Brazilian Journal of Probability and Statistics
2014, Vol. 28, No. 4, 467–482
DOI: 10.1214/13-BJPS218

Bayesian skew-probit regression for binary response data

Jorge L. Bazán[a], José S. Romeo[b] and Josemar Rodrigues[a]

[a]*Universidade de São Paulo*
[b]*Universidad de Santiago de Chile*

Abstract. Since many authors have emphasized the need of asymmetric link functions to fit binary regression models, we propose in this work two new skew-probit link functions for the binary response variables. These link functions will be named power probit and reciprocal power probit due to the relation between them including the probit link as a special case. Also, the probit regressions are special cases of the models considered in this work. A Bayesian inference approach using MCMC is developed for real data suggesting that the link functions proposed here are more appropriate than other link functions used in the literature. In addition, simulation study show that the use of probit model will lead to biased estimate of the regression coefficient.

1 Introduction

Binary regression is an important special case of the generalized linear models for which Bayesian inference has been developed in the literature (see, Dey, Ghosh and Mallick, 1999). For example, the Probit model (PM) is a generalized linear model obtained by considering $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)'$ a $n \times 1$ vector of n independent dichotomous random variables with probability $p_i = P[Y_i = 1]$, and $\mathbf{x}_i = (x_{i1}, \dots, x_{ik})'$ a $k \times 1$ vector of covariates, where x_{i1} may equals 1, corresponding to an intercept, $i = 1, \dots, n$. Also, $\mathbf{X}$ denotes the $n \times k$ design matrix with rows $\mathbf{x}_i'$, and $\boldsymbol{\beta} = (\beta_1, \dots, \beta_k)'$ is a $k \times 1$ vector of regression coefficients. For this model, the probability of binary response variable is given by

$$p_i = E(Y_i|\mathbf{x}_i) = \Phi(\boldsymbol{\eta}_i) = \Phi(\mathbf{x}_i'\boldsymbol{\beta}), \qquad i = 1, \dots, n, \tag{1.1}$$

where $\Phi(\cdot)$ denotes a cumulative distribution function (c.d.f.) of the standard normal distribution. The inverse of the function Φ, namely Φ^{-1}, is called link function and $\boldsymbol{\eta}_i = \mathbf{x}_i'\boldsymbol{\beta}$, the linear predictor. The graphic considering p_i as a function of η_i is called response curve or probability of success and has a symmetric form centered in 0.5. Other symmetric link functions are obtained when Φ is replaced by the c.d.f. of a distribution in the class of the elliptical distributions, such as, the logistic, Student-t, double exponential and Cauchy distributions (Albert and Chib, 1993). However, as Chen, Dey and Shao (1999) have been argued, when

Key words and phrases. Skew-probit links, binary regression, Bayesian estimation, power normal distribution, reciprocal power normal distribution.
Received April 2012; accepted January 2013.

the probability of a given binary response variable approaches 0 at different rates than it approaches 1, symmetric link functions may be not useful to fit binary data and asymmetric link functions must be considered. Many asymmetric link functions, including the logit or probit link as special case, have been considered by Prentice (1976) (available in the Stata software), Guerrero and Johnson (1982), Stukel (1988), Czado and Santner (1992), Czado (1994), Chen, Dey and Shao (1999, 2001), Bazán, Branco and Bolfarine (2006), Bazán, Bolfarine and Branco (2010) and recently Wang and Dey (2010).

Moreover, some textbooks (see, e.g., Collet, 2003) have reported that an asymmetric link function may be more appropriate than a symmetric one for some specific situations.

In this paper, we propose two new asymmetric links. In both cases, the probit link function is a particular case of them. The skew link functions proposed here introduce a parameter that controls the rate of increase (or decrease) of the probability of success (failure) of the binary response variables. One is based on the c.d.f. of the power-normal (PN) distribution given by Gupta and Gupta (2008), Kundu and Gupta (2013) and the other is based on the c.d.f. of named reciprocal power-normal (RPN) distribution introduced here. In this context, the probability of success is obtained from a c.d.f. evaluated at the linear predictor. An asymmetry parameter associated with these c.d.f.'s is also introduced independently of the linear predictor and a latent linear structure will be not necessary for this link approach. The most important aspect of the modeling in this setting is the potential improvement to model fit that is gained by using this particular class of asymmetric link functions since that some data simply cannot be modeled appropriately with symmetric link functions as showed in this paper.

This work is organized as follows. In Section 2, the PN and RPN distributions are presented. In Section 3, two new skew-probit models for binary responses variables are formulated. In Section 4, a Bayesian estimation approach is developed using a selection models criteria and MCMC output. In Section 5, applications are showed, one by considering simulated data and other considering real data. Finally, a discussion and extensions of the link functions proposed in this paper are considered in Section 6.

2 A power normal and reciprocal power normal distributions

The probability density function (p.d.f.) of the PN distribution introduced by Gupta and Gupta (2008) is given by

$$g_1(r) = \frac{\lambda}{\sigma}\left[\Phi\left(\frac{r-\mu}{\sigma}\right)\right]^{\lambda-1}\phi\left(\frac{r-\mu}{\sigma}\right), \tag{2.1}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ denote, respectively, the density and cumulative distribution functions of the standard normal distribution.

The notation considered is $R \sim \mathrm{PN}(\theta)$ with $\theta = (\mu, \sigma^2, \lambda)$, where $\mu \in \mathbb{R}$ is a location parameter, $\sigma^2 > 0$ is a scale parameter and $\lambda > 0$ is a shape parameter.

If $\lambda = 1$, the density of R in (2.1) reduces to the density of the $N(\mu, \sigma^2)$. The special case $\mu = 0$ and $\sigma^2 = 1$ is called the *standard PN* distribution which will be denoted by $S \sim \mathrm{PN}(\lambda)$ and its corresponding p.d.f. and c.d.f. are given by

$$f_1(s) = \lambda\big[\Phi(s)\big]^{\lambda-1}\phi(s), \qquad F_1(s) = \big[\Phi(s)\big]^{\lambda}, \tag{2.2}$$

respectively. Some properties of these distributions are presented in the Appendix.

A new random variable can be obtained by the following c.d.f. or p.d.f.

$$F_2(s) = 1 - \big[\Phi(-s)\big]^{\lambda}, \qquad f_2(s) = \lambda\big[\Phi(-s)\big]^{\lambda-1}\phi(s). \tag{2.3}$$

In this case we write $S \sim \mathrm{RPN}(\lambda)$, to denote the *standard reciprocal PN* distribution. The name is justified since $F_2(y) = 1 - F_1(-y)$ and, so, the standard PN and the standard RPN distributions are distinct, although, closely related since one is the reflection of the other. That is also justified since that if $X \sim \mathrm{PN}(\mu, \sigma^2, \lambda)$ then $-X \sim \mathrm{RPN}(\mu, \sigma^2, \lambda)$. Also, note that $F_1(-y) \neq 1 - F_1(y)$ or $F_2(-y) \neq 1 - F_2(y)$ and then F_1 and F_2 are not symmetric.

A location-scale version of the RPN distribution with parameter $\theta = (\mu, \sigma^2, \lambda)$ is given by

$$g_2(r) = \frac{\lambda}{\sigma}\left[\Phi\left(-\left(\frac{r-\mu}{\sigma}\right)\right)\right]^{\lambda-1}\phi\left(\frac{r-\mu}{\sigma}\right). \tag{2.4}$$

If $\lambda = 1$, the density of R in (2.4) reduces to the density of the $N(\mu, \sigma^2)$.

As suggested by Gupta and Gupta (2008), the PN density is an unimodal density which is skewed to the right if $\lambda > 1$ and to the left if $0 < \lambda < 1$. On the other hand, by considering the reciprocal formulation the RPN density is also unimodal density which is skewed to the left if $\lambda > 1$ and to the right if $0 < \lambda < 1$. Thus, the density in (2.2) and (2.3) are weighted normal densities with the weight function $w_1(s) = \lambda[\Phi(s)]^{\lambda-1}$ and $w_2(s) = \lambda[\Phi(-s)]^{\lambda-1}$, respectively, given by

$$f_i(s) = \frac{w_i(s)\phi(s)}{E[w_i(S)]}, \qquad i = 1, 2. \tag{2.5}$$

Figure 1 displays probability density fuctions for various values of λ in both cases.

3 The skew-probit link models

Two new models for binary data, called here as Skew-Probit models (SPM), can be obtained by replacing $\Phi(\cdot)$ in (1.1) by the standard PN c.d.f. or the standard RPN, that is, for both cases we have:

$$p_i = E(Y_i|\mathbf{x}_i) = F_\lambda(\eta_i) = F_\lambda\big(\mathbf{x}_i'\boldsymbol{\beta}\big), \tag{3.1}$$

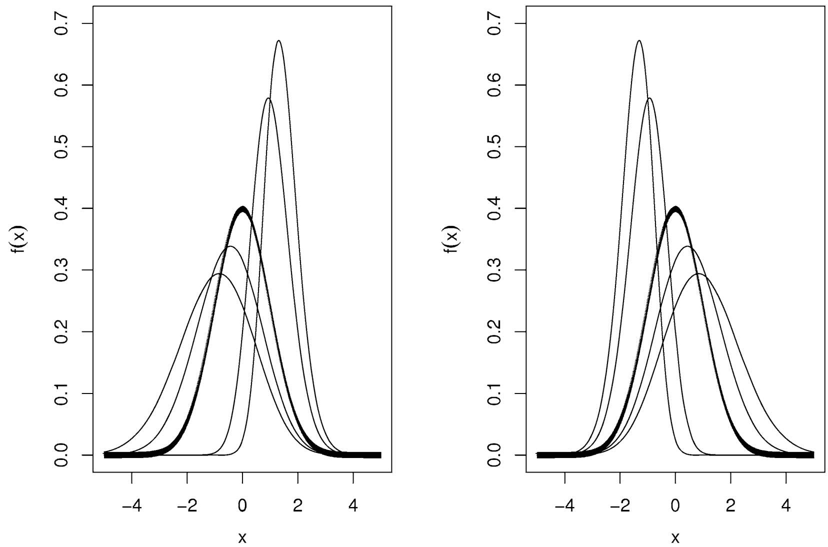

Figure 1 *Probability density functions*: *Power Normal* (*left*) *for* $\lambda = \{0.4, 0.6, 1, 4, 8\}$ *and Reciprocal Power Normal* (*right*) *for* $\lambda = \{8, 4, 1, 0.6, 0.4\}$ *with curves showed from left to right in both figures*.

where F_λ denotes the c.d.f. $F_1(\cdot)$ (or $F_2(\cdot)$) in (2.2) (or (2.3)).

The formulations in (2.2) and (2.3) imply that the two new models, named, Power Probit (PP) and Reciprocal Power Probit (RPP) models have the usual probit model in (1.1) as particular case and nested in the class of SPM.

Figure 2 depicts different probability curves for the PP and RPP models by using different values for λ and η. For $\lambda = 1$, the PP and the RPP models correspond to the curve of the probit model. For $\lambda < 1$ (or $\lambda > 1$), the PP curve is generally above (below) to the probit curve for a range of η values. Also, for each value of λ, RPP curve is a reflection of the probit curve and thus for $\lambda < 1$ (or $\lambda > 1$) the corresponding curve is generally below (above) the corresponding curve for the probit model.

The likelihood function corresponding for the SPM model indexed by λ is given by

$$L(\boldsymbol{\beta}, \lambda|\mathbf{y}, \mathbf{x}) = \prod_{i=1}^{n} [F_\lambda(\eta_i)]^{y_i} [1 - F_\lambda(\eta_i)]^{1-y_i}. \tag{3.2}$$

Note that in the SP models the parameters β and λ have quite different meaning. On one hand, λ is a structural parameter associated with the choice of the link function. On the other hand, traditional β parameters are a vector of structural parameters inherent to the observed data and not depending on model choice (for a discussion, see, e.g., Taylor and Siqueira, 1996). So, two scenarios can be considered. The first scenario is one in which the traditional β parameters and λ

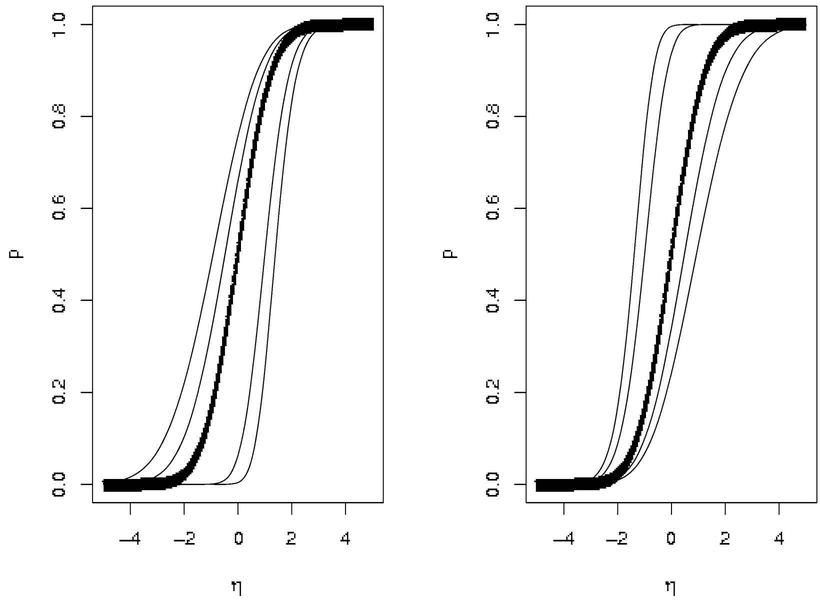

Figure 2 *Response curves*: *Power Probit* (*left*) *for* $\lambda = \{0.4, 0.6, 1, 4, 8\}$ *and Reciprocal Power Probit* (*right*) *for* $\lambda = \{8, 4, 1, 0.6, 0.4\}$ *with different values of the linear predictor* η_i *and curves showed from left to right in both figures.*

are jointly estimated; in the second scenario only the traditional parameters β are allowed to vary and λ is fixed at its "true" value λ_0. As in Taylor and Siqueira (1996), we shall refer to these two scenarios as the unconditional and conditional on λ, respectively. Inference under the conditional scenario for λ is easier to be implemented, because it corresponds to a particular PM model with a fixed value for the parameter λ. In this paper, we are interested in the unconditional approach for λ from the Bayesian paradigm.

4 Bayesian estimation

In order to simplify the Bayesian computation, we introduce the δ-transformation $\delta = \ln(\lambda)$ in the SPM model. Under this parametrization or the parameter λ in the context of a Bayesian analysis, it is necessary to specify a prior distributions for $\boldsymbol{\beta}$ and λ or $\boldsymbol{\beta}$ and δ. It is usual to assume independent priors as

$$\pi(\boldsymbol{\beta}, \delta) = \pi_1(\boldsymbol{\beta})\pi_2(\delta). \tag{4.1}$$

Thus, for $\boldsymbol{\beta}$, we can use the typical priors considered with the probit model (see, e.g., Zellner and Rossi, 1984), including a normal prior ($\beta_j \sim N(\mu_{\beta_j}, \sigma^2_{\beta_j})$) where in the common situation where little prior information is available about this parameters, one can chose $\sigma^2_{\beta_j}$ to be a large value. Since $\delta \in \Re$, we can adopt a normal prior for δ, that is, $\delta \sim N(\mu_\delta, \sigma^2_\delta)$.

Considering the likelihood function in (3.2) and a general prior specification given in (4.1), the Bayesian estimation can be implemented via Metropolis–Hasting providing a simple and efficient sampling from the marginal posterior distributions. Another possibility by considering an augmented likelihood version as the presented in Bolfarine and Bazan (2010) can be obtained using Gibbs Sampling. We do not implement this procedure due to it is slow to converge given the auxiliary latent variables introduced. Besides, it is required to generate values of the PN and RPN distributions which are not implemented directly in SAS software.

In this case, the prior hierarchical structure using the δ-parametrization is as follows:

$$Y_i|\beta, \lambda \sim \text{Ber}\big(F_\delta(\eta_i)\big),$$
$$\boldsymbol{\beta} \sim \pi_1(\boldsymbol{\beta})$$

and

$$\delta \sim \pi_2(\delta).$$

This hierarchical structure can be easily implemented in the WinBugs or SAS software. Further, when $\delta = 0$ $(\lambda = 0)$, the hierarchical structure of the likelihood of the *PM* model follows by eliminating the third line in the above model.

It is crucial to mention that selection model is an essential part of any statistical analysis, thus for comparison among alternative models for binary regression, different selection procedures have been proposed in the literature in the MCMC scenary. The main selection procedures used here are the Deviance Information Criterion (*DIC*) proposed by Spiegelhalter et al. (2002), the Expected Information Criteria (Akaike (*EAIC*) and the Schwarz and Bayesian (*EBIC*)) which can seen in Carlin and Louis (2000) and Brooks (2002). These criteria are based on the *Posterior Mean of the Deviance*, $E[D(\boldsymbol{\beta}, \lambda)]$, which is a measure of fit that can be approximated via the MCMC output by

$$Dbar = \frac{1}{G}\sum_{g=1}^{G} D\big(\boldsymbol{\beta}^{(g)}, \lambda^{(g)}\big),$$

where the index (g) represent the gth realization of a total of G realizations, and

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

is the *Bayesian deviance*.

EAIC, EBIC and DIC can be estimated using MCMC output by considering

$$\widehat{EAIC} = Dbar + 2p,$$
$$\widehat{EBIC} = Dbar + p\log N$$

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 given by

$$\rho_D = E\big[D(\boldsymbol{\beta}, \lambda)\big] - D\big[E(\boldsymbol{\beta}), E(\lambda)\big],$$

where $D[E(\boldsymbol{\beta}), E(\boldsymbol{\lambda})]$ is the *deviance of posterior mean*. It is obtained by considering the mean of the values generated from the posterior distribution as

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

Given two alternative models, the model that fits better to a data set is the one with the smallest value of the Posterior Mean of the Deviance, *DIC*, *EBIC* and *EAIC*.

Additionally, checking through examination of individual observations can be obtained by considering posterior mean of the standardized residual $\frac{(y_i - E(Y_i))}{\sqrt{v(Y_i)}}$. Also, another measures of global fit by considering posterior mean of unstandardized residuals $e_i = y_i - E(Y_i)$ can be defined as the sum of squared residuals (*SSR*) and sum of absolute residuals (*SAR*), that is, $SSR = \sum_{i=1}^{n} e_i^2$ and $SAR = \sum_{i=1}^{n} |e_i|$.

In the following section, we illustrate the Bayesian approaches developed in this work under the *SPM* model by considering two simulation studies: one for the parameter recovery and other for compare model selection criteria. In addition, for a well-known data set from the literature, we improve the fitting in the Binary Regression Model, by using a Skew-Probit Link function, when compared with other links functions in the literature. All the models considered here were implemented using proc mcmc of SAS 9.2 software (SAS Institute Inc., 2009). A code is showed in Appendix B.

5 Applications

5.1 Simulated data

A simulation study is carried out to evaluate the relative performance of the estimation procedure in terms of parameter recovery and the PP model by considering: (a) three sample sizes $n = \{50, 100, 200\}$ and (b) the use of posterior mean or posterior median as measure of summary for λ.

For this purpose, a dataset was simulated following a similar procedures as in Chen, Dey and Shao (1999), that is, by considering the value of λ fixed following the strategy: We independently generate $x_i \sim N(1, 3)$, $i = 1, 2, \ldots, n$. Then using

Table 1 *Study on parameter recovery of the power probit model with skewed to the left* ($\lambda = 0.25$) *under three different sizes of the sample with* 100 *generated datasets*

Parameter	Estimator	Skew probit model			Probit model		
		$n = 50$	$n = 100$	$n = 200$	$n = 50$	$n = 100$	$n = 200$
beta	mean	1.190	1.117	1.094	0.585	0.536	0.522
	se	0.036	0.025	0.020	0.016	0.009	0.007
	bias of mean beta	0.190	0.117	0.094	−0.415	−0.464	−0.478
	mse of mean beta	0.162	0.077	0.050	0.199	0.224	0.233
lambda	mean	0.272	0.255	0.246	–	–	–
	se	0.006	0.006	0.005	–	–	–
	median	0.247	0.240	0.239	–	–	–
	se	0.006	0.006	0.005	–	–	–
	bias of mean lambda	0.022	0.005	−0.004	–	–	–
	bias of median lambda	−0.003	−0.010	−0.011	–	–	–
	mse of mean lambda	0.0041	0.0035	0.0028	–	–	–
	mse of median lambda	0.0036	0.0037	0.0031	–	–	–
DIC	mean	25.974	50.676	99.598	38.590	76.951	152.293
	se	0.658	0.930	1.317	0.685	1.010	1.534
EAIC	mean	28.712	53.217	101.956	39.632	77.971	153.299
	se	0.635	0.908	1.299	0.684	1.010	1.534
EBIC	mean	32.536	58.427	108.553	41.544	80.576	156.598
	se	0.635	0.908	1.299	0.684	1.010	1.534

the same set of x_i values, we independently generate 100 datasets, so that for each one, n independent Bernoulli response variables y_i are obtained for the PP model with one covariate x_i and the true value of $\beta = 1$ and $\lambda = 0.25$, with no intercept.

For each generated data set, we fit the corresponding true model. For each simulated sample, we use the estimation procedure described in Section 4. We burned 4000 in the 100,000 values of chain and thin of 50. The effective sample is 2000. Efficiency by using the Effective Sample Sizes above of 0.9 were obtained. The results are presented in Table 1.

From the simulation study, we found in the PP model as expected there is an improvement in the accuracy (bias and MSE decrease) of the estimation of β and λ parameters as sample size increases. In addition we found that the best summary measures for λ is the posterior median.

In addition, in Table 1, we present a comparative study of the skew-probit model ("true model") with the probit model ("false model") by using model selection criteria presented in Section 4. Our main goal is to compare the above mentioned model selection criteria to find the suitable one. For this purpose, we examine the relative performance of these procedure to select the best underlying model. The data of size 50, 100 and 200 are the same considered above. For each one of the 100 generated data set in each case, we fit the corresponding true model (skew probit) and the alternative model (probit).

Table 1 also presents the mean and standard errors of the different selection criteria for the simulated samples. In all cases DIC, EAIC and EBIC show evidence in favor of the skew probit as best model for the data set generated, confirming an adequate performance of these model comparison criteria. The skew probit model is correctly selected considering all selection criteria in (100%) of the times. This results show that the use of an incorrect specification of the true model by considering probit model will lead to biased coefficients and specific underestimation of the regression coefficient is observed.

In addition the results of bias and MSE for PP and P models show the performance of the skew-probit model as the sample size increases, as well as the relative performance of it respect to the standard probit model. Only in the skew probit model the estimated values are very close of the real values. However, considering a additional studies suggested by a referee, since the parameter space associated with the $\delta = \ln(\lambda)$ parameter is the whole real line, when values above of $\delta = 3$ (approx. $\lambda = 20$) or below $\delta = -3$ (approx. $\lambda = 0.05$) are considered, a biased estimator is obtained for this parameter but the bias decreases when the sample size increases.

5.2 Real data

We illustrate the Bayesian approaches developed in this paper by using a popular dataset from the literature for the asymmetrical link function in the binary regression. We improve the fitting in Binary Regression Model, by using a Skew-Probit Link function, when compared with other links functions in the literature. All the models considered were implemented using proc mcmc of SAS 9.2 software (SAS Institute Inc., 2009). The code is shown in Appendix B.

Milicer and Szczotka (1966) analyzed the occurrence of menarche as a function of age in a sample of 3918 Warsaw girls. The data are from a study conducted on pre-teen and teenage girls in Warsaw. The subjects were classified into 25 age categories. The number of girls in each group (sample size: n) and the number that reached menarche at the time of the study were recorded (y). The age for a group corresponds to the midpoint for the age interval (x). The datasets are showed in Table 3 and they are available in Finney (1971) and Stukel (1988).

In order to illustrate the usefulness of the links functions proposed in this work with the Warsaw girl dataset, we performed some comparisons between the skew probit link functions and some link functions from the literature, namely logit, generalized logit (or scobit), power logit, cloglog, loglog and gev (Wang and Dey, 2010).

In all cases, an effective sample size of 2000 was considered discarding the 4000 initial iterations and considering thin of 50. The time of execution and thin values were also evaluated for model comparison. In addition, many procedures based on the BOA and CODA packages were used to evaluate the convergence of the chain. Running Mean Plots for each chain, for all the links functions considered, provide strong indication of the convergence in all cases.

Table 2 *Performance of the selection criteria for different models with the Warsaw data*

Models	Dbar	Dmean	p_D	p	DIC	EAIC	EBIC	time (sec)
L	112.78	110.76	2.02	2	114.81	116.78	129.33	6.43
P	108.96	106.94	2.02	2	110.98	112.96	125.51	8.00
CLL	204.82	202.88	1.94	2	206.76	208.82	221.37	6.28
LL	120.66	118.69	1.96	2	122.62	124.66	137.20	6.37
PL	106.10	103.26	2.83	3	108.93	112.10	130.92	9.03
RPL	109.85	106.96	2.89	3	112.74	115.85	134.67	8.81
PP	102.29	100.75	1.54	3	103.83	108.29	127.11	8.53
RPP	100.99	99.22	1.77	3	102.75	106.99	125.81	8.12
GEV	101.49	98.52	2.97	3	104.46	107.49	126.31	12.76

Note: L: logit, P: probit, CLL: Cloglog, LL: loglog, PL: power logit, RPL: reciprocal power logit, PP: power probit, RPP: reciprocal power probit, GEV: generalized extreme value link.

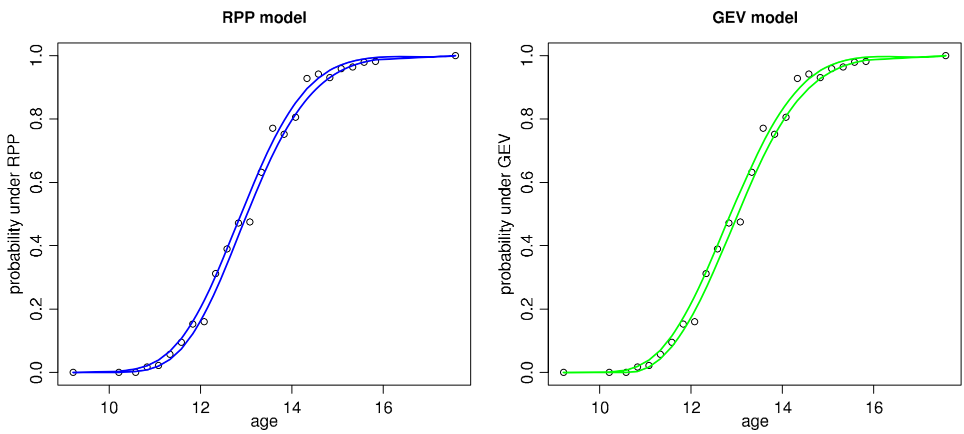

Figure 3 95% *Credible Interval for proportions under RPP and GEV links.*

By considering small values of the three selection procedures in Table 2, GEV, PP and RPP are the better models for fit Warsaw data. Figure 3 shows an 95% credible interval for proportions to dataset analyzed under RPP and GEV link functions which present maybe a similar performance in the prediction, but by considering all criteria, RPP model is the best model with small convergence time among them. Also, by considering residuals for the 25 age categories we found that standardized residuals in RPP model varying between -1.404 ($age = 13.08$) and 1.458 ($age = 14.33$) and in the GEV model varying between -1.407 ($age = 13.08$) and 1.567 ($age = 14.33$) with $SSR = 184.079$ and $SAR = 47.0715$ for RPN model and $SSR = 192.879$ and $SAR = 44.152$ for GEV model confirming that the RPN model is the best model for this data. Posterior Values estimated of the number that reached menarche at the time of the study $\hat{y}_i = E(\hat{y_i}) = n\hat{p}_i$ by con-

Table 3 *Number that reached menarche ($\hat{y}$) and probability $\hat{p}$ estimated in* 2869 *Warsaw girls under the RPN model*

Obs.	y	n	x	$\hat{p}$	$E(\hat{y})$	$v(\hat{y})$	st. res.	Obs.	y	n	x	$\hat{p}$	$E(\hat{y})$	$v(\hat{y})$	st. res.
1	0	376	9.21	0.000	0.005	0.005	−0.069	14	67	106	13.33	0.630	66.809	24.701	0.038
2	0	200	10.21	0.001	0.293	0.293	−0.542	15	81	105	13.58	0.708	74.350	21.703	1.427
3	0	93	10.58	0.006	0.565	0.562	−0.754	16	88	117	13.83	0.776	90.815	20.325	−0.624
4	2	120	10.83	0.014	1.670	1.646	0.257	17	79	98	14.08	0.833	81.676	13.605	−0.725
5	2	90	11.08	0.029	2.574	2.501	−0.363	18	90	97	14.33	0.880	85.328	10.267	1.458
6	5	88	11.33	0.053	4.677	4.429	0.153	19	113	120	14.58	0.916	109.878	9.268	1.025
7	10	105	11.58	0.090	9.461	8.609	0.184	20	95	102	14.83	0.943	96.148	5.516	−0.489
8	17	111	11.83	0.141	15.638	13.435	0.372	21	117	122	15.08	0.962	117.380	4.445	−0.180
9	16	100	12.08	0.205	20.531	16.316	−1.122	22	107	111	15.33	0.976	108.308	2.627	−0.807
10	29	93	12.33	0.282	26.185	18.812	0.649	23	92	94	15.58	0.985	92.583	1.396	−0.493
11	39	100	12.58	0.366	36.639	23.215	0.490	24	112	114	15.83	0.991	112.963	1.028	−0.950
12	51	108	12.83	0.456	49.214	26.788	0.345	25	1049	1049	17.58	1.000	1048.877	0.123	0.351
13	47	99	13.08	0.545	53.957	24.549	−1.404								

Table 4 *Posterior summaries of the parameters of the RPP model under the Warsaw data*

Parameters	mean	sd	$P_{2.5}$	median	$P_{97.5}$
α	−18.753	2.099	−20.386	−19.085	−17.376
β	1.591	0.200	1.462	1.625	1.747
δ	−1.582	0.340	−1.851	−1.668	−1.393
λ	0.220	0.091	0.157	0.189	0.248

sidering the RPN model is showed in Table 3. In addition the posterior estimates of variability associated $v(\hat{y}_i) = n_i \hat{p}_i(1 - \hat{p}_i)$ and standardized residual are also included. As expected, between the ages of 12.58 and 13.83 (when p_i is around 0.5) the maximum variance is obtained. Also for each age value note that all standardized residuals are between -1.404 and 1.458 indicating that the fit is adequate.

Since the Figure 3, the observed proportions, which are bounded between zero and one, have a lazy S-shape (a sigmoidal function) when plotted against age. The change in the observed proportions for a given change in age is much smaller when the proportion is near 0 or 1 than when the proportion is near 1/2. This phenomenon is captured by the RPN model.

Finally, Table 4 shows the posterior summaries of RPP model for the data. Note that the shape parameter λ can be considered as different from zero, since the 95% credibility interval does not include it. Also, note that the values $\hat{\alpha} = -18.753$ and $\hat{\beta} = 1.591$ can be interpreted similarly as the values obtained in the probit model (-21.182 and 1.63, resp.). Since $\hat{\lambda} = 0.22$, in the RPP model can estimate the estimated probabilities through $\hat{p} = 1 - \Phi(-(-18.753 + 1.591x))^{0.22}$.

6 Extensions and discussion

This paper was introduced by new asymmetrical link functions for the binary response variables by considering the cumulative distribution of the power-normal distribution (Gupta and Gupta, 2008) and the reciprocal power normal link function. These two skew-probit link functions have as particular cases the probit link function. In these models, we introduce a parameter for the asymmetry of the response curves. This parameter is associated with the selected distribution and is independent of the linear predictor. Also, it defines a class of skew link functions that can control the rate of increasing (or decreasing) of the probability of success (failure) of the binary responses variables.

An attractive aspect of the model is that can be easily implemented via MCMC by using the software WinBugs or proc MCMC in SAS with common proper but non informative priors.

A second version of the likelihood function by considering a latent linear structure for the model similar to Albert and Chib (1995) can be formulated and studied.

The simulation study was carried out to evaluate the relative performance of the procedure of estimation in terms of parameter recovery. Also, the selection model criteria for the comparison of symmetrical and asymmetrical models as 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) were considered. For study of parameter recovery we found in the PP model, estimated values of the parameters of the model were very close of the real value and acceptable accuracy in the estimation of these parameters. In addition, by considering the skew probit as "true model", this model was correctly selected considering all selection criteria in (100%) of the times.

In addition, by using a known dataset from the literature, we improve the fitting in Binary Regression Model, by using a Skew-Probit Link function, when compared with other links functions in the literature.

In our paper, we focus on Bayesian estimation and given the obtained results from the simulation study we believe that these new links can be considered in analyzing binary response data. In addition, we found some difficulties in the implementation of ML estimation and then comparison with Bayesian estimation was not considered. Specifically, we found that estimates of an equivalent model (Power Logistic) are not close with the corresponding estimates obtained in Stata (StataCorp, 2009). Thus, bias correction methods for the estimates of these models and Probit model in ML estimation, (e.g., Firth, 1993 and Cordeiro and McCullagh, 1991, which are compared in Maiti and Pradhan, 2008), should be considered in order to show a convenient comparison. An interesting algorithm proposed by Devidas and George (1999) can be explored in order to implement ML estimation of PP and RPP model but additional simulation studies should be considered.

However, note that the bias observed in the estimation of the regression coefficients under the probit model reflects the fact that this model is not the true model.

Extensions of the methods developed in this paper for dichotomous responses variables to ordinal response one (Albert and Chib, 1993 and Johnson and Albert, 1999) are possible if we model in terms of cumulative probabilities. So, the conditional probability of a response in the category c is obtained as the difference among two conditional accumulative probabilities:

$$P(Y_i = c|\mathbf{x}_i) = F_\lambda(\eta_{i,c}) - F_\lambda(\eta_{i,c-1}),$$

where $\eta_{i,c} = \gamma_c - \mathbf{x}_i'\boldsymbol{\beta}$, and $\gamma_c \geq \gamma_{c-1}$.

By considering $G(\cdot) = \Phi(\cdot)$ the c.d.f. of the standard normal distribution $F_1(\cdot)$ is named Power Normal distribution (Gupta and Gupta, 2008) and $F_2(\cdot)$ is a new distribution in the literature and was named here as Reciprocal Power Normal distribution. Both distributions are particular cases of the named Beta-Normal distribution (Eugene, Lee and Famoye, 2002) and can be generalized to another exponentiated models as showed by Achcar et al. (2013).

Also, one possible unification of the models proposed in this work can be obtained from the Kumaraswamy distribution (Cordeiro and de Castro, 2011) which will be our future research.

Applications in many areas, where the symmetric links functions are not justified, can be obtained with the proposed models. It includes binomial models, epidemiological studies, longitudinal data analysis, meta-analysis, measurement error models, calibration model and mixture models in survival analysis and item response models. Also expert prior elicitation as suggests by Bedrick et al. (1996) can be explored with the models proposed in this work. Also, extensions for binary regression mixed models as seen in Longford (1994) are direct.

Appendix A: Properties of the standard power-normal distribution

Considering $Z \sim \mathrm{PN}(\lambda)$, the following properties are readily established (see Gupta and Gupta, 2008 and Kundu and Gupta, 2013): The mean and variance of Z are given, respectively, by

$$E[Z] = \frac{(\lambda)(\lambda-1)}{2\pi} L(\lambda-2, 1/\sqrt{2}) \quad \text{and} \quad \operatorname{Var}[Z] = E[Z^2] - E[Z]^2,$$

where

$$E[Z^2] = 1 + \frac{(\lambda)(\lambda-1)(\lambda-2)}{4\pi\sqrt{3}} L(\lambda-3, 1/\sqrt{3})$$

with

$$L(n, \lambda) = \int_{-\infty}^{-\infty} [\Phi(\lambda z)]^n \phi(z)\, dz.$$

In addition, the skewness is give by

$$\gamma = \frac{E(Z^3) - 3E(Z^2)E(Z) + 2(E(Z))^3}{(\mathrm{Var}(Z))^{3/2}}.$$

The moments also can be obtained by numerical integration by considering

$$E[Z^k] = \lambda \int_{-\infty}^{-\infty} z^k [\Phi(z)]^n \phi(z)\, dz = \lambda \int_0^1 [\Phi^{-1}(u)]^k u^{\lambda-1}\, du$$
$$\approx \lambda \sum_{j=1}^{M} [\Phi^{-1}(u_j)]^k u_j^{\lambda-1}$$

for $M \to \infty$ and $u_j \sim \mathrm{Uniform}(0, 1)$.

Since that $-Z \sim \mathrm{RPN}(\lambda)$, analogous expressions for moments, mean, variance and skewness for the RPN distribution can be easily derived.

Appendix B: Program

The SAS program, by considering the proc mcmc, used to implement the regression link model proposed in this work is described as follows:

```
title 'POWER-NORMAL';
proc mcmc data=meninas ntu=4000 nmc=104000 nthin=50 propcov=quanew diag=(mcse ess autocorr)
    outpost=unoout monitor=(alphastar beta delta lambda alpha) seed=246810 dic;
        ods select PostSummaries PostIntervals mcse ess TADpanel dic;
        parms (alphastar beta delta) 0;
        beginprior;
                prior alphastar ~ normal(0, var=1000);
                prior beta ~ normal(0, var=1000);
                prior delta ~ normal(0, var=100);
                alpha=alphastar-beta*meanx;
                lambda=exp(delta);
        endprior;
        eta=alphastar+beta*xm;
        p=cdf('normal',eta,0,1)**lambda;
        model y ~ binomial(n,p);
run;
```

Acknowledgments

The authors acknowledges the partial financial support from FAPESP and DGI-PUCP for this first author. The second author was partially supported by Grants from FONDECYT 11080007.

References

Achcar, J., Coelho-Barros, E. A. and Cordeiro, G. M. (2013). Beta generalized distributions and related exponentiated models: A Bayesian approach. *Brazilian Journal of Probability and Statistics* **27**, 1–19. MR2991775

Albert, J. H. and Chib, S. (1993). Bayesian analysis of binary and polytomous response data. *Journal of the American Statistical Association* **88**, 669–679. MR1224394

Albert, J. H. and Chib, S. (1995). Bayesian residual analysis for binary response regression models. *Biometrika* **82**, 747–759. MR1380812

Bazán, J. L., Bolfarine, H. and Branco, M. D. (2010). A framework for skew-probit links in Binary regression. *Communications in Statistics—Theory and Methods* **39**, 678–697. MR2745312

Bazán, J. L., Branco, M. D. and Bolfarine, H. (2006). A skew item response model. *Bayesian Analysis* **1**, 861–892. MR2282209

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. MR1439085

Bolfarine, H. and Bazán, J. L. (2010). Bayesian estimation of the logistic positive exponent IRT model. *Journal of Educational and Behavioral Statistics* **35**, 693–713.

Brooks, S. P. (2002). Discussion on "Bayesian measures of model complexity and fit" by D. J. Spiegelhalter, N. G. Best, B. P. Carlin and A. van der Linde. *Journal of the Royal Statistical Society, Ser. B* **64**, 616–618.

Carlin, B. P. and Louis, T. A. (2000). *Bayes and Empirical Bayes Methods for Data Analysis*, 2nd ed. Boca Raton, FL: Chapman & Hall. MR1427749

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. MR1731481

Chen, M. H., Dey, D. K. and Shao, Q.-M. (2001). Bayesian analysis of binary data using skewed logit models. *Calcutta Statistical Association Bulletin* **51**, 201–202. MR1892782

Collet, D. (2003). *Modelling Binary Data*, 2nd ed. Boca Raton, FL: Chapman & Hall/CRC. MR1999899

Cordeiro, G. and de Castro, M. (2011). A new family of generalized distributions. *Journal of Statistical Computation and Simulation* **81**, 883–898. MR2806932

Cordeiro, G. M. and McCullagh P. (1991). Bias reduction in generalized linear models. *Journal of the Royal Statistical Society, Ser. B* **53**, 629–643. MR1125720

Czado, C. (1994). Bayesian inference of binary regression models with parametric link. *Journal of Statistical Planning and Inference* **41**, 121–140. MR1292151

Czado, C. and Santner, T. J. (1992). The effect of link misspecification on binary regression inference. *Journal of Statistical Planning and Inference* **33**, 213–231. MR1190622

Devidas, M. and George, E. O. (1999). Monotonic algorithms for maximum likelihood estimation in generalized linear models. *Sankhyā, Ser. B* **61**, 382–396. MR1745742

Dey, D. K., Ghosh, S. K. and Mallick, B. K. (1999). *Generalized Linear Models: A Bayesian Perspective*. New York: Marcel Dekker. MR1893779

Eugene, N., Lee, C. and Famoye, F. (2002). Beta Normal distribution and its applications. *Communication and Statistics—Theory and Methods* **31**, 497–512. MR1902307

Finney, D. J. (1971). *Probit Analysis*. London: Cambridge Univ. Press. MR0347013

Firth D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika* **80**, 27–38. MR1225212

Guerrero, V. M. and Johnson, R. A. (1982). Use of the Box–Cox transformation with binary response models. *Biometrika* **69**, 309–314. MR0671968

Gupta, R. D. and Gupta, R. C. (2008). Analyzing skewed data by power normal model. *Test* **17**, 197–210. MR2393359

Johnson, V. and Albert, J. (1999). *Ordinal Data Modeling*. New York: Springer-Verlag. MR1683018

Kundu, D. and Gupta, R. D. (2013). Power-normal distribution. *Statistics* **47**, 110–125. MR3023018

Longford, N. T. (1994). Logistic regression with random coefficients. *Computational Statistics Data Analysis* **17**, 1–15.

Maiti, T. and Pradhan, V. (2008). A comparative study of the bias corrected estimates in logistic regression. *Statistical Methods in Medical Research* **17**, 621–634. MR2654668

Milicer, H. and Szczotka, F. (1966). Age at menarche in Warsaw girls in 1965. *Human Biology* **38**, 199–203.

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

SAS Institute Inc. (2009). *SAS/STAT® 9.2 User's Guide*, 2nd ed. Cary, NC: SAS Institute Inc.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Ser. B* **64**, 583–639. MR1979380

StataCorp (2009). *Stata: Release 11. Statistical Software*. College Station, TX: StataCorp LP.

Stukel, T. (1988). Generalized logistic models. *Journal of the American Statistical Association* **83**, 426–431. MR0971368

Taylor, J. and Siqueira, A. (1996). The cost of adding parameters to a model. *Journal of the Royal Statistical Society* **58**, 593–607. MR1394369

Wang, X. and Dey, D. (2010). Generalized extreme value regression for binary response data: An application to B2B electronic payments system adoption. *The Annals of Applied Statistics* **4**, 2000–2023. MR2829944

Zellner, A. and Rossi, P. E. (1984). Bayesian analysis of dichotomous quantal response models. *Journal of Econometrics* **25**, 365–393. MR0752447

J. L. Bazán
J. Rodrigues
Instituto de Ciências Matemática e de Computação
Universidade de São Paulo
Avenida Trabalhador São Carlense, 400
CEP13560-970
São Carlos, SP
Brasil
E-mail: jlbazan@icmc.usp.br

J. S. Romeo
Departamento de Matemática y Ciencia de la Computación
Universidad de Santiago de Chile
Chile