on auxiliary variables are equivalent in that they yield identical probabilities for the observed binary random variables and are therefore not identifiable from the data. For example, the top row of Figure 1 shows two different auxiliary variable specifications that satisfy (4) and therefore produce equivalent conditional probabilities at the data level. In this example involving a single covariate, the auxiliary variables in Figure 1(a) are asymmetrically distributed around a linearly varying trend, while the auxiliary variables in Figure 1(b) are symmetrically distributed but have a non-linear relationship with the covariate.

Figure 1(c) and 1(d) shows the resulting equivalent probability function and probability gradient. The bottom row of Figure 1 shows another pair of auxiliary variable specifications that result in the same right-skewed probability function, with plots arranged analogously. Figure 1(e) shows an example of auxiliary variables that are normally distributed and vary linearly and heteroskedastically with the covariate. Figure 1(f) shows normally distributed, homoskedastic auxiliary variables that have a non-linear relationship with the covariate.

All four model specifications shown in Figure 1 can also be represented using the three components of the GLM framework. In the top row, Figure 1(a) corresponds to a binary regression model with a linear systematic component and asymmetric inverse link function defined by the CDF of an asymmetric Laplace distribution. Figure 1(b) corresponds to probit regression with a non-linear systematic component. In the bottom row, Figure 1(e) corresponds to a linear systematic component with an asymmetric inverse link function given by the CDF of a Gaussian distribution with standard deviation equal to a linear function of the covariate (i.e. $g^{-1}(\eta(x)) = \Phi\left(\frac{\eta(x)}{\gamma_0 + \gamma_1 x})\right)$. Figure 1(f) is another instance of a probit link paired with a non-linear systematic component.

The important conclusion conveyed by these pairs of equivalent models is that asymmetry in a probability function can arise for a variety of reasons including asymmetry, heteroskedasticity or non-linearity (or some combination thereof) in the auxiliary variables. Moreover, certain characteristics of the auxiliary variables such as asymmetry and heteroskedasticity cannot be identified from the data without making strong assumptions about their relationship with the covariates.

The link function and systematic component do not operate independently on the probability function. Indeed, provided sufficient flexibility is permitted for the systematic component, $\eta(\mathbf{x})$, the particular choice of the link function has no impact on the flexibility of the resulting

probability map p. Thus, without scientific knowledge that provides the basis for modelling assumptions, one approach has been to pair a convenient probit link function and a non-parametric systematic component (e.g. Choudhuri *et al*., 2007).

The examples in Figure 1 illustrate that relaxing assumptions about the conditional distribution of the auxiliary variables can be equivalent to relaxing assumptions of linearity. As the data are unable to help the researcher choose between the models associated with the first and second columns, the only reason to prefer one equivalent model over another would be to accommodate existing scientific knowledge about the data-generating process. The example of ecological presence/absence data mentioned in Section 1.2 presents one such case where acknowledging that the binary data arise because of thresholded observations of a discrete count process motivates the choice of a cloglog link (Poisson auxiliary variable) over a logistic one.

The univariate examples in Figure 1 reveal the interplay between different model components, but they are a notable simplification compared with most real applications. As the number of predictors in the systematic component grows, highly flexible functional spaces become increasingly impractical to implement, and constraints must be introduced that may make it possible to identify both the systematic component and the conditional distribution of auxiliary variables or link function. Nevertheless, even though complete non-identifiability may not be a practical concern, these simple univariate examples show that it will always be important to consider the impacts of assumptions about each model component holistically.

2.2 Quantile Regression for Binary Data

One approach for developing a flexible model for binary data uses quantile regression at the level of the auxiliary variable in a hierarchical model specification (Benoit & Van den Poel, 2012). When used in a continuous-response setting, quantile regression relaxes distributional assumptions about the way the response varies around the linear trend. It is assumed that each quantile of the response distribution varies linearly with the covariates, but the linear effect of each covariate is permitted to vary across quantiles. The approach represents an effective way to model data exhibiting heteroskedastic and/or non-Gaussian residuals (Koenker & Bassett, 1978; Koenker, 2005). An example of a heteroskedastic conditional random variable is shown in the bottom left plot of Figure 1. The grey lines in Figure 1 show quantile curves for a selection of quantile levels. The parameters of interest in binary quantile regression are the slopes of these grey lines.

A generalised form of quantile regression has been proposed as a model for binary data by assuming that the responses, y_i, are generated from the auxiliary variable model in (3) (Benoit & Van den Poel, 2012). The quantiles of the auxiliary variables, z_i, are assumed to be globally linear, such that $Q(z_i|\boldsymbol{\beta}, \tau) = \mathbf{x}_i'\boldsymbol{\beta}(\tau)$ for all $\tau \in (0, 1)$, where $Q(z_i|\tau)$ denotes the τ-th quantile of z_i (Figure 1a and 1e). Analogous extensions have also been proposed for count-valued responses (Machado & Santos Silva, 2005; Lee & Neocleous, 2010). In traditional logistic and probit regression, quantiles are implicitly defined by specifying either a logistic or Gaussian density, respectively, and both cases assume homoskedasticity for the auxiliary variables. Quantile regression for binary data represents an attempt to estimate $\boldsymbol{\beta}(\tau)$ for a specific set of τ, rather than define quantile functions through the specification of the density, $f_Z(z_i|\boldsymbol{\beta})$. A package for the R statistical programming language called `bayesQR` (Benoit & Van den Poel, 2017) was recently developed that aims to provide practitioners with a tool to fit quantile regression models to data, including binary-valued data.

Quantile regression coefficients for a fixed quantile, τ, are identifiable only up to a multiplicative constant when the data are binary (Manski, 1985). However, the meaning of model parameters in binary quantile regression can easily be misinterpreted and may not provide

useful scientific learning even when interpreted correctly. We present a geometric perspective to quantile regression that permits intuitive visualisations for the simplified cases of one or two covariates.

2.2.1 *Parameter identifiability: single predictor*

Consider the probability function for a hypothetical binary response that depends on an intercept and a single covariate, x, shown in Figure 2(c), such that $\boldsymbol{\beta}(\tau) = (\beta_0(\tau), \beta_1(\tau))'$. A selection of nine quantiles are depicted as points along the curve. By definition, the value of x at which the probability function achieves a particular quantile, τ, corresponds to the value of x for which $Q(z|x, \beta_0, \beta_1, 1 - \tau) = 0$. That is, the proportion of the probability mass of the conditional random variable $z|x$ that occurs in the region $z > 0$ is τ. Thus, a particular probability function evaluated at a set of quantiles corresponds to an equal-sized set of points in the xz-plane at the locations where the quantile curves $Q(z|x, \beta_0, \beta_1, 1 - \tau)$ intersect the x-axis. Figure 2(a) and 2(b) shows two different possible distributions for $z|x$ that both have quantile functions intersecting the x-axis at the same locations, and both satisfy global linearity. Thus, the quantile regression parameters are not identifiable from any amount of binary data.

The auxiliary variables in Figure 2(a) exhibit heterogeneity in the covariate effects across quantiles, while the auxiliary variables in Figure 2(b) do not. Hence, even under the global linearity assumption, it is not possible to identify the parameters $\boldsymbol{\beta}(\tau)$ from a binary response. It is not even possible to determine whether, under a global linearity assumption, the auxiliary variables exhibit heteroskedasticity of any kind. Geometrically, this is equivalent to stating that one cannot estimate the slope of the lines in the xz-plane; one can only estimate where they intersect the x-axis. For each fixed value of τ, the non-identifiability of the slope is exactly the limitation pointed out in Manski (1985). The set of functions defined by $z = k\beta_0(\tau) + k\beta_1(\tau)x$ for $k \in \mathbb{R}$ coincides with the set of lines that pass through a common x-intercept, $\dfrac{-\beta_0(\tau)}{\beta_1(\tau)}$.

The estimation procedure in `bayesQR` appears to provide stable estimates of $\boldsymbol{\beta}(\tau)$ because it provides approximate inference about quantile-level effects by fitting a suite of sub-models to the data. In each sub-model, the auxiliary variables are assumed to arise from an asymmetric Laplace distribution with unit variance and known skewness parameter corresponding to one particular quantile of interest. Although all sub-models are likely misspecified, the procedure leads to valid posterior distributions for the covariate effects for continuous-response data in the sense that, for fixed $\boldsymbol{\beta}(\tau)$, posterior distributions constructed using this procedure converge

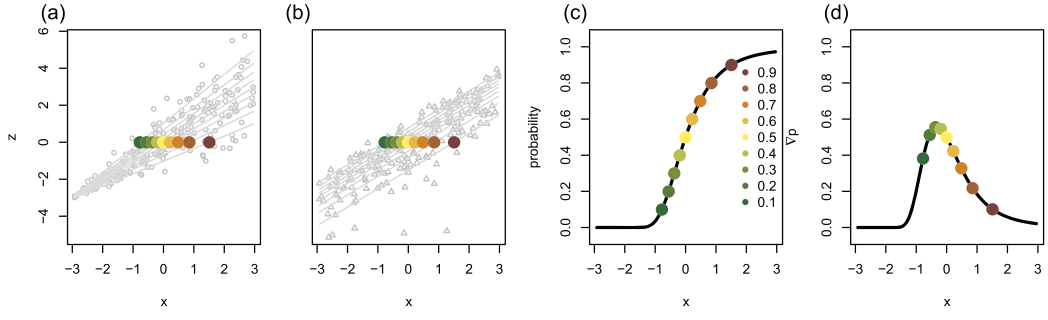

Figure 2. *Projection of quantiles onto intercepts of quantile functions for z|x. Plots (a) and (b) show realisations from two different auxiliary variable distributions that give rise to equivalent probability functions (c) and probability function gradients (d)*

asymptotically to point masses at the fixed levels (Sriram *et al*., 2013; Yang *et al*., 2016). However, posterior validity has not been established for the case of binary data.

By introducing constraints on the vector $\boldsymbol{\beta}(\tau)$, such as $||\boldsymbol{\beta}(\tau)||_2^2 = 1$ as suggested by Manski (1985), it is possible to identify certain functions of the quantile regression coefficients. The second plot in Figure 2 corresponds to a constraint that $\beta_1(\tau) = 1$ for all τ. After a constraint is enforced to ensure identifiability, it is possible to use quantile regression methods to make inference about the probability function (Kordas, 2006), although Bayesian methods that provide valid estimates of uncertainty have not yet been developed.

2.2.2 Parameter identifiability: multiple predictors

Typical regression analyses consider several covariates of interest. Figure 3 shows how quantile surfaces for an auxiliary variable exhibiting global linearity with respect to two covariates (a) results in a particular probability surface (b) and corresponding gradient surface (c). The locations where the quantile functions intersect the two covariate axes are uniquely determined by the probability function, as they were for the univariate case (Figure 2). However, the slopes of the planes in the left plot are not fully identified without further constraints.

For the case of multiple predictors, a potentially useful parameter identifiable from the data is the relative effect of factor j given by $\beta_j^*(\tau) = \beta_j(\tau)/|\boldsymbol{\beta}_{-0}(\tau)|$, where $|\boldsymbol{\beta}_{-0}(\tau)|$ is a suitable vector norm of all effects besides the intercept. This quantity can be interpreted as the contribution of the j-th covariate on the τ-th quantile of the auxiliary variable relative to the total contributions of all covariates together. Thus, $\beta_j^*(\tau_1) < \beta_j^*(\tau_2)$ means that the relative contribution of the j-th factor is greater for quantile τ_2 than for quantile τ_1.

However, inequality between normalised effects tells us nothing about the change in the magnitude of an effect across different quantiles; it may be that $\beta_j(\tau_1) > \beta_j(\tau_2)$. Commonly produced 'forest plots' that show how the value of $\beta_j(\tau)$ varies with τ are nonsensical in the context of a binary response because they present non-identifiable quantities. One could, in principle, examine $\beta_j^*(\tau)$ for a range of quantiles, τ. However, the scientific learning afforded by measuring the relative contribution of a particular linear effect across quantiles is unclear. Lacking a coherent interpretive motivation, quantile regression for binary data reduces to another method like non-parametric probit regression that permits a high degree of flexibility in the shape of the probability surface. Other methods exist for estimating probability functions that offer

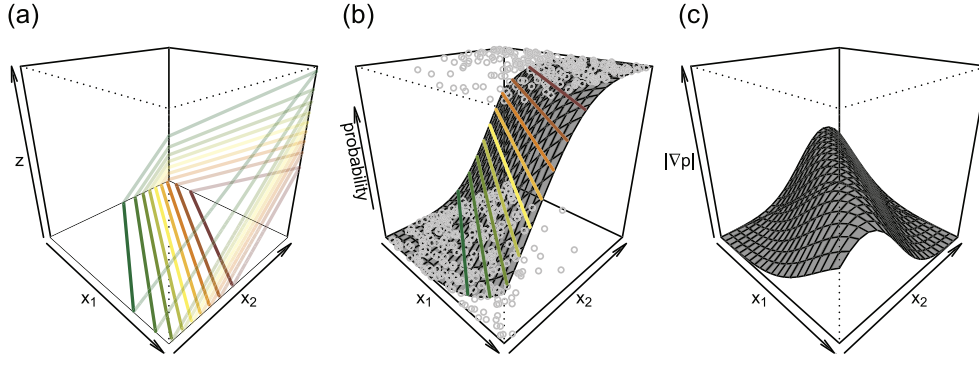

Figure 3. *Projection of quantiles onto intercepts of quantile functions for z. Plot (a) shows an example of auxiliary variables exhibiting global linearity that give rise to the probability surface in (b) and probability surface gradient, the magnitude of which is shown in (c)*

equivalent levels of flexibility yet do not rely on approximate inferential procedures (e.g. Wood, 2017). The necessity of quantile regression methods for binary data therefore remains to be established.

3 Modifying the Random Component

A third way to introduce flexibility in the probability surface is to relax assumptions about the third component of a traditional GLM: the random component. For binary data, the marginal distribution for the observations, y_i, is Bernoulli. However, the tacit assumption made in traditional GLMs that the distributions of y_i are conditionally independent can be relaxed through the introduction of random effects in the systematic component. For example, a simple extension of the probit regression model might allow for a random intercept in the systematic component, indexed by some underlying group structure, written as $y_i|\boldsymbol{\beta}, \zeta_i \sim \text{Bern}(y_i|p_i(\mathbf{x}_i, \boldsymbol{\beta}, \zeta_i))$; $\Phi^{-1}(p_i) = \mathbf{x}_i'\boldsymbol{\beta} + \zeta_i$; $\zeta_i \sim N(\zeta_i|m(i), \sigma_\zeta^2)$, where $m(i)$ is the intercept associated with the group to which datum i belongs, and the intercept in the systematic component, β_0, is fixed at 0 to ensure parameter identifiability. Researchers are often uninterested in ζ_i. Including the random intercept is performed to ensure that any group-level effects that may be confounded with the covariates of interest are accounted for so that valid uncertainty estimates are produced for $\boldsymbol{\beta}$ (see Larsen *et al*., 2000, for a discussion of parameter interpretation in the analogous logistic model with random effects). A hierarchical formulation of this so-called generalised linear mixed model (GLMM) is given by $y_i|z_i = 1_{z_i > 0}$; $z_i|\boldsymbol{\beta}, \zeta_i \sim N(z_i|\mathbf{x}_i'\boldsymbol{\beta} + \zeta_i, 1)$; $\zeta_i \sim N(\zeta_i|m(i), \sigma_\zeta^2)$.

Conditioned on the group random effect, $\boldsymbol{\zeta} = (\zeta_1, \ldots, \zeta_n)'$, the data are independent as before, but integrating out the random effects induces positive dependence between random variables y_i and y_j when i and j belong to the same group (i.e. $m(i) = m(j)$). Thus, the random component has been modified in its joint structure, although the marginal distribution remains Bernoulli.

Random effects associated with other indices such as spatio-temporal information induce dependence for proximate responses. In Section 4, we demonstrate how a spatial random effect may be used to account for residual spatial structure in the occupancy pattern of European red squirrels (*S. vulgaris*) in Switzerland. In the example involving random intercepts, identifiability is achievable if the number of groups is much less than the number of observations. For spatial random effects, an analogous requirement is that the range of spatial dependence not be too short relative to the distances between spatial locations.

4 Application to Occupancy Status of Red Squirrels

We demonstrate how the concepts presented earlier on binary regression modelling may be used in practice with an application to the distribution of the European red squirrel in Switzerland. The data in this example are binary responses indicating whether any red squirrels were observed during visit j to site i, and our goal was to develop a model for the occurrence of this species that allows us to infer the relationship between occurrence and relevant landscape covariates, as well as predict the occupancy status of new sites. The data were collected as part of the Swiss breeding bird survey (Monitoring Häufige Brutvögel, MHB Schmid *et al*., 2004) carried out by the Swiss Ornithological Institute. We developed two new models for occupancy that account for imperfect detection of the species and are parameterised by interpretable quantities related to species abundance. We compared the predictive performance of our proposed models with a baseline model that assumes constant occurrence and detection rates across sites

and visits, in addition to a reanalysis based on the approach taken in Kéry & Royle (2015). All model fitting was performed using the package NIMBLE (de Valpine *et al*., 2017) for the R statistical programming environment.

The 2007 data set is composed of detections/non-detections (1/0) of the red squirrel in two hundred sixty-five 1-km^2 survey quadrats in Switzerland. Each site was visited on two or three separate occasions for a total of 747 observations, 194 of which were detections. Landscape-level predictors hypothesised to covary with occurrence probability ($\mathscr{X}$) are elevation and per cent forest cover (Figure 1 of Supporting Information A). Predictors measured on each visit and hypothesised to covary with the detectability of red squirrels ($\mathscr{W}$) are the date and duration of the observation procedure, which ranged from 13 April to 27 July and 1.5 to 9.5 h, respectively.

Setting aside the issue of detectability, in the presence/absence setting, a logistically or normally distributed auxiliary variable may not correspond to any observable natural phenomenon. However, as suggested in Section 1.2, a non-negative, integer-valued auxiliary variable could be interpreted as the true number of individuals present at a given site (first proposed in Royle & Nichols, 2003). If we let z_i denote such a random variable for site i, then the probability of observing an individual under a scenario of perfect detection would be $\Pr(z_i > 0)$. In two of the models we consider in this section, z_i is modelled as a Poisson-distributed random variable with $\mathbb{E}z_i|\boldsymbol{\beta} = \lambda(\mathbf{x}_i) = e^{\mathbf{x}'_i\boldsymbol{\beta}}$.

To address the reality of imperfect detection, the probability of observing the study species at site i on occasion j is typically decomposed into the probability that the site is truly occupied, often denoted ψ_i, and the conditional probability that a species is detected, given the site is occupied, which we denote by r_{ij}. Allowing for imperfect detection modifies the model for the data such that $Pr(y_{ij} = 1|\psi_i, r_{ij}) = \psi_i r_{ij}$. A hierarchical formulation for such an imperfect detection is given by $y_{ij}|z_i, u_{ij} = 1_{z_i > 0}1_{u_{ij} > 0}$, $z_i|\boldsymbol{\beta} \sim f_Z(z_i|\mathbf{x}_i, \boldsymbol{\beta})$, and $u_{ij}|\boldsymbol{\alpha} \sim f_U(u_{ij}|\mathbf{w}_{ij}, \boldsymbol{\alpha})$, where u_{ij} is a latent variable controlling the detection process. The occupancy probability ψ_i is implicitly defined as the probability that z_i is positive, and the conditional probability of detection, r_{ij}, is the probability that u_{ij} is positive. A large body of literature has focused specifically on the development of statistical methods for occupancy data (e.g. MacKenzie *et al*., 2002, 2018), with emphasis on capturing temporal dynamics (e.g. Buckland & Elston, 1993; MacKenzie *et al*., 2003; Royle & Kéry, 2007) and residual spatial dependence (e.g. Buckland & Elston, 1993; Hooten *et al*., 2003; Johnson *et al*., 2013).

There is clear motivation for specifying a discrete, non-negative distribution for z_i such as the Poisson distribution. However, a natural motivation for the specification of f_U is more elusive. In the absence of a naturally arising distribution for f_U, we specified a standard normal distribution with mean $\mathbf{w}'_{ij}\boldsymbol{\alpha}$, where $\mathbf{w}_{ij}$ is a vector of covariates related to detectability at site i on visit j (as in Dorazio & Rodríguez, 2012).

We considered a total of four model formulations outlined in the succeeding text. Prior specifications and implementation details, including code to reproduce all analyses, are available in Supporting Information A and B. The most basic ('naive') model we considered assumes that occupancy is constant across all visits and the probabilities of occupancy and detection (given a site is occupied) are constant in space such that $y_{ij} \sim \text{Bern}(y_{ij}|rz_i)$, $z_i \sim \text{Bern}(\psi)$, $r \sim \text{Beta}(r|a_r, b_r)$, and $\psi \sim \text{Beta}(\psi|a_\psi, b_\psi)$. A second model extends traditional logistic regression for the case of occupancy data with imperfect detection as described by Kéry & Royle (2015): $y_{ij}|z_i, u_{ij} = 1_{z_i > 0}1_{u_{ij} > 0}$, $z_i|\boldsymbol{\beta} \sim \text{Logistic}(z_i|\mathbf{x}'_i\boldsymbol{\beta}, 1)$, and $u_{ij}|\boldsymbol{\alpha} \sim \text{Logistic}(u_{ij}|\mathbf{w}'_{ij}\boldsymbol{\alpha}, 1)$. A third model assumes Poisson-distributed auxiliary variables related to occupancy, and normally distributed auxiliary variables for detection: $z_i|\boldsymbol{\beta} \sim \text{Pois}(z_i|\lambda(\mathbf{x}_i) = e^{\mathbf{x}_i\boldsymbol{\beta}})$, and $u_{ij}|\boldsymbol{\alpha} \sim N(u_{ij}|\mathbf{w}'_{ij}\boldsymbol{\alpha}, 1)$. Finally, a fourth model extends the

Poisson model by introducing a spatial random effect: $z_i|\boldsymbol{\beta}, \zeta_i \sim \text{Pois}(z_i|\lambda(\mathbf{x}_i) = e^{\mathbf{x}_i'\boldsymbol{\beta} + \zeta_i})$. The spatial random effect, $\boldsymbol{\zeta} = (\zeta_1, \ldots \zeta_n)'$, is modelled as a Gaussian random vector with conditionally Markovian dependence structure. For the purposes of illustration, we considered Gaussian Markov random fields with a known neighbourhood structure (Rue & Held, 2005).

We specified an intrinsic conditional autoregressive distribution for $\boldsymbol{\zeta}$ (Besag & Kooperberg, 1995; Ver Hoef *et al*., 2018), such that the precision matrix is defined as $\mathbf{Q} = \tau^{-2}(\mathbf{D} - \mathbf{A})$, where $\mathbf{A}$ is a binary adjacency matrix representing a known neighbourhood structure and $\mathbf{D}$ is diagonal with entries equal to the sum of the neighbours for each site (i.e. the row sums of $\mathbf{A}$). We defined $\mathbf{A}$ as the adjacency matrix of sites in the unique Delaunay triangulation of the observation locations. The spatial random effect could be introduced at one of several locations in the hierarchical structure, but the most natural place is at the same level as the occupancy covariates (Johnson *et al*., 2013; Schmidt *et al*., 2015). Introducing the random effect in this way is convenient because ζ_i has support given by the entire real line, and e^{ζ_i} may be interpreted as the extra multiplicative effect on the expected abundance at site i due to unobserved covariates.

For each of the Kèry–Royle (KR), Poisson (P) and spatial Poisson (SP) models, we considered two possible collections of covariates for both $\mathscr{X}$ and $\mathscr{W}$, shown in Table 1. The first collection (L) includes multiple first-order effects of landscape covariates on occupancy (z_i) and detection (u_{ij}). The second collection (Q) was selected based on the analysis of Kéry & Royle (2015) and includes all the first-order effects as well as multiple quadratic effects and interactions. Matching the proposed model structures with the two collections of covariates results in seven predictive models for the occurrence of red squirrels: the naive model (N), two Kéry–Royle models (KRL and KRQ), two Poisson models (PL and PQ) and two spatial Poisson models (SPL and SPQ).

Before discussing the results of each proposed model/covariate combination, we briefly summarise what practical and scientific considerations might be useful to select from among these possibilities, setting aside the naive model as an overly simplistic approach used only as a baseline against which we compare the performance of more plausible models. The KR approach represents the most traditional approach for analysing data of this type. Points in favour of the KR approach are the existence of vetted model fitting software and the ability to include non-linear effects of covariates on both occupancy and detection probabilities, yielding flexible probability functions for the data. The primary limitations of the KR models are that neither the auxiliary variables nor the effects parameters, $\boldsymbol{\beta}$, are directly interpretable. If the primary scientific goals include only prediction and estimation of the probability functions for occupancy and detection, then the KR approach represents a valuable and appropriate statistical tool.

Table 1. *Collections of covariates.*

Occupancy ($\mathscr{X}$)			*Detection ($\mathscr{W}$)*		
Covariate	*L*	*Q*	*Covariate*	*L*	*Q*
Intercept	✓	✓	Intercept	✓	✓
Elevation	✓	✓	Date	✓	✓
Forest	✓	✓	Duration	✓	✓
Elevation × Forest		✓	Duration^2		✓
Elevation^2		✓			
Forest^2		✓			
Elevation^2 × Forest		✓			
Elevation × Forest^2		✓			

The Poisson-based approaches provide the same level of flexibility for the occupancy and detection probability functions, with the additional benefit of interpretable auxiliary variables. Using one of these approaches, one can obtain an estimate with uncertainty of the abundance of red squirrels at any of the visited sites or predict abundance at a new site. Such abundance estimates can depend strongly on the particular non-negative discrete-valued distribution specified for $z_i|\boldsymbol{\beta}$, (*sensu* Barker *et al*., 2018), but nevertheless provide preliminary estimates useful, for example, in developing future study designs. The spatial Poisson approach provides even greater flexibility for the probability functions of occupancy and detection and allows for the possibility that important landscape-level covariates may be missing from the analysis.

4.1 Results

Figure 4(a)–4(d) shows the marginal probability curves for occupancy by elevation (a) and per cent forest cover (b) for all seven model–covariate combinations, and the marginal probability curves for detection as a functions of visit date (c) and duration (d). The overall shapes of the curves are consistent across the six non-naive models, with the possible exception of detectability as a function of survey duration for which models incorporating quadratic covariates show a decreasing effect of survey duration on detectability above 6 h, albeit with a considerable increase in uncertainty compared with the linear covariates. Figure 4(e)–4(h) shows the marginal probability gradients, arranged analogously to Figure 4(a)–4(d).

The Poisson-based approaches allow for abundance estimates to be made at both the surveyed sites and new locations. Figure 5 shows the posterior median of the abundance intensity, λ, across Switzerland at a resolution of 1 km^2. A notable difference between the linear and quadratic covariate combinations is smoothness of the predicted abundance surface.

The spatial random effect in the SPL and SPQ models provides insight into where the measured covariates alone may not be able to adequately explain variability in the data. Figure 6

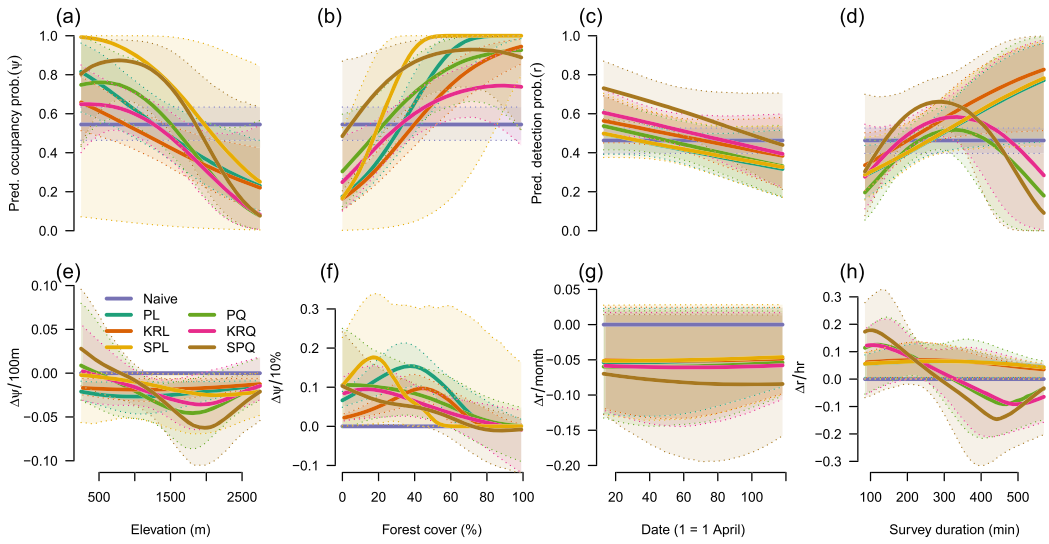

Figure 4. *Top row: marginal probability curves for occupancy as a function of elevation (a) and per cent forest cover (b), and conditional probability of detection curves as a function of survey date (c) and survey duration (d). Solid lines represent pointwise posterior medians from each model, and dotted lines give pointwise equal-tailed 95% credible intervals. Bottom row (e)-(h): derivatives of marginal probability curves in top row*

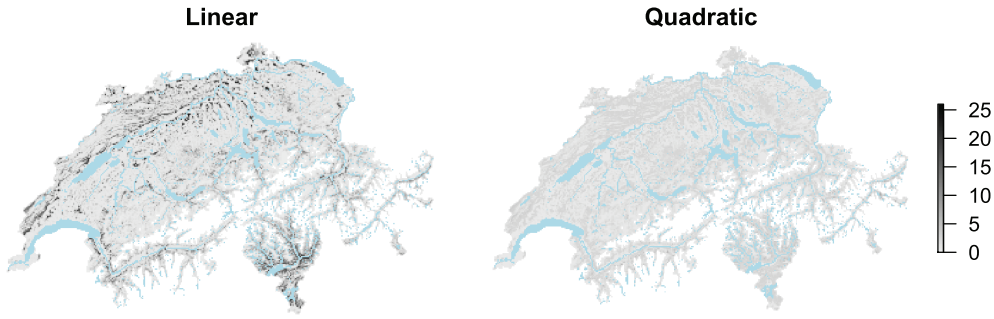

Figure 5. *Posterior median of abundance intensity, λ, for the PL and PQ models. Light blue represents major bodies of water*

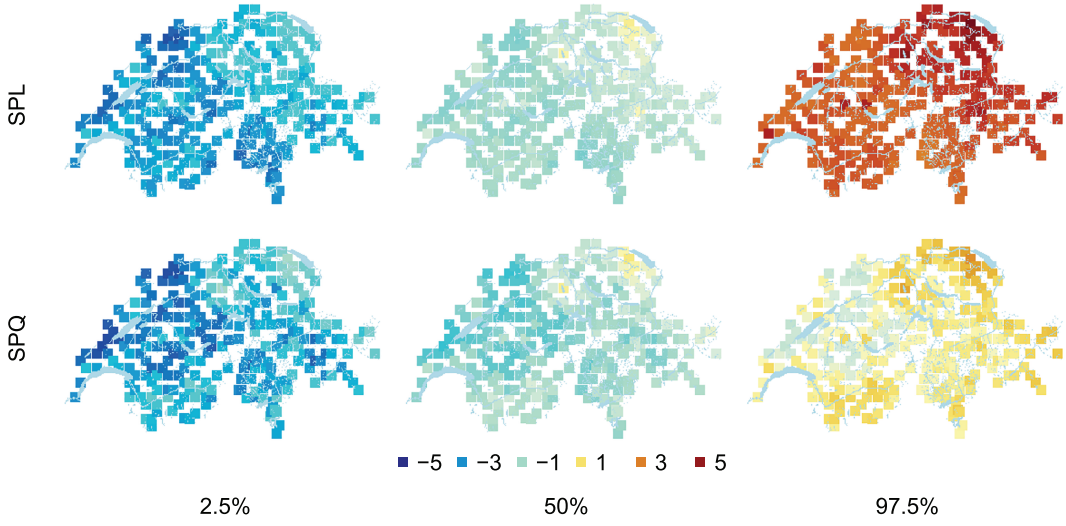

Figure 6. *Posterior summary of spatial random effect, ζ, for the SPL and SPQ models*

shows three quantiles (0.025, 0.5 and 0.975) summarising the posterior distribution of the spatial random effect, ζ, at each surveyed location. The quantiles for SPQ (Figure 6, bottom row) suggest that something about the western part of Switzerland, not captured by the covariates, is associated with a reduction in occupancy.

To investigate goodness of fit, we made partial residual box plots aggregated into a discrete number of bins for each covariate (Figure 4 of Supporting Information A). In general, models using only the linear predictors showed some systematic over-prediction of occupancy for moderately low forest cover.

We investigated the predictive performance of each of the seven models using a *K*-fold cross-validation approach (Section 3 of Supporting Information A). The only clearly inferior model was the naive one. When several models are indistinguishable based on their predictive score, model selection should be performed using all available scientific information relevant to the specific study. We advocate for selecting models whose mechanisms best align with prior knowledge (see also Ver Hoef & Boveng, 2015, for arguments in favour of selecting interpretable models even when predictive scores are worse).

5 Discussion

We showed how new and existing binary regression models can be constructed using auxiliary variables. Hierarchical binary regression models specified using auxiliary variables can facilitate interpretable parameterisations when the auxiliary variables correspond to a real but unobserved natural process that gives rise to the observed binary response. Because binary data can often be understood as summaries of natural processes through thresholding or censoring, interpretable auxiliary variables are often intuitive to define.

A necessary step in the development of any new statistical methodology, establishing parameter identifiability in a proposed binary regression model, is of critical importance and not always straightforward. In the pursuit of flexible models for data that require minimal assumptions, it is possible to inadvertently introduce non-identifiability issues. Approximate model fitting procedures can sometimes mask issues where fully Bayesian inference might be more revealing. Unfortunately, exact inferential procedures are not always available.

Generalisations to traditional binary regression models can be usefully characterised by the ways in which modifications are made to the three components of GLMs (random component, systematic component and link function). Through an application to the occurrence of red squirrels in Switzerland, we demonstrated how to construct several potential hierarchical models for binary data. The occupancy status of an areal unit can be thought of as a thresholded observation of abundance at 0, which motivated the specification of a Poisson-distributed auxiliary variable corresponding to site abundance. This auxiliary variable distribution implied a modification of the traditional symmetric link function to the complementary log–log link function. We also proposed a further generalisation of this latent abundance model by including a spatial random effect, which implied a modification of the traditional assumption of conditional independence in the random component.

One avenue that could yield more reliable quantile regression inference for binary data would be to extend one of the more holistic Bayesian approaches of (Kottas & Krnjajić, 2009; Taddy & Kottas, 2010; Reich *et al*., 2011; Tokdar & Kadaney, 2012; Yang & Tokdar, 2017) to the case of binary data. Although many of the same identifiability issues that appear in existing quantile regression approaches for binary data would remain, these alternative methods do not rely on approximate inferential procedures. Fully Bayesian inference would aid in the detection of identifiability issues and obviate the need for theoretical guarantees about approximation error.

ACKNOWLEDGEMENTS

Any use of trade, firm or product names is for descriptive purposes only and does not imply endorsement by the US government.

References

Agresti, A. 2002. *Categorical Data Analysis*, 2. John Wiley & Sons: Hoboken, NJ.

Albert, J.H. & Chib, S. 1993. Bayesian analysis of binary and polychotomous response data. *J. Am. Stat. Assoc.*, **88** (422), 669–679.

Augustin, N.H., Mugglestone, M.A. & Buckland, S.T. 1996. An autologistic model for the spatial distribution of wildlife. *J. Appl. Ecol.*, **33**(2), 339–347.

Barker, R.J., Schofield, M.R., Link, W.A. & Sauer, J.R. 2018. On the reliability of N-mixture models for count data. *Biometrics*, **74**(1), 369–377. https://doi.org/10.1111/biom.12734

Bazán, J.L., Bolfarine, H. & Branco, M.D. 2010. A framework for skew-probit links in binary regression. *Commun. Stat. – Theory Methods*, **39**(4), 678–697.

Benoit, D.F. & Van den Poel, D. 2012. Binary quantile regression: A Bayesian approach based on the asymmetric Laplace distribution. *J. Appl. Econ.*, **27**(7), 1174–1188.

Benoit, D.F. & Van den Poel, D. 2017. bayesQR: A Bayesian approach to quantile regression. *J. Stat. Softw.*, **76**(7), 1–32.
Besag, J. & Kooperberg, C. 1995. On conditional and intrinsic autoregressions. *Biometrika*, **82**(4), 733–746.
Buckland, S.T. & Elston, D.A. 1993. Empirical models for the spatial distribution of wildlife. *J. Appl. Ecol.*, **30**(3), 478–495.
Chen, M.H., Dey, D.K. & Shao, Q.M. 1999. A new skewed link model for dichotomous quantal response data. *J. Am. Stat. Assoc.*, **94**(448), 1172–1186.
Choudhuri, N., Ghosal, S. & Roy, A. 2007. Nonparametric binary regression using a Gaussian process prior. *Stat. Methodol.*, **4**(2), 227–243.
Conn, P.B., Johnson, D.S., Williams, P.J., Melin, S.R. & Hooten, M.B. 2018. A guide to Bayesian model checking for ecologists. *Ecol. Monogr.*, **88**(4), 526–542.
de Valpine, P., Turek, D., Paciorek, C.J., Lang, D.T. & Bodik, R. 2017. Programming with models: Writing statistical algorithms for general model structures with NIMBLE. *J. Comput. Graph. Stat.*, **26**, 403–417, available at arXiv:1505.05093v1
Diggle, P.J., Tawn, A.J. & Moyeed, R.A. 1998. Model-based geostatistics. *J. R. Stat. Soc.: Ser. C (Appl. Stat.)*, **47**(3), 299–350.
Dorazio, R.M. & Rodríguez, D.T. 2012. A Gibbs sampler for Bayesian analysis of site-occupancy data. *Methods Ecol. Evol.*, **3**(6), 1093–1098.
Fahrmeir, L. & Raach, A. 2007. A Bayesian semiparametric latent variable model for mixed responses. *Psychometrika*, **72**(3), 327–346.
Goldenberg, A., Zheng, A.X., Fienberg, S.E. & Airoldi, E.M. 2010. A survey of statistical network models. *Found. Trends® Mach. Learn.*, **2**(2), 129–233.
Hastie, T. & Tibshirani, R. 1986. Generalized additive models. *Stat. Sci.*, **1**(3), 297–310.
Hefley, T.J., Broms, K.M., Brost, B.M., Buderman, F.E., Kay, S.L., Scharf, H.R., Tipton, J.R., Williams, P.J. & Hooten, M.B. 2017. The basis function approach for modeling autocorrelation in ecological data. *Ecology*, **98**(3), 632–646. https://doi.org/10.1002/ecy.1674
Hooten, M.B., Larsen, D.R. & Wikle, C.K. 2003. Predicting the spatial distribution of ground flora on large domains using a hierarchical Bayesian model. *Landsc. Ecol.*, **18**(5), 487–502.
Johnson, D.S., Conn, P.B., Hooten, M.B., Ray, J.C. & Pond, B.A. 2013. Spatial occupancy models for large data sets. *Ecology*, **94**(4), 801–808.
Kéry, M. & Royle, J.A. 2015. Modeling static occurrence and species distributions using site-occupancy models. In *Applied Hierarchical Modeling in Ecology*, Academic Press: San Diego, CA, USA, pp. 551–629.
Khan, S. 2013. Distribution free estimation of heteroskedastic binary response models using probit/logit criterion functions. *J. Econ.*, **172**(1), 168–182. https://doi.org/10.1016/j.jeconom.2012.08.002
Koenker, R. 2005. *Quantile Regression*. Cambridge University Press: New York, NY, USA.
Koenker, R. & Bassett, G.W.Jr. 1978. Regression quantiles. *Econometrica*, **46**(1), 33–50.
Komori, O., Eguchi, S., Ikeda, S., Okamura, H., Ichinokawa, M. & Nakayama, S. 2016. An asymmetric logistic regression model for ecological data. *Methods Ecol. Evol.*, **7**(2), 249–260.
Kordas, G. 2006. Smoothed binary regression quantiles. *J. Appl. Econ.*, **21**(3), 387–407.
Kottas, A. & Krnjajić, M. 2009. Bayesian semiparametric modelling in quantile regression. *Scandinavian J. Stat.*, **36**(2), 297–319.
Larsen, K., Petersen, J.H., Budtz-Jørgensen, E. & Endahl, L. 2000. Interpreting parameters in the logistic regression model with random effects. *Biometrics*, **56**(3), 909–914.
Lee, D. & Neocleous, T. 2010. Bayesian quantile regression for count data with application to environmental epidemiology. *J. R. Stat. Soc.: Ser. C (Appl. Stat.)*, **59**(5), 905–920.
Lemonte, A.J. & Bazán, J.L. 2018. New links for binary regression: An application to coca cultivation in Peru. *TEST*, **27**(3), 597–617.
Maalouf, M. & Trafalis, T.B. 2011. Robust weighted kernel logistic regression in imbalanced and rare events data. *Comput. Stat. Data Anal.*, **55**(1), 168–183. https://doi.org/10.1016/j.csda.2010.06.014
MacKenzie, D.I., Nichols, J.D., Hines, J.E., Knutson, M.G. & Franklin, A.B. 2003. Estimating site occupancy, colonization, and local extinction when a species is detected imperfectly. *Ecology*, **84**(8), 2200–2207.
MacKenzie, D.I., Nichols, J.D., Lachman, G.B., Droege, S., Royle, J.A. & Langtimm, C.A. 2002. Estimating site occupancy rates when detection probabilities are less than one. *Ecology*, **83**(8), 2248–2255.
MacKenzie, D.I., Nichols, J.D., Royle, J.A., Pollock, K.H., Bailey, L.L. & Hines, J.E. 2018. *Occupancy Estimation and Modeling*. Elsevier.
Machado, J.A.F. & Santos Silva, J.M.C. 2005. Quantiles for counts. *J. Am. Stat. Assoc.*, **100**(472), 1226–1237.
Manski, C.F. 1985. Semiparametric analysis of discrete response: Asymptotic properties of the maximum score estimator. *J. Econ.*, **27**(3), 313–333.

McCullagh, P. & Nelder, J.A. 1983. *Generalized Linear Models*. Chapman and Hall: London, UK.
Nelder, J.A. & Wedderburn, R.W.M. 1972. Generalized linear models. *J. R. Stat. Soc.: Ser. A (General)*, **135**(3), 370–384.
Padellini, T. & Rue, H. 2019. Model-aware quantile regression for discrete data. *arXiv*, **1804**(v2), 1–18.
Prentice, R.L. 1976. A generalization of the probit and logit methods for dose response curves. *Biometrics*, **32**(4), 761–768.
Reich, B.J., Fuentes, M. & Dunson, D.B. 2011. Bayesian spatial quantile regression. *J. Am. Stat. Assoc.*, **106**(493), 6–20.
Royle, J.A. & Kéry, M. 2007. A Bayesian state-space formulation of dynamic occupancy models. *Ecology*, **88**(7), 1813–1823.
Royle, J.A. & Nichols, J.D. 2003. Estimating abundance from repeated presence–absence data or point counts. *Ecology*, **84**(3), 777–790.
Rue, H. & Held, L. 2005. Continuous time random walks. In *Gaussian Markov Random Fields: Theory and Application*, Chapman and Hall/CRC, pp. 123–130.
Schmid, H., Zbinden, N. & Keller, V. 2004. *Überwachung der Bestandsentwicklung häufiger Brutvögel in der Schweiz*. Schweizerische Vogelwarte: Sempach.
Schmidt, A.M., Rodríguez, M.A. & Capistrano, E.D. 2015. Population counts along elliptical habitat contours: Hierarchical modelling using Poisson-lognormal mixtures with nonstationary spatial structure. *Ann. Appl. Stat.*, **9**(3), 1372–1393.
Sriram, K., Ramamoorthi, R.V. & Ghosh, P. 2013. Posterior consistency of Bayesian quantile regression based on the misspecied asymmetric Laplace density. *Bayesian Anal.*, **8**(2), 479–504.
Taddy, M.A. & Kottas, A. 2010. A Bayesian nonparametric approach to inference for quantile regression. *J. Bus. Econ. Stat.*, **28**(3), 357–369.
Tokdar, S.T. & Kadaney, J.B. 2012. Simultaneous linear quantile regression: A semiparametric Bayesian approach. *Bayesian Anal.*, **7**(1), 51–72.
Ver Hoef, J.M. & Boveng, P.L. 2015. Iterating on a single model is a viable alternative to multimodel inference. *J. Wildlife Manag.*, **79**(5), 719–729.
Ver Hoef, J.M., Peterson, E.E., Hooten, M.B., Hanks, E.M. & Fortin, M.J. 2018. Spatial autoregressive models for statistical inference from ecological data. *Ecol. Monogr.*, **88**(1), 36–59.
Wang, X. & Dey, D.K. 2010. Generalized extreme value regression for binary response data: An application to B2B electronic payments system adoption. *Ann. Appl. Stat.*, **4**(4), 2000–2023.
Wood, S.N. 2017. *Generalized Additive Models: An Introduction with R*, 2. Chapman and Hall/CRC: Boca Raton, FL, USA.
Wright, W.J., Irvine, K.M. & Higgs, M.D. 2019. Identifying occupancy model inadequacies: Can residuals separately assess detection and presence?. *Ecology*, **0**(0), e02703.
Xing, J.-J. & Qian, X.-Y. 2017. Bayesian expectile regression with asymmetric normal distribution. *Commun. Stat. – Theory Methods*, **46**(9), 4545–4555.
Yang, Y. & Tokdar, S.T. 2017. Joint estimation of quantile planes over arbitrary predictor spaces. *J. Am. Stat. Assoc.*, **112**(519), 1107–1120. https://doi.org/10.1080/01621459.2016.1192545
Yang, Y., Wang, H.J. & He, X. 2016. Posterior inference in Bayesian quantile regression with asymmetric Laplace likelihood. *Int. Stat. Rev.*, **84**(3), 327–344.

[*Received July 2020; accepted December 2021*]