A generalized skew probit class link for binary regression

Jorge L. Bazán[*], Heleno Bolfarine[†], Márcia D. Branco[†]

May, 09, 2006

Abstract

We introduce a generalized skew probit (*gsp*) class of links for the modeling of binary regression giving some properties and conditions for the existence of the maximum likelihood estimator and of the posterior distributions of the parameters of the model when improper uniform priors are established. As shown, asymmetric links already proposed in the literature are special cases of the general family of links proposed. A Bayesian inference approach using MCMC is developed and implementations of the approach are facilitated by considering the augmented likelihood proposed. Several models comparison criteria are introduced and confirm that the *gsp* class are more adequate for the biological data sets analyzed than other asymmetrical and symmetrical links in the literature. Moreover, extensions of the methods developed in this paper for dichotomous responses to ordinal response data and mixed models for binary response data are indicated.

Keywords: asymmetrical link, binary response, Bayesian estimation, skew-normal distribution, model comparison .

[1]Department of Sciences. Catholic University. Perú.
[2]Department of Statistics. University of São Paulo. Brasil. Caixa postal 66281, CEP 05315-970.

1 Introduction

Data are considered dichotomous when each observed response falls into one of two possible categories, such as success or failure, positive or negative, correct or incorrect. These kind of data are common in several areas of applications such as, the social, medical, agricultural, genetics and behavioral sciences. Typically, binary regression using probit and logit links, deals with dichotomous data as a response variable and explore its relationship with a set of other explanatory variables combined as a linear predictor. However, Nagler (1994), argues that, with both, logit and probit links, it is assumed that individuals with a .5 probability of success are more sensitive to changes in the independent variables, e.g., a 1 unit change in X will have a greater effect on someone who has a .5 probability of success than it will have on someone with a .3 or .7 probability of success. Nagler (1994) argues that this isn't necessarily the case, e.g., somebody with a .4 probability of success may be more affected by a 1 unit change in X than somebody who has a .5 probability of success. If so, the distribution is "skewed" - things aren't symmetric about .5 and then asymmetrical links are justified.

The probability of success is obtained by considering a cumulative distribution function (cdf) evaluated at the linear predictor and typically, the cdf´s used are the logistic (logit model) and the standard normal (probit model) distribution functions. But Chen *et al.* (1999) also argues that when the probability of a given binary response approaches 0 at a different rate than it approaches 1, symmetric links may be inadequate to fit binary data and it is necessary consider asymmetric links.

Binary regression is an important special case of the generalized linear models (GLM) for which Bayesian inference is well documented. This paper is devoted to propose a new flexible parametric family of asymmetric links to the probability of success. In particular, we present an unified approach for skew probit links presented early in the literature, by considering a generalized skew probit link for binary regression. Chen *et al.* (1999) and Bazán *et al.* (2005) skew probit links and the usual probit link follow as special cases. Further, others skew-probit links not yet in the literature can be derived from this approach.

The skew link proposed here introduces a parameter that controls the rate of increasing (or decreasing) of the probability of success (failure) of the binary response and is based on the cdf of the skew normal distribution given by Azzalini (1985). For a review about the skew normal distribution see Dalla Valle (2004). Some properties used in this paper are presented in the Appendix.

The asymmetric binary regression model proposed here is characterized by: i) the probability of success is obtained by considering a cumulative distribution function (cdf) evaluated at the linear predictor and ii) an asymmetry parameter associated with these cdfs is introduced independently of the linear predictor and, in addition, iii) a latent linear structure is not necessary for the link formulation. However, as we will see later, such structure will be considered for computational purposes.

Another aim of this paper is to show that the model proposed can be easily implemented using the Bayesian software WinBugs. Further, a discussion on Bayesian criteria for model comparison is also conducted in two data set.

The paper is organized as follows. Section 2 revise some of the symmetric and asymmetric links in the literature at the present time. To define a general class of skew probit links in Section 4, we give a special representation for the cdf of the skew normal distribution in Section 3. In Section 5, inference aspects are discussed, considering results on the existence of maximum likelihood estimators and propriety of the posterior distributions when improper uniform priors are used. A data augmented likelihood is considered in Section 6 and in Section 7 a MCMC Bayesian estimation approach is developed. In Section 8, two applications to a real data set are presented, including a discussion on models comparison and variable selection criteria. Finally extensions and discussion of the *gsp* class introduced is showed in Section 9.

2 Symmetric and Asymmetric links

We consider $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)'$ a $n \times 1$ vector of n independent dichotomous random variables, assuming that $y_i = 1$ with probability p_i and $y_i = 0$ with probability $1 - p_i$, and $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{in})'$ a $k \times 1$ vector of covariates, where x_{i1} may equals 1, corresponding to an intercept, $i = 1, \ldots, n$. Moreover, $\boldsymbol{X}$ denotes the $n \times k$ design matrix with rows x_i', and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)'$ is a $k \times 1$ vector of regression coefficients. Binary regression models assume that

$$p_i = F(\eta_i(\boldsymbol{\beta})) = F(\boldsymbol{x_i'\beta}), \quad i = 1, \ldots, n, \tag{1}$$

where $F(.)$ denotes a cumulative distribution function (cdf). The inverse of the function F namely F^{-1}, is typically called link function and $\eta_i = \boldsymbol{x}_i'\boldsymbol{\beta}$, the linear predictor. The graphic considering p_i as a function of η_i is called response curve. When F is a cdf of a symmetric distribution, the response curve has symmetric form about 0.5. Examples are obtained when F is in the class of the elliptical distributions as, for example, standard normal, logistic, Student-t, double exponential and Cauchy distributions.

In the literature several asymmetric links have been considered, among others, by Prentice (1976), Guerrero and Johnson (1982), Stukel (1988), Czado and Santner (1992), Chen *et al.* (1999), Basu and Mukhopadhyay (2000) and Haro-López,*et al.* (2000). Moreover, examples are listed in different textbooks (see, for example, Collet, 2003) reporting situations where an asymmetric link may be more appropriate than a symmetric one.

There are several ways to obtain asymmetric links by considering the model structure defined in (1). The most important ones follow by:
(a) taking F as the cdf of an asymmetric distribution; (b) considering a modification of the linear predictor η_i and (c) considering F in a very general parametric class of probability distributions, for example, in a class of mixtures of distributions.

A very popular example of situation (a) is the complementary log-log link, where the cdf of the Gumbel distribution is considered, but the cdf of the Weibull and log-normal distributions can also be considered. In these cases, the cdf is completely specified and it does not depend of any unknown parameter and no relationship between them and the usual symmetric links are established. Other models are obtained when considering the following cdfs:

$$F(\eta_i) = 1 - (1 + e^{\eta_i})^{-\lambda} \qquad \text{and} \qquad F(\eta_i) = (1 + e^{-\eta_i})^{-\lambda}, \quad \lambda > 0.$$

In spite of the fact that there is no well agreed name for the first cdf, Achen (2002) has named it as the scobit distribution. The second cdf corresponds to the Burr type II distribution. Corresponding links using these distributions were proposed in Prentice (1976) and were popularized in the statistical literature by Aranda-Ordaz (1981) and in the economic literature by Nagler (1994) and Achen (2002). These links are skewed logit and are here termed *scobit* and *power logit*, respectively, and include the logit link as special case by considering the parameter $\lambda = 1$.

Case (b) keeps F as a symmetric distribution and considers a modification of the linear predictor η_i by $m(\eta_i, \lambda)$, where $m(.)$ is a nonlinear and continuous function, and λ is the parameter that controls asymmetry. Guerrero and Johnson (1982), and Stukel (1988) use the logistic distribution and Czado (1994) use the standard normal distribution. Hence, p_i is obtained considering $p_i = F(m(\eta_i, \lambda))$ for some specification range for η_i. For example, in Stukel's model (Stukel, 1988), the specification of $m(\eta_i, \lambda)$ is different for $\eta_i \geq 0$ than it is for $\eta_i \leq 0$. That is, $m(\eta_i, \lambda)$ is a partitioned function for different ranges of η_i and λ. For $\eta_i \geq 0$, it is considered that

$$m(\eta_i, \lambda) = \begin{cases} \lambda^{-1}(exp(\lambda\eta_i) - 1), & \lambda > 0; \\ \eta, & \lambda = 0; \\ -\lambda^{-1}log(1 - \lambda\eta_i), & \lambda < 0, \end{cases}$$

and for $\eta_i \leq 0$ it is considered that

$$m(\eta_i, \lambda) = \begin{cases} -\lambda^{-1}(exp(-\lambda\eta_i) - 1), & \lambda > 0 \\ \eta, & \lambda = 0 \\ \lambda^{-1}log(1 + \lambda\eta_i), & \lambda < 0. \end{cases}$$

Another possibility in case (b), is obtained when the linear predictor is replaced by a polynomial expression, usually quadratic or cubic (see, for example, Collet, 2003).

An example of case (c) is given when F is in the class of the elliptical scale mixtures of distributions given by $\mathcal{F} = \{F(.) = \int_{[0,\infty]} H(.|\nu)dG(\nu)\}$, where G is a cdf on $[0, \infty]$ and H is an elliptical distribution (see Basu and Mukhopadhyay, 2000, Haro-López, *et al.*, 2000). Another recent formulation, using an asymmetric probit link appears in Chen *et al.* (1999), considering a class of mixtures of normal distributions, where the mixture measure is the half normal distribution. The half normal distribution is denoted by $HN(0, 1)$ and has probability density function (pdf) given by $g(x) = 2\phi(x),\ x > 0$, where $\phi(.)$ is the corresponding pdf of the normal distribution.

Using auxiliary latent variables the skew probit model proposed by Chen *et al.* (1999), from now on denoted by *CDS sp*, is given by

$$y_i = \begin{cases} 1, & s_i > 0 \\ 0, & s_i \leq 0, \end{cases} \tag{2}$$

where

$$s_i = \eta_i + \lambda v_i + w_i, \tag{3}$$

with v_i independent of w_i,

$$v_i \sim HN(0, 1) \text{ and } \quad w_i \sim N(0, 1). \tag{4}$$

Therefore,

$$p_i = \int_0^\infty \Phi(\eta_i + \lambda v_i)g(v_i)dv_i, \tag{5}$$

where $\Phi(.)$ denotes the cdf of a standard normal distribution and $g(.)$ is the HN pdf.

Expression (5) follows from the stochastic representation of the skew normal distribution given by Chen *et al.* (1999) and Branco and Dey (2002), which is a particular case of the asymmetric model in Sahu *et al.* (2003) (see expression in Table 1), where $\lambda \in \mathbb{R}$ is a skewness parameter.

Another skew-probit link was proposed by Bazán *et al.* (2005) considering F as a cdf of the standard skew-normal distribution given in Azzalini (1985). Without using latent variable structure, they consider

$$p_i = F_\lambda(\eta_i) = 2\Phi_2\left(\begin{pmatrix} \eta_i \\ 0 \end{pmatrix}; \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & -\delta \\ -\delta & 1 \end{pmatrix} \right), \tag{6}$$

where $F_\lambda(.)$ denotes the cdf of the standard skew-normal distribution with asymmetry parameter $\lambda \in \mathbb{R}$, and $\Phi_2(.)$ denoting the cdf of the bivariate standard normal distribution with correlation coefficient $-\delta$, where $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}} \in [-1, 1]$. The latter skew probit link has been used in the item response theory (IRT) context by Bazán *et al.* (2005) and, we denote this model by *BBB sp*. We can also formulate the model in (6) using a latent structure as in (3-4) by considering the stochastic representation of the skew normal distribution to s_i given by Henze (1986) (see expression in Table 1).

A common aspect of the *CDS sp* and *BBB sp* links is that the asymmetry parameter is associated with $F(.)$ and is independent of the linear predictor. To emphasize this, we write $F_\lambda(.)$ instead of $F(.)$. The main difference is that with *BBB sp,* it is not necessary to consider a latent linear structure representation as in *CDS sp*, but it is sufficient specifying p_i as a cdf of an asymmetric distribution and an analytical expression to (5) can be obtained as in (6). However, as we show in the next section, both links are indeed a special case of a more general class of links.

3 The cumulative density function of the skew normal distribution

A random variable R follows a skew-normal distribution (Dalla Valle, 2004), with parameter $\theta = (\mu, \sigma^2, \lambda)$ where $\mu \in \mathbb{R}$ is a location parameter and $\sigma^2 > 0$ is a scale parameter and $\lambda \in \mathbb{R}$ is an asymmetry parameter, if the probability density function (pdf) of R is given by

$$f_\theta(r) = \frac{2}{\sigma}\phi\left(\frac{r-\mu}{\sigma}\right)\Phi\left(\lambda\frac{r-\mu}{\sigma}\right), \tag{7}$$

where $\phi(.)$ and $\Phi(.)$ denote, respectively, the density and distribution functions of the standard normal distribution. The notation considered is $R \sim SN(\theta)$ where $\theta = (\mu, \sigma^2, \lambda)$. If $\lambda = 0$, the density of R in (7) reduces to the density of the $N(\mu, \sigma^2)$. The special case of $\mu = 0$ and $\sigma^2 = 1$ is called the standard skew-normal distribution as given by Azzalini (1985) with the following pdf:

$$f_\lambda^A(s) = 2\phi(s)\Phi(\lambda s).$$

In this case we write $S \sim SN_A(\lambda)$, denoting, the standard Azzalini´s skew normal distribution. Moreover, in the special case of $\mu = 0$ and $\sigma^2 = 1 + \lambda^2$, the pdf of R in (7) is the standard skew normal given in Chen *et al.* (1999) and Branco and Dey (2002), which is a particular case of the skew normal distribution due to Sahu *et al.* (2003). In this case, we write $S \sim SN_S(\lambda)$ to denote this distribution, with the corresponding pdf:

$$f_\lambda^S(s) = \frac{2}{\sqrt{1+\lambda^2}}\,\phi\left(\frac{s}{\sqrt{1+\lambda^2}}\right)\Phi\left(\lambda\frac{s}{\sqrt{1+\lambda^2}}\right).$$

To define a general class of skew probit links we give next a special representation for the cdf of the skew normal distribution. See the Appendix for the notation introduced, some properties and the proofs of the results presented next.

Proposition 1. *Let $R \sim SN(\theta)$. Then, the cdf of R can be written as*

$$F_\theta(r) = \int_0^\infty g(v)\Phi\Big(\frac{r-\mu-\sigma\delta v}{\sigma\sqrt{1-\delta^2}}\Big)dv = \int_0^\infty g(v)\Phi\Big(\frac{r-\mu}{\sigma}\sqrt{1+\lambda^2}-\lambda v\Big)dv. \quad (8)$$

An alternative representation for the cumulative distribution function is provided next.

Proposition 2. *The distribution function of the $SN(\mu, \sigma^2, \lambda)$ can also be written as*

$$F_\theta(r) = 2\Phi_2\Big(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}\Big), \quad (9)$$

where $\boldsymbol{x} = (r, 0)$, and $\Phi_2(.)$ denote the cdf of the bivariate normal distribution, with parameters $\boldsymbol{\mu} = (\mu, 0)'$ and $\boldsymbol{\Sigma} = \begin{pmatrix} \sigma^2 & -\delta \\ -\delta & 1 \end{pmatrix}$, and $-\delta$ is the correlation coefficient.

The representation of the skew normal cdf in Proposition 1 is similar to a result given in Chen *et al.* (1999) and it will be used to prove the propositions in Section 5. The representation of the skew normal cdf in Proposition 2 is similar to a result given in Bazán *et al.* (2005) and it indicates that the distribution of the skew-normal distribution evaluated at a point r is also obtained by considering a bivariate normal distribution with mean vector $\boldsymbol{\mu}$ and variance-covariance matrix $\boldsymbol{\Sigma}$ evaluated at the point $(r, 0)$. This result is important since several efficient computational algorithms are available for computing integrals related to the bivariate normal distribution. Another algorithm for evaluating the cdf of the skew normal distribution is based in the use of Owen´s function (Azzalini, 1985; Dalla Valle, 2004) and is available for R and Matlab programs. Additionally, the distribution of R is obtained considering $(U, W) \sim N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ defined above and defining $R = U|W > 0$.

As a direct consequence of the above result, we have

Corollary 1. *Corresponding to the standard Azzalini´s skew normal distribution, we have that*

$$F_\lambda^A(s) = \int_0^\infty g(v)\Phi\Big(s\sqrt{1+\lambda^2}-\lambda v\Big)dv = 2\Phi_2\Big(\begin{pmatrix} s \\ 0 \end{pmatrix}; \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & -\delta \\ -\delta & 1 \end{pmatrix}\Big).$$

Corresponding to the standard Sahu´s et al. skew normal distribution, we have that

$$F_\lambda^S(s) = \int_0^\infty g(v)\Phi\Big(s-\lambda v\Big)dv = 2\Phi_2\Big(\begin{pmatrix} s \\ 0 \end{pmatrix}; \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{1}{1-\delta^2} & -\delta \\ -\delta & 1 \end{pmatrix}\Big).$$

By considering Properties 1, 2 and 3 in the Appendix and recalling that the skewness index γ is given by $\gamma = \mu_2^3 = E(\frac{R-E[R]}{Var[R]^{1/2}})^3$, we have some interesting expressions for Azzalini's and Sahu´s et al. (2003) standard skew normal distributions in Table 1.

For Azzalini´s skew normal distribution we have that $E[S] \in [-0.80, 0.80]$, $Var[S] \in [0.36, 1]$ and $\gamma_A \in [-0.995, 0.995]$ but for Sahu´s et al. skew normal distribution we have that $E[S] \in [-0.80|\lambda|, 0.80|\lambda|]$, $Var[S] \in [1, 1 + 0.36\lambda^2]$ and $\gamma \in [-0.843, 0.843]$. Then, $E[S]$ and $Var[S]$ take lower values in the SN_A case. The stochastic representation for Azzalini´s skew normal distribution is known as Henze's representation and a similar

Table 1: Comparisons of Azzalini's and Sahu´s et al. standard skew normal distributions

	$S \sim SN_S(\lambda)$	$S \sim SN_A(\lambda)$
$M_S(t)$	$2exp\left(\frac{1}{2}t^2\right)\Phi(\delta t)$	$2exp\left(\frac{1}{2}(1+\lambda^2)t^2\right)\Phi(\lambda t)$
$E[S]$	$(\frac{2}{\pi})^{1/2}\lambda$	$(\frac{2}{\pi})^{1/2}\delta$
$Var[S]$	$1+(1-\frac{2}{\pi})\lambda^2$	$1-\frac{2}{\pi}\delta^2$
γ	$(\frac{2}{\pi})^{3/2}(1-\frac{2}{\pi})sig(\lambda)\frac{\lambda^3}{(1+(1-\frac{2}{\pi})\lambda^2)^{3/2}}$ $(\frac{2}{\pi})^{3/2}(1-\frac{2}{\pi})sig(\delta)\frac{\delta^3}{(1-\frac{2}{\pi}\delta^2)^{3/2}}$	$(\frac{2}{\pi})^{3/2}(2-\frac{\pi}{2})sig(\lambda)\frac{\lambda^3}{(1+(1-\frac{2}{\pi})\lambda^2)^{3/2}}$ $(\frac{2}{\pi})^{3/2}(2-\frac{\pi}{2})sig(\delta)\frac{\delta^3}{(1-\frac{2}{\pi}\delta^2)^{3/2}}$
Stochastical representation	$S=\delta V+(1-\delta^2)^{1/2}W \sim SN_A(\lambda)$ $S\|V=v \sim N(\delta v, 1-\delta^2)$ $v \sim HN(0,1)$	$S=\lambda V+W \sim SN_S(\lambda)$ $S\|V=v \sim N(\lambda v, 1)$ $v \sim HN(0,1)$

$sig(.)$ is a signal function which equals 1 when its argument is positive and equals -1 otherwise.

representation for Sahu´s et al. skew normal distribution is due to Chen *et al.* (1999) and Branco and Dey (2002). In addition, it is possible to establish a relationship between Azzalini´s and Sahu´s et al. standard skew normal distributions by considering that given $S \sim SN_A(\lambda)$ then $S^* = \sqrt{1+\lambda^2}S \sim SN_S(\lambda)$.

4 The generalized skew probit model

From the results in the previous section, a generalized skew probit (*gsp*) class of models is obtained considering

$$p_i = F_\theta(\eta_i) = F_\theta(\boldsymbol{x_i'\beta}), \quad i=1,\ldots,n, \tag{10}$$

where $F_\theta(.)$ is the cdf of the skew normal distribution given in (8) or (9) with parameter vector $\theta = (\mu, \sigma^2, \lambda)$.

Corollary 2. *Some special links of the gsp class follows from (10) and are as given in the following:*

- *If* $\theta = (0,1,0)$ *then the probit link follows.*
- *If* $\theta = (0, 1+\lambda^2, -\lambda)$ *then the CDS sp link given in (5) follows.*
- *If* $\theta = (0,1,\lambda)$ *then the BBB sp link given in (6) follows.*

Notice also that a multitude of different skew probit links can be formulated by considering other values for μ and σ^2 in (10). The more general one follows by considering the parameters μ and σ^2 to be estimated from the data and is named the *complete* sp link. Another interesting skew probit link can be obtained by considering the skew normal distribution with $\mu = -\frac{\sqrt{2}\delta}{\sqrt{\pi - 2\delta^2}}$ and $\sigma^2 = \frac{\pi}{\pi - 2\delta^2}$. In this case it follows, by considering in (7) Property 2 in the Appendix, that $E(R) = 0$ and $V(R) = 1$. This link is named the standard skew probit link, which is denoted as *standard sp* link.

The formulation of the *gsp* class assumptions essentially imply that the probit model is nested in any of the skew-probit models described above. Figure 1 depicts different probability curves of the *CDS sp* and *BBS sp* models by using different values for η_i. Note that the *CDS sp* link presents probabilities of success that grow slower than the ones corresponding to the *BBB sp* link. Further, *CDS sp* seems to be more adequate if the range of η_i is wider.

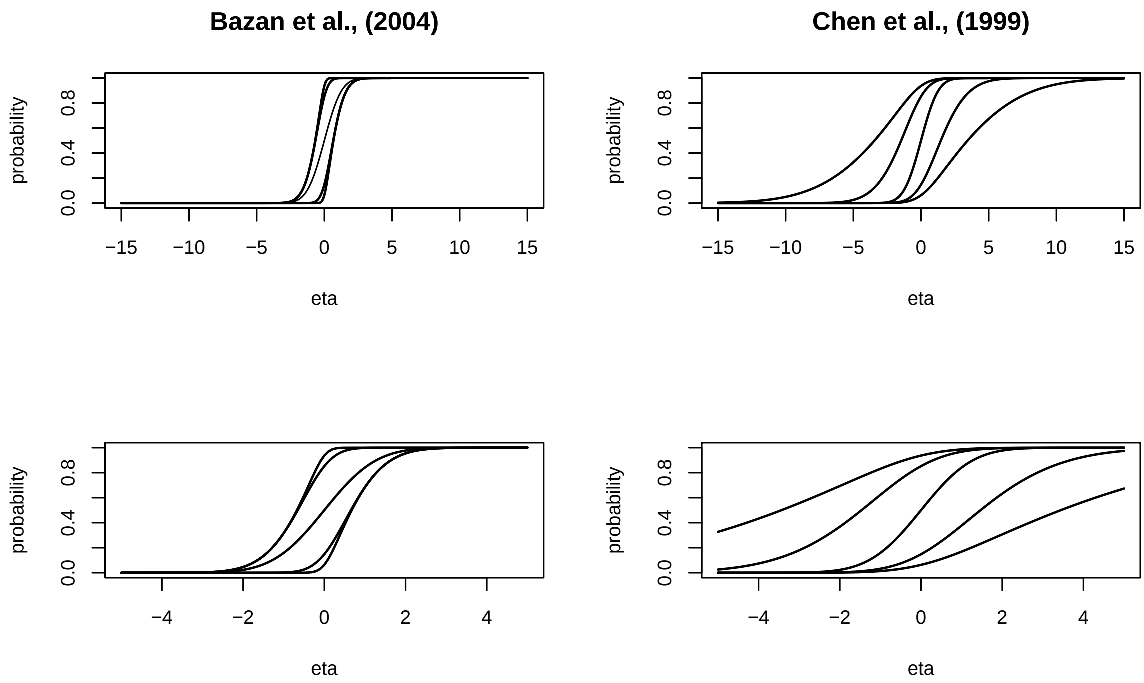

Figure 1: Probability curves for $\lambda = -5, -2, 0, 2, 5$ in *CDS and BBB sp* models considering different ranges for η_i

The likelihood function for the *gsp* class of models is given by

$$L(\boldsymbol{\beta}, \theta | \boldsymbol{y}, \boldsymbol{X}) = \prod_{i=1}^{n} [F_\theta(\eta_i)]^{y_i} [1 - F_\theta(\eta_i)]^{1-y_i}, \tag{11}$$

where η_i is the linear predictor, $\boldsymbol{X}$ is the design matrix as defined in Section 1 and $F_\theta(\eta_i)$ is given in (10).

Another version of the likelihood function for the *gsp* class of models, similar to the one given by Chen *et al.* (1999) is obtained by considering augmented data. Using

$v_i \sim HN(0,1)$, $i = 1, \ldots, n$, as auxiliary latent variables and denoting $\eta_i^* = \frac{\eta_i - \mu}{\sigma}$ with (8), we obtain

$$L(\boldsymbol{\beta}, \theta | \boldsymbol{y}, \boldsymbol{X}, \boldsymbol{v}) = \prod_{i=1}^{n} \left[\Phi(\eta_i^* \sqrt{1+\lambda^2} - \lambda v_i)\right]^{y_i} \left[1 - \Phi(\eta_i^* \sqrt{1+\lambda^2} - \lambda v_i)\right]^{1-y_i} g(v_i), \quad (12)$$

leading to the marginal likelihood function for the *gsp* class

$$L(\boldsymbol{\beta}, \theta | \boldsymbol{y}, \boldsymbol{X}) = \prod_{i=1}^{n} \int_0^{\infty} \left[\Phi(\eta_i^* \sqrt{1+\lambda^2} - \lambda v_i)\right]^{y_i} \left[1 - \Phi(\eta_i^*) \sqrt{1+\lambda^2} - \lambda v_i)\right]^{1-y_i} g(v_i) dv_i. \quad (13)$$

To obtain the versions of the likelihood function for the *BBB* and *CDS sp* links in (11) and (13), is only necessary to consider the corresponding values of θ in Corollary 2 to $F_\theta(\eta_i)$, $i = 1, \ldots, n$.

5 Inference

Note that in the *gsp* class, the parameters θ and $\boldsymbol{\beta}$ have quite different meaning. Whereas θ is a vector of structural parameters associated with the choice of the link function, the parameter $\boldsymbol{\beta}$ is a vector of parameters inherent to the observed data and not depending on the model choice (for a discussion see Taylor and Siqueira). By considering this fact, two scenarios can be considered. The first scenario is one in which $\boldsymbol{\beta}$ and θ are estimated; in the second scenario only $\boldsymbol{\beta}$ is allowed to vary and θ is fixed at its "true" value θ_0. As in Taylor and Siqueira (1996), we shall refer to these two scenarios as the unconditional and conditional ones, respectively.

Inference in the conditional scenario is easier to be implemented from both maximum likelihood (ML) and Bayesian approaches. However, conditions shall be imposed under the design matrix for the existence of the ML estimators and the posterior distribution of $\boldsymbol{\beta}$ under improper uniform priors. The following proposition is a direct consequence of results in Chen and Shao (2000).

Proposition 3. *By considering the conditional approach for the gsp class, that is, θ is known or fixed and $\boldsymbol{\beta}$ is the only parameter of interest, and letting $t_i = 1$ if $y_i = 0$, and $t_i = -1$ if $y = 1$, $\boldsymbol{X}$ the $n \times k$ known design matrix with rows $\boldsymbol{x}_i'$ and define $\boldsymbol{X^*}$ as the matrix with rows $t_i \boldsymbol{x_i}'$. Then, under the conditions*
(C1) $\boldsymbol{X}$ is of full rank;
(C2) There exists a positive vector $\boldsymbol{a} = (a_1, \ldots, a_n)'$ such that $\boldsymbol{X^{*\prime}} \boldsymbol{a} = \boldsymbol{0}$,
it follows that

- *for an improper uniform prior for $\boldsymbol{\beta}$, i.e., $\pi(\boldsymbol{\beta}) \propto 1$, the posterior distribution of $\boldsymbol{\beta}$ is proper, i.e., $\int_{R^k} L(\boldsymbol{\beta} | \boldsymbol{y}, \boldsymbol{X}, \theta) d\boldsymbol{\beta} < \infty$;*
- *the maximum likelihood estimate (MLE) of $\boldsymbol{\beta}$ exists.*

In the unconditional approach, computing the ML estimators using the versions of the likelihood functions given in Section 4 is not simple and is necessary to develop new conditions for the existence of the ML estimators. Additionally, we study the propriety of the posterior distribution under improper uniform priors for $\boldsymbol{\beta}$ and λ but considering proper priors for μ and σ^2. These scenarios consider independence between the priors, such that

$$\pi(\boldsymbol{\beta}, \theta) = \pi(\boldsymbol{\beta})\pi(\theta), \quad \text{where} \quad \pi(\theta) = \pi(\lambda)\pi(\mu)\pi(\sigma^2). \tag{14}$$

Proposition 4. *Consider the unconditional approach for the generalized skew probit link, that is, θ and $\boldsymbol{\beta}$ are the parameters of interest, and let $S_0 = \{i : y_i = 0\}$, $S_1 = \{j : y_j = 1\}$, $m_0 = card(S_0)$, $m_1 = card(S_1)$ and $\boldsymbol{X}_0$ and $\boldsymbol{X}_1$ are the matrices with rows $\boldsymbol{x}_i'$, $i \in S_0$ and $\boldsymbol{x}_j'$, $j \in S_1$. Then, the conditions*
(C1') $\boldsymbol{X}_0$ and $\boldsymbol{X}_1$ are of full rank;
(C2') There exists positive vectors $\boldsymbol{a}_0 = (a_{01}, \ldots, a_{m_0})' \in R^{m_0}$ and $\boldsymbol{a}_1 = (a_{11}, \ldots, a_{1m_1})' \in R^{m_1}$ such that $\boldsymbol{X}_0'\boldsymbol{a}_0 = \boldsymbol{0}$ and $\boldsymbol{X}_1'\boldsymbol{a}_1 = \boldsymbol{0}$,
imply that

- *for the prior $\pi(\boldsymbol{\beta}, \lambda) \propto 1$, and $\pi(\mu)$ and $\pi(\sigma^2)$ being proper priors, the joint posterior distribution of $(\boldsymbol{\beta}, \theta)$, where $\theta = (\mu, \sigma^2, \lambda)$, is proper, i.e.,*

$$\int_0^\infty \int_{-\infty}^\infty L(\mu, \sigma^2 | \boldsymbol{y}, \boldsymbol{X})\pi(\mu)\pi(\sigma^2) d\mu d\sigma^2 < \infty$$

  where $L(\mu, \sigma^2 | \boldsymbol{y}, \boldsymbol{X}) = \int_{-\infty}^\infty \int_{R^k} L(\boldsymbol{\beta}, \theta | \boldsymbol{y}, \boldsymbol{X}) d\boldsymbol{\beta} d\lambda$;

- *the maximum likelihood estimate (MLE) of $\boldsymbol{\beta}$ exists.*

6 Augmented Likelihoods

In this section we present two complete-data likelihood functions for the *gsp* class model, so that we start with an important alternative representation for the *gsp* model.

Proposition 5. *The gsp model, as defined before, is equivalent to considering that*

$$y_i = I(s_i > 0) = \begin{cases} 1, & s_i > 0; \\ 0, & s_i \leq 0, \end{cases}, \quad i = 1, \ldots, n \tag{15}$$

where $s_i \sim SN(\theta)$, with $\theta = (\eta_i - \mu\,, \sigma^2, -\lambda)$, and $I(.)$ is the usual indicator function.

The latent variables s_i´s are introduced to avoid working with Bernoulli type likelihoods and this representation shows a latent linear structure which produces an equivalent model with the *gsp* class. Furthermore, notice that the asymmetry parameter with the auxiliary latent variable is the negative of the value of the asymmetry parameter specified in (10).
Therefore, the *first complete-data likelihood function* for the *gsp* class is given by

$$L(\boldsymbol{\beta}, \theta | s, y) = \prod_{i=1}^n f_\theta(s_i) p(y_i \mid s_i), \tag{16}$$

where $p(y_i \mid s_i) = I(s_i, y_i) = I(s_i > 0)I(y_i = 1) + I(s_i \leq 0)I(y_i = 0)$, $i = 1, \ldots, n$. When $\mu = \eta_i$, $\sigma^2 = 1$ and $\lambda = 0$, the corresponding result in Albert and Chib (1993) follows the probit link.

An alternative data augmentation approach for the *gsp* class follows by considering

$$s_i = \eta_i + e_i, \quad e_i \sim SN(\theta), \tag{17}$$

that is, e_i is the equation error, and by considering the stochastic representation given in the Property 3 of the Appendix, we have that

$$e_i = -\mu + \sigma(-\delta v_i - (1-\delta^2) w_i), \quad i = 1, \ldots, n \tag{18}$$

where $v_i \sim HN(0,1)$ and $w_i \sim N(0,1)$. It follows that the conditional distribution $e_i|v_i$ is a normal distribution with mean $-\sigma\delta v_i - \mu$ and variance $(1-\delta^2)\sigma^2$ and then we have that $s_i^* \doteq s_i|v_i \sim N(\eta_i - \mu - \sigma\delta v_i, (1-\delta^2)\sigma^2)$. According to this result, simulation of s_i in the linear structure (17) can be performed in two steps. First, simulate $v_i \sim HN(0,1)$ and then simulate $s_i^* = s_i|v_i \sim N(\eta_i - \mu - \sigma\delta v_i, (1-\delta^2)\sigma^2)$.
By considering these results, the *second complete-data likelihood function* for the *gsp* class is given by

$$L(\boldsymbol{\beta}, \theta | s^*, v, y) = \prod_{i=1}^{n} \phi(s_i^*) g(v_i) p(y_i \mid s_i), \tag{19}$$

where $\phi(.)$ is the pdf of the normal distribution. The density $p(y_i \mid s_i)$ can be substituted for the corresponding MCMC estimated version $p(y_i \mid s_{i^*})$ because $I(S_i, y_i)$ implies $I(S_{i^*}, y_i)$.

Note that the errors e_i are independent and are "latent data' residuals (Albert and Chib, 1995) when considering $e_i = s_i - \eta_i$ which is estimated using the generated data. They can be used for model checking. To understand how the observations y_i change the distribution of these residuals, consider the posterior distribution of e_i conditional on $\boldsymbol{\beta}$, $\boldsymbol{\theta}$, s_i and v_i, that is, $e_i^* = e_i|\beta, \theta, y_i, s_i, v_i$.

7 MCMC Bayesian estimation

In the context of a Bayesian analysis, it is required to specify prior distributions for $(\boldsymbol{\beta}, \theta)$. By considering the results in Section 5, and assuming independent priors as given in (14), we can use for $\boldsymbol{\beta}$, the typical priors considered in the probit model (see, for example, Zellner and Rossi, 1984), including a normal prior ($\beta_j \sim N(\mu_{\beta_j}, \sigma^2_{\beta_j})$) or the uniform prior ($\pi(\boldsymbol{\beta}) = 1$). Jeffreys priors, as has been considered in Ibrahim and Laud (1991), can also be considered.

Since we consider $\pi(\theta) = \pi(\mu)\pi(\sigma^2)\pi(\lambda)$, it is possible to consider for μ and σ^2 priors usually considered for the normal model, such as, $\mu \sim N(\mu_0, \tau_0^2)$ and $\sigma^2 \sim Inv-\chi^2(\omega, \kappa)$ the scaled inverse-chi-square distribution with $\omega > 0$, s degrees of freedom and scale parameter $\kappa > 0$. We can also work with the precision defined as $1/\sigma^2$, in which case the $Gamma(\omega, \kappa)$ prior distribution can be specified.

Finally, for the λ parameter is possible to use a non-informative prior or a Student-t prior. When the model is parameterized using λ, it is called the "lambda parameterization" but when we can consider the *gsp* class in terms of δ, we call it as the "delta parameterization" and a non informative prior follows by considering that $\delta \sim U(-1,1)$ or, equivalently, $\lambda \sim T(0, 0.5, 2)$, where $T(a, b, \nu)$ denotes the Student-t distribution with location a, scale b and ν degrees of freedom, (Bazán, et al., 2005).

Note that the uniform prior in the "delta parameterization" is non informative but proper, however, the Student-t prior in the "lambda parameterization" is informative. By considering this fact, an invariant non informative prior that can be used for λ is the Jefrey´s prior derived in Liseo and Loperfido (2005).

Hence, to any priors indicated above, by considering the results in Propositions 3 and 4, it is possible to obtain proper posterior distributions and implement a Bayesian estimation procedure using the likelihood function in (11) or (13) and the prior distribution in (14). However, such an approach is complicated since the integrals involved to obtain the marginal posterior distributions are difficult to deal with. Two approaches based on data augmentation as considered in Albert and Chib (1993) were introduced in Section 6. The approaches given in (16) and (19) allow the implementation of MCMC methods which simplify efficient sampling from the marginal posterior distributions.

By considering the *first complete-data likelihood function* in (16) and the latent structure in (15), the full conditionals for the *gsp* class, to implement Bayesian inference using MCMC can be obtained as in Albert and Chib (1993). However, some of the full conditionals can not be directly sampled from, requiring more complex algorithms such as the Metropolis-Hastings. To overcome the difficulties described above we can use the *second complete-data likelihood function* in (19).
In the remainder of this section we develop a computational procedure for the *gsp* class based in this second augmented likelihood function. The hierarchical structure specification for the *delta parameterization* can be obtained by considering:

$$S_i^*|v_i\, x_i, y_i, \boldsymbol{\beta}, \mu, \sigma^2, \delta \sim N\left(\eta_i - \mu - \sigma\delta v_i\,,\,(1-\delta^2)\sigma^2\right) I(S_i^*, y_i),$$

$$V_i \sim HN(0,1),$$

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

$$\mu \sim \pi(\mu),$$

$$\sigma^2 \sim \pi(\sigma^2),$$

and

$$\delta \sim \pi(\delta).$$

where $\pi(\boldsymbol{\beta})$, $\pi(\mu)$, $\pi(\sigma^2)$ and $\pi(\delta)$ can be taken of the distributions proposed above. For the hierarchical structure specification for the *lambda parameterization* it is only necessary to specify prior distributions for λ and then use the transformation $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}}$.

The above hierarchical structure can be easily implemented in the WinBugs software. If μ and σ^2 are considered to be estimated, a general probit-normal model is obtained, but when considering the *CDS sp* and *BBB sp* links, the fourth and fifth lines in the hierarchy are eliminated (in the *CDS sp* link the fifth line is eliminated since ($\sigma^2 = 1+\lambda^2$)). Further, notice that when $\delta = 0$ or $\lambda = 0$, the hierarchical structure of the augmented likelihood corresponding to the probit-normal model follows by eliminating additionally the sixth line in the hierarchy.

8 Applications

We illustrate the Bayesian approaches developed in this paper for the *gsp* class using two data sets from the literature.

8.1 Beetles dataset

The beetle mortality data set (Collet, 2003) records the number of adult flour beetles killed after five hours of exposure to gaseous carbon disulphide at various concentrations levels. It is established that the log of the concentration differences of the poison explain the proportion of adult flour beetles killed and a binary regression model is adequate for analyzing the data set. This well-known data was also analyzed among others by

Prentice (1976), Czado (1994) and Stukel (1988) concluding that an asymmetric link is more convenient and significantly improves symmetric links such as the logit and probit links.

In order to illustrate the usefulness of the links proposed, we performed comparisons between *gsp* class links, in particular *complete, CDS, BBB, standard sp* and probit links with several links in the literature. We compare with cloglog, logit, generalized logit or scobit, power logit and logit using a quadratic term when fitted to the beetle data set. The generalized logistic link given by Stukel (1998) and the generalized probit link given by Czado (1994) were not considered since the first leads to an improper posterior distribution when improper uniform priors are considered (Chen *et al.*, 1999) and the second is not favored over the logit link using a quadratic term, as reported in Czado and Raftery (2003), using approximate Bayes factors (Kass and Raftery, 1995).

In all cases noninformative priors for the regression parameters are used. To facilitate model comparisons, as in Spiegelhalter et al. (1996), we consider the following reparameterization and priors for the regression parameters:

$$\eta_i = \beta_0^* + \beta_1(x_i - \overline{x}), \qquad \text{with} \qquad \beta_0 = \beta_0^* - \beta_1 \overline{x}, \qquad \beta_1 \sim N(0, 1000), \qquad \beta_0^* \sim N(0, 1000).$$

For all links in the *gsp* class considered it is assumed that $\delta \sim U(-1, 1)$ when considered the delta parameterization (in *generalized*, *standard* and *BBB sp*) and $\lambda \sim N(0, 100)$ when considered the lambda parameterization (in *CDS sp*). For the parameters associated with the *generalized* sp link it is assumed that μ and σ^2 are estimated from the data set, hence diffuse priors are considered. That is, $\mu \sim N(0, 1000)$ and $\tau \sim Gamma(0.001, 0.001)$ where $\tau = 1/\sigma^2$ is the precision parameter.

All the above models were implemented in WINBugs and in all cases an effective sample size of 2000 was considered. Since that presence of autocorrelations between chain values is expected due to the augmentation scheme (Chen *et al.*, 2000), for comparison, the decision on the number of the iterations and thin values used is defined by considering the first autocorrelations in the chains to be smaller than 0.5. In addition, several criteria computed using the BOA and CODA package, including the one proposed by Geweke (1993) was used to evaluate convergence. Running Mean Plots for each chain, for all the links considered, provide strong indication of chain convergence in all cases. Thin values used and time of execution (time in seconds to perform 10000 iterations on a Pentium IV with 3000 MHZ and 512 Ram) to the links considered are shown in Table 2.

A variety of methodologies exist to compare alternative bayesian model fits but the principal criterions used in those works are the Deviance Information Criterion (DIC) proposed by Spiegelhalter *et al.* (2002), and Expected Information Criteria corresponding to Akaike ($EAIC$), Schwarz and Bayesian ($EBIC$) as proposed in Carlin e Louis (2000) and Brooks (2002). The criteria are based in the *Posterior mean of the deviance*: $E\Big[D(\boldsymbol{\beta}, \boldsymbol{\theta})\Big]$ which is also a measure of fit that can be approximated by using the MCMC algorithm, considering the value of $Dbar = \frac{1}{G}\sum_{i=1}^{G} D(\boldsymbol{\beta}^g, \boldsymbol{\theta}^g)$, where the index g represent the g-ith realization of a total of G realizations, where $D(\boldsymbol{\beta}, \boldsymbol{\theta}) = -2ln(p(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{\theta},)) = -2\sum_{i=1}^{n} lnP(Y_i = y_i|\boldsymbol{\beta}, \boldsymbol{\theta})$, is the *bayesian deviance* (Dempster, 1977).

$EAIC$, $EBIC$ and DIC can be estimated using MCMC by considering $\widehat{EAIC} = Dbar + 2p$, $\widehat{EBIC} = Dbar + plogN$ and $\widehat{DIC} = Dbar + \widehat{\rho_D} = 2Dbar - Dhat$ 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 defined as $\rho_D = E\Big[D(\boldsymbol{\beta}, \boldsymbol{\theta})\Big] - D\Big[E(\boldsymbol{\beta}), E(\boldsymbol{\theta})\Big]$ and $Dhat = D\Big(\frac{1}{G}\sum_{i=1}^{G}\boldsymbol{\beta}^g, \frac{1}{G}\sum_{i=1}^{G}\boldsymbol{\theta}^g, \Big)$ is an estimate of $D\Big[E(\boldsymbol{\beta}), E(\boldsymbol{\theta}), E(\boldsymbol{u}))\Big]$

the *deviance of posterior mean* obtained when considering the means values of the generated posterior means of the model parameters.

In the hierarchical modelling representation to the *gsp* class, we use the complete-data likelihood function with observed and auxiliary latent variables (as fixed and random effects respectively) to obtain the posterior distributions to the parameter of interesse. In these cases, we consider marginal DIC to observed variables because the focus of the analysis is $p(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{\theta})$. In addition, to the *gsp* class only, we propose using the median of *Sum-of-squares of posterior latent residual* ($SSPLR = \sum_{i=1}^{n} e_i^*$) as a global measure of discrepancy to model comparison. This late approach is only applicable to the *gsp* class of links since that latent residuals are available only in this case. These values are shown in Table 2.

Given the comparison of two alternative models, the model that fits better a data set is the model with the smallest value of the posterior mean of the deviance, DIC, $EBIC$, $EAIC$ and $SSPLR$. In the presence of auxiliary latent variables, $EAIC$ and $EBIC$ are easily implemented because $2p$ and $plogN$ are fixed to penalize *posterior mean deviance* but in hierarchical models, it is not easy to define p and N. Moreover, there is no consensus in the use of the *Deviance of the posterior mean* (see discussion in Spiegelhalter *et al.*, 2003). By considering these aspects, the use of more than one criteria seems more appropriate to models comparison.

Table 2: Model comparison for the beetle mortality data

class	models	thin	time (sec.)	$Dbar$	DIC	$EBIC$	$EAIC$	$SSPLR$
non-*gsp*	logit	5	37	374.40	376.4	386.8	378.4	
	cloglog	5	41	366.7	368.7	379.1	370.7	
	scobit	65	48	369.2	369.5	381.5	373.2	
	power-logit	40	48	367.6	368.5	379.9	371.6	
	quadratic logit	15000	57	373.8	375.8	392.3	379.8	
gsp	probit	5	39	373.4	375.4	385.8	377.4	480.8
	CDS sp	610	199	160.4	273.4	178.9	166.4	479.8
	BBB sp	470	165	295.2	277.5	313.7	301.2	242.7
	standard sp	200	83	268.5	242.2	287.0	274.5	232.2
	complete sp	400	64	269.3	110.0	300.2	279.3	815.5

The model checking criteria used in Table 2 confirms results in Collet (2003, p. 149), where it is noted (using deviance and frequentist residuals) that a cloglog link fits this data as well as the polynomial logistic link using a quadratic term and both present better fit than the logistic model. This can also be easily verified by using, for example, GLM in SPlus. Note that the scobit and power logit link are also adequate in relation to symmetric links. However, from Table 2, it is clear that the asymmetric links in the *gsp* class are more adequate for the data set analyzed than the *non-gsp* links. Note that *standard sp* is better when considering $SSPLR$ and that *CDS sp* is better when $Dbar$, $EBIC$ and $EAIC$ are used. Also, *complete sp* is better when DIC is used, which is explained by the fact that $Dhat$ has high value as consequence of the variability associate with the estimation of parameters μ and σ^2 and, consequently, the $SSPLR$ estimate is also high. It is interesting to observe the good performance of the *standard sp* link using DIC and $SSPLR$, which can be considered the best of the asymmetric links with three parameters, since the good performance of *CDS sp* with *EBIC* and *EAIC* criteria seems to be consequence of the low value of the the Posterior Expected Deviance (*Dbar*) obtained with this link, which is reflected in *EBIC* and *EAIC*.

Moreover, by considering time of execution for the different models in Table 2, we note that in the *gsp* class of links, the time of execution is high as is also observed in Bazán *et al.* (2005) for the *BBB sp* when applied to Item Response Theory models. By considering this fact, thin values up to 400 are recommended and consequently a large number of iterations are necessary for inference based on the joint posterior density of the *gsp* class. Note also that the time required to simulate the chains with the cloglog links is smaller since that presence of high autocorrelations are not observed. On the other hand, the time of execution with the scobit link, and especially, with the quadratic logit are somewhat higher so that large thin values are necessary to diminish the first autocorrelations. In the future methods such as joint updating (Holmes and Held, 2006) can be considered to guarantee reduction of autocorrelations in the *gsp* class.

In addition, we consider that any skew-probit link in the *gsp* class is also more convenient for this data set since the cloglog link has no parameters controlling the rate of increasing (or decreasing) of the probability of success (failure) of the binary response and other parameters associated with this probability. We also consider that the *gsp* class is more adequate for this data set than other asymmetric links since the *gsp* class is easily implemented and it is unnecessary range specification for their implementation as discussed in the Introduction to other asymmetrical links. Moreover, as shown in Section 5, the posterior considered are proper for the generalized skew probit link, which is not the case for other asymmetric links, as is the case with the one due to Stukel (Chen *et al.*, 1999).

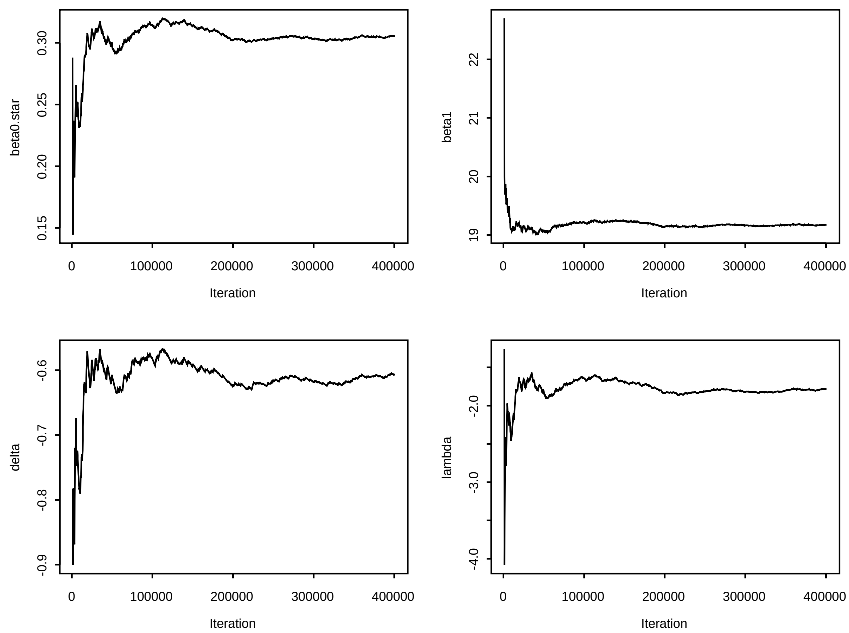

Figure 2: Running means for each parameter with the standard *gsp* link for the beetle data set

Finally, Table 3 reports, posterior statistics for the asymmetry parameter in the skew-probit models. As is expected, the values of the asymmetry parameters have different signs in the *CDS sp* and *BBB sp* links. Note also that for this data set, the 95% HPD interval to the asymmetry parameter in *BBB, standard* and *complete sp* links include zero: however, this is not the case for the *CDS sp* model and therefore the later model can be more adequate.

Table 3: Posterior summary for asymmetry parameters of the *gsp* class models with beetle mortality data

models	parameter	mean	median	95 %	HPD	MCerror
				Lower Bound	Upper Bound	
CDS sp	λ	4.19	4.02	1.43	7.82	0.07
BBB sp	δ	-0.48	-0.70	-0.99	0.46	0.02
	λ	-1.33	-0.99	-4.46	0.87	0.05
standard sp	δ	-0.61	-0.86	-0.99	0.46	0.02
	λ	-1.78	-1.69	-497.3	0.98	0.07
complete sp	δ	-0.59	-0.85	-1.0	0.45	0.02
	λ	-1.72	-1.58	-517.9	0.81	0.07
	μ	-0.02	-0.29	-4.33	456.5	0.57
	σ	9.93	6.38	0.03	317.1	0.59

Figure 3 depicts estimated posterior densities of each parameter in the *standard gsp* link. Note that the posterior densities for the δ (λ) parameters are somewhat skew so that the posterior mean seems not to be an adequate summary statistics.

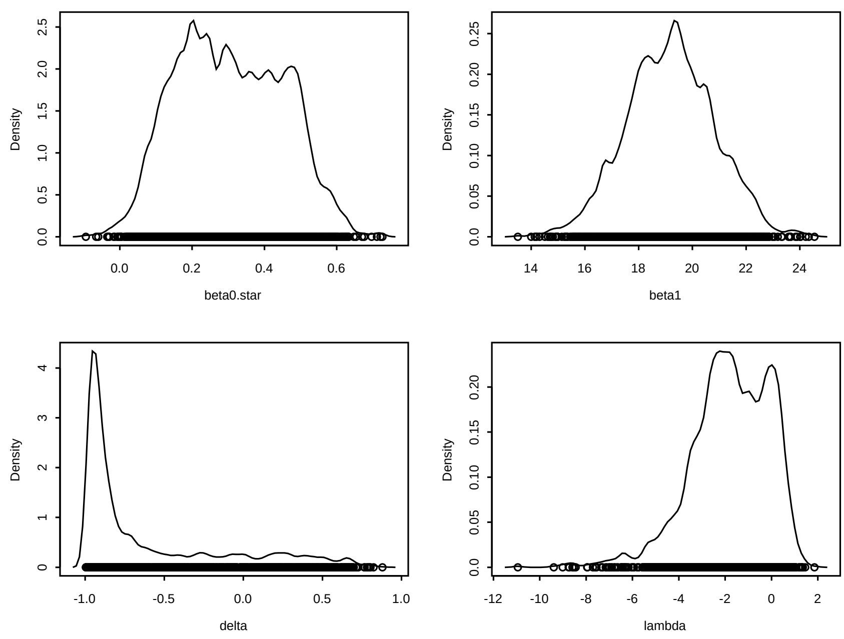

Figure 3: Estimated posterior densities for each parameter in the *standard gsp* link for the Beetle data set

8.2 Prostatic data set

Chib (1995), Collet (2003) and Congdon (2001) consider alternative models for the presence (or not) of prostatic nodal involvement in a sample of 53 cancer patients. The data set includes a binary response variable y that takes the value 1 if cancer had spread to the surrounding lymph nodes and value zero otherwise. Five explanatory variables are considered: age of the patient in years at diagnosis (x_1); level of serum acid phosphate (x_2); the result of an X-ray examination, coded 0 if negative and 1 if positive (x_3); the size of the tumor, coded 0 if small and 1 if large (x_4); and the pathological grade of the tumor, code 0 if less serious and 1 if more serious (x_5). The cited authors considered nine possible models (see Table 4), and showed the improvement obtained over the constant model (model 1) by the model involving the constant and x_3 (model 4), and by a more complex model involving the constant, logx_2, x_3 and x_4 (model 8). Finally, the most complex model considered adds x_5 to model 8 (model 9).

By considering binary probit regression and different criteria of the fit, Chib (1995), Collet (2003) and Congdon (2001) conclude that the model 4 fits the data set better than model 1, and model 8 fits better than model 9. Congdon (2001) argues that there seems to be little difference between models 8 and 9.

We evaluate the potential use of Probit, *CDS*, *BBB* and *standard sp* links for the nine models considered by Chib (1995) and introduce the models 10, 11, 12 and 13 by considering the x_2 and not logx_2 in the models 3, 7, 8 and 9, respectively. However, the MCMC algorithms seems not to converge when using the *CDS sp* link, as can be depicted from Figure 4, considering 1000000 iterations and thin values between 50 to 500. In the other cases, chain convergence was found for 100000 iterations after bur-in and a thin value of 50.

To compare the fits for the models entertained, we considered the same criteria used in Section 8.1 (DIC, $EAIC$, $EBIC$, *Dbar* and $SSPLR$). The priors considered to β are the ones given by Chib (1995) and Congdon (2001). That is, we consider β_2 $N(0, 25)$, β_1 $N(0, 100)$ and β_j $N(0.75, 25), j = 3, 4, 5, 6$, $\delta \sim U(-1, 1)$ (for *standard* and *BBB sp* links) and $\lambda \sim N(0, 100)$ (*CDS sp* link). The results are showed in Table 4.

For the probit link, as expected, model 4 shows improvement over model 1 (using *Dbar*, DIC, $EAIC$ and $EBIC$ criteria), and model 8 over model 9 (using $SSRL$, DIC, $EAIC$ and $EBIC$ criteria). Moreover, models 10, 11, 12 and 13 are worst than models 3, 7, 8 and 9, respectively. Then, the inclusion of the logx_2 is justified. A similar conduct is observed when considering the *standard sp* link. Using *BBB sp* link, the conduct is also similar with exception that model 8 is better than 9 only using DIC and $SSPLR$, but not using $DBAR$, $EAIC$ and $EBIC$ criteria.

In general, most criteria indicate models 8, 9, 13 and 7, as most adequate. However, order of preference seems to depend on the link function. The above ordering is given for the probit link. For the *standard sp* link, the order is 8, 7, 9 and 13; and 9,8, 13 and 7 for the *BBB sp* link. The other models seems not appropriate.

By comparing models using different *gsp* links, we find that the models under *BBB* and *Standard sp* are better than the corresponding models using probit link.

We also obtained estimates for the percentage of correct predictions for the best models. It is obtained by evaluation the posterior probabilities for each model, by considering the posterior summary of β and λ (mean and median, respectively) and the values of the explanatory variable. Using the probit link, the values are: $69.8\%, 73.6\%, 75.5\%$ and 79.2% for the models 7, 8, 9 and 13, respectively, but invoking the principle of parsimony, model 8 should be considered.

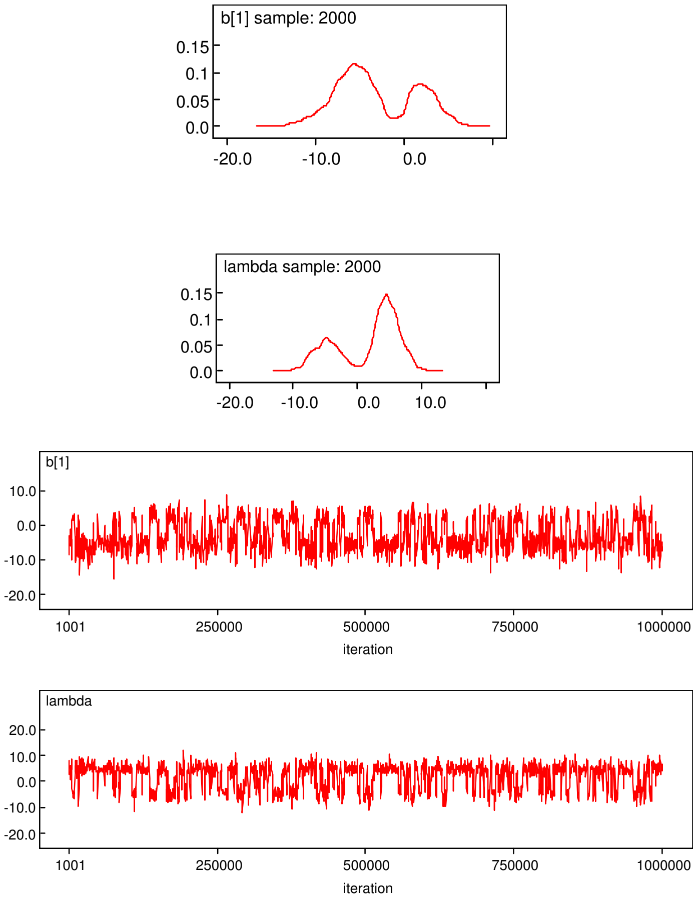

Figure 4: Chain history and posterior densities for the parameters in Model 8 under *CDS link* for the prostatic data set

For the *standard* and *BBB sp* links, the values are $75.5\%, 75.5\%, 75.5\%, 79.2\%$ and $69.8\%, 75.5\%, 75.5\%, 79.2\%$, respectively. We observe that models 8, 9 and 13 (in this order) present better prediction than model 7, but again invoking the principle of parsimony, model 8 should be considered with both links.

Finally, by considering both criteria, we indicate that model 8 using the *standard sp* link is an alternative to model 8, using probit link to the data set considered. Another alternative model is model 9 using *BBB sp*.

9 Extensions and Discussion

This article proposes a new asymmetrical link for binary response by considering the cumulative distribution of the skew-normal distribution (Dalla-Valle, 2004). The generalized skew-probit class link introduced has as particular cases the skew-probit link due to Chen *et al.* (1999), the skew-probit link due Bazán *et al.* (2005) and the probit link. This model introduces a parameter for the asymmetry of the response curves that is easily interpreted and defines a class of asymmetric links that controls the rate of increasing (or decreasing) of the probability of success (failure) of the binary responses.

Furthermore, corresponding to this link, the probability of success is obtained by considering the cdf of a distribution function evaluated at the linear predictor. The asymmetry parameter is associated with the distribution chosen and is independent of the linear predictor and the latent linear structure is not necessary for model formulation. With this, the formulation of the generalized skew probit model seems more appropriate than the formulation due to Chen *et al.* (1999) and they mention that the model can be easily generalized (see Chen, 2004).

Moreover, the *gsp* class is a fixed effects model, assuming that all observations are independent of each other, and is not appropriate for the analysis of several types of correlated data structures, in particular, for clustered and/or longitudinal data and more generally, in multilevel models. In this case mixed models for binary response can easily proposed by extending the *gsp* class . Assume there are $i = 1, \ldots, n$, subjects (level-2 units), and $j = 1, \ldots, n_i$, repeated observations (level-1 units), nested within each subject. Denote z_{ij} as the $r \times 1$ vector of variables having random effects (a column of ones is usually included for the random intercept) and b_i a vector of random effect that can be assumed to follow a multivariate normal distribution with mean vector $\mathbf{0}$ and variance-covariance matrix $\boldsymbol{\Sigma_b}$ or a multivariate skew-normal distribution (see Arellano *et al*, 2005). In this case the model in (1) is now written as

$$\eta_{ij} = \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{b}_i,$$

$j = 1, \ldots, n_i$, $i = 1, \ldots, n$. Note that p_{ij} is now specified as $E(y_{ij}|\boldsymbol{b}_i, \boldsymbol{x}_{ij})$, namely, in terms of the vector of random effects. The results shown in the paper regarding the augmented likelihood are valid and can be easily applied to this new model since hardly a hierarchical structure is added to the hierarchical structure in Section 6 corresponding to the random effects model introduced, $\boldsymbol{b_i} \sim N_r(\mathbf{0}, \boldsymbol{\Sigma_b})$ or $\boldsymbol{b_i} \sim SN_r(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Delta})$ (see notation in Arellano *et al.*, 2005). In a particular case of that, if the vector of random effects is not considered, then we have the multivariate skew probit model as the multivariate probit model (Chib and Greenberg, 1998).

Moreover, extensions of the methods developed in this paper for dichotomous responses to ordinal response data (Albert and Chib, 1993) follow, by considering this model in terms of cumulative probabilities, so that the conditional probability of a response in category c is obtained as the difference of two conditional cumulative probabilities:

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

where

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

with $c-1$ strictly increasing model thresholds γ_c. Thus, we have a random-effects ordinal regression model for multilevel analysis.

Applications in several areas where symmetric links are not justified can be obtained with the proposed model. It includes binomial models, epidemiological studies, multilevel modelling, longitudinal data analysis, meta-analysis and item response theory. Applications of the *gsp* class link proposed to item response models can seen in Bazán *et al.* (2005). For others application in regression see Chen *et al.* (1999).

Main results about the existence of Maximum Likelihood estimators and propriety of the posterior distributions of the parameters of the model when improper uniform priors are chosen are given. A Bayesian estimation approach is developed and implementations of the approach can be easily obtained by considering the versions of the augmented likelihood proposed, particularly the version introduced in Section 5. That is an attractive aspect of the model that can be implemented via MCMC by using software as WinBugs with uniform priors.

However, as Johnson and Albert (1999) mention, the specification of a prior density in binary regression can be a difficult task due to the indirect effect that the regression parameters may exert on the success probabilities. Paulino *et al.* (2003) mentions that, in a subjective bayesian analysis, the introduction of further regression parameters leads to a potentially serious problem because these parameters do not relate directly to the data as in the case when competing choices of the link function must be considered, which seems, happened in the developed application. They stick to expert prior elicitation and suggests that methods as the one proposed by Bedrick *et al.* (1996) can be used. In this class of distributions, prior beliefs about the location of the success probabilities p_i are assessed for particular values of the covariates x_i, and this information is used to construct a prior for the regression parameter vector β. By considering the estimation of the success probabilities for several models in the Table 3, the expert prior elicitation of p_i can be important for the choice of the link.

Finally other aspects in the paper is the comparison of symmetrical and asymmetrical models by using the Deviance Information Criterion (DIC) described in Spiegelhalter *et al.* (2002), the Expected Akaike Information Criterion ($EAIC$) and the Expected Bayesian Information (Schwarz) Criterion ($EBIC$) proposed in Brooks (2002). We also introduce latent residuals for the models and global discrepancy measures as the posterior sum-of-squares of latent residuals ($SSPLR$), which can be used for model comparison including different choices of the *gsp* class. This criterion seems to result adequate in the example used and can be used in accordance with other approaches as the development to see the relationship between the skew probit and probit.

APPENDIX

Properties of the skew normal distribution

We present next some important properties of the skew normal distribution.

P1. The moments generating function is given by

$$M_R(t) = E(e^{tR}) = 2exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right)\Phi(\sigma\delta t)$$

P2. In particular, as consequence of P1., the mean and variance are gives by: $E[R] = \mu + \sqrt{\frac{2}{\pi}}\delta\sigma$ and $Var[R] = (1 - \frac{2}{\pi}\delta^2)\sigma^2$, where $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}} \in [-1, 1]$.

Approximating to two decimals, if $\mid \delta \mid$ increases, then $E[R] \in [\mu - 0.8\sigma, \mu + 0.8\sigma]$ increases and $Var[R] \in [0.36\sigma^2, \sigma^2]$ decreases.

P3. Consider $W \sim N(0, 1)$ and $V \sim HN(0, 1)$ independent random variables with $|\delta| < 1$. Then, $R = \mu + \sigma\big[\delta V + (1 - \delta^2)^{1/2} W\big] \sim SN(\theta)$, where $\theta = (\mu, \sigma^2, \lambda)$, and $R|V \sim N(\mu + \sigma\delta v, (1 - \delta^2)\sigma^2)$.

P4. $S = (R - \mu)/\sigma \sim SN_A(\lambda)$. Also $R^* = a + bR \sim SN(a + b\mu, b^2\sigma^2, sign(b)\lambda)$ where $sign(.)$ is the sign function which is equal to 1 if the argument is positive and is -1 otherwise.

P5. In particular, given $S \sim SN_A(\lambda)$, then $S^* = a+bS \sim SN(a, b^2, sign(b)\lambda)$ and follows that $R = \mu + \sigma S \sim SN(\mu, \sigma^2, \lambda)$ and $R = \mu - \sigma S \sim SN(\mu, \sigma^2, -\lambda)$. Also given $S \sim SN_S(\lambda)$, then $R = \mu + \sigma(1 - \delta^2)S \sim SN(\mu, \sigma^2, \lambda)$.

Proofs of the Propositions

Proof of Proposition 1. From the stochastic representation given in Table 1, it follows that:

$$f_\lambda(s) = \int_0^\infty \phi(\frac{s - \delta v}{\sqrt{1 - \delta^2}})g(v)dv$$

Hence, the cdf of S can be obtained as

$$F_\lambda(s) = \int_{-\infty}^{s}\int_0^\infty \phi\Big(\frac{t - \delta v}{\sqrt{1 - \delta^2}}\Big)g(v)dvdt = \int_0^\infty \Big[\int_{-\infty}^{s} \phi\Big(\frac{t - dv}{\sqrt{1 - \delta^2}}\Big)dt\Big]g(v)dv$$

$$F_\lambda(s) = \int_0^\infty \Phi\Big(\frac{s - \delta v}{\sqrt{1 - \delta^2}}\Big)g(v)dv$$

Defining $R = \mu + \sigma S$, it follows that $F_\theta(r) = P(R \leq r) = P(\mu + \sigma S \leq r) = P(S \leq \frac{r-\mu}{\sigma})$ and we have, therefore,

$$F_\theta(r) = \int_0^\infty \Phi\Big(\frac{\frac{r-\mu}{\sigma} - dv}{\sqrt{1 - \delta^2}}\Big)2\phi(v)dv.\square$$

Proof of Proposition 2. Given $S \sim SN_A(\lambda)$ we can write the corresponding cdf as

$$F_\theta(s) = \int_{-\infty}^{s} 2\phi(t)\Phi(\lambda t)dt = \int_{-\infty}^{s} 2\phi(t)\Phi(\frac{-\delta t}{\sqrt{1-\delta^2}})dt = 2\Phi_2(s, 0; -\delta).$$

The last equality follows from Parrish and Bergmann (1981), which shows that the cumulative distribution function of the standard bivariate normal distribution with correlation ρ evaluated at (h, k) can be written as

$$\Phi_2(h, k; \rho) = \int_{-\infty}^{h} \phi(w)\Phi\Big(\frac{k - \rho w}{\sqrt{1-\rho^2}}\Big).$$

Hence, the desired result follows by considering the transformation $R = \mu + \sigma S$, $F_\theta(r) = 2\Phi_2\Big(\frac{r-\mu}{\sigma}, 0, -\delta\Big)$, and Property 5 in Appendix. □

Proof of Proposition 3. When θ is known, the likelihood function in (11) is reduced to $L(\boldsymbol{\beta}|\boldsymbol{y}, \boldsymbol{X}, \theta)$ and for simplicity we write $L(\boldsymbol{\beta}|\boldsymbol{y})$. Considering the prior $\pi(\boldsymbol{\beta}) \propto 1$, the posterior is proper if $\int_{R^k} L(\boldsymbol{\beta}|\boldsymbol{y})d\boldsymbol{\beta} < \infty$. To see that conditions (*C1*) and (*C2*) are sufficient, apply Theorems 2.3 and 3.1 given by Chen and Shao (2000). Since the skew normal distribution is a continuous distribution such that $M_R(t) = E(e^{tR}) = \int_{-\infty}^{\infty} e^{tr}dF_\theta(r) < \infty$, and, consequently, $E(|u^k|) = \frac{d^k M_R(t)}{dt^k}|_{t=0} < \infty$; it has finite moments (this can be easily verified from Property 1).

Conditions *(C1)* and *(C2)* are also sufficient for the existence of the ML estimator which follows by considering Theorem 3.1 given in Chen and Shao (2000), since the cdf of the skew normal distribution is continuous. □

Proof of Proposition 4. Given improper uniform priors for λ and $\boldsymbol{\beta}$, the problem to guarantee proper posteriors was studied by Chen *et al.*(1999) for the *CDS* sp link. In these cases, since the pdf of the half normal distribution take positives values in $\mathbb{R}$ and the skew normal distribution has finite moments, by considering their Theorem 2, (C1') and (C2') are necessary conditions which guarantee that the posterior distribution of $(\boldsymbol{\beta}, \lambda)$ are proper. As $p(\boldsymbol{\beta}, \lambda|\boldsymbol{y}, \boldsymbol{X}) \propto L(\boldsymbol{\beta}, \lambda|\boldsymbol{y}, \boldsymbol{X})\pi(\boldsymbol{\beta}, \lambda)$, and $\pi(\boldsymbol{\beta}, \lambda) \propto 1$, we have that

$$\int_{-\infty}^{\infty}\int_{0}^{\infty} L(\boldsymbol{\beta}, \lambda|\boldsymbol{y}, \boldsymbol{X})d\boldsymbol{\beta}d\lambda < \infty$$

where, by (18),

$$L(\boldsymbol{\beta}, \lambda \mid \boldsymbol{y}, \boldsymbol{X}) = \prod_{i=1}^{n} \int_{0}^{\infty} [\Phi(\eta_i + \lambda v_i)]^{y_i} \Big[1 - \Phi(\eta_i + \lambda v_i)\Big]^{1-y_i} g(v_i)dv_i.$$

In these cases, the representation given in (2)-(4) is a linear structural subjacent to the likelihood function above, and by considering (3) $S_i = \eta_i + \lambda v_i + u_i \sim SN_S(\lambda)$. Moreover, the fact that any skew normal distribution is continuous (in particular the standard Sahu´s skew normal distribution), and by considering that the conditions (C1') and (C2') being satisfied, Theorem 5 given by Chen *et al.*(1999) imply that the ML estimators of $(\boldsymbol{\beta}, \lambda)$ exist.

To extend the results above for the skew normal distribution $SN(\mu, \sigma^2, \lambda)$ where μ and σ^2 are given, consider Property 4 in this Appendix. That is, let $R = \mu + \sigma(1-\delta^2)S \sim SN(\mu, \sigma^2, \lambda)$. In this case, we have that

$$L(\mu, \sigma^2|\boldsymbol{y}, \boldsymbol{X}) = \int_{-\infty}^{\infty}\int_{R^k} L(\boldsymbol{\beta}, \theta|\boldsymbol{y}, \boldsymbol{X}) d\boldsymbol{\beta} d\lambda < \infty$$

where $L(\boldsymbol{\beta}, \theta|\boldsymbol{y}, \boldsymbol{X})$ is given in (13). By considering this result, when proper priors are specified for μ and σ^2, it follows directly that

$$\int_0^{\infty}\int_{-\infty}^{\infty} L(\mu, \sigma^2|\boldsymbol{y}, \boldsymbol{X})\pi(\mu)\pi(\sigma^2) d\mu d\sigma^2 < \infty$$

Hence, we have that the posterior distribution of $(\boldsymbol{\beta}, \theta)$ is proper since $p(\boldsymbol{\beta}, \theta|\boldsymbol{y}, \boldsymbol{X}) \propto L(\boldsymbol{\beta}, \theta|\boldsymbol{y}, \boldsymbol{X})\pi(\boldsymbol{\beta}, \lambda)\pi(\mu), \pi(\sigma^2) = L(\mu, \sigma^2|\boldsymbol{y}, \boldsymbol{X})\pi(\mu)\pi(\sigma^2)$. □

Proof of Proposition 5. Consider $s_i = r + e$, where $e \sim SN(-\mu, \sigma^2, -\lambda)$, then

$$p_i = P(Y_i = 1) = P(s_i > 0) = P(e + r > 0) = P(e > -r) = 1 - P(e \leq -r)$$

and it follow that

$$p = 1 - P(\frac{e+\mu}{\sigma} \leq \frac{-r+\mu}{\sigma}) = 1 - F_{-\lambda}^{A}(\frac{-r+\mu}{\sigma}) = F_{\lambda}^{A}(\frac{r-\mu}{\sigma})$$

where the final expressions are obtained from Property 4 in Appendix and considering that $F_{\lambda}^{A}(z) = 1 - F_{-\lambda}^{A}(-z)$ (Azzalini, 1985). Finally, by considering Property 5, we have that $p = F_{\lambda}^{A}(\frac{r-\mu}{\sigma}, \lambda) = F_{\theta}(r)$, where $\theta = (\mu, \sigma^2, \lambda)$.□

References

Achen, C. H. (2002). Toward a new political methodology: microfoundations and art. *Annual Review of Political Science*, *5*, 423-450.

Albert, J. H. and Chib, S. (1993). Bayesian analysis of binary and polytomous response data. *Journal of the American Statistical Association*, *88*, 669-679.

Albert, J. H. and Chib, S. (1995). Bayesian residual Analysis for Binary Response Regression Models. *Biometrika*, *82*, 747-759.

Aranda-Ordaz (1981). On two families of transformations to additivity for binary response data. *Biometrika*, *68*, 357-363.

Arellano-Valle, R.B., Bolfarine, H. and Lachos, V.H. (2005). Skew-normal linear mixed models. *Journal of Data Science*, 3, 415-438.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scand. J. Statistical*, *12*, 171-178.

Basu, S. and Mukhopadhyay, S. (2000). Binary response regression with normal scale mixtures links, in *Generalized Linear Models: A Bayesian Perspective*,eds. D.K. Dey, S.K. Ghosh, and B.K: Mallick, New York: Marcel Dekker.

Bazán, J., Branco, M.D., Bolfarine, H. (2005). A skew item response model. *Accepted in Bayesian Analysis.*

Bedrick, E. J., Christensen, R. and Johnson, W (1996). A new perspective on priors for generalized linear models. *Journal of the American Statistical Association*, *91*, 1450-1460.

Branco, M.D. and Dey, D.K.,(2002). Regression model under skew elliptical error distribution. *The Journal of Mathematical Sciences*, Delhi, New Series, *1*, 151-169.

Brooks, S. P. (2002). Discussion on the paper by Spiegelhalter, Best, Carlin, and van de Linde (2002). *Journal of the Royal Statistical Society Series B*, *64*, 3,616-618.

Chen, M. H. (2004). The skewed link models for categorical response data, in *Skew-elliptical distributions and their applications: a journey beyond normality*, ed. Marc. G. Genton, Boca Raton: Chapman & Hall/CRC.

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*, 448, 1172-1186.

Chen, M. H. and Shao, Q. M. (2000). Property of posterior distribution for dichotomous quantal response model. *Proceedings of the American Mathematical Society.* 129,1,293-302.

Chen, M. H., Shao, Q. M, and Ibrahim, J. G (2000). *Monte Carlo methods in Bayesian computation.* New York: Springer Verlag.

Chib, S. and Greenberg, E. (1995). Undestanding the Metropolis-Hastings algorithm. *The American Statistician*, *49*, 327-335.

Chib, S. and Greenberg, E. (1998). Bayesian analysis of multivariate probit models. *Biometrika*, *85*, 347-361.

Collet, D. (2003). *Modelling binary data.* Chapman & Hall/CRC, Second Edition, Boca Raton, USA.

Czado, C., and Santner, T. J. (1992). The effect of link misspecifiaction on binary regression inference. *Journal of Statistical Planning and Inference*, *33*, 213-231.

Czado, C. (1994). Bayesian inference of binary regression models with parametric link. *Journal of Statistical Planning and Inference*, *41*, 121-140.

Czado, C. e Raftery, A. E. (2003). Choosing the link function and accounting for link uncertainty in generalized linear models using Bayes factors. Working Paper no. 18. Center for Statistics and the Social Sciences University of Washington. Disponível na web.

Dalla Valle (2004). The Skew-Normal Distribution, in *Skew-elliptical distributions and their applications: a journey beyond normality*,ed. Marc. G. Genton, Boca Raton: Chapman & Hall/CRC.

Geweke, J. (1992). Evaluating the accuraty of sampling-based approaches to the calculation of posterior moments. In J. M. Bernardo, J. O. Berger, A. P. Dawid, & A. F. M. Smith (Eds.). *Bayesian Statistics*, *4*, (pp 169-193). Oxford: Oxford University Press.

Guerrero, V. M., and Johnson, R. A. (1982). Use of the Box-Cox transformation With binary response models. *Biometrika*, *69*, 309-314.

Haro-López, R.A., Mallick, B. K., and Smith, A.F.M. (2000). Binary regression using data adaptative robust link functions, in *Generalized Linear Models: A Bayesian Perspective*,eds. D.K. Dey, S.K. Ghosh, and B.K: Mallick, New York: Marcel Dekker.

Holmes, C. C. and Held, L. (2006). Bayesian Auxiliary Variable Models for Binary and Multinomial Regression. *Bayesian Analysis*, *1*, 145-168.

Henze, N. (1986). A probabilistic representation of the "skew-normal" distribution. *Scandinavian Journal Statistical.* *13*, 271-275.

Ibrahim, J. G. and Laud, P. W. (1991). On Bayesian analysis of generalized linear models using Jeffrey's prior. *Journal of the American Statistical Association.* *93*, 1282-1293.

Johnson, V., and Albert, J. (1999). *Ordinal data modeling.* New York, MA: Springer-Verlag.

Kass, R. E., and Raftery, A. E.(1995). Bayes factors. *Journal of the American Statistical Association*, *93*, 1282-1293.

Liseo, B. and Loperfido, N. (2006). A note on reference priors for the scalar skew-normal distribution. *Journal of Statistical Planning and Inference.* In press.

Nagler J. (1994). Scobit: an alternative estimator to logit and probit. *American Journal Political Science*, *38*, 230-255.

Paulino, C., Soares, P. and Neuhaus, J. (2003). Binomial regression with misclassification. *Biometrics*, 59,670-675.

Prentice, R. L. (1976). A Generalization of the probit and logit methods for dose-response curves. *Biometrika*, 32,761-768.

Sahu, S. K., Dey, D.K., Branco, M.D. (2003). A new class of multivariate skew distributions with applications to Bayesian regression models. *The Canadian Journal of Statistics*, Vol. 29,217-232.

Spiegelhalter, D. J., Thomas, A., Best, N. G., and Gilks, W.R.(1996). *BUGS 0.5 examples* (Vol. 1 Version i). Cambrigde, UK: University of Cambride.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal Royal Statistical Society, Series B*, *64*,3, 583-639.

Stukel, T. (1988). Generalized logistic models. *Journal of the American Statistical Association*, *83*,426-431.

Taylor, J. and Siqueira, A. (1996). The cost of adding parameters to a model. *Journal of the Royal Statistical Society*, *58*,593-607.

Zellner, A. and Rossi, P. E. (1984). Bayesian analysis of dichotomous quantal response models. *Journal of Econometrics*, *25*,365-393.

Table 4: Criteria of Comparison of models to probit, *standard* and *BBB sp* links for the Prostatic data set

		Probit					*Standard sp*					*BBB sp*				
Model	Description	*SSRL*	*Dbar*	*DIC*	*EAIC*	*EBIC*	*SSPLR*	*Dbar*	*DIC*	*EAIC*	*EBIC*	*SSPLR*	*Dbar*	*DIC*	*EAIC*	*EBIC*
1	c	52.8	71.3	72.3	75.2	73.3	44.3	62.9	55.5	70.8	66.9	46.0	63.8	57.3	63.8	63.8
2	$c + x_1$	54.3	71.2	73.1	79.1	75.2	43.4	61.5	53.1	73.4	67.5	40.8	62.9	56.9	62.9	62.9
3	$c + log(x_2)$	54.0	66.9	69.0	74.9	70.9	44.8	58.4	52.1	70.4	64.4	41.9	61.4	57.8	61.4	61.4
4	$c + x_3$	54.9	63.2	65.3	71.2	67.2	45.2	56.1	51.0	68.0	62.1	42.7	57.9	54.6	58.0	58.0
5	$c + x_4$	54.7	64.6	66.6	72.5	68.6	45.2	56.8	51.0	68.7	62.8	41.5	58.7	54.9	58.7	58.6
6	$c + x_5$	55.0	68.2	70.2	76.2	72.2	45.8	60.5	54.8	72.4	66.5	42.3	62.7	59.1	62.6	62.6
7	$c + log(x_2) + x_4$	55.3	59.4	62.4	71.3	65.4	44.9	52.0	47.7	67.9	60.0	45.0	55.2	54.0	55.3	55.3
8	$c + log(x_2) + x_3 + x_4$	56.2	55.5	59.4	71.4	63.5	46.6	48.4	45.3	68.2	58.4	46.3	52.3	53.0	52.2	52.2
9	$c + log(x_2) + x_3 + x_4 + x_5$	57.4	54.6	59.6	74.5	64.6	47.3	48.1	46.6	71.9	60.1	48.0	51.7	53.8	51.6	51.6
10	$c + x_2$	54.6	69.2	71.2	77.1	73.2	44.8	60.8	54.4	72.7	66.8	42.4	63.3	59.5	63.4	63.5
11	$c + x_2 + x_4$	54.7	61.5	64.5	73.4	67.5	46.0	53.8	49.2	69.7	61.8	44.7	57.2	55.8	57.2	57.2
12	$c + x_2 + x_3 + x_4$	54.6	69.2	71.2	85.1	77.2	46.6	50.4	47.6	70.2	60.4	46.3	53.6	53.9	53.5	53.5
13	$c + x_2 + x_3 + x_4 + x_5$	56.9	56.6	61.6	76.5	66.6	48.1	49.9	48.2	73.7	61.9	48.6	53.7	55.9	53.7	53.7