Skewed link-based regression models for misclassified binary data

Article *in* Revista de la Real Academia de Ciencias Exactas, Físicas y Naturales. Serie A, Matemáticas · August 2018

DOI: 10.1007/s13398-018-0571-3

3 authors:

Lizbeth Naranjo
Universidad Nacional Autónoma de México
35 PUBLICATIONS **376** CITATIONS

Carlos J. Pérez
Universidad de Extremadura
125 PUBLICATIONS **1,870** CITATIONS

Jacinto Martín
Universidad de Extremadura
92 PUBLICATIONS **1,929** CITATIONS

Skewed link-based regression models for misclassified binary data

Lizbeth Naranjo · Carlos J. Pérez · Jacinto Martín

Abstract In this paper, we propose flexible Bayesian approaches for binary regression models in the presence of misclassified data. These approaches consider asymmetric links based on the skew-normal and the asymmetric exponential power distributions. The computational difficulties have been avoided by using blue data augmentation schemes. The idea of using data augmentation schemes with two types of latent variables is exploited to derive efficient MCMC algorithms. A simulation study and an application illustrate the model performance in comparison with the standard methods that do not consider skewness and/or which do not consider misclassification.

Keywords Asymmetric link function · Bayesian inference · Binary regression · Data augmentation · Markov chain Monte Carlo method · Misclassification.

1 Introduction

Data-generating processes are sometimes not error-free when data are collected in the real world. In this context, noise parameters are necessary to correct the bias yielded by the use of data that has been measured with error. If the noise in a data generating process is not properly modelled, then the information may be

Lizbeth Naranjo
Departamento de Matemáticas, Facultad de Ciencias, Universidad Nacional Autónoma de México, Avenida de la Universidad 3000, Circuito Exterior s/n, Delegación Coyoacán, C.P. 04510, Ciudad Universitaria, Ciudad de México, Mexico
Tel.: +52 55 5622 3899, Ext.: 45735
E-mail: lizbethna@ciencias.unam.mx

Carlos J. Pérez
Departamento de Matemáticas, Facultad de Veterinaria, Universidad de Extremadura, Avenida de la Universidad s/n, Cáceres 10071, Spain
E-mail: carper@unex.es

Jacinto Martín
Departamento de Matemáticas, Facultad de Ciencias, Universidad de Extremadura, Avenida de Elvas s/n, Badajoz 06006, Spain
E-mail: jrmartin@unex.es

perceived as being more accurate than it actually is. In many cases, this can lead to a non optimal decision making. Therefore, statistical models should address the problem of measurement error (see Gustafson (2003), Carroll et al. (2006) and Buonaccorsi (2010)). For categorical data, this is named misclassification.

Generalised linear models have been used to describe the dependence of data on explanatory variables when the binary outcome is subject to misclassification. For example, Paulino et al. (2003) used Bayesian methods and misclassification probabilities to characterise measurement error. McInturff et al. (2004) presented methods for binomial regression when the outcome is determined using the results of a single diagnostic test with imperfect sensitivity and specificity. Later, Achcar et al. (2004) proposed a Bayesian logistic regression model, concentrating on the sensitivity and specificity of medical tests in the presence of misclassification. Paulino et al. (2005) presented a Bayesian analysis of misclassified binary data under the framework of a logistic regression model with random effects. Finally, Naranjo et al. (2014b) described the dependence of data on explanatory variables when the binary outcome is subject to misclassification, considering both probit and t-link regressions under Bayesian methodology. These authors used a data augmentation framework to derive efficient Gibbs sampling and expectation-maximization, i.e., EM algorithm.

Several link functions have been considered when modelling binary response data. The most popular models are the logistic and probit models. However, in some applications, the overall fit can be improved by using asymmetric links. To describe a link, Chen et al. (1999) considered the rates at which the probabilities of a given binary response approaches 1 or 0. Under this notion, a link is symmetric if the rates are the same, otherwise the link is skew or asymmetric. A skew link can be characterised as positively skew if the rate approaching 1 is faster than the rate approaching 0, otherwise it is negatively skew.

The analysis of asymmetric links for binary response models has been extensive. For example, Aranda-Ordaz (1981) introduced two families of power transformations to model symmetric and asymmetric departures from the logistic model. Stukel (1988) proposed a class of generalised logistic models. Later, Czado (1994) extended the binary regression models, such as logistic or probit regression, to include links to parametric transformation families. These binary regression models with parametric links are designed to avoid possible link miss-specification and to improve the fit of some data sets. By using the latent variable approach that was developed by Albert and Chib (1993), Chen et al. (1999) proposed a class of skew link-based models, where the underlying latent variable has a model structure with mixed effects. Liu (2004) proposed a robust symmetric t-link, which is called a robit link, in which the normal distribution of the probit model is replaced by a t distribution with unknown degrees of freedom. Kim et al. (2008) introduced a link that is based on a generalised t distribution. Later, Bazán et al. (2010) reviewed several asymmetrical links for binary regression models and presented a unified approach for two skew-probit links proposed in the literature. Finally, Naranjo et al. (2015) analyzed a binary regression model by using the inverse of the asymmetric exponential power cumulative distribution function (cdf) as the link function.

In this paper, asymmetric link-based regression models for misclassified binary data are developed under the Bayesian methodology. The skew-normal (SN) and the asymmetric exponential power (AEP) distributions are considered. The proposed approaches are extensions, which address misclassifications, of the error-free regression models proposed by Chen et al. (1999) and Naranjo et al. (2015). We integrate the method that is used in Naranjo et al. (2014b) to address misclassifications (for both SN and AEP) and the idea of using a scale mixture of uniform representation of the AEP distribution in Naranjo et al. (2015) into a new framework. Computational difficulties have been avoided by using data augmentation schemes, which have allowed us to derive efficient Markov chain Monte Carlo (MCMC) algorithms (see Gamerman and Lopes (2006)). Although the augmented models increase the dimensionality, the generation processes are easier and the convergence is improved. A simulation shows the potential of the proposed approaches. To the best of the authors' knowledge, the approaches that we propose here are the first to address misclassification at the same time that flexible asymmetric link functions are considered for the involved binary regression models.

The potential applicability of the proposed approaches to many fields of knowledge makes this proposal interesting. For example, Paulino et al. (2005) were motivated by data gathered in a study of human papillomavirus (HPV) infection. Their purpose was to analyze the association of several potential risk factors with HPV cervical infection. However, the available test for HPV infection was limited to one subtype or a group of subtypes, which meant that a certain number of infections were missed. Therefore, the response variable is prone to be affected by misclassification, producing some false negative results and less probable, false positive results due to sample contamination and other reasons associated with laboratory work. Another motivating problem was studied by Rekaya et al. (2001), who analyzed a threshold model for misclassified binary responses with applications to animal production. In this case, the genetic evaluation for fertility is based on a test that is also subject to misclassification. It was found that skewness in the link function could help in misclassified binary data analysis when the success probability is approaching 1 or 0 at a different rate.

The rest of this paper is structured as follows. Section 2 describes how misclassification is addressed in the proposed binary regression models. Section 3 explores the posterior distributions. Section 4 shows the performance of the proposed approaches through a simulation example. Section 5 considers a caries experience data set to illustrate the proposed approaches. Finally, Section 6 presents the conclusion. Some technical details are given in the Appendix.

2 Addressing misclassification in binary regression models

Suppose that we observe n independent binary random variables $Y_1, \ldots, Y_n$, where Y_i is distributed as a Bernoulli with probability of success $\mathrm{p}(Y_i = 1) = \tau_i$. These probabilities are defined as $\tau_i = p_i(1 - \lambda_{10}) + (1 - p_i)\lambda_{01}$, where p_i is the true positive probability for the ith observation, λ_{10} is the false negative probability, and λ_{01} is the false positive probability. The parameters p_i are related to a set of covariates $\mathbf{x}_i = (x_{i1}, \ldots, x_{ik})^T$ through a binary regression model.

We assume a binary response model with $p_i = \Psi(\mathbf{x}_i^T \boldsymbol{\beta})$, where Ψ is a cdf. First, we have considered a skew-normal cdf with location parameter $\mu = 0$, scale parameter $\sigma = 1$ and skew parameter δ, as denoted by $\mathrm{SN}(\mu, \sigma, \delta)$, whose probability density function (pdf) is given by

$$\psi_{SN}(w|\mu,\sigma,\delta) = \frac{2}{\sigma}\phi\left(\frac{w-\mu}{\sigma}\right)\Phi\left(\frac{\delta(w-\mu)}{\sigma}\right).$$

Second, we have considered the asymmetric exponential power distribution denoted by AEP$(\mu,\sigma,\alpha,\theta_1,\theta_2)$, which was studied in Zhu and Zinde-Walsh (2009) and applied to binary regression models in Naranjo et al. (2015). The pdf is given by

$$\psi_{AEP}(w|\mu,\sigma,\alpha,\theta_1,\theta_2) = \begin{cases} \frac{1}{\sigma}\exp\left(-\left|\frac{w-\mu}{\alpha\sigma/\Gamma(1+1/\theta_1)}\right|^{\theta_1}\right) & \text{if } w \leq \mu \\ \frac{1}{\sigma}\exp\left(-\left|\frac{w-\mu}{(1-\alpha)\sigma/\Gamma(1+1/\theta_2)}\right|^{\theta_2}\right) & \text{if } w > \mu \end{cases},$$

where $\mu = 0$ is the location parameter, $\sigma = 1$ is the scale parameter, $\alpha \in (0,1)$ is the skew parameter, and $\theta_1 > 0$ and $\theta_2 > 0$ are the left and right tail parameters (i.e., shape parameters or power parameters controlling kurtosis).

Although the posterior distributions of the parameters are intractable for direct generation, it is possible to generate samples from them by augmenting the models with latent variables. The first type of latent variable to be introduced is related with the misclassification (see Naranjo et al. (2014b)). Binary latent variables c^i_{hk}, $h,k = 0,1$, are defined, where h represents the index for the true value and k represents the index for the observed value. When the latent variable takes a value of one, it denotes the group where the observation i has been assigned: true positive ($c^i_{11} = 1$), false negative ($c^i_{10} = 1$), false positive ($c^i_{01} = 1$), or true negative ($c^i_{00} = 1$). Note that $c^i_{11}+c^i_{01} = 1$ (when $y_i = 1$) or $c^i_{10}+c^i_{00} = 1$ (when $y_i = 0$). The latent vectors $\mathbf{c}^i = (c^i_{11}, c^i_{10}, c^i_{01}, c^i_{00})^T$ and the latent matrix $\mathbf{c} = (\mathbf{c}^1, \mathbf{c}^2, \ldots, \mathbf{c}^n)^T$ are then defined.

The second type of latent variable is defined in a similar way as in Chen et al. (1999) and Naranjo et al. (2015) for the SN and AEP links, respectively. This is based on the idea of data augmentation that was developed by Albert and Chib (1993); that is, n independent latent variables $w_1, \ldots, w_n$ are considered such that $c^i_{11} + c^i_{10} = 1$ if $w_i > 0$ and $c^i_{01} + c^i_{00} = 1$ if $w_i \leq 0$, where, for the SN link,

$$w_i = \mathbf{x}_i^T\beta + \delta z_i + \varepsilon_i, \quad z_i \sim \mathcal{G}, \quad \varepsilon_i \sim \mathcal{F}, \tag{1}$$

and, for the AEP link (see Naranjo et al. (2015)),

$$w_i = \mathbf{x}_i^T\beta + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{H}, \tag{2}$$

where z_i and ε_i are independent, $\delta \in (-\infty,\infty)$ is a skew parameter, $\mathcal{G}$ is the half standard normal distribution defined on $(0,\infty)$, $\mathcal{F}$ is the standard normal distribution, and $\mathcal{H}$ is an AEP distribution.

In Equation (1), when the intercept is known, the model is positively skew if $\delta > 0$ and it is negatively skew if $\delta < 0$. When $\delta = 0$, the skew link-based model becomes the probit model. Equation (2) provides a flexible link function, which allows us to model symmetric/asymmetric and lighter/heavier tails. General symmetric link cases can be obtained with an AEP$(0,1,0.5,\theta,\theta)$ distribution. As symmetric particular cases, the probit model is obtained by using AEP$(0,\sqrt{2\pi},0.5,2,2)$) and the exponential power link by using AEP$(0,2^{\theta/2+1}\Gamma(\theta/2+1),0.5,2/\theta,2/\theta)$ (see Naranjo et al. (2014a)).

Further details of the SN and AEP models are presented in the following subsections.

2.1 Skew-normal link

To define the SN link-based model from (1), the conditional distribution of w_i given z_i is N $\left(\mathbf{x}_i^T\boldsymbol{\beta} + \delta z_i, 1\right)$, and the marginal distribution of w_i is SN $\left(\mathbf{x}_i^T\boldsymbol{\beta}, \sqrt{1+\delta^2}, \delta\right)$.

The skew-normal model defined by Chen et al. (1999) has the limitation that the intercept and the skew parameter δ are confounded with each other. To avoid this problem, it is possible to exclude the intercept from the model or to fix the intercept. Another option is to use a proper informative distribution for δ.

2.2 Asymmetric exponential power link

When the AEP link-based model is assumed, then $p_i = \Psi(\mathbf{x}_i^T\boldsymbol{\beta})$, where Ψ is the cdf of the distribution AEP$(0, 1, \alpha, \theta_1, \theta_2)$. Independent latent variables $w_1, \ldots, w_n$ are introduced, where w_i given $\boldsymbol{\beta}$, α, θ_1 and θ_2 is distributed as AEP$(\mathbf{x}_i^T\boldsymbol{\beta}, 1, \alpha, \theta_1, \theta_2)$. The mixture representation of the AEP defined by Naranjo et al. (2015) is used by including latent variables u_{1i} and u_{2i}, leading to the following expressions

$$\begin{aligned} & w_i|\mathbf{x}, \mathbf{u}_1, \mathbf{u}_2, \boldsymbol{\beta}, \alpha, \theta_1, \theta_2 \\ & \sim \begin{cases} \mathrm{U}\left(\mathbf{x}_i^T\boldsymbol{\beta} - \frac{\alpha}{\Gamma(1+1/\theta_1)} u_{1i}^{1/\theta_1} \ , \ \mathbf{x}_i^T\boldsymbol{\beta}\right) & \text{with probability } \alpha \\ \mathrm{U}\left(\mathbf{x}_i^T\boldsymbol{\beta} \ , \ \mathbf{x}_i^T\boldsymbol{\beta} + \frac{(1-\alpha)}{\Gamma(1+1/\theta_2)} u_{2i}^{1/\theta_2}\right) & \text{with probability } 1-\alpha \end{cases}, \end{aligned}$$

and

$$u_{1i}|\theta_1 \sim \mathrm{Ga}\left(1 + 1/\theta_1, 1\right), \quad u_{2i}|\theta_2 \sim \mathrm{Ga}\left(1 + 1/\theta_2, 1\right).$$

Note that, if $w_i \sim$ AEP$(\mathbf{x}_i^T\boldsymbol{\beta}, 1, \alpha, \theta_1, \theta_2)$, then the following equality holds: if $0 < \mathbf{x}_i^T\boldsymbol{\beta}$, then $\mathrm{p}(w_i < 0) = \mathrm{p}(w_i^* < 0)$, where $w_i^* \sim$ AEP$(\mathbf{x}_i^T\boldsymbol{\beta}^*, 1, 1/2, \theta_1, \theta_2)$ and $\boldsymbol{\beta}^* = \boldsymbol{\beta}/(\alpha/2)$, and if $\mathbf{x}_i^T\boldsymbol{\beta} < 0$, then $\mathrm{p}(w_i > 0) = \mathrm{p}(w_i^* > 0)$, where $w_i^* \sim$ AEP$(\mathbf{x}_i^T\boldsymbol{\beta}^*, 1, 1/2, \theta_1, \theta_2)$ and $\boldsymbol{\beta}^* = \boldsymbol{\beta}/((1-\alpha)/2)$. This means that parameters α and β are not identifiable. Without loss of generality, the intercept parameter can be set as $\beta_1 = 1$. Therefore, the shape of the tail related with the observations equal to zero is modeled with the parameter θ_1, and the shape of the tail related with the observations equal to one is modeled with the parameter θ_2. If the observed data are grouped, then the model can be used by ungrouping the data, and the parameters θ_1 and θ_2 are related with proportions lower than α and greater than α, respectively.

3 Exploring the posterior distributions

The posterior distribution is explored in this section. Given the data $\mathbf{D} = \{\mathbf{x}, \mathbf{y}\}$, the joint posterior distribution of the unobservables $\mathbf{c}$, $\boldsymbol{\beta}$, $\boldsymbol{\lambda}$, and Θ is

$$\begin{aligned} & \pi(\boldsymbol{\beta}, \Theta, \mathbf{c}, \boldsymbol{\lambda}|\mathbf{D}) \qquad (3) \\ & \propto \pi(\boldsymbol{\beta})\pi(\Theta)\pi(\boldsymbol{\lambda}) \\ & \times \prod_{i=1}^{n} \Bigg[\{p_i(1-\lambda_{10})\}^{c_{11}^i} \{p_i\lambda_{10}\}^{c_{10}^i} \{(1-p_i)\lambda_{01}\}^{c_{01}^i} \{(1-p_i)(1-\lambda_{01})\}^{c_{00}^i} \\ & \times \Big(I[y_i = 1] I[c_{11}^i + c_{01}^i = 1] + I[y_i = 0] I[c_{10}^i + c_{00}^i = 1] \Big) \Bigg], \end{aligned}$$

where Θ is the set of parameters for each link, specifically, $\Theta = \{\delta\}$ for the SN link and $\Theta = \{\alpha, \theta_1, \theta_2\}$ for the AEP link.

The prior distribution for the regression parameters vector is a multivariate normal one $\mathrm{N}_k(\mathbf{b}, \mathbf{B})$; that is, $\pi(\boldsymbol{\beta}) \propto \exp\left\{-\frac{1}{2}(\boldsymbol{\beta}-\mathbf{b})^T\mathbf{B}^{-1}(\boldsymbol{\beta}-\mathbf{b})\right\}$. Beta distributions are assumed for the misclassification parameters. This is the natural choice for modelling the uncertainty about probabilities; that is, $\mathrm{Be}(a_{10}, b_{10})$ and $\mathrm{Be}(a_{01}, b_{01})$. Although, non-informative prior distributions can be used in any Bayesian context, in this context, the model performance will improve when informative prior distribution for the misclassification parameters is considered. The prior distribution for the misclassification parameter can be elicited based on historical information, expert opinion and/or validation datasets. This information is useful to correct the bias produced by the misclassified data. Without this information, models that address misclassification may not have an advantage over error-free models. The other prior distributions are: $\delta \sim \mathrm{Normal}(\mu_\delta, \sigma_\delta^2)$, $\alpha \sim \pi(\alpha)$, $\theta_1 \sim \pi(\theta_1)$, and $\theta_2 \sim \pi(\theta_2)$. Care must be taken when using improper prior distributions for the shape parameters of the AEP distribution (see Rubio (2015)).

To apply a Gibbs sampling algorithm for the joint posterior distribution (3), the full conditional distributions must be derived. The full conditional distributions for $\mathbf{c}$ and $\boldsymbol{\lambda}$ are easy to obtain

$$\mathbf{c}^i|\boldsymbol{\beta}, \Theta, \boldsymbol{\lambda}, \mathbf{D} \sim \mathrm{Multinomial}\left(1, \pi_{c^i}\left(c_{11}^i, c_{10}^i, c_{01}^i, c_{00}^i\right)\right), \tag{4}$$

$$\lambda_{10}|\boldsymbol{\beta}, \Theta, \mathbf{c}, \mathbf{D} \sim \mathrm{Be}\left(a_{10} + \sum_{i=1}^n c_{10}^i,\ b_{10} + \sum_{i=1}^n c_{11}^i\right),$$

$$\lambda_{01}|\boldsymbol{\beta}, \Theta, \mathbf{c}, \mathbf{D} \sim \mathrm{Be}\left(a_{01} + \sum_{i=1}^n c_{01}^i,\ b_{01} + \sum_{i=1}^n c_{00}^i\right), \tag{5}$$

where $\tau_i = p_i(1-\lambda_{10}) + (1-p_i)\lambda_{01}$ and

$$\begin{aligned}
\pi_{c^i}(1,0,0,0) &= p_i(1-\lambda_{10})I[y_i=1]/\tau_i,\\
\pi_{c^i}(0,1,0,0) &= p_i\lambda_{10}I[y_i=0]/(1-\tau_i),\\
\pi_{c^i}(0,0,1,0) &= (1-p_i)\lambda_{01}I[y_i=1]/\tau_i,\\
\pi_{c^i}(0,0,0,1) &= (1-p_i)(1-\lambda_{01})I[y_i=0]/(1-\tau_i).
\end{aligned}$$

However, the full conditional distributions $\pi(\boldsymbol{\beta}|\Theta, \mathbf{c}, \boldsymbol{\lambda}, \mathbf{D})$ and $\pi(\Theta|\boldsymbol{\beta}, \mathbf{c}, \boldsymbol{\lambda}, \mathbf{D})$ do not have closed expressions that would allow us to easily generate samples. Although generating samples from these distributions could be addressed by using a Metropolis-Hastings algorithm, a Gibbs-within-Gibbs or Metropolis-Hastings steps-within-Gibbs algorithm is more efficient and easier to implement by considering the introduction of latent variables in $\pi(\boldsymbol{\beta}, \Theta|\mathbf{c}, \boldsymbol{\lambda}, \mathbf{D})$.

The proposed framework is general for both SN and AEP link-based binary regression models that address misclassification. We are able to derive the Gibbs sampling algorithms because of the way that the latent variables have been included in the models allows. Further details of the full conditional distributions of the parameters for SN link are provided in Subappendix A.1 and for AEP in Subappendix A.2.

4 Simulation-based example

To compare several competing models, the total absolute error (TAE) is proposed to measure the discrepancy between the real probabilities and the estimated probabilities. This is defined as $TAE = \sum_{i=1}^{n} |\vartheta_i - \hat{\vartheta}_i|$. However, this is only applicable when the data are simulated or when a gold standard is available and, therefore, the true classification is known. Here ϑ and $\widehat{\vartheta}$ are the generated probability and the fitted probability, respectively, based on the model under consideration. For the error free models $\vartheta = p$ and thus $\widehat{\vartheta} = \hat{p}$. For the models considering misclassification, we consider both fitted probabilities; this means that the true probabilities $\vartheta = p$ and then $\widehat{\vartheta} = \hat{p}$ (TAE), and the probabilities addressing misclassification $\vartheta = \tau$ and so $\widehat{\vartheta} = \hat{\tau}$ (TAE mis).

In this section we use simulation to show that when data are generated from skew link models without misclassification and errors are introduced in the responses, the skew models considering misclassification perform better than both the symmetric link model (considering misclassification) and the standard skew models (error free). This suggests that the proposed models can be used as an alternative for data when the data generation process is unknown and it can possibly be related to misclassified binary regression and skew links. Next, we will present an illustrative example.

Datasets are generated under the SN and AEP models by using two covariate sets, which are generated from $x_{i1} \sim$ U$(0, 2)$, and $x_{i2} \sim$ U$(0, 2)$, for $i = 1, \ldots, n$, where $n = 200$, $\delta = 5$, $\alpha = 0.3$, $\theta_1 = 1$, $\theta_2 = 2$, and the probabilities are obtained for the error free model by $p_i = \Psi(\mathbf{x}_i^T \boldsymbol{\beta})$, where $\boldsymbol{\beta} = (1, -5, 2)^T$ and $\boldsymbol{\beta} = (1, 1, -7)^T$, and Ψ is the cdf of SN$(0, \sqrt{1 + \delta^2}, \delta)$ and AEP$(0, 1, \alpha, \theta_1, \theta_2)$.

The true binary dependent variable y^{true} is randomly generated by using the following process: (i) generate $u_i \sim$ U$(0, 1)$; (ii) if $p_i \geq u_i$, then $y_i^{true} = 1$, else $y_i^{true} = 0$. Then, different misclassification parameters are used, specifically, $(\lambda_{10}, \lambda_{01}) = \{(0.1, 0.1), (0.01, 0.2), (0.2, 0.01), (0.05, 0.15)\}$; that is, the outcomes are randomly misclassified according to the following process: (iii) generate $v_i \sim$ U$(0, 1)$; (iv) if $y_i^{true} = 1$ and $v_i \leq \lambda_{10}$ then $y_i = 0$, but if $y_i^{true} = 0$ and $v_i \leq \lambda_{01}$ then $y_i = 1$, else $y_i = y_i^{true}$. Thus, the new response variable y remains equal to y^{true} for the non-misclassified outcomes. Figure 1 shows two misclassified covariate datasets that have been randomly chosen from a model with SN link having $\delta = 5$ and from a model with AEP link having $\alpha = 0.3$, $\theta_1 = 1$ and $\theta_2 = 2$. In the left-hand plots, the straight lines represent the cut points at 0.5 of the binary regression models. In the right-hand plots, the graphics show the effect of asymmetric link functions, the points are the response variables and the curves represent the corresponding probability of success $p_i = \Psi(\mathbf{x}_i^T \boldsymbol{\beta})$. Note that the rates approaching 1 and 0 are not the same.

The main objective is to compare the performance of the proposed models with the standard models, both without considering misclassification assumption and without using skew links. This simulation-based scenario allows us to compare the predictive outcomes with the real outcomes instead of the observed outcomes and, therefore, to know which model performs better.

We have considered four error free models, two of them with symmetric link functions –the probit model (Normal) defined by Albert and Chib (1993) and the exponential power link model (EP) defined by Naranjo et al. (2014a)– and another two with asymmetric link functions –the skew probit link model (SN) defined by

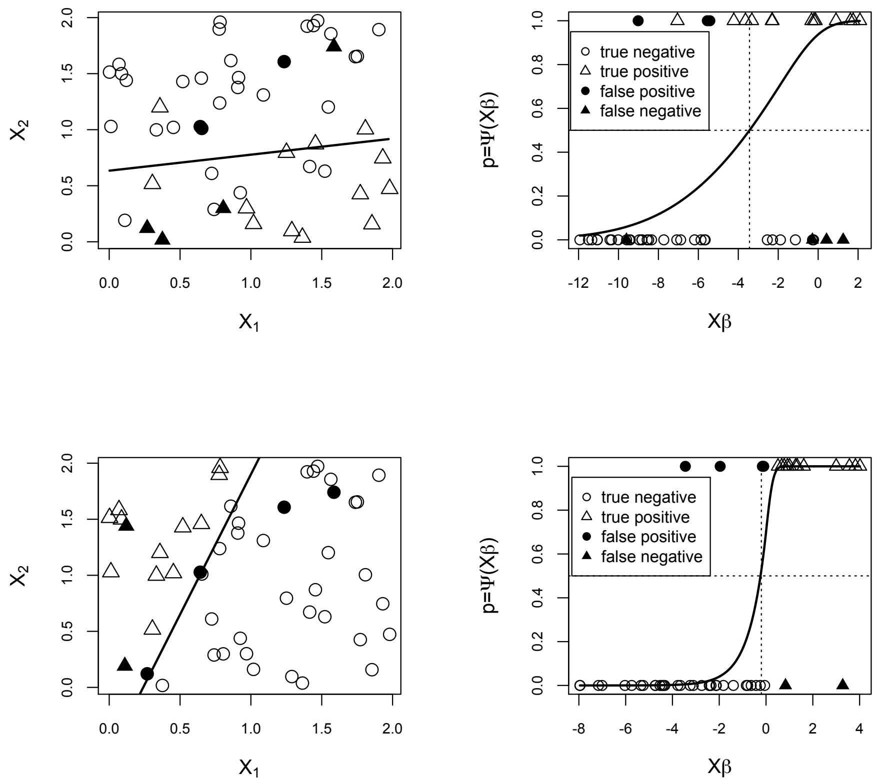

Fig. 1 Datasets with misclassification for SN link with $\delta = 5$ (Top) and for AEP link with $\alpha = 0.3$, $\theta_1 = 1$ and $\theta_2 = 2$ (Bottom). Left plots: the straight lines represents the cut points at 0.5. Right plots: the points are the response variable and the curves represent the corresponding probability of success.

Chen et al. (1999) and the asymmetric exponential power link model (AEP) defined by Naranjo et al. (2015). Moreover, we have applied four models considering misclassification, two of them have symmetric link functions –the probit model (Normal Mis) defined by Naranjo et al. (2014b) and the exponential power link model of Naranjo et al. (2014a) addressing misclassification (EP Mis)– and, finally, the two models with misclassification errors and skew links proposed in this paper –SN Mis and AEP Mis.

Prior distributions for β and λ are given by $\boldsymbol{\beta} \sim \mathrm{N}_k(\mathbf{b}, \mathbf{B})$, where $\mathbf{b}^T = (0, 0, 0)$ and $\mathbf{B} = \mathrm{diag}(100, 100, 100)$, $\lambda_{10} \sim \mathrm{Be}(a_{10}, b_{10})$ and $\lambda_{01} \sim \mathrm{Be}(a_{01}, b_{01})$, where a_{10}, b_{10}, a_{01} and b_{01} are defined such that $\lambda_{10} = \frac{a_{10}}{a_{10}+b_{10}}$, $\lambda_{01} = \frac{a_{01}}{a_{01}+b_{01}}$, $a_{10} + b_{10} = 50$ and $a_{01} + b_{01} = 50$. Besides, for the SN models $\delta \sim \mathrm{Normal}(0, 100)$, and for the AEP models $\alpha \sim \mathrm{U}(0, 1)$, $\theta_1 \sim \mathrm{U}(1, 2)$ and $\theta_2 \sim \mathrm{U}(1, 2)$.

The algorithm has been implemented in `R` software, using a 2.5GHz Intel Core i7 Processor with 16GB 1600 MHz DDR3 RAM. The `BOA` package has been used to analyze the convergence (see Smith (2007)). Specifically, Raftery and Lewis, and Heidelberger and Welch convergence diagnostic techniques have been used. For the AEP link, a total of one million iterations of MCMC were generated by using Gibbs sampling algorithms; then, it is considered a burn-in of $500,000$. For the SN link, a total of $20,000$ iterations of MCMC were generated by using Gibbs sampling algorithms; then, it is considered a burn-in of $10,000$. In these specifications, the chains seem to have converged. Although the introduction of latent variables increases the autocorrelation and longer chains are needed, the convergence is satisfactory.

The estimated TAEs for each model are given in Tables 1 and 2. The models addressing misclassification produce the best performance and they have the lowest TAE estimates.

Table 1 Estimated mean (SD) TAEs for SN data with different misclassification probabilities.

Model	TAE	TAE mis
Data: SN with misclassification ($\lambda_{10} = 0.1$, $\lambda_{01} = 0.1$)		
Normal	18.299 (4.623)	—
Normal Mis	16.740 (7.461)	11.035 (4.381)
SN	16.342 (5.118)	—
SN Mis	13.215 (5.117)	9.528 (3.127)
Data: SN with misclassification ($\lambda_{10} = 0.01$, $\lambda_{01} = 0.2$)		
Normal	30.086 (5.573)	—
Normal Mis	17.799 (7.947)	11.021 (4.285)
SN	27.058 (6.137)	—
SN Mis	13.913(4.950)	10.014 (2.670)
Data: SN with misclassification ($\lambda_{10} = 0.2$, $\lambda_{01} = 0.01$)		
Normal	16.736 (4.790)	—
Normal Mis	13.712 (7.078)	9.795 (4.588)
SN	16.923 (4.987)	—
SN Mis	11.264 (3.301)	8.241 (2.672)
Data: SN with misclassification ($\lambda_{10} = 0.05$, $\lambda_{01} = 0.15$)		
Normal	23.760 (5.005)	—
Normal Mis	17.566 (7.192)	11.204 (4.167)
SN	21.298 (5.561)	—
SN Mis	13.254 (4.339)	9.621 (2.657)

5 Application to caries data

The Signal-Tandmobiel® (ST) study is a longitudinal oral health intervention project conducted in Flanders (North of Belgium) between 1996 and 2001 (see, Vanobbergen et al. (2000)). A total of 4468 children (2315 boys and 2153 girls) were examined on a yearly basis during their primary school time (between 7 and 12 years of age) by one of sixteen trained dentists. The clinical examinations were based on visual and tactile observations and took place in a mobile dental clinic, with a standard chair and artificial dental light, and without radiographs. Data on oral hygiene and dietary habits were obtained through questionnaires completed by the parents.

Diagnosing CE is difficult due to several reasons. For instance, the view of the dental examiner can be hampered, because composite materials can imitate

Table 2 Estimated mean (SD)TAEs for AEP data with different misclassification probabilities.

Model	TAE	TAE mis
Data: AEP with misclassification ($\lambda_{10} = 0.1$, $\lambda_{01} = 0.1$)		
EP	29.607 (5.236)	—
EP Mis	5.284 (2.201)	6.533 (2.262)
AEP	22.873 (4.177)	—
AEP Mis	4.088 (1.985)	5.625 (2.229)
Data: AEP with misclassification ($\lambda_{10} = 0.01$, $\lambda_{01} = 0.2$)		
EP	42.549 (5.815)	—
EP Mis	5.155 (2.530)	6.620 (2.624)
AEP	34.475 (9.625)	—
AEP Mis	3.868 (1.733)	5.675 (2.375)
Data: AEP with misclassification ($\lambda_{10} = 0.2$, $\lambda_{01} = 0.01$)		
EP	18.378 (3.981)	—
EP Mis	5.020 (2.162)	5.184 (1.701)
AEP	13.094 (3.858)	—
AEP Mis	3.844 (1.824)	4.375 (1.420)
Data: AEP with misclassification ($\lambda_{10} = 0.05$, $\lambda_{01} = 0.15$)		
EP	35.494 (5.545)	—
EP Mis	4.908 (2.149)	6.351 (2.397)
AEP	31.379 (9.055)	—
AEP Mis	4.243 (2.163)	5.777 (2.312)

the natural enamel so well that it is difficult to spot a restored lesion or the location of the cavity, far back in the mouth. Besides, the dental examiner could also classify discolorations as CE. So, diagnosing CE involves misclassifications. Therefore, there exists no infallible scorer for CE. The best one can do is to take a very experienced dental examiner, called benchmark (see Wacholder et al. (1993)), who is assumed to be error-free or is nearly so.

Calibration exercises were performed by the 16 examiners according to the guidelines of training and calibration published by the British Association for the Study of Community Dentistry (BASCD, Pitts et al. (1997)). The calibration of the dental examiners was performed by comparing their scores on the tooth surfaces of a group of children to those of a benchmark examiner. In order to maintain a high level of intra- and inter-examiner reliability, calibration exercises were carried out twice a year for all examiners involved. A contingency table of dental examiners and the benchmark examiner was determined, yielding Table 3 with misclassified scores presented in Lesaffre and Lawson (2012). Data of the three calibration exercises were combined into one validation dataset, and also examiners' data were combined into one.

Table 3 Misclassification in the ST study.

		Benchmark	
		0	1
Examiners	0	4684	146
	1	87	428

Lesaffre and Lawson (2012) considered data collected in 2001 from 100 children randomly selected. The purpose was to search for predictors of caries experience (CE), being the covariates used: `age` at examination, `gender`, and `dentition type` (deciduous or permanent tooth). A subset of this dataset for 50 children is freely available and will be used in this section.

One of the main problems of these data is the imbalance in the response, having 75 teeth with CE (1's) and 1125 teeth free of CE (0's). Moreover, the misclassification problem using data coming from ST study has been extensively studied by using different models considering misclassification, see e.g. Mwalili et al. (2007), García-Zattera et al. (2012) and Mutsvari et al. (2013). This suggests that using an asymmetric link function in a misclassification-based framework could provide good results.

The proposed model is defined as follows. Let Y_i be the binary CE outcome which is prone to misclassification, distributed as a Bernoulli(τ_i), where $\tau_i = p_i(1-\lambda_{10}) + (1-p_i)\lambda_{01}$, p_i is the true probability, and λ_{01} and λ_{10} are the misclassification probabilities. The relation between the true probabilities and the covariates is given by

$$\Psi^{-1}(p_i) = \beta_1 + \beta_{gender} \times \texttt{gender}_i + \beta_{age} \times \texttt{age}_i + \beta_{type} \times \texttt{type}_i.$$

Since misclassification has been proved to exist in these data, symmetric (probit and EP) and asymmetric (SN and AEP) link functions are considered for models based on misclassifications. Validation data, historical data and/or experts' information can be considered to elicit the prior distribution for the misclassification parameters. In this case, from the misclassification scores in Table 3, the following informative prior distributions for the misclassification parameters are used here: $\lambda_{01} \sim \text{Beta}(87, 4684)$ and $\lambda_{10} \sim \text{Beta}(146, 428)$.

In order to avoid parameter identifiability problems, for the skew probit link model considering misclassification (SN Mis) and for the asymmetric exponential power link model considering misclassification (AEP Mis), intercept parameters have been set to $\beta_1 = 0$, and in the case of the AEP Mis model $\alpha = 0.5$. MCMC specifications were chosen to achieve convergence with the four models. Convergence checking was performed as described in the previous section.

Table 4 presents the estimated parameters for the four considered models. Note that, for all the models, the misclassification probabilities are very well estimated, basically they are around $\widehat{\lambda_{01}} = 0.018$ and $\widehat{\lambda_{10}} = 0.255$, very close to the estimated values from the validation data, that are 0.0182351 and 0.2543554, respectively. These accurate results are obtained because informative prior distributions have been used for them.

Table 4 Estimated mean (SD) for model parameters.

Parameters	Probit Mis	EP Mis	SN Mis	AEP Mis
Intercept	-0.8230 (0.1540)	-1.0691 (0.3193)	—	—
Gender (girls)	0.4858 (0.2339)	0.6749 (0.4478)	0.6354 (0.3196)	0.7617 (0.4224)
Age	-0.0292 (0.1082)	-0.0412 (0.1092)	-0.0512 (0.1319)	-0.0789 (0.1374)
Dentition type	-1.4405 (0.2530)	-1.8729 (1.2068)	-1.6853 (0.3024)	-1.8374 (0.6690)
λ_{01}	0.0187 (0.0020)	0.0185 (0.0019)	0.0187 (0.0020)	0.0185 (0.0020)
λ_{10}	0.2551 (0.0182)	0.2539 (0.0186)	0.2546 (0.0183)	0.2536 (0.0180)
θ	—	1.1418 (0.4419)	—	—
δ	—	—	-1.4687 (0.4144)	—
θ_1	—	—	—	0.2574 (0.0233)
θ_2	—	—	—	1.0476 (0.4128)
DIC	411.6	406.6	410.3	397.4

The deviance information criterion (DIC) proposed by Spiegelhalter et al. (2002) has been also presented in Table 4. DIC represents a goodness-of-fit criterion. The lower its value is, the better the fitting is. SN Miss provides a better data-fitting than Probit Miss. The same happens to AEP Miss with EP Miss. This means that models with asymmetric links should be preferred for these data. Therefore, this shows how misclassification-based models with asymmetric links can be competitive with the symmetric-based ones in applications based on real data.

6 Conclusion

Misclassified data are found in many studies and they have many different causes. The impact that misclassified data may produce on inferences can be considerable. Consequently, it is recommended to build statistical models that allow us to address misclassification. Therefore, it is important consider the inclusion of noise parameters to correct the bias arising from the misclassified data.

This paper proposes a Bayesian analysis of skew link-based regression models when the binary outcome is subject to misclassification. We have considered links based on skew-normal and the asymmetric exponential power. The use of two types of latent variables enables us to avoid computational difficulties, even by increasing the problem dimension. A simulation study shows the advantages of addressing misclassification and using skew links. To the best of the authors' knowledge, the approaches that are proposed here are the first to address misclassification at the same time that flexible asymmetric link functions are considered for the involved binary regression models.

We have shown that when data are generated from skew link-based models and errors are introduced into the responses, the skew link-based models considering misclassification perform better than those with symmetric links, and they also perform better than the error-free skew link-based models. This suggests that skew link-based regressions can be used as an alternative for processes generating misclassified data and when the rates at which the probabilities of a given binary response approaches 1 and 0 are different.

A Full conditional posterior distributions

A.1 Skew-normal link

To sample from the distribution $\pi(\boldsymbol{\beta}, \delta|\mathbf{c}, \boldsymbol{\lambda}, \mathbf{D})$, the model is augmented, including the latent variables $w_1, \ldots, w_n$ related to the model by equation (1) and defined in Section 2. The new distribution of interest is

$$\pi(\mathbf{z}, \mathbf{w}, \boldsymbol{\beta}, \delta|\mathbf{c}, \boldsymbol{\lambda}, \mathbf{D}) \propto \pi(\boldsymbol{\beta})\pi(\delta) \prod_{i=1}^{n} \Big\{ \phi(w_i; \mathbf{x}_i^T \boldsymbol{\beta} + \delta z_i, 1)\phi(z_i) I[z_i > 0] \\ \times \left(I[w_i > 0] I[c_{11}^i + c_{10}^i = 1] + I[w_i \leq 0] I[c_{01}^i + c_{00}^i = 1] \right) \Big\}.$$

Now, the four full conditional distributions are easily derived:

$$z_i|\cdots \sim \mathrm{N}\left((w_i - \mathbf{x}_i^T\boldsymbol{\beta})\frac{\delta}{1+\delta^2}, \frac{1}{1+\delta^2}\right) I[z_i > 0], \tag{6}$$

$$w_i|\cdots \sim \begin{cases} \mathrm{N}(\mathbf{x}_i^T\boldsymbol{\beta} + \delta z_i, 1) I[w_i > 0] & \text{if } c_{11}^i + c_{10}^i = 1 \\ \mathrm{N}(\mathbf{x}_i^T\boldsymbol{\beta} + \delta z_i, 1) I[w_i \le 0] & \text{if } c_{01}^i + c_{00}^i = 1 \end{cases}, \tag{7}$$

$$\boldsymbol{\beta}|\cdots \sim \mathrm{N}_k\left(\mathbf{b}^*, \mathbf{B}^*\right), \tag{8}$$

$$\delta|\cdots \sim \mathrm{N}\left(\frac{\mathbf{w}^T\mathbf{z} - \boldsymbol{\beta}^T\mathbf{x}^T\mathbf{z} + \mu_\delta/\sigma_\delta^2}{\mathbf{z}^T\mathbf{z} + 1/\sigma_\delta^2}, \frac{1}{\mathbf{z}^T\mathbf{z} + 1/\sigma_\delta^2}\right), \tag{9}$$

where

$$\mathbf{b}^* = \mathbf{B}^*(\mathbf{x}^T(\mathbf{w} - \delta\mathbf{z}) + \mathbf{B}^{-1}\mathbf{b}), \qquad \mathbf{B}^* = (\mathbf{x}^T\mathbf{x} + \mathbf{B}^{-1})^{-1}.$$

The final algorithm consists of choosing initial values $\mathbf{w}^{(0)}$, $\boldsymbol{\beta}^{(0)}$, $\delta^{(0)}$, $\mathbf{c}^{(0)}$ and $\boldsymbol{\lambda}^{(0)}$, and generating iteratively from the full conditional distributions. Note that generating from the full conditional distributions is easy. The distributions are standard, and generating from them is trivial and efficient. The following order is proposed: $\mathbf{z}^{(j)}$, $\mathbf{w}^{(j)}$, $\boldsymbol{\beta}^{(j)}$, $\delta^{(j)}$, $\mathbf{c}^{(j)}$ and $\boldsymbol{\lambda}^{(j)}$ using (6), (7), (8), (9), (4) and (5), respectively.

A.2 Asymmetric exponential power link

To sample from the distribution $\pi(\boldsymbol{\beta}, \alpha, \theta_1, \theta_2|\mathbf{c}, \boldsymbol{\lambda}, \mathbf{D})$, the model is augmented, including the latent variables $w_1, \ldots, w_n$ related to the model by equation (2) and defined in Section 2. The joint posterior density of interest is

$$\begin{aligned} &\mathrm{p}(\mathbf{w}, \mathbf{u}_1, \mathbf{u}_2, \boldsymbol{\beta}, \alpha, \theta_1, \theta_2|\mathbf{c}, \boldsymbol{\lambda}, \mathbf{D}) \propto \pi(\boldsymbol{\beta})\pi(\theta_1)\pi(\theta_2)\pi(\alpha) \\ &\times \prod_{i=1}^n \Bigg\{ \Bigg(\exp(-u_{1i}) I\left[\mathbf{x}_i^T\boldsymbol{\beta} - \tfrac{\alpha}{\Gamma(1+1/\theta_1)} u_{1i}^{1/\theta_1} < w_i \le \mathbf{x}_i^T\boldsymbol{\beta}\right] \\ &+ \exp(-u_{2i}) I\left[\mathbf{x}_i^T\boldsymbol{\beta} < w_i < \mathbf{x}_i^T\boldsymbol{\beta} + \tfrac{(1-\alpha)}{\Gamma(1+1/\theta_2)} u_{2i}^{1/\theta_2}\right] \Bigg) \\ &\times \left(I[w_i > 0] I[c_{11}^i + c_{10}^i = 1] + I[w_i \le 0] I[c_{01}^i + c_{00}^i = 1] \right) \Bigg\}. \end{aligned}$$

The full conditional distributions are as follows

$$u_{1i}|\cdots \sim \mathrm{Exp}(1) I\left[u_{1i} > \left(\tfrac{\max\{0\ ,\ \mathbf{x}_i^T\boldsymbol{\beta} - w_i\}}{\alpha\Gamma(1+1/\theta_1)}\right)^{\theta_1}\right], \tag{10}$$

$$u_{2i}|\cdots \sim \mathrm{Exp}(1) I\left[u_{2i} > \left(\tfrac{\max\{0\ ,\ w_i - \mathbf{x}_i^T\boldsymbol{\beta}\}}{(1-\alpha)\Gamma(1+1/\theta_2)}\right)^{\theta_2}\right], \tag{11}$$

$$w_i|\cdots \sim \begin{cases} \mathrm{U}\Big(\max\left\{0, \mathbf{x}_i^T\boldsymbol{\beta} - \tfrac{\alpha}{\Gamma(1+1/\theta_1)} u_{1i}^{1/\theta_1}\right\}, \\ \qquad \max\left\{0, \mathbf{x}_i^T\boldsymbol{\beta} + \tfrac{(1-\alpha)}{\Gamma(1+1/\theta_2)} u_{2i}^{1/\theta_2}\right\} \Big) \text{ if } c_{11}^i + c_{10}^i = 1, \\ \mathrm{U}\Big(\min\left\{0, \mathbf{x}_i^T\boldsymbol{\beta} - \tfrac{\alpha}{\Gamma(1+1/\theta_1)} u_{1i}^{1/\theta_1}\right\}, \\ \qquad \min\left\{0, \mathbf{x}_i^T\boldsymbol{\beta} + \tfrac{(1-\alpha)}{\Gamma(1+1/\theta_2)} u_{2i}^{1/\theta_2}\right\} \Big) \text{ if } c_{01}^i + c_{00}^i = 1, \end{cases} \tag{12}$$

$$\beta_j|\cdots \sim \mathrm{N}(\mathrm{b}_j^*\ ,\ \mathrm{B}_j^*) I\left[\beta_j \in \left(\underline{\beta}_j\ ,\ \overline{\beta}_j\right)\right], \tag{13}$$

$$\pi(\alpha|\cdots) \propto \pi(\alpha) I\left[\underline{\alpha} < \alpha < \overline{\alpha}\right], \tag{14}$$

$$\pi(\theta_1|\cdots) \propto \pi(\theta_1) I\left[\theta_1 \in \bigcap_{\{i:u_{1i}>0\}} \Theta_{1i}\right], \tag{15}$$

$$\pi(\theta_2|\cdots) \propto \pi(\theta_2) I\left[\theta_2 \in \bigcap_{\{i:u_{2i}>0\}} \Theta_{2i}\right], \tag{16}$$

where Exp(1) denotes the exponential distribution with parameter equal to 1, and

$$\begin{aligned}
\mathrm{b}_j^* &= \mathbf{b}_j - \mathbf{B}_{j(-j)} \mathbf{B}_{(-j)(-j)}^{-1} \left(\boldsymbol{\beta}_{(-j)} - \mathbf{b}_{(-j)}\right), \\
\mathrm{B}_j^* &= \mathbf{B}_{jj} - \mathrm{B}_{j(-j)} \mathbf{B}_{(-j)(-j)}^{-1} \mathbf{B}_{(-j)j}, \\
\underline{\beta}_j &= \max\left\{ \max_{\{i:u_{1i}>0, x_{ij}<0\}} \left\{ \frac{w_i - \mathbf{x}_{i(-j)}^T \boldsymbol{\beta}_{(-j)}}{x_{ij}} + \left(\frac{\frac{\alpha}{\Gamma(1+1/\theta_1)} u_{1i}^{1/\theta_1}}{x_{ij}} \right) \right\}, \right. \\
&\qquad \left. \max_{\{i:u_{2i}>0, x_{ij}>0\}} \left\{ \frac{w_i - \mathbf{x}_{i(-j)}^T \boldsymbol{\beta}_{(-j)}}{x_{ij}} - \left(\frac{\frac{(1-\alpha)}{\Gamma(1+1/\theta_2)} u_{2i}^{1/\theta_2}}{x_{ij}} \right) \right\} \right\}, \\
\overline{\beta}_j &= \min\left\{ \min_{\{i:u_{1i}>0, x_{ij}>0\}} \left\{ \frac{w_i - \mathbf{x}_{i(-j)}^T \boldsymbol{\beta}_{(-j)}}{x_{ij}} + \left(\frac{\frac{\alpha}{\Gamma(1+1/\theta_1)} u_{1i}^{1/\theta_1}}{x_{ij}} \right) \right\}, \right. \\
&\qquad \left. \min_{\{i:u_{2i}>0, x_{ij}<0\}} \left\{ \frac{w_i - \mathbf{x}_{i(-j)}^T \boldsymbol{\beta}_{(-j)}}{x_{ij}} - \left(\frac{\frac{(1-\alpha)}{\Gamma(1+1/\theta_2)} u_{2i}^{1/\theta_2}}{x_{ij}} \right) \right\} \right\},
\end{aligned}$$

$$\begin{aligned}
\underline{\alpha} &= \max\left\{0, \max_{\{i:u_{1i}>0\}} \left\{ \frac{(\mathbf{x}_i^T \boldsymbol{\beta} - w_i)\Gamma(1+1/\theta_1)}{u_{1i}^{1/\theta_1}} \right\} \right\}, \\
\overline{\alpha} &= \min\left\{0, \min_{\{i:u_{2i}>0\}} \left\{ 1 - \frac{(w_i - \mathbf{x}_i^T \boldsymbol{\beta})\Gamma(1+1/\theta_2)}{u_{2i}^{1/\theta_2}} \right\} \right\}, \\
\Theta_{1i} &= \left\{ \theta_1 : \frac{\mathbf{x}_i^T \boldsymbol{\beta} - w_i}{\alpha} < \frac{1}{\Gamma(1+1/\theta_1)} u_{1i}^{1/\theta_1} \right\}, \\
\Theta_{2i} &= \left\{ \theta_2 : \frac{w_i - \mathbf{x}_i^T \boldsymbol{\beta}}{(1-\alpha)} < \frac{1}{\Gamma(1+1/\theta_2)} u_{2i}^{1/\theta_2} \right\}.
\end{aligned}$$

Given that the prior distribution of $\boldsymbol{\beta}$ is multivariate normal, then the conditional distribution of β_j given $\boldsymbol{\beta}_{(-j)}$ is normal, where $\boldsymbol{\beta}_{(-j)}^T = (\beta_1, \ldots, \beta_{j-1}, \beta_{j+1}, \ldots, \beta_k)$. The subscript $(-j)$ denotes that the jth element has been removed.

Note that the distributions given in (10), (11) and (12) are standard and their sampling is straightforward. The final algorithm consists of choosing initial values $\mathbf{w}^{(0)}$, $\boldsymbol{\beta}^{(0)}$, $\alpha^{(0)}$, $\theta_1^{(0)}$ and $\theta_2^{(0)}$, and iteratively sampling $\mathbf{u}_1^{(j)}$, $\mathbf{u}_2^{(j)}$, $\mathbf{w}^{(j)}$, $\boldsymbol{\beta}^{(j)}$, $\alpha^{(j)}$, $\theta_1^{(j)}$, $\theta_2^{(j)}$, $\mathbf{c}^{(j)}$ and $\boldsymbol{\lambda}^{(j)}$ from the full conditional distributions (10), (11), (12), (13), (14), (15), (16), (4) and (5), respectively.

Acknowledgements This research has been supported by *Ministerio de Economía y Competitividad*, Spain (Project MTM2014-56949-C3-3-R), *Gobierno de Extremadura*, Spain (Project GRU18108) and *European Union* (European Regional Development Funds). The authors wish thank the editor and the two anonymous referees for comments and suggestions, which have improved the content and readability of the paper.

References

Achcar JA, Martínez EZ, Louzada-Neto F (2004) Binary data in the presence of misclassifications. In: Antoch J (ed) Proceedings of the Computational Statistics Conference, Physica-Verlag/Springer, Prague, Czech Republic, pp 581–588

Albert J, Chib S (1993) Bayesian analysis of binary and polychotomous response data. Journal of the American Statistical Association 88(422):669–679
Aranda-Ordaz FJ (1981) On two families of transformations to additivity for binary response data. Biometrika 68(2):357–363
Bazán JL, Bolfarine H, Branco MD (2010) A framework for skew-probit links in binary regression. Communications in Statistics - Theory and Methods 39(4):678–697
Buonaccorsi JP (2010) Measurement Error. Chapman & Hall/CRC, London
Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu CM (2006) Measurement Error in Nonlinear Models: A Modern Perspective, 2nd edn. Chapman & Hall/CRC, Boca Raton, Florida
Chen MH, Dey DK, Shao QM (1999) A new skewed link model for dichotomous quantal response data. Journal of the American Statistical Association 94(448):1172–1186
Czado C (1994) Parametric link modification of both tails in binary regression. Statistical Papers 35:189–201
Gamerman D, Lopes HF (2006) Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference, 2nd edn. Chapman & Hall/CRC, Boca Raton, Florida
García-Zattera MJ, Jara A, Lesaffre E, Marshall G (2012) Modeling of multivariate monotone disease processes in presence of misclassification. Journal of the American Statistical Association 107(499):976–989
Gustafson P (2003) Measurement error and misclassification in statistics and epidemiology: Impacts and Bayesian adjustments. Chapman and Hall, Boca Raton, Florida
Kim S, Chen MH, Dey DK (2008) Flexible generalized t-link models for binary response data. Biometrika 95(1):93–106
Lesaffre E, Lawson AB (2012) Bayesian Biostatistics. John Wiley & Sons, New York
Liu C (2004) Robit regression: A simple robust alternative to logistic and probit regression. In: Gelman A, Meng XL (eds) Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives, John Wiley and Sons, New York, pp 227–238
McInturff P, Johnson WO, Cowling D, Gardner IA (2004) Modelling risk when binary outcomes are subject to error. Statistics in Medicine 23:1095–1109
Mutsvari T, Bandyopadhyay D, Declerck D, Lesaffre E (2013) A multilevel model for spatially correlated binary data in the presence of misclassification: an application in oral health research. Statistics in Medicine 32(30):5241–5259
Mwalili SM, Lesaffre E, Declerck D (2007) The zero-inflated negative binomial regression model with correction for misclassification: an example in caries research. Statistical Methods in Medical Research pp 1–17
Naranjo L, Martín J, Pérez CJ (2014a) Bayesian binary regression with exponential power link. Computational Statistics and Data Analysis 71:464–476
Naranjo L, Martín J, Pérez CJ, Rufo MJ (2014b) Addressing misclassification for binary data: probit and t-link regressions. Journal of Statistical Computation and Simulation 84(10):2187–2213
Naranjo L, Pérez CJ, Martín J (2015) Bayesian analysis of some models that use the asymmetric exponential power distribution. Statistics and Computing 25(3):497–514
Paulino CD, Soares P, Neuhaus J (2003) Binomial regression with misclassification. Biometrics 59:670–675
Paulino CD, Silva G, Achcar JA (2005) Bayesian analysis of correlated misclassified binary data. Computational Statistics and Data Analysis 49:1120–1131
Pitts NB, Evans DJ, Pine CM (1997) British association for the study of community dentistry (BASCD) diagnostic criteria for caries prevalence surveys-1996/97. Community Dent Health 14(Suppl 1):6–9
Rekaya R, Weigel KA, Gianola D (2001) Threshold model for misclassified binary responses with applications to animal breeding. Biometrics 57(4):1123–1129
Rubio FJ (2015) Letter to the Editor: on the use of improper priors for the shape parameters of asymmetric exponential power models. Statistics and Computing 25(6):1281–1287
Smith BJ (2007) BOA: an R package for MCMC output convergence assessment and posterior inference. Journal of Statistical Software 21(11):1–37
Spiegelhalter D, Best N, Carlin B, van der Linde A (2002) Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society, Series B 64:583–639
Stukel TA (1988) Generalized logistic models. Journal of the American Statistical Association 83:426–431
Vanobbergen J, Martens L, Lesaffre E, Declerck D (2000) The Signal-Tandmobiel® project, a longitudinal intervention health promotion study in Flanders (Belgium): baseline and first

year results. European Journal of Paediatric Dentistry 2:87–96
Wacholder S, Armstrong B, Hartge P (1993) Validation studies using an alloyed gold standard. American Journal of Epidemiology 137(11):1251–1258
Zhu D, Zinde-Walsh V (2009) Properties and estimation of asymmetric exponential power distribution. Journal of Econometrics 148(1):86–99