Research Article

Applied Stochastic Models in Business and Industry

Received 6 October 2015, Revised 13 August 2016, Accepted 24 October 2016 Published online in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/asmb.2215

Power and reversal power links for binary regressions: An application for motor insurance policyholders

J.L. Bazán[a*†], F. Torres-Avilés[a,b], A.K. Suzuki[a] and F. Louzada[a]

In binary regression, symmetric links such as logit and probit are usually considered as standard. However, in the presence of unbalancing of ones and zeros, these links can be inappropriate and inflexible to fit the skewness in the response curve and likely to lead to misspecification. This is the case of covering some type of insurance, where it can be observed that the probability of a given binary response variable approaches zero at different rates than it approaches one. Furthermore, when usual links are considered, there is not a skewness parameter associated with the distribution chosen that, regardless of the linear predictor, is easily interpreted. In order to overcome such problems, a proposal for the construction of a set of new skew links is developed in this paper, where some of their properties are discussed. In this context, power links and their reversal versions are presented. A Bayesian inference approach using MCMC is developed for the presented models. The methodology is illustrated considering a sample of motor insurance policyholders selected randomly by gender. Results suggest that the proposed link functions are more appropriate than other alternative link functions commonly used in the literature. Copyright © 2016 John Wiley & Sons, Ltd.

Keywords: binary regression; Bayesian approach; MCMC; power links; skewed link

1. Introduction

Historically, binary models have been widely used by financial and insurance companies as a major tool of vital importance for rating applicants/customers on granting credit to clients and ensuring integrity of financial operations. According to [1], granting credit facilities and ensuring the integrity of financial operations started to be more important for the profitability of companies in the financial sector, becoming one of the main sources of revenue for financial and insurance institutions in general.

In this context, the immediate example is the logit model (L), which is obtained by considering $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_n)'$, a $n \times 1$ vector of n independent dichotomous random variables with probability $p_i = P[Y_i = 1]$, and $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ik})'$ a $k \times 1$ vector of covariates, where x_{i1} may equal 1, this leads to assume an intercept, $i = 1, \ldots, n$. Therefore, $\boldsymbol{X}$ will denote a $n \times k$ design matrix with rows $\boldsymbol{x}_i'$, and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)'$ a $k \times 1$ vector of regression coefficients. For this model, the probability of the binary response variable will be given by

$$p_i = E(Y_i|\boldsymbol{x}_i) = L(\boldsymbol{\eta}_i) = L(\boldsymbol{x}_i'\boldsymbol{\beta}) = \frac{exp(\boldsymbol{x}_i'\boldsymbol{\beta})}{1 + exp(\boldsymbol{x}_i'\boldsymbol{\beta})}, \quad i = 1, \ldots, n, \tag{1}$$

where $L(.)$ denotes a cumulative distribution function (cdf) of the standard logistic distribution. The inverse of the function L, namely, L^{-1}, is called link function, and $\boldsymbol{\eta}_i = \boldsymbol{x}_i'\boldsymbol{\beta}$ is the corresponding linear predictor. The graphic considering p_i as a function of η_i is called response curve or probability of success, and has a symmetric form centred in 0.5.

[a] *Departamento de Matemática Aplicada e Estatística, Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, Sáo Carlos-SP, Brazil*
[b] *Departamento de Matemática y Ciencia de la Computación, Universidad de Santiago de Chile, Santiago, Chile*
**Correspondence to: Jorge Bazan, Departamento de Matemática Aplicada e Estatística, Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, Avenida Trabalhador São-carlense, 400 - Centro CEP: 13566–590, São Carlos-SP, Brazil.*
† *E-mail: jlbazan@icmc.usp.br*

Bayesian approach for binary regression models has been widely developed in the literature since the 1990s [2]. In finance and insurance, credit scoring (e.g., [3–7] and [8], only to mention a few) and fraud detection (e.g., [9–12] and [13], only to mention a few) are some of the applications where logistic regression is considered as a standard procedure.

Other known symmetric link functions are obtained when L is replaced by the cdf of the standard normal, $\Phi(.)$, and standard Cauchy distribution, $C(.)$, [14], raising the probit, $p_i = \Phi(\boldsymbol{x}_i'\boldsymbol{\beta})$, and cauchit, $p_i = C(\boldsymbol{x}_i'\boldsymbol{\beta}) = 0.5 + \pi^{-1} arctan(\boldsymbol{x}_i'\boldsymbol{\beta})$, link functions, respectively. On the other hand, some textbooks (e.g., [15]) have reported that an asymmetric link function may be more appropriate than a symmetric one for some specific situations. Thus, the loglog, $p_i = Gu(\boldsymbol{x}_i'\boldsymbol{\beta}) = exp\{-exp(-\boldsymbol{x}_i'\boldsymbol{\beta})\}$, and cloglog, $p_i = RGu(\boldsymbol{x}_i'\boldsymbol{\beta}) = 1 - exp\{-exp(\boldsymbol{x}_i'\boldsymbol{\beta})\}$, are other known links, which can be obtained by considering the cdf of the gumbel $Gu(.)$, and the called reversed gumbel distribution $RGu(.)$, respectively. This name will be justified in the next section.

Several authors have argued, for example, [16], that when the probability of a given binary response variable approaches 0 at different rates, then it approaches 1. This is the case of unbalancing of ones or zeros, in which symmetric link functions may be not useful to fit binary data, and asymmetric link functions must be considered. Unbalancing may occur for data concerning credit scoring, fraud, attrition, churning and insurance claims. Consequently, many asymmetric link functions have been proposed in the literature. Some examples include [17–25] and [26].

In this paper, we propose a general approach for asymmetric link functions in binary regression. We consider the relation between the Gumbel and reversal Gumbel distributions, named here as its reversal property, to extend this to a class of asymmetrical link functions that are different from the classical cloglog and loglog links, because of the inclusion of an additional parameter that is associated with the asymmetry of the distribution and used in the construction of the link function, which controls the rate of increasing (or decreasing) the probability of success (failure) of the binary response variables.

The link functions introduced here have, as particular cases, some known links mostly used in the literature, and the asymmetry parameter associated with these cdfs is also introduced regardless of the linear predictor, and the latent linear structure will not be necessary for this link approach. The most important aspect of the modelling stage under this setting is the potential improvement to model the fit in binary regression that can be obtained considering a particular class of asymmetric link functions when data cannot be modelled appropriately with usual links as shown in this paper.

This work is organised as follows. In Section 2, the reversal property is presented and several distributions satisfying this property are shown emphasising power and reversal power distributions. In Section 3, the modelling for binary response variables considering a class of asymmetric link functions based on power and reversal power distributions are formulated. In Section 4, a Bayesian estimation approach is developed including a discussion on model comparison criteria obtained from the MCMC output. In Section 5, an illustrative example is presented through a real application by considering real data already analysed in the literature specialised in insurance. Finally, a discussion and extensions of the link functions proposed in this paper are considered in Section 6.

2. Reversal property and power and reversal power distributions

In a general framework, consider F^{-1} link functions for binary regression with $F(.)$ and $f(.)$ the correspondent cdf and pdf respectively of a random variable S. In this section, we define the reversal property and introduce the power and reversal power distributions.

Definition 1
Let $S \sim F(.)$ (or $S \sim f(.)$), we say that the distribution of S satisfies the reversal property if the cdf of $-S$ is a different distribution that can be written as $-S \sim G(.) \equiv 1 - F(-.)$ (or $g(s) \equiv f(-s)$). In this case, the distribution of $G(.)$ is called the reversal distribution of $F(.)$.

Result 1
Consider $S \sim F(.)$ (or $S \sim f(.)$) a random variable (r.v.) following a symmetric distribution, that is, a pdf $f(.)$ is said to be symmetrical if and only if there is a value x_0 such that $f(x_0 - \delta) = f(x_0 + \delta)$ for all real numbers δ. Then, $F(.)$ does not satisfy the reversal property. That is, S and $-S$ have the same distribution and $F(s) = 1 - F(-s), f(s) = f(-s)$. Examples of this are Logistic, Normal and Cauchy distributions.

Result 2
Consider $S \sim Gumbel(\mu, \beta)$ with the pdf given by $f(s;\mu,\beta) = \frac{1}{\beta}e^{-(z+e^{-z})}$, where $z = \frac{s-\mu}{\beta}$. Then, it is easy to see that S satisfies the reversal property and only $-S \sim reversed\ Gumbel(\mu, \beta)$ with the pdf given by $g(s;\mu,\beta) = \frac{1}{\beta}e^{-(z+e^{-z})}$ where $z = \frac{-s-\mu}{\beta}$. In addition, note also that the reversed Gumbel distribution satisfies the reversal property and its reversal distribution is only the Gumbel distribution. In this case, the correspondents cdf satisfy $F(s) + G(-s) = 1$ and $f(s) = g(-s)$ and S and $-S$ have

the Gumbel and reversed Gumbel distributions, respectively. Thus, the latter two distributions are not symmetrical in this case, but are closely related where they are called mirror images of one another.

Remarks 1
Some examples of distributions considered in binary regression that do not satisfy the reversal property are the scale mixtures of normal distributions that were considered by [27] and the generalised t considered by [28]. On the other hand, examples of distributions considered to propose link functions in binary regression that satisfy the reversal property are the skew normal, skew t, which were used by [16, 18], and [29], respectively, to propose asymmetric links for binary regression. Therefore, the case addressed by [25], which use of the generalised extreme value distribution, also satisfies the property.

Definition 2
A univariate random variable S is said to follow a power distribution, $S \sim \mathcal{P}(\mu, \sigma^2, \lambda)$, with location, scale and shape parameters given by $-\infty < \mu < \infty$, $\sigma^2 > 0$ and $\lambda > 0$, respectively, if the density of such distribution has the form

$$f_p(s \mid \mu, \sigma^2, \lambda) = \frac{\lambda}{\sigma} g\left(\frac{s-\mu}{\sigma}\right)\left[G\left(\frac{s-\mu}{\sigma}\right)\right]^{\lambda-1}, \tag{2}$$

with $G(.)$ denoting any absolute continuous cdf with real support and $g(.)$ a unimodal and log concave pdf with support in the real line $< -\infty, \infty >$ called the *baseline distribution*.

If $\lambda = 1$, the density of S in (2) reduces to the density of the $g(\mu, \sigma^2)$ baseline distribution. The *standard* power distribution is obtained from Equation 2 with $\mu = 0$ and $\sigma^2 = 1$. It is denoted by $Z \sim \mathcal{P}(\lambda)$ with the corresponding pdf and cdf given by

$$f_p(z \mid \lambda) = \lambda g(z)[G(z)]^{\lambda-1} \tag{3}$$

and

$$F_p(z) = [G(z)]^{\lambda}. \tag{4}$$

Result 3
Consider S a r.v. following a standard power distribution with cdf $F_p(s) = G(s)^{\lambda}$, where $G(.)$ is the cdf of a baseline distribution and $\lambda > 0$ is an additional parameter. Then $F_p(.)$ satisfies the reversal property with $F_{rp}(s) = 1 - G(-s)^{\lambda}$ being its corresponding reversal cdf distribution. Similarly, consider S a r.v. with cdf $F_{rp}(s) = 1 - G(-s)^{\lambda}$, called standard reversal power, where $G(.)$ is the cdf of a baseline distribution and $\lambda > 0$ an additional parameter. Then $F_{rp}(.)$ satisfies the reversal property with $F_p(s) = G(s)^{\lambda}$ being its correspondent reversal cdf distribution.

Remarks 2
Some remarks are as follows:

- $F_p(-s) \neq 1 - F_p(s)$ or $F_{rp}(-s) \neq 1 - F_{rp}(s)$ and then F_p and F_{rp} are not point-symmetric, for $\lambda \neq 1$.
- $F_p(\pm s) + F_{rp}(\mp s) = 1$ and thus, both distributions are distinct, though closely related.
- When $\lambda = 1$ then $F_p(s) = G(s) = F_{rp}(s)$ and thus $G(.)$ is a particular case of both distributions.
- It is easy to observe that if S follows an $F_p(.)$ distribution, then $-S$ follows an $F_{rp}(.)$ distribution.

Remarks 3
Some examples of power and reversal power distributions are as follows:

- By considering Result 3 and taking $G(.) = L(.)$, the cdf of the standard logistic distribution $F_p(.)$ is called proportional reversed hazard logistic distribution [30] and is also known as the Type-I generalised logistic distribution [31]. Here, we called this distribution power logistic. Additionally, $F_{rp}(.)$ is named scobit distribution [32] and is also known as the Type-II generalised logistic distribution [31], which is called here reversal power logistic. Both distributions are particular cases of the named beta-logistic distribution or log F distributions [33], or Type-IV generalised logistic distribution [31].
- By considering Result 3 and taking $G(.) = \Phi(.)$, the cdf of the standard normal distribution $F_p(.)$ is called power normal distribution [34, 35]. Moreover, $F_{rp}(.)$ is a new distribution in the literature and can be named reversal power normal distribution. Both distributions are particular cases of the named beta-normal distribution [36].
- Other examples of power distributions are the power cloglog and power loglog distribution, recently proposed by [37]. Considering this, both distributions should be called here power reversal gumbel and power gumbel distribution because these distributions are the correspondent base distribution based on the distribution of the extreme minimum and extreme maximum, respectively. In addition, we propose the correspondent reversal versions of each one, that is,

reversal power cloglog and reversal power loglog. Note that these distributions are not respectively the reversal version of the other, and also power cloglog is not reversal power loglog, and power loglog is not reversal power cloglog.

- In addition, two new distributions: the power cauchy distribution and the reversal power cauchy (RPC) distribution are proposed here by considering the Cauchy distribution as the baseline.

3. On the new class of asymmetric link functions in binary regression

A new class of models for binary data can be obtained by considering that the probability of success is given by

$$p_i = E(y_i|\boldsymbol{x}_i) = F_\lambda(\eta_i) = F_\lambda(\boldsymbol{x}_i'\boldsymbol{\beta}), \tag{5}$$

where F_λ can be the cdfs of a standard power or a standard reversal power distribution indexed with $\lambda > 0$, that is,

$$F_p(y) = G(y)^\lambda \quad \text{or} \quad F_{rp}(y) = 1 - G(-y)^\lambda, \tag{6}$$

with $G(.)$ denoting any cdf of baseline distribution with support in the real line. We named this new class of links of binary regression models the *positive exponent links*.

Note that the preceding formulation for new links of binary regression is rather general. Therefore, various cases can be obtained from the new class as discussed subsequently. Also note that λ parameter has quite a different meaning to $\boldsymbol{\beta}$ parameters because it is a structural parameter associated with the choice of the link function. On the other hand, $\boldsymbol{\beta}$ parameters are a vector of structural parameters inherent to the observed data and do not depend on model choice.

In binary regression, by considering $G(.) = \Phi(.)$, the cdf of the standard normal distribution $F_p^{-1}(.)$ is a power probit link and $F_{rp}^{-1}(.)$ is its reversal power probit, and the probit link will be a particular case. Both cases of links were proposed by [26]. By considering $G(.) = L(.)$, the cdf of the standard logistic distribution $F_p^{-1}(.)$ is a power logit link and $F_{rp}^{-1}(.)$ is its reversal power logit, and the logit link is a particular case again. Both are particular cases of the link proposed by [23]. $F_p^{-1}(.)$ is usually called power logit link and $F_{rp}^{-1}(.)$ is also called scobit link [32]. Table I shows the pdf of some positive exponent links in the literature, including some based on the power distributions and their reversal versions.

Figure 1 depicts different response curves for the power and reversal power models by using different values for λ, for different values of linear predictor η considering different baseline distributions as baseline links (logit, probit, loglog, cloglog and cauchit). For $\lambda = 1$, the correspondent power and reversal power links correspond to the baseline link and for $\lambda < 1$ (or $\lambda > 1$) the curve corresponding to the power model is generally above (below) the curve corresponding to the baseline link model within a range of η values. Note also that for each value of λ, reversal power curve is a reflection

Table I. Some baseline, power and reversal power distributions, and your correspondent links and positive exponent links to binary regression.

Type	Name of distributions	pdf	Name of link	Abbreviature
Baseline	logistic	$L(z) = \frac{1}{1+e^{-z}}$.	logit	L
	normal	$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-t^2/2}\, dt$	probit	P
	cauchit	$C(Z) = \frac{1}{\pi} \arctan(z) + \frac{1}{2}$	cauchit	C
	gumbel	$Gu(z) = e^{-e^{-z}}$.	loglog	Ll
	reversal-gumbel	$Rgu(z) = 1 - e^{-e^{z}}$.	cloglog	Cll
power	type I generalised logistic	$L(z)^\lambda$	power logit	PL
	power normal	$\Phi(z)^\lambda$	power probit	PP
	power cauchy	$C(z)^\lambda$	power cauchit	PC
	power gumbel	$Gu(z)^\lambda$	power loglog	PLl
	power reversal-gumbel	$Rgu(z)^\lambda$	power cloglog	PCll
reversal power	type II generalised logistic	$1 - L(-z)^\lambda$	reversal power logit	RPL
	reversal power normal	$1 - \Phi(-z)^\lambda$	reversal power probit	RPP
	reversal power cauchy	$1 - C(-z)^\lambda$	reversal power cauchit	RPC
	reversal power gumbel	$1 - Gu(-z)^\lambda$	reversal power loglog	RPLl
	reversal power reversal-Gumbel	$1 - Rgu(-z)^\lambda$	reversal power cloglog	RPCll

*where $z = (x - \mu)/\sigma$.

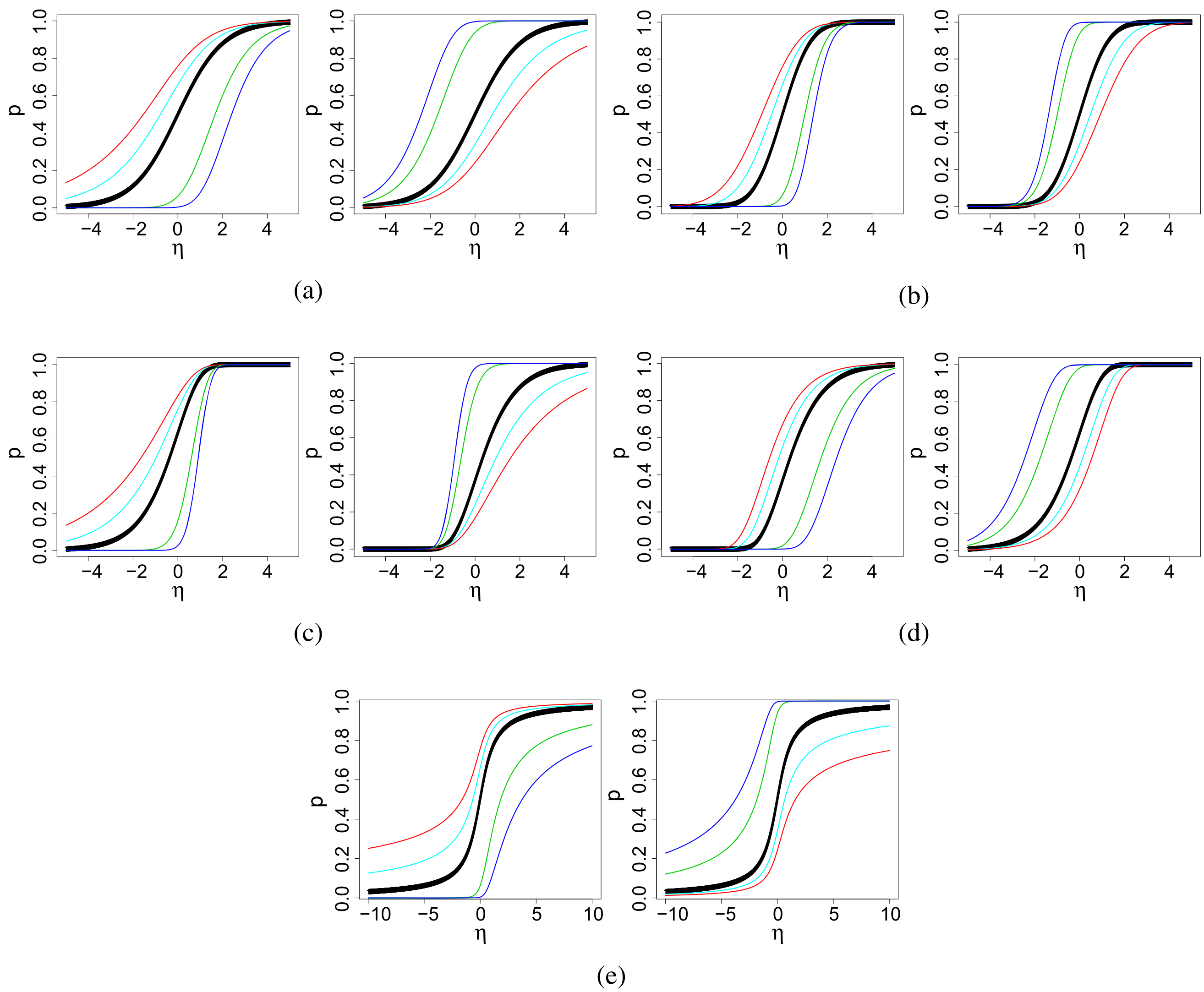

Figure 1. CDFs comparison for λ=0.4 (red), 0.6 (grey), 1.0 (black), 4.0 (green), 8.0 (blue): (a) power logit versus reversal power logit; (b) power probit versus reversal power probit; (c) power loglog versus reversal power loglog, (d) power cloglog versus reversal power cloglog, (e) power cauchit versus reversal power cauchit.

of the correspondent power curve and thus for $\lambda < 1$ (or $\lambda > 1$) the corresponding curve is generally below (above) the corresponding curve for the baseline curve.

Specifically, by example, consider the different probability curves for the power cauchit and reversal power cauchit in Figure 1(e). For $\lambda = 1$, both links correspond to the curve of the cauchit model. For the power cauchit link, if $\lambda < 1$, the correspondent curve is generally above the cauchit curve and then its probability of success is greater than the correspondent probability under the cauchit link. Thus, λ is a *bonus* parameter. Conversely, if $\lambda > 1$, the power cauchit curve is generally below the cauchit curve and then its probability of success is less than the correspondent probability under the cauchit link. Thus, λ is a *penalty parameter*. For the reversal power cauchit, the interpretation is inverse, that is, when $\lambda < 1$, it is a *penalty parameter* and if $\lambda > 1$ then it is a *bonus parameter*.

4. Bayesian inference

For inference, we consider a full Bayesian approach. The likelihood function to the binary regression model considering the new class of positive exponent links is given by

$$L(\boldsymbol{\beta}, \lambda | y, X) = \prod_{i=1}^{n} [F_{\lambda}(\boldsymbol{x}_i' \boldsymbol{\beta}]^{y_i} [1 - F_{\lambda}(\boldsymbol{x}_i' \boldsymbol{\beta})]^{1-y_i}. \tag{7}$$

In addition, introducing the latent variables

$$y_i = I(s_i > 0) = \begin{cases} 1, & s_i > 0, \\ 0, & s_i \leqslant 0, \end{cases}, \quad i = 1, \dots, n,$$

we can obtain an alternative and complete-data likelihood function that is given by

$$L(\boldsymbol{\beta}, \lambda | s, y, X) = \prod_{i=1}^{n} f_\lambda^*(s_i) p(y_i | s_i),$$

with $p(y_i|s_i) = I(s_i, y_i) = I(s_i > 0)I(y_i = 1) + I(s_i \leqslant 0)I(y_i = 0)$ and f_λ^* is the pdf of the distribution corresponding to the reversal distribution of F_λ considered in (3).

As prior distributions, we are assuming independence such that

$$\pi(\boldsymbol{\beta}, \delta) = \pi_1(\boldsymbol{\beta})\pi_2(\delta) \tag{8}$$

with $\boldsymbol{\beta} \sim \pi_1(.)$ and $\delta \sim \pi_2(.)$, where $\delta = \log(\lambda)$. For $\pi_1(.)$, usual priors for regression as $\beta_j \sim N\left(\mu_\beta, \sigma_\beta^2\right), j = 1, 2, \dots, k$ [38] can be considered with $\mu_\beta = 0$, and $\sigma_\beta^2 = 100$ to assure non information about them. For $\pi_2(.)$, we tested two prior distributions for δ parameter. The first one was $N(0, 1)$, and the second one a $Uniform(-2, 2)$ because values out of the range $[e^{-2}, e^2]$ have much less or non-probability of occurrence.

4.1. MCMC scheme

Considering the likelihood function in (7) and a general prior specification given (8), the Bayesian estimation can be easily implemented via Metropolis-Hasting providing a simple and efficient sampling from the marginal posterior distributions that will be used for inference about the parameters of the model.

Therefore, the hierarchical structure of the binary regression model considering positive exponent links using the δ-parametrization is as follows:

$$Y_i | \boldsymbol{\beta}, \delta \sim Bernoulli\Big(F_\delta\left(\boldsymbol{x}_i' \boldsymbol{\beta}\right)\Big)$$

$$\boldsymbol{\beta} \sim \pi_1(\boldsymbol{\beta}),$$

and

$$\delta \sim \pi_2(\delta),$$

Note that when $\delta = 0$ ($\lambda = 1$), the hierarchical structure of the likelihood of the baseline model based on the $G(.)$ distribution in (6) follows by eliminating the third line in the preceding model.

Furthermore, note that this hierarchical structure can be easily implemented in several packages such as WinBugs, JAGS, Rstan or SAS software. In this paper, all the models considered in the application were implemented using PROC MCMC of SAS 9.3 software [39]. In all cases, an effective sample size of 4000 was considered discarding the 5000 initial iterations and considering jumps of 50.

An evaluation of the mean plot, first-order autocorrelations of the posterior samples, the effective sample size of each parameter and the standard errors of the posterior mean estimate provide strong indications of the convergence of MCMC output in all cases. In addition, Geweke, Heidelberger-Welch and Raftery-Lewis Diagnostics as implemented in the PROC MCMC (see more details in [39]) were calculated. All of them show that the convergence was reached. The results have been omitted to save space.

4.2. Model comparison criteria

In this paper, several measures have been evaluated for model comparison purposes (some traditional and others recently proposed) to assess the potential fit. One issue to consider in our comparison analysis is that each of them are deviance based, easily computed after the MCMC samples were obtained. Thus, we started with the main selection procedures such as the deviance information criterion (*DIC*) proposed by [40], the expected Akaike's information criteria (*EAIC*) and its Schwarz's Bayesian version (*EBIC*), all of them considered in [41] and [42]. As it was mentioned previously, these criteria are based on the *Posterior Mean of the Deviance*, $E\Big[D(\boldsymbol{\beta}, \lambda)\Big]$, and it is a measure of fit that can be approximated using the MCMC output by

$$Dbar = \frac{1}{M}\sum_{k=1}^{M} D(\boldsymbol{\beta}^{(k)}, \lambda^{(k)}),$$

where the index (k) represents the kth realisation of a total of M realisations, and

$$D(\boldsymbol{\beta}, \lambda) = -2\ln(p(\mathbf{y}|\boldsymbol{\beta}, \lambda)) = -2\sum_{i=1}^{n} \ln P(Y_i = y_i|\boldsymbol{\beta}, \lambda),$$

is the *Bayesian deviance*.

Hence, the measures $EAIC$, $EBIC$ and DIC can be estimated using the MCMC output through $EAIC = Dbar + 2p$, $EBIC = Dbar + plog(N)$ and $DIC = Dbar + \widehat{\rho_D} = 2Dbar - Dhat$, respectively, where p is the number of parameters in the model, N is the total number of observations and ρ_D, namely, the *effective number of parameters*, is given by $\rho_D = E\left[D(\boldsymbol{\beta}, \lambda)\right] - D\left[E(\boldsymbol{\beta}), E(\lambda)\right]$ where $D\left[E(\boldsymbol{\beta}), E(\lambda)\right]$ is the *deviance of posterior mean*. It is obtained by the means of the mean of the values generated from the posterior distribution as

$$Dhat = D\left(\frac{1}{G}\sum_{g=1}^{G} \boldsymbol{\beta}^{(g)}, \frac{1}{G}\sum_{i=1}^{G} \lambda^{(g)}, \right).$$

Additionally, we used other recent proposals to assess the model fit. Among them, two other criteria were the proposals given by [43,44], called Watanabe's information criterion (denoted by WAIC) and Bayesian predictive information criterion (or simply IC), respectively. Wanatabe's proposal can be viewed as an approximation to cross-validation [45], and the only difference is the computation of the parameter of complexity p_{WAIC}. In this paper, we will use the variance version to approximate this parameter given its stability properties. Hence, in this case,

$$p_{WAIC} = \sum_{i=1}^{N} Var\left(\ln(p(y_i|\boldsymbol{\beta}, \lambda))\right).$$

Therefore,

$$WAIC = Dhat + 2p_{WAIC}.$$

On the other hand, [46] proposed a model comparison alternative to obtain a measure that represents the predictive behaviour of the model. Furthermore, the author states that this measure will avoid the over-fitting problems that measures such as DIC can cause. This measure is simply approximated from the already computed measures by

$$IC = Dbar + 2p_D.$$

Given all the alternative models, the model that fits better to a data set is that one with the smallest value of each of the already computed measures.

5. An application

In this section, we will illustrate the performance of our models using the data Full Coverage available in [47] (see chapter 3). The term Full Coverage is referred to the contract that covers the costs associated to the damage of the insured vehicle produced by any accident or collision, car theft and vandalism, damage caused by a storm or flood or any impact made by an inanimate object. The data set consists of a sample with 4000 motor insurance policyholders selected randomly by gender. The response variable is whether the client bought a full coverage plan (success coded as 1) or not (failure coded as 0). Potential predictors to explain the coverage choice are given by individual characteristics, such as SEX (M=1, F=0), DRIVING AREA (Urban=1, Other=0), VEHICLE USE (Particular=1, Other=0), MARITAL STATUS of the policyholder (Single, Married and Others), AGE and SENIORITY in the company. It is expected that most of these features can identify a potential client of this plan. More details about the data set can be obtained from the chapter of the book. Our aim is to propose and compare different regression models to predict the probability that a policy holder chooses FULL COVERAGE, considering this set of predictors. The guidelines of our analysis are those exposed by the author, with the difference that we will discuss our results in terms of obtaining the best model fit.

Based on the results from [48] considering the L model, the covariate MARITAL STATUS does not seem to influence the explanation of buying a full vehicle coverage plan, thus we decided to compare two groups of models, a set of complete

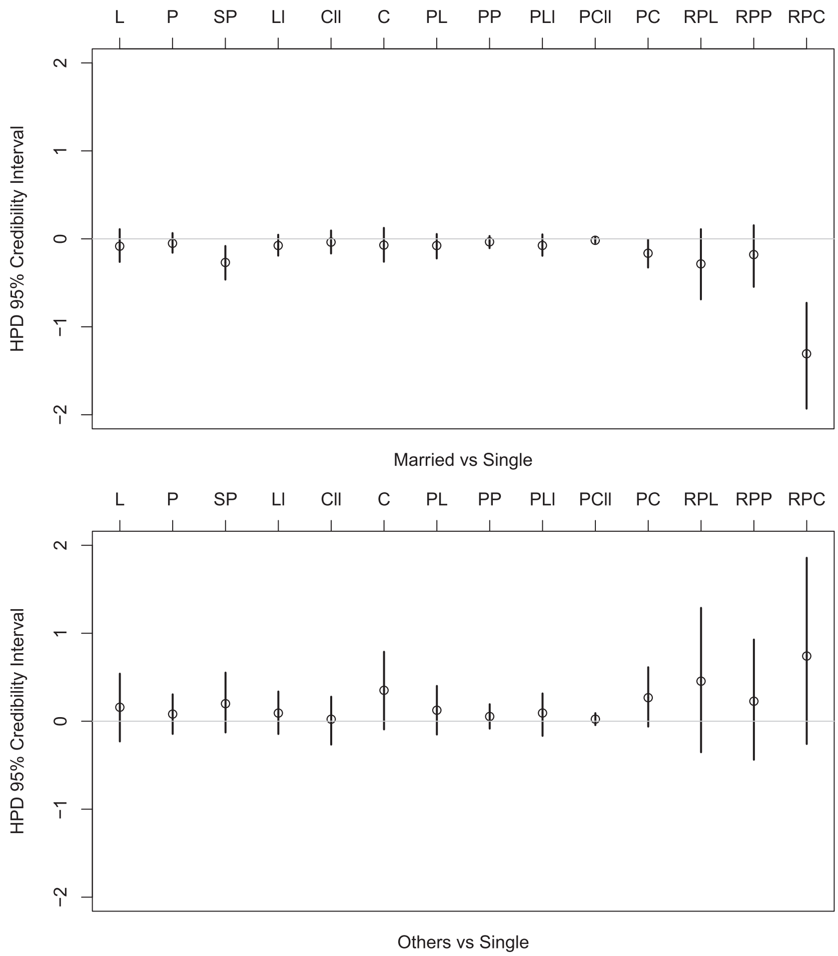

Figure 2. 95% HPD interval. Upper panel: Married vs Single for the fitted models. Lower panel: Other vs Single for the fitted models. L, logit; P, probit; SP, skew-probit; Ll, loglog; Cll, cloglog; C, cauchit; PL, power logit; PP, power probit; PLl, power loglog; PCll, power cloglog; PC, power cauchit; RPL, reversal power logit; RPP, reversal power probit; RPC, reversal power cauchit.

models where the covariate MARITAL STATUS was considered (Complete model) and a second set of models, where this variable was dropped out from the analysis (Reduced model). We started the analysis considering the Complete model to the different links we are exposing in this work, specifically where fit L: logit, P: probit, Ll: loglog, Cll: cloglog, C: cauchit and their correspondent power versions and reversal power versions of the logit, probit and cauchit were considered. For the sake of comparison, the skew probit model proposed by [18] was also considered.

Based on the results from our gold-standard, the outcomes from [48] for the L model were verified, that is, the covariate MARITAL STATUS does not seem to influence the explanation of buying a vehicle full coverage plan. Thus, as shown in Figure 2, the 95% HPD interval for the regression coefficients of categories or dummy variables of the MARITAL STATUS (Married vs Single and Others vs Single) for the Complete L model include the zero point. Moreover, considering this figure, all fitted Complete models reveal a non-influence of the categories of this variable with the exception of the skew probit and RPC models, where only the coefficient of regression associated with the comparison of Married vs Single is significant. Thus, we decided to support the decision of excluding this variable that could be important in the specialised literature, but this was not verified in the analysed data.

Furthermore, different model comparison criteria discussed in Section 3 were obtained for all the models to fit in each group of models: Complete and Reduced. The results are presented in the lines of Table II.

By considering the different measures of model comparison criteria, Complete and Reduce versions present a similar fit with the exception of the RPC model that presents the best result for the Complete version. Based on our aforementioned results, we decided for the reduced version of this model to continue with our analysis, because of the inconsistency that the inclusion of the marital status could produce in the final results. Considering reduced versions of the model, it was observed that all models, with the exception of the cloglog model and the cauchy model, are better models than the L model. In addition, the measures of comparing the models point out the power versions of the cauchit link as the best fit,

Table II. Model choice criteria for the full coverage dataset.

Link Model*	Complete model Dbar	DIC	WAIC2	EAIC	EBIC	IC	Reduced model Dbar	DIC	WAIC2	EAIC	EBIC	IC
L	4295.8	4303.7	4304.6	4311.8	4324.6	4311.6	4296.0	4302.0	4302.6	4308.0	4317.6	4308.0
P	4294.0	4301.9	4302.8	4310.0	4322.8	4309.7	4294.0	4300.0	4300.7	4306.0	4315.6	4306.0
SP	4149.1	4155.8	4161.1	4167.1	4181.5	4162.4	4161.6	4164.5	4176.0	4175.6	4186.8	4167.4
Ll	4250.8	4258.8	4260.2	4266.8	4279.6	4266.8	4252.1	4258.1	4259.4	4264.1	4273.7	4264.0
Cll	4387.8	4395.7	4397.1	4403.8	4416.6	4403.7	4386.2	4392.2	4393.0	4398.2	4407.8	4398.1
C	4341.2	4348.9	4351.0	4357.2	4370.1	4356.5	4343.5	4349.2	4350.7	4355.5	4365.1	4354.8
PL	4249.5	4258.3	4259.2	4267.5	4281.9	4267.0	4250.8	4257.5	4258.5	4264.8	4276.0	4264.3
PP	4270.4	4273.6	4279.8	4288.4	4302.8	4276.7	4269.6	4271.4	4276.7	4283.6	4294.8	4273.3
PLl	4250.9	4258.9	4260.4	4268.9	4283.3	4267.0	4252.0	4258.0	4259.4	4266.0	4277.2	4263.9
PCll	4280.8	4286.4	4290.0	4299.0	4313.2	4292.1	4279.8	4283.5	4287.0	4293.8	4305.0	4287.1
PC	4131.8	4140.8	4141.2	4149.8	4164.2	4149.9	4140.4	4147.3	4147.8	4154.4	4165.6	4154.3
RPL	4206.6	4215.6	4216.8	4224.6	4239.0	4224.6	4209.6	4216.6	4217.8	4223.6	4234.8	4223.5
RPP	4273.1	4276.0	4282.8	4291.1	4305.5	4278.9	4273.4	4276.1	4281.0	4287.4	4298.7	4278.7
RPC	3998.2	4007.2	4007.9	4016.2	4030.6	4016.3	4032.1	4039.1	4040.0	4046.1	4057.3	4045.1

Comparison among complete model considering marital status and a reduced version without this covariate.
*L, logit; P, probit; SP, skew-probit; Ll, loglog; Cll, cloglog; C, cauchit; PL, power logit; PP, power probit; PLl, power loglog; PCll, power cloglog; PC, power cauchit; RPL, reversal power logit; RPN, reversal power probit; RPC, reversal power cauchit.

Table III. Parameter estimates for the logit link model and the two best choices: power cauchit and reversal power cauchit.

Link	logit			Power cauchit			Reversal power cauchit		
Parameter	Mean	St. Dev.	95% HPD	Mean	St. Dev.	95% HPD	Mean	St. Dev.	95% HPD
Intercept	−0.29	0.49	(−1.31, 0.60)	0.92	0.45	(−0.02, 1.74)	1.80	2.03	(−2.13, 5.24)
MEN	−0.97	0.08	(−1.14, −0.81)	−1.01	0.08	(−1.18, −0.86)	−3.48	0.39	(−4.23, −2.73)
URBAN	1.18	0.08	(1.03, 1.33)	1.26	0.08	(1.09, 1.42)	4.64	0.50	(3.66, 5.59)
PRIVATE	1.10	0.48	(0.24, 2.09)	0.97	0.42	(0.18, 1.83)	4.35	2.01	(0.83, 7.97)
AGE	−0.06	0.004	(−0.07, −0.05)	−0.06	0.004	(−0.06, −0.05)	−0.20	0.02	(−0.24, −0.16)
SENIORITY	0.13	0.007	(0.12, 0.15)	0.14	0.008	(0.12, 0.15)	0.48	0.05	(0.39, 0.58)
λ				2.30	0.14	(2.05, 2.57)	0.35	0.02	(0.32, 0.38)

St. Dev., standard deviation.

where the RPC model obtained the lowest values in all six criteria. Although we prefer the reduced versions of this model as an alternative to our gold-standard model based on the exclusion of MARITAL STATUS by [48], note that the complete RPC model can also be a good alternative.

Parameter estimates for the L model and the two best choices: PC and RPC models considering the reduced version are presented in Table III.

It can also be observed that the parameter associated with variable AGE continues being negative for all models, reaching a lower value for the RPC link, that is, it can be argued that the older the policyholder is, the lower the chances are of him/her buying a full coverage plan. The parameter for the selected link model for MEN considers a negative and the lowest value for all the models. Therefore, as expected, men would be less prone to buying full coverage than women. On the other hand, customers driving in an urban area have higher probabilities of buying a full coverage compared with those who drive in rural areas, obtaining a higher influence in the reversal version of the power cauchit link. Indeed, the SENIORITY in the company has a positive influence.

Its parameter continues with a positive effect across the models; therefore, customers who have been in the company for a longer period of time have a higher probability of buying full coverage than recent or new customers. Finally, the variable PRIVATE is also of interest to this class of problems. It can be observed that vehicles for private use are more likely for a COMPLETE COVERAGE plan than vehicles for commercial use. As argued by [48], these effects are clearly good predictors for buying a full coverage plan. Even more, our proposal seems to reinforce covariates with higher and lower coefficient estimates than those covariates included in their model.

On the other hand, because we are interested in comparing the predictive assessment of our link models, we will use standard measures, such as rate of good classification, sensitivity, specificity and area under the curve using the empirical and bi-normal approximations [49]. Predictive measures for the logit link model and the two best choices, power cauchit and reversal power cauchit, are shown in Table IV.

Table IV. Predictive measures for the logit link model and the two best choices: power cauchit and reversal power cauchit.

Link	Predicted	Observed		Rate of good classification	Sensitivity	Specificity	Empirical AUC	Binormal AUC
logit		**0**	**1**					
	0	2446	1150	0.67	0.17	0.94	0.79	0.76
	1	167	237					
Power		**0**	**1**					
cauchit	**0**	2079	607	0.71	0.56	0.80	0.79	0.78
	1	534	780					
Reversal		**0**	1					
power	**0**	1966	452	0.72	0.67	0.75	0.79	0.80
cauchit	**1**	647	935					

AUC, area under the curve.

For the models under analysis (L, PC and RPC models), a 2×2 subtable of frequencies between predicted value and the observed value was elaborated as it is showed in Table IV. For each subtable, the code **1** denotes the observed or predicted success (the client bought a full coverage plan), and the code **0** denotes the observed or predicted failure (the client not bought a full coverage plan). For each subtable, the frequencies in the diagonal cells represent the two types of correct prediction: true positives (TP) or correctly identified and true negatives (TN) or correctly rejected. Other cells in the subtable represent the two types of incorrect prediction: false positives (FP) or incorrectly identified and false negatives (FN) or incorrectly rejected. Therefore, the *rate of good classification* or accuracy is defined as $(TP + TN)/(TP + FP + FN + TN)$, the *Sensitivity* or the true positive rate is defined as $TP/(TP + FN)$, and the *Specificity* or true negative rate is defined as $TN/(TN + FP)$.

For each model, the rate of good classification measures the proportion of clients that bought (or not) a full coverage plan that is correctly identified as such. Sensitivity measures the proportion of clients that bought a full coverage plan that is correctly identified as such and specificity measures the proportion of clients that did not buy a full coverage plan that is correctly identified as such. Finally, as mentioned by [48], the ROC analysis provides a way to select possible optimal models and to discard suboptimal ones based on classification performance at various threshold levels. A subproduct of this analysis, the area under the curve, was used in this work to identify the best performance. As mentioned in the literature, a model with the largest area under the curve should be preferred to other possible models.

According to our computations (Table IV), it can be observed that the best results are exposed by the reversal power cauchit link, considering that it improves the rate of good classification, the sensitivity of the prediction and the area under the curve, when we are interested in predicting clients who will buy a FULL COVERAGE plan. It is interesting to see the increment in all of the measures while we evaluate their power assuming the different models. Specificity clearly has the opposite behaviour because of the unbalanced quantities of successes and failures of the sample, 34% and 66%, respectively, and showing a reduction of its measure while its sensitivity improves.

This effect is common in most of the predictive models working with unbalanced samples: a classification method with a high specificity will have a very low rate of false alarms, caused by classifying normal events (absence of the rare event) as abnormal (rare event) [50], that is, when the event in the response variable is rare, the ROC curve will be dominated by the minority class and thus insensitive to the change of true positive rate and consequently give high values of specificity, which provides little information for model diagnosis as is the case of the logit model here.

Finally, by considering the analysis conducted, the reversal power cauchit link model presents the best performance to fit the data full coverage, where the probabilities estimated are obtained considering

$$\hat{p}_i = 1 - \left(0.5 + \frac{1}{\pi} arctan\Big(-(1.80 - 3.48\text{MEN} + 4.64\text{URBAN} + 4.35\text{PRIVATE} - 0.20\text{AGE} + 0.48\text{SENIORITY})\Big)\right)^{0.35}$$

where particular cases can be obtained substituting the values of covariates.

We consider that this final model ensures a good fit and covariates with well-behaved parameters, from a Bayesian point of view, and fulfilled their significant roles in the analysis. Moreover, considering this model, all parameters associated with the predictors were significant because the HPD intervals excluding zero indicate a significant covariate effect on the response variable. Furthermore, their performance in terms of predictive modelling seems to be better than the other choices.

Specifically, we found that men would be less prone to buying full coverage than women because the parameter associated with the variable MEN is negative (−3.48) and significant. It can be observed that the parameter associated with the variable AGE is also negative (−0.20) and significant indicating that the older the policyholder is, the lower the chances of him/her buying full coverage. Likewise, customers driving in an urban area have a higher probability of buying full coverage compared with those who drive in rural areas. This is because the parameter for URBAN is positive (4.64) and significant. Furthermore, PRIVATE is positive (4.35) and significant indicating that policyholders who own a private use vehicle have a larger probability of buying FULL COVERAGE than those insured with a vehicle for commercial use. In addition, SENIORITY in the company has a positive influence. Its parameter is positive (0.48) and significant. This means that those customers who have been in the company for a longer period of time have a higher probability of buying full coverage than recent or new customers. Marital status of the policyholder, which separates single, married and other individuals, does not seem to have a clear effect on this case as observed in Figure 2. Finally, we found a penalty parameter (0.35) indicating and significant indicating that the probability of buying full coverage presents a penalisation factor associated with the predictors.

6. Final comments

In this paper, we introduced power and reversal power as new links for binary regression that need asymmetrical links as is the case of unbalanced data. In these links, the λ parameter is a shape parameter of the considered cdf $F(.)$, which is independent of the interval of variation of the linear predictor. In addition, known links of generalised linear models are particular cases. Two versions of the likelihood function were presented, and classical and Bayesian estimation can be implemented by considering them. We introduced a Bayesian approach, which can be easily implemented, for instance in proc mcmc of SAS, WinBUGS, OpenBUGS or RJAGS.

Moreover, finance and insurance applications that use logistic regression as a default procedure must also be analysed by considering the links introduced in this paper in order to verify which one is the most suitable for fitting data. For instance, in the insurance dataset considered here, the best results are obtained by an RPC link, leading to improvements in the rate of good classification, sensitivity of the prediction and area under the curve. This is especially so when we are interested in predicting clients who will buy a full coverage plan. Other applications include Big data in the industry such as in [51] or consumer heterogeneity such as in [52].

Extensions of the methods developed in this paper for mixed ordinal response data followed by considering this model in terms of cumulative probabilities in such a way that the conditional probability of a response in category c can be obtained as the difference of two conditional cumulative probabilities,

$$P(Y_{ij} = c|\, \boldsymbol{b}_i, \boldsymbol{x}_{ij}, z_{ij}) = F_\lambda(\eta_{ijc}) - F_\lambda(\eta_{ijc-1})$$

where

$$\eta_{ijc} = \gamma_c - \left[\boldsymbol{x}'_{ij}\boldsymbol{\beta} + z'_{ij}\boldsymbol{b}_i\right],$$

and we have $c-1$ strictly increasing thresholds γ_c. This yields a random-effects ordinal regression model for multilevel analysis. Furthermore, model multivariate binary data based on copulas using the links proposed here can be defined as topics for future research. In addition, applications in various areas of interest such as binomial models, multilevel models, item response theory, calibration model, dose response, and mixture models in survival analysis should be suggested. Moreover, the use of criteria for elicitation of priors for the probabilities pi can be introduced. Finally, by considering another baseline pdf g(.) defined in the real line with a unique mode, continuous and log concave, other links can be obtained, for example, using the links based on the generalised extreme value function [52].

Acknowledgements

The authors would like to thank the Editor, Associate Editor, and an anonymous referee for the valuable comments that considerably improved the first version of the manuscript. J.L. Bazán and A. K. Suzuki are funded by FAPESP, Brazil, F. Torres-Avilés was funded with a post-doctoral fellowship by Becas Chile and F. Louzada is funded by CNPq and FAPESP.

References

1. Altman EI, Saunders A. Credit risk measurement developments over the last 20 years. *Journal of Banking and Finance* 1998; **21**:1721–1742.
2. Dey DK, Ghosh SK, Mallick BK. *Generalized Linear Models: A Bayesian Perspective*. Marcel Dekker: New York, 1999.
3. Steenackers A, Goovaerts MJ. A credit scoring model for personal loans. *Insurance: Mathematics and Economic* 1989; **8**:31–34.
4. Hardle W, Mammen E, Muller M. Testing parametric versus semiparametric modeling in generalized linear models. *Journal of the American Statistical Association* 1998; **93**(44):1461–1474.
5. Finlay SM. Towards profitability a utility approach to the credit scoring problem. *The Journal of the Operational Research Society* 2008; **59**(7):921–931.

6. Kočenda E, Vojtekb M. Default predictors in retail credit scoring: evidence from Czech banking data. *Emerging Markets Finance and Trade* 2011; **47**(6):80–98.
7. Louzada F, Ferreira-Silva PH, Diniz CAR. On the impact of disproportional samples in credit scoring models: an application to a Brazilian bank data. *Expert Systems with Applications* 2012; **39**(9):8071–8078.
8. Wu DD, Olson DL, Luo C. A decision support approach for accounts receivable risk management. *IEEE Transactions on Systems Man and Cybernetics Systems* 2014; **44**(12):1624–1632.
9. Bermúdez LI, Pérez JM, Ayuso M, Gómez E, Vásquez FJ. A Bayesian dichotomous model with asymmetric link for fraud in insurance. *Insurance: Mathematics and Economics* 2008; **42**(2):779–786.
10. Jha SA, Guillen MB, Westland JC. Employing transaction aggregation strategy to detect credit card fraud. *Expert Systems with Applications* 2012; **39**(16):12650–12657.
11. Louzada F, Ara A. Bagging k-dependence probabilistic networks an alternative powerful fraud detection tool. *Expert Systems with Applications* 2012; **39**(14):11583–11592.
12. Owen AB. Infinitely imbalanced logistic regression. *Journal of Machine Learning Research* 2007; **8**:761–773.
13. Li Z. Research on distinguish the accounting information distortion based on the principle components-logistic regression model. *International Journal of Security and its Applications* 2014; **8**:37–50.
14. Albert JH, Chib S. Bayesian analysis of binary and polytomous response data. *Journal of the American Statistical Association* 1993; **88**(422): 669–679.
15. Collet D. *Modelling Binary Data* 2nd ed. Chapman & Hall/CRC: Boca Raton, USA, 2003.
16. Chen MH, Dey DK, Shao QM. A new skewed link model for dichotomous quantal response data. *Journal of the American Statistical Association* 1999; **94**(448):1172–1186.
17. Bazán JL, Branco MD, Bolfarine H. A skew item response model. *Bayesian Analysis* 2006; **1**:861–892.
18. Bazán JL, Bolfarine H, Branco MD. A framework for skew-probit links in binary regression. *Communications in Statistics - Theory and Methods* 2010; **39**(4):678–697.
19. Chen MH, Dey DK, Shao QM. Bayesian analysis of binary data using skewed logit models. *Calcutta Statistical Association Bulletin* 2001; **51**:201–202.
20. Czado C, Santner TJ. The effect of link misspecification on binary regression inference. *Journal of Statistical Planning and Inference* 1992; **33**(2):213–231.
21. Czado C. Bayesian inference of binary regression models with parametric link. *Journal of Statistical Planning and Inference* 1994; **41**(2): 121–140.
22. Guerrero VM, Johnson RA. Use of the Box-Cox transformation with binary response models. *Biometrika* 1982; **69**(2):309–314.
23. Prentice RL. A Generalization of the probit and logit methods for dose-response curves. *Biometrika* 1976; **32**:761–768.
24. Stukel T. Generalized logistic models. *Journal of the American Statistical Association* 1988; **83**(402):426–431.
25. Bazán JL, Romeo JS, Rodrigues J. Bayesian skew-probit regression for binary response data. *Brazilian Journal of Probability and Statistics* 2014; **28**(4):467–482.
26. Basu S, Mukhopadhyay S. Binary response regression with normal scale mixtures links. In *Generalized Linear Models: A Bayesian Perspective*, Dey DK, Ghosh SK, Mallick BK (eds). Marcel Dekker: New York, 2000; 231–242.
27. Kim S, Chen MH, Dey DK. Flexible generalized t-link models for binary response data. *Biometrika* 2008; **95**(1):93–106.
28. Kim HJ. Binary regression with a class of skewed t link models. *Communications in Statistics: Theory Methods* 2002; **31**(10):1863–1886.
29. Nadarajah S. The skew logistic distribution. *AStA Advances in Statistical Analysis* 2009; **93**:187–203.
30. Johnson NL, Kotz S, Balakrishnan N. *Continuous Univariate Distributions*, vol. 2. Wiley: New York, 1995.
31. Nagler J. Scobit: an alternative estimator to logit and probit. *American Journal Political Science* 1994; **38**(1):230–255.
32. Jones MC. Family of distributions arising from distribution of order statistics. *Test* 2004; **13**(1):1–43.
33. Gupta RD, Gupta RC. Analysing skewed data by power normal model. *Test* 2008; **17**:197–210.
34. Kundu D, Gupta RD. Power-normal distribution. *Statistics* 2013; **47**(1):110–125.
35. Eugene N, Lee C, Famoye F. Beta Normal distribution and its applications. *Communication and Statistics: Theory and Methods* 2002; **31**(4): 497–512.
36. Abanto-Valle CA, Bazán JL, Smith AC. State space mixed models for binary responses with skewed inverse links using jags. *Technical Report N 278.* Departamento de Métodos Estatísticos UFRJ. Brazil. 2015.
37. Zellner A, Rossi PE. Bayesian analysis of dichotomous quantal response models. *Journal of Econometrics* 1984; **25**(3):365–393.
38. SAS Institute Inc. 2015. SAS/STAT 9.3®. *User´s Guide*, Chapter 73 MCMC Procedure. SAS Institute Inc: Cary, NC, 2015.
39. Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. Bayesian measures of model complexity and fit. *Journal Royal Statistical Society, Series B* 2002; **64**(3):583–639.
40. Carlin BP, Louis TA. *Bayes and Empirical Bayes Methods for Data Analysis* 2nd ed. Chapman and Hall: Boca Raton, FL, 2000.
41. Brooks SP. Discussion on the paper by Spiegelhalter, Best, Carlin, and van der Linde. *Journal of the Royal Statistical Society Series B* 2002; **64**(3):616–618.
42. Watanabe S. Asymtotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research* 2010; **14**:867–897.
43. Watanabe S. A widely applicable Bayesian information criterion. *Journal of Machine Learning Research* 2013; **14**:867–897.
44. Gelman A, Hwang J, Vehtari A. Understanding predictive information criteria for Bayesian models. *Statistics and Computing* 2014; **24**: 997–1016.
45. Ando T. Predictive bayesian model selection. *American Journal of Mathematical and Management Sciences* 2011; **31**(1-2):13–38.
46. Frees E, Derrig RA, Meyers G. *Predictive Modeling Applications in Actuarial Science*. Cambridge University Press: New York, 2014.
47. Guillén M. Regression with categorical dependent variables. In *Predictive Modeling Handbook for Actuarial Science*, Frees EW, Derrig R, Meyer G (eds), vol. I. Cambridge University Press: New York, 2014; 65–86.
48. Gonen M. *Analyzing Receiver Operating Characteristic Curves Using SAS*. SAS Institute: Cary, NC, 2007.
49. Marques de Sá JP. *Applied Statistics Using SPSS, STATISTICA, MATLAB and R*. Springer-Verlag Berlin Heidelberg: New York, 2007.

50. Kruegera DC, Montgomery DC. Modeling and analyzing semiconductor yield with generalized linear mixed models. *Applied Stochastic Models in Business and Industry* 2014; **30**(6):691–707.
51. Pancras J, Dey DK. A comparison of generalized multinomial logit and latent class approaches to studying consumer heterogeneity with some extensions of the generalized multinomial logit model. *Applied Stochastic Models in Business and Industry* 2011; **27**:567–578.
52. Wang X, Dey DK. Generalized extreme value regression for binary response data an application to B2B electronic payments system adoption. *The Annals of Applied Statistics* 2010; **4**(4):2000–2023.