Brazilian Journal of Probability and Statistics
2012, Vol. 26, No. 4, 344–357
DOI: 10.1214/11-BJPS143

Latent residual analysis in binary regression with skewed link

Rafael B. A. Farias and Marcia D. Branco

Universidade de São Paulo

Abstract. Model diagnostics is an integral part of model determination and an important part of the model diagnostics is residual analysis. We adapt and implement residuals considered in the literature for the probit, logistic and skew-probit links under binary regression. New latent residuals for the skew-probit link are proposed here. We have detected the presence of outliers using the residuals proposed here for different models in a simulated dataset and a real medical dataset.

1 Introduction

Diagnostic techniques are indispensable tools to check the fit of the models. In particular, residuals are used to check whether or not the model assumptions are satisfied by the data. Moreover, residuals are also useful to help in outlier detection, which can provide disproportional interference in inferential results. In the classical approach for binary regression models, outlier detection is usually based on the ordinary residual $y_i - \hat{p}_i$, where $\hat{p}_i = F(\mathbf{x}^t\widehat{\boldsymbol{\beta}})$ is the ith fitted observation, $F(\cdot)$ is a known link function, x_i is a vector of covariates and $\widehat{\boldsymbol{\beta}}$ is the maximum likelihood estimator (MLE) of $\boldsymbol{\beta}$. There are several types of residuals in literature, among them, Pearson, deviance and Anscombe residuals (see, e.g., McCullagh and Nelder (1989)). However, these residuals have unknown sampling distributions due to the discrete nature of the response variable. This affects the interpretation of the residual plots and the outlying detection. Albert and Chib (1995) proposed a Bayesian latent residual in binary regression which has continuous distribution.

In this work we describe some types of residuals used in Bayesian binary regression model framework. Among them, we present latent residuals obtained through scale mixture of normals. For skew-probit model (Chen et al. (1999)) we propose two types of latent residuals. The first generalize the latent residual proposed by Albert and Chib (1995) for symmetric models. In the skew-probit setting, this residual can have expectation different from zero. Therefore, we propose the second residual, which is based on the stochastic representation of the skew-normal distribution given in Sahu, Dey and Branco (2003). When the shape parameter is unknown in the skew-probit link we suggest using a residual that does not depend on unknown parameters and has uniform distribution.

Key words and phrases. Binary regression, MCMC algorithm, residual analysis, skew-probit models.

Received January 2011; accepted February 2011.

This paper is organized as follows. First, we review some symmetrical and skewed link functions for binary response models, with special attention to the skew-probit model. Section 4 presents all residuals developed in this work. In Section 5 we present a residual analysis in a simulated dataset. An application to a real dataset is presented and discussed in Section 6. Finally, some conclusions are presented in Section 7.

2 Binary regression

Consider $\mathbf{y} = (y_1, \ldots, y_n)^t$ a set of binary responses (0/1), where $y_1, \ldots, y_n$ are independent random variables. Consider also that $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})^t$ is a set of previous fixed quantities associated with y_i, where x_{i1} can be equal to 1 (that corresponds to the intercept). The binary regression model with independent responses is given by

$$p_i = \mathbb{P}(y_i = 1) = F(\mathbf{x}_i^t \boldsymbol{\beta}), \tag{2.1}$$

where F^{-1} is a known function that linearizes the relationship between the success probability and the covariates, and $\boldsymbol{\beta}$ is a p dimensional vector of regression coefficients. In the generalized linear models theory (GLM), the function F^{-1} is called link function. This link function can depend on additional parameters.

The most popular binary regression models are probit and logistic models, adequate when we do not have an evidence that the probability of sucess increases in a different rate than it decreases.

A large class of continuous, unimodal and symmetrical distributions with support on the real line is given by

$$\mathcal{F}_S = \left\{ F(\cdot) = \int_{[0,\infty)} \Phi\left(\frac{\cdot}{\delta}\right) dG(\delta), G \text{ is a cdf on } [0, \infty) \right\}, \tag{2.2}$$

where G is a cdf on interval $[0, \infty)$. Therefore, each member in this class are symmetric distributions when δ has a continuous distribution. Some particular cases are: probit model: when δ has degenerated distribution on point 1; Student-t model: when δ follows a *gamma* distribution; logistic model: when $\delta = 4\psi^2$ and ψ follows a Kolmororov–Smirnov distribution (Devroye (1986)).

Considering $F \in \mathcal{F}_S$ in model (2.1) and using a vector of auxiliary variables $\mathbf{z} = (z_1, \ldots, z_n)^t$, it is obtained the following representation for binary regression with symmetrical link belonging to a normal mixture of scale:

$$y_i = \begin{cases} 1 & \text{if } z_i > 0, \\ 0 & \text{otherwise,} \end{cases} \tag{2.3}$$

with

$$z_i | \boldsymbol{\beta}, \delta \sim \mathcal{N}(\mathbf{x}_i^t \boldsymbol{\beta}, \delta(\psi)) \quad \text{and} \quad \psi \sim g(\psi), \tag{2.4}$$

where $\delta(\psi) > 0$ for all $\psi > 0$, $\delta(\cdot)$ is a bijective function and $g(\cdot)$ is a density of continuous mixture.

3 Skewed models using auxiliary variables

Symmetric links can be inappropriate when the probability of sucess approaches zero at a different rate than it approaches one. Skewed links can be obtained when the link function is the inverse of the cdf of an asymmetrical distribution.

Chen (2004) carried out a simulation study to investigate the importance of the choice of a link function in binary response variables prediction. He considered two simulation schemes; (i) the dataset were generated according to the probit model; and (ii) the data were generated according to the C log–log model. In both situations, the probit, logit and C log–log models were fitted. The author observed that when the true link function is probit, there are a few differences between the probit and logit models. However, the C log–log model is inadequate. On the other hand, when the true link function is C log–log, the symmetric models were clearly inadequate, and the logit model had a better behavior than the probit model in this case. This happened because for small probabilities of sucess, the logit and C log–log links become very close, decreasing to zero faster than the probit link. The author concluded in this empirical study that the choice of the link function is very important, and in case it is badly specified, it can provide poor predictions.

3.1 Skew-probit regression

We considered the following class of distributions

$$\mathcal{F}_{\mathcal{A}} = \left\{ F_\lambda(z) = \int_{[0,\infty)} F(z - \lambda w)\, dG(w) \right\}, \tag{3.1}$$

where $\lambda \in \mathbb{R}$, F is a cdf of a symmetrical distribution around zero with support on the real line, and G is a cdf of an asymmetrical distribution on $[0, \infty)$. The model defined in (3.1) has some attractive properties: (a) when $\lambda = 0$ or G is a degenerated distribution, the model reduces to the model with a symmetrical link; (b) the skewness of the link function can be characterized by λ and G; and (c) heavy and light tails for F_λ can be obtained according to the choice of F.

A skewed binary regression model belonging to class (3.1) can be defined by

$$\begin{aligned} y_i &= \begin{cases} 1 & \text{if } z_i > 0, \\ 0 & \text{otherwise,} \end{cases} \\ z_i &= \mathbf{x}_i^\top \boldsymbol{\beta} + \varepsilon_i^*, \end{aligned} \tag{3.2}$$

where

$$\varepsilon_i^* = -\lambda w_i + \varepsilon_i, \qquad \varepsilon_i \sim F \quad \text{and} \quad w_i \sim G.$$

A particular case of this model is obtained when we consider that F is the cdf of a normal distribution and G is the cdf of a half-normal distribution. This results in the skew-probit model of Chen and Dey (1998) and Chen et al. (1999).

4 Residual analysis

Residuals in regression models are based mainly on comparison between the ith response variable and its expectation. The first binary residual considered here will be the residual $r_i = y_i - p_i$, where $p_i = F(\mathbf{x}_i^t\boldsymbol{\beta})$. This residual has continuous distribution and it is a function of the vector of parameters $\boldsymbol{\beta}$. Then, the posterior knowledge that we have about $\boldsymbol{\beta}$ will be reflected on its posterior residual distribution. The support of the posterior distribution of r_i is the interval $(y_i - 1, y_i)$. Then, if the value of y_i and the posterior distribution of p_i are in conflict, the posterior distribution of this residual will be concentrated towards extreme values. Therefore, the observation $y_i = 0$ can be suspicious of being an outlier if the posterior distribution of r_i is concentrated around -1. On the other hand, the observation $y_i = 1$ can be suspicious of being an outlier if the posterior distribution of r_i will be concentrated towards the endpoint 1.

Farias and Branco (2011) describes some ways to obtain a sample from the posterior distribution of the vector of parameters $\boldsymbol{\theta}$ for probit and logistic models ($\boldsymbol{\theta} = \boldsymbol{\beta}$) and for the skew-probit model $[\boldsymbol{\theta} = (\boldsymbol{\beta}, \lambda)^t]$. For example, let $\{\boldsymbol{\theta}^{(t)}\}_{t=1}^T$ be a sample of size T from the posterior distribution of $\boldsymbol{\theta}$, since y_i and x_i are known in order to obtain a sample from these residuals it is enough to consider $r_i^{(t)} = y_i - F_{\theta^{(t)}}(\mathbf{x}_i^t\boldsymbol{\beta}^{(t)})$, $t \leq T$. From this sample we can obtain descriptive measures and estimated densities from the posterior distribution of the residuals. Albert and Chib (1995) proposed using graphics boxplots to represent the samples generated from posterior distributions of the residuals against fitted probabilities. These graphics jointly considered can help us to view unusual residuals.

4.1 Latent residuals

An alternative way to defining Bayesian residuals is based on the use of latent variables. The model given in (3.2) provides a general representation for a class of the binary regression models that encloses probit, logit and skew-probit models. We can define several types of latent residuals under the class of link function (3.1). The first latent residual considered here is

$$\varepsilon_i(z_i, \boldsymbol{\beta}) = z_i - \mathbf{x}_i^t\boldsymbol{\beta}. \tag{4.1}$$

This residual was defined by Albert and Chib (1995) in the binary response framework for outlying detection in symmetrical models. The latent variable z_i here can represent, for example, an insect's tolerance to a pesticide in the bioassay setting. However, it can also be useful for diagnostic in skewed links.

For symmetrical links, the residual given in (4.1) is distributed with the same distribution used to define the link function. For example, when the model considered is the probit, the residual ε_i is normally distributed a priori. On the other hand, when the model is the skew-probit given in (3.2), the residual (4.1) has a skew-normal prior distribution. However, in this case the shape parameter of the residual distribution has an opposite sign of the distribution used to define the link function. Note that the distribution of the residual ε_i is a function of the shape parameter λ in the skew-probit model. If λ is unknown, then the distribution of this residual is a function of the prior distribution of the shape parameter. For example, when λ is normally distributed a priori then the prior distribution of ε_i is unknown and the interpretation of the size of the residuals is affected. A solution for this problem is to use a point estimate for this parameter or to define residuals which do not depend on prior parameters. In this way, we use the stochastic representation for skew-normal distribution given in Sahu, Dey and Branco (2003) and define the following residual for the skew-probit case

$$\varepsilon_i^*(z_i, w_i, \boldsymbol{\beta}, \lambda) = z_i - (\mathbf{x}_i^t \boldsymbol{\beta} - \lambda w_i). \tag{4.2}$$

This residual has symmetrical distribution F given in (3.1) with support on the real line. If the adopted model is symmetrical, the skewness parameter λ is equal to zero and the residual given in (4.2) reduces to the residual defined in (4.1).

In symmetrical models the residuals $\varepsilon_i(z_i, \boldsymbol{\beta}), i = 1, \dots, n$, can have prior variances different from one. In order to have variance one, for the scale mixture of normal models given in (2.2), we define a standard residual

$$\tau_i(\boldsymbol{\beta}, \delta_i) = \frac{z_i - \mathbf{x}_i^t \boldsymbol{\beta}}{\sqrt{\delta(\psi)}}, \tag{4.3}$$

where $\delta(\psi)$ and ψ are given in (2.4). The model (4.3) conditioned on $\{\boldsymbol{\beta}, \delta\}$ has a standard normal distribution. Therefore, we can use this residual and assume that an observation is an outlier by using the same criteria for the probit and skew-probit setting in the residuals ε and ε^*, respectively. Namely, we assume that an observation can be an outlier if the probability of this residual being bigger in absolute value than 1.64 is high. This value is chosen due to the normality of the prior distribution, where this probability is around 0.10.

4.2 Latent residual uniformly distributed

The latent residual ε_i presented in (4.1) can depend on unknown parameters, as it is the case of the skew-probit model which depends on the shape parameter λ. Then, the prior distribution of this residual depends on the prior distribution of λ. Then, these values of residuals can not be compared directly across different models, because they have different prior distributions. In order to built residuals with the same probability distribution it is used the inverse probability integral transformation method on the previous latent residuals.

The first uniformly distributed residual proposed here is related with ε in (4.1) and given by

$$u_i = F^{-1}(z_i - \mathbf{x}_i^t \boldsymbol{\beta}), \tag{4.4}$$

where F is the cumulative distribution function which provides the link function for the binary model and can depend on unknown parameters. The function F depends on the shape parameter of the skew-probit model. This residual is uniformly distributed on the interval (0, 1).

We also apply the inverse probability integral transform method in the others residuals showed in previous section. For logit model we define the residual

$$u_i^* = \Phi(\tau_i; 0, 1), \tag{4.5}$$

where τ_i is given by (4.3), and for the skew-probit model we define one more residual as follows

$$u_i^* = \Phi(\varepsilon_i^*; 0, 1) \tag{4.6}$$

where ε_i^* is given by (4.2). An observation is suspicious of being an outlier by using this residual if its posterior residual distribution is significantly different from its prior residual distribution (uniform distribution). A way to check that is comparing tail probabilities in each distribution. Computing the posterior probabilities of the residual u_i is out of the interval $(\alpha/2, 1-\alpha/2)$, where α is the prior probability of residual u_i being outside the interval $(\alpha/2, 1-\alpha/2)$.

5 A simulated dataset

In order to illustrate the behavior of the residuals presented here, we work with a dataset simulated from a model that induces a few outliers. Consider the success probability given by $p_i = F(\mathbf{x}_i^t \boldsymbol{\beta} | \mu = 0, \sigma = 1, \lambda, \nu)$, where F is a cdf of a skew-t distribution. The skew-t density function is

$$f(z) = \frac{2}{\sqrt{1+\lambda^2}} t_\nu\left(\frac{z}{\sqrt{1+\lambda^2}}\right) T_{\nu+1}\left[\lambda z \sqrt{\frac{\nu+1}{(1+\lambda^2)\nu + z^2}}\right], \tag{5.1}$$

where t_α and T_α are, respectively, the pdf and cdf of a standard Student-t distribution with ν degrees of freedom. It has the propriety to be skewed and heavy tailed. For more details about this distribution see Azzalini and Capitanio (2003).

Denote by $\boldsymbol{\theta} = (\boldsymbol{\beta}, \nu, \lambda)^t$, where $\boldsymbol{\beta} = (\beta_0, \beta_1)^t$ are the regression parameters, ν the degrees of freedom and λ the skewness parameter.

The simulated dataset contains 600 observations divided by 10 categories with 60 observations each, using the following values for the parameters $\boldsymbol{\beta} = (1, 3)^t$, $\nu = 5$ and $\lambda = -3$. It is possible to select the success probabilities on several ways. Chen (2004) suggests choosing the p_i's equally spaced. However, in order to have an artificial data that reflects properties of real ones, we decided for not choosing the p_i's equally spaced. As it is typically observed in dose–response model, the covariates $x_1, \dots, x_n$ are increase and equally spaced. This results, in general, in not equally spaced p_i's. In order to achieve this goal, we proceed following these

Table 1 *Simulated dataset*

Category	x_i	$y_{i\bullet}$	n_i
1	−1.1901	5	60
2	−1.0883	4	60
3	−0.9866	8	60
4	−0.8848	12	60
5	−0.7831	13	60
6	−0.6813	20	60
7	−0.5796	28	60
8	−0.4779	36	60
9	−0.3761	49	60
10	−0.2744	56	60

steps: (i) we choose the smallest and the largest success probabilities, denoted here by p_1 and p_n, respectively; (ii) we calculate the initial values $x_1^0 = F^{-1}(p_1)$ and $x_n^0 = F^{-1}(p_n)$; (iii) we calculate the remaining initial values the following way, $x_i^0 = x_{i-1}^0 + (x_n^0 - x_1^0)/(n-1), i = 2, \dots, n-1$, and finally; (iv) we calculate the success probabilities $p_i = F(x_i^0), i = 1, \dots, n$. The smallest and the largest success probabilities chosen were $p_1 = 0.05$ and $p_{30} = 0.95$. Finaly, we obtained $x_1, \dots, x_n$ by

$$x_i = \frac{F^{-1}(p_i) - \beta_0}{\beta_1}.$$

The simulated dataset is presented in Table 1.

Considering the following prior distribution for $\boldsymbol{\beta} \sim \mathcal{N}_2(\mathbf{0}, 1000\mathbf{I}_n)$ and $\lambda \sim \mathcal{N}(0, 1000)$, we fit the probit, logit and skew-probit models for this dataset. The results are presented in Table 2 and Figure 1.

Table 2 *Posterior inference for regression parameters for the simulated dataset*

Model	Parameters	Mean	Median	95% HPD Lower	95% HPD Upper
Probit	β_0	1.91	1.91	1.57	2.20
	β_1	3.15	3.17	2.68	3.56
Logit	β_0	3.26	3.04	2.76	3.87
	β_1	5.47	5.14	4.56	6.22
Skew-probit	β_0	2.35	2.25	1.37	3.74
	β_1	8.96	8.78	4.51	13.72
	λ	−4.05	−4.01	−6.50	−1.67

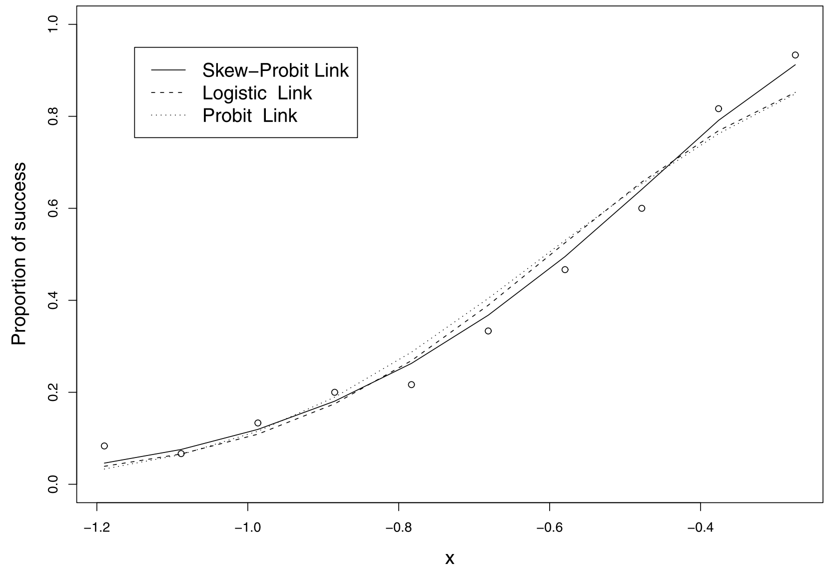

Figure 1 *Proportion of success and fitted curves for the simulated dataset.*

In order to compare the models we obtained the values of the Deviance information criterion (DIC), the Bayes factor (Kass and Raftery (1995)) and the pseudo-Bayes factor (Geisser and Eddy (1979)). These values are presented in Table 3.

Note that the skew-probit model outperforms logistic and probit models for all criteria used. This result is an indicative of possible outliers in probit and logistic models. Then, a residual analysis is carried out to try to detect possible outliers in each fitted model. Figure 2 graph the boxplots of posterior distribution of the latent residual ε_i, τ_i and ε_i^* for the probit, logit and skew-probit models, respectively, by using parallel box plots of residual distributions against fitted probabilities. We fitted the probabilities through the posterior mean of the success probabilities. These residuals are prior normally distributed for the probit, logit and skew-probit models.

This graphical representation through boxplot is useful to inform about the variability and symmetry of the distribution. The central box of the boxplots presented

Table 3 *DIC, Bayes factor and pseudo-Bayes factor for the simulated dataset*

		Bayes factor			Pseudo-Bayes factor		
	DIC	$\mathcal{M}_1$	$\mathcal{M}_2$	$\mathcal{M}_3$	$\mathcal{M}_1$	$\mathcal{M}_2$	$\mathcal{M}_3$
Logistic ($\mathcal{M}_1$)	575.267	—	3.069	0.032	—	3.067	0.039
Probit ($\mathcal{M}_2$)	577.478	0.326	—	0.011	0.326	—	0.013
Skew-probit ($\mathcal{M}_3$)	569.080	30.732	94.337	—	25.459	78.083	—

correspond to quartiles, and the extreme values of the hatched lines correspond to 5th and 95th percentiles of the distribution.

An observation y_i is a candidate for an outlier for these models if the posterior distribution of its respective residual is far from its prior median. Alternatively, another way to check if an observation y_i is an outlier, it is calculating the probability of the event $R = \{k_1 < \varepsilon_i < k_2\}^c$, where k_1 and k_2 are such that the prior probability of this event to happen is small, around 0.10. The values k_1 and k_2 are chosen such that $\text{Prob}(R|y)$ is approximately 0.10, which is given by $k_1 = -k_2 = \Phi^{-1}(0.95) = 1.64$ in the probit model and $k_1 = -k_2 = 2.94$ in the logit model (because the tails of logistic distribution are heavier than the normal distribution). Following these ideas, for the skew-probit model we want to have k_1 and k_2 such that, the probability of R^c is approximately 0.90. Because the asymmetry of this distribution, there are many ways to do it. We consider here equal tails interval. Another alternative was to choose an HPD interval (high density). We added parallel lines in the boxplots to show the points k_1 and k_2. Moreover, the posterior probabilities of a residual exceeding these values are presented in Table 4 for all three models considered.

Figure 2 and Table 4 show that the observations with large probabilities to be outliers considering the residual ε_i for all models is the first dosage ($x_i = -1.1901$), where it was obtained 5 success in 60 outcomes. Moreover, since the residual ε_i is a function of a shape parameter in the skew-probit, that can affect the residual analysis when this parameter is unknown because of the residual ε_i have an unknown prior distribution. An alternative is using the residual $\varepsilon_i^* = z_i - (\mathbf{x}_i^t \boldsymbol{\beta} - \lambda w_i)$ on the skew-probit case. This residual does not depend on prior parameters and it is normally distributed. That allows us to detect outliers in the same considered for the residual ε_i^* in the probit case. Another residual which is normally distributed a priori, it is the residual $\tau_i = (z_i - \mathbf{x}_i^t \boldsymbol{\beta})/\sqrt{\delta_i}$ in the logit model, that happens because the logistic model can be obtained through scale mixture of normals. The fourth and sixth column of Table 4 show the posterior probabilities of these residuals in absolute value are bigger than 1.64 in the logit and skew-probit models, respectively.

Figure 2 and Table 4 show that the observations with large probabilities of being outliers considering the residual ε for probit and τ_i for logit are the same. Since there is no unknown parameter for skew-probit model, the residual ε^* does not present any observation with high probability of being outlier, different from the residual ε. The residual ε^* removes the dependence on λ of the skew-probit model. Concluding, the residuals analysis shows that there are no observations with a large probability of being outliers for the skew-probit model.

6 Application

We illustrate our procedures using the dataset presented in Christensen (1997). It consists of a randomly selected subset of 300 patients admitted to the University of

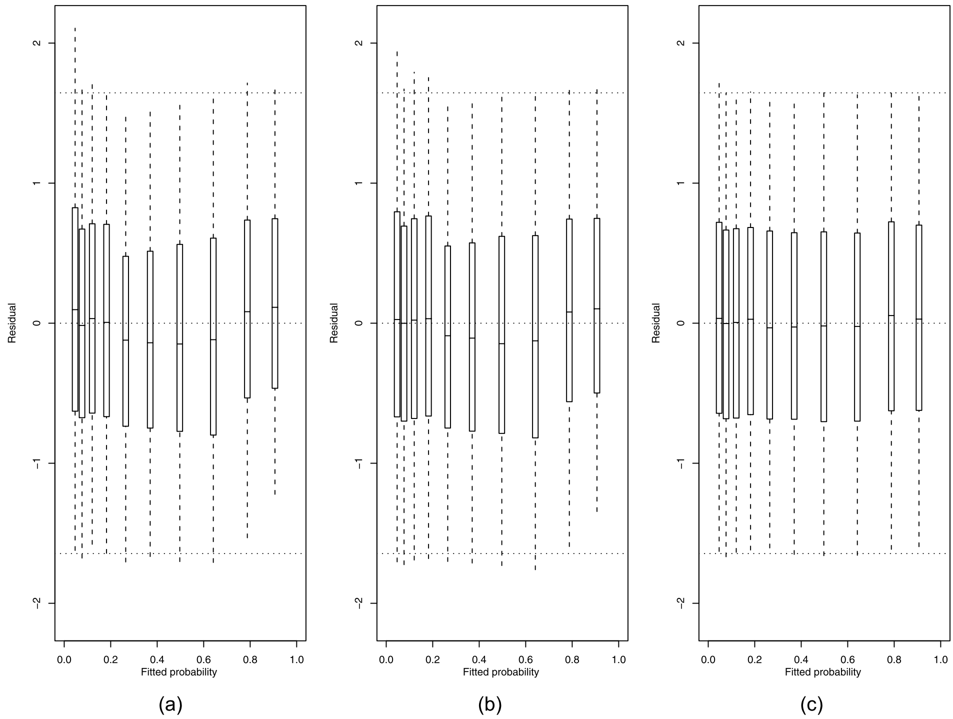

Figure 2 *Boxplots of posterior distribution of the latent residuals* $\varepsilon_i = z_i - \mathbf{x}_i^t \boldsymbol{\beta}$, $\tau_i = (z_i - \mathbf{x}_i^t \boldsymbol{\beta})/\sqrt{\delta_i}$ *and* $\varepsilon_i^* = z_i - (\mathbf{x}_i^t \boldsymbol{\beta} - \lambda w_i)$, *respectively, against the fitted probabilities* $\mathbb{E}(p_i|\mathbf{y})$ *for the fitting of the probit* (a), *logit* (b) *and skew-probit* (c) *model.*

Table 4 *Outlying probabilities for the models fot the simulated dataset*

	Model				
	Probit	Logit		Skew-probit	
Dosage	ε	ε	τ	ε	ε^*
1	0.146*	0.140*	0.142*	0.141*	0.106
2	0.107	0.109	0.105	0.114	0.102
3	0.101	0.121	0.114	0.125	0.098
4	0.102	0.117	0.108	0.120	0.098
5	0.094	0.100	0.093	0.114	0.092
6	0.093	0.101	0.100	0.114	0.096
7	0.102	0.110	0.101	0.123	0.102
8	0.105	0.113	0.104	0.122	0.100
9	0.098	0.101	0.090	0.120	0.098
10	0.078	0.082	0.075	0.106	0.096

New Mexico Trauma Center between the years 1991 and 1994. Of these, 22 died. One of the objectives of this study was to explain the probability of the patient eventually died due to the injuries by using binary regression model and considering the following explanatory variables: injury severity score (ISS), revised trauma score (RTS), patient's age (AGE) and the type of injuries (TI), that is, whether they were blunt (TI = 0) or penetrating (TI = 1). The response variable are 1 if the patient died and 0 if the patient survives.

The data considered here has been analyzed by means of binary regression model assuming different link functions, such as logistic, probit, complementary log–log models and skew-probit link. Christensen (1997) compared logistic, probit, complementary log–log models through of Bayes Factor and suggest against complementary log–log model, but there are not a serious preference between logistic and probit models. However, there exists significant difference between the observed number of 0's (278 survivors) and 1's (22 fatalities) on the dataset, that indicates a skewed link. Thus, Farias and Branco (2011) proposed a skew-probit link to analyze this dataset. The skew-probit link is able to fit positively and negatively skewed data. They fitted a model with null intercept, the predictors ISS, RTS, AGE, TI and the interaction AGE and TI. Furthermore, it was compared logistic, probit and skew-probit models through of several Bayesian criteria and concluded that the skew-probit model seems to be more appropriate to fit the Trauma dataset than the logistic and probit models.

For each regression parameter we considered independent normal diffuse prior with mean 0 and variance 100. We check the convergence of the MCMC method using several diagnostic procedures, such as the graphs of the ergodic averages and the Geweke statistic. These diagnostic procedures showed that convergence had been achieved. Finally, the Monte Carlo sample size was taken to be $M = 3000$ in all calculations.

Firstly, to compare the models we present in Table 5 the values of the DIC, the Bayes factor and the pseudo-Bayes factor.

The skew-probit link outperforms logistic and probit models for all criteria used. This result is an indicative of possible outliers in probit and logistic models. Then, a residual analysis is carried out to try to detect possible outliers in each fitted model.

Table 5 *Values of DIC, Bayes factor and pseudo-Bayes factor for Trauma data*

		Bayes factor			Pseudo-Bayes factor		
	DIC	$\mathcal{M}_1$	$\mathcal{M}_2$	$\mathcal{M}_3$	$\mathcal{M}_1$	$\mathcal{M}_2$	$\mathcal{M}_3$
Logistic ($\mathcal{M}_1$)	112.998	—	0.269	0.159	—	0.551	0.191
Probit ($\mathcal{M}_2$)	112.032	1.689	—	0.591	1.813	—	0.347
Skew-probit ($\mathcal{M}_3$)	109.781	6.275	3.715	—	5.224	2.881	—

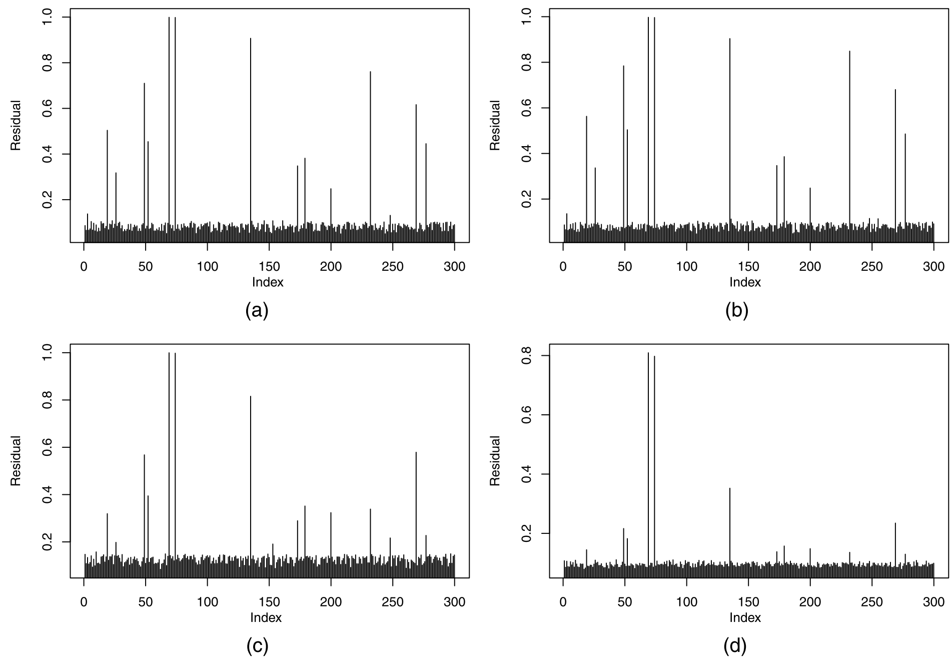

Figure 3 *Posterior probabilities of the latent residuals u_i being outliers for probit* (a) *and logit* (b), *skew-probit u* (c) *and u^** (d) *for Trauma dataset.*

Figure 3 shows the posterior probabilities of the latent residuals $u_i = \Phi(\varepsilon_i)$ for the probit, logit and skew-probit models. For the skew-probit we also show the posterior probabilities of the latent residuals u_i^*.

Table 6 shows that there are eight observations with a large probability of being outliers for probit and logit models. They are observations 19, 49, 52, 69, 74,

Table 6 *Outlying probabilities for the probit, logit and skew-probit models for Trauma dataset by using uniformly distributed residuals*

	Model (residual)				
Index	Probit (u_i)	Logit (u_i^*)	Logit (u_i)	Skew-probit (u_i)	Skew-probit (u_i^*)
69	0.999*	0.997*	0.867*	0.999*	0.809*
74	0.998*	0.996*	0.841*	0.998*	0.797*
135	0.907*	0.904*	0.692*	0.816*	0.352
232	0.761*	0.849*	0.693*	0.338	0.138
49	0.710*	0.784*	0.593*	0.568*	0.216
269	0.616*	0.608*	0.542*	0.579*	0.235
19	0.504*	0.563*	0.507*	0.319	0.146
52	0.454*	0.504*	0.445*	0.395	0.182

Table 7 *Inference summaries for Trauma data*

Parameter	Posterior mean	Posterior SD	95% HPD interval	
			Lower	Upper
β_1	0.088	0.032	0.029	0.151
β_2	−0.553	0.137	−0.823	−0.303
β_3	0.053	0.021	0.0183	0.098
β_4	0.722	1.272	−1.771	3.263
β_5	0.006	0.034	−0.063	0.074
λ	−4.652	1.988	−8.924	−1.204

135, 232 and 269, all with posterior probabilities bigger than 0.4. However, for the skew-probit model, the residual u_i shows 5 observations suspicious of being outliers and the residual u_i^* shows only two observations with outlying probability bigger than 0.4. That result was expected since the skew-probit models outperforms logistic and probit models for all model comparasion criteria used in Farias and Branco (2011) and presented in Table 5. Therefore, our residuals analysis confirm that the skew-probit link should be prefered than logistic and probit models.

Finally, Table 7 presents posterior summaries of the parameters for skew-probit model, where SD and HDP represent the standard deviation from the posterior distributions and the 95% highest posterior density interval, respectively.

7 Conclusion

In this work we described the use of latent variables in the Bayesian binary regression model framework. The introduction of these latent variables has the goal to obtain known forms for the full conditional posterior distributions, which make easy to implement the Gibbs sampling algorithm. Moreover, different kinds of residuals based on these latent variables were defined and those are useful for outlying detection.

We implemented severals latent residuals for probit, logit and skew-probit links for a simulated dataset and a medical dataset. For the simulated data, we detected the same outliers under the probit and logit models which is associated to the first category. The same outlier was found under the residual ε_i for the skew-probit model. However, when we remove the dependence on λ in skew-probit model by using the residual ε_i^*, there is no outlier observation. For the medical dataset the uniform residuals show the same outliers under the symmetrical models and less outliers under the skew-probit model. Then, the latent residuals proposed in this paper have showed to be able to detect the difference between symmetrical and asymmetrical models. Finally, we suggest the use of the uniform residuals when the goal is to compare directly the residuals values between different models.

Acknowledgments

We acknowledge helpful comments and suggestions from Professor Dipak Dey and an anonymous referee. During the development of this paper the first author was partially supported by CNPq and FAPESP (Grant 2007/03598-3).

References

Albert, J. H. and Chib, S. (1995). Bayesian residual analysis for binary response regression models. *Biometrika* **82**, 747–956. MR1380812

Azzalini, A. and Capitanio (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew-t distribution. *Journal of the Royal Statistical Society, Series B* **65**, 367–389. MR1983753

Chen, M.-H. and Dey, D. K. (1998). Bayesian modelling of correlated binary responses via scale mixture of multivariate normal link functions. *Sankhyã, Ser. A* **60**, 322–343. MR1718864

Chen, M.-H., Dey, D. K. and Shao, Q.-M. (1999). A new skewed link model for dichotomous quantal response data. *Journal of the American Statistical Association* **94**, 1172–1186.

Chen, M.-H. (2004). The skewed link models for categorical response data. In *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality* (M. G. Genton, ed.). Boca Raton: Chapman & Hall/CRC. MR2155327

Christensen, R. (1997). *Log-Linear Models and Logistic Regression*, 2nd ed. New York: Springer-Verlag. MR1633357

Devroye, L. (1986). *Non-Uniform Random Variate Generation*. New York: Springer-Verlag. MR0836973

Farias, R. B. A. and Branco, M. D. (2011). Efficient algorithms for Bayesian binary regression model with skew-probit link. In *Recent Advances in Biostatistics* (M. Bhattacharjee, S. K. Dhar and S. Subramanian, eds.) 143–168. Newark, NJ: World Scientific.

Geisser, S. and Eddy, W. (1979). A predictive approach to model selection. *Journal of the American Statistical Association* **74**, 153–160. MR0529531

Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association* **90**, 773–795.

McCullagh, P. and Nelder, J. (1989). *Generalized Linear Models*, 2nd ed. London: Chapman & Hall. MR0727836

Sahu, S. K., Dey, D. K. and Branco, M. D. (2003). A new class of multivariate skew distributions with applications to Bayesian regression models. *The Canadian Journal of Statistics* **31**, 129–150. MR2016224

Departamento de Estatística
Universidade de São Paulo
CP 66.281
ZIP 05315-970 São Paulo–SP
Brazil
E-mail: rfarias@ime.usp.br
mbranco@ime.usp.br