State space mixed models for binary responses with skewed inverse links using JAGS

Carlos A. Abanto-Valle[†] Jorge L. Bazán[‡] and Anne C. Smith[§,⋆]

[†] Department of Statistics, Federal University of Rio de Janeiro,
Caixa Postal 68530, CEP: 21945-970, RJ, Brazil

[‡] Departamento de Matematica Aplicada e Estatística, Instituto de Ciencias Matemáticas e de Computação,
Universidade de São Paulo, São Carlos, SP, Brazil.

[§] Evelyn F. McKnight Research Institute, University of Arizona,Tucson, AZ 85724

⋆ Visiting Scientist, Department of Brain and Cognitive Science,
Massachusetts Institute of Technology, Cambridge, MA 02139

Abstract

State space mixed models (SSMM) for binary time series were the inverse link function is modeled to be a cumulative distribution function of a new class of asymmetric links is introduced. The resulting class consider the power of the normal, power logistic and power cloglog distributions and it include the probit, logit, cloglog and loglog links as a particular cases. We showed that these links are potentially useful in order to choose alternative candidate links. They are flexible in fitting skewness in the response curve using a free shape parameter. Markov chain Monte Carlo (MCMC) methods for Bayesian analysis of SSMM with these power links are implemented using the JAGS package, a freely available software. The flexibility of the proposed links is illustrated to measure effects of deep brain stimulation (DBS) on attention of a macaque monkey performing a reaction-time task (Smith et al., 2009). Empirical results showed that the links introduced fits better compared to the more commly-used probit and logit inverse links.

Keywords: Binary time series, cloglog link, logit link, Markov chain Monte Carlo, power distribution, probit link, state space models.

1 Introduction

Binary responses can be described by generalized linear models (McCullagh and Nelder, 1989). However, if serial correlation is present as is the case in time series or if the observations are overdispersed, these models may not be adequate, and several alternate approaches can be taken. Extensions such as generalized linear state space models address those problems and are treated in a paper by West et al. (1985) in a conjugate Bayesian setting. They have been subject to further research by Fahrmeir (1992), Song (2000), Carlin and Polson (1992), Czado and Song (2008) and Abanto-Valle and Dey (2014) among others.

In modeling binary time series data, the choice of the link function is a critical issue. Different choices of the link function lead to different models. In the context of binary regression problems symmetric links are widely used in the literature (Albert and Chib, 1993; Basu and Mukhopadhyay, 2000*a*;*b*). However, as Chen et al. (1999) have argued, when the probability of a given binary response approaches to 0 at a different rate than it approaches 1, symmetric link functions may not be as successful in fitting binary data. For example, if the link function is misspecified, there can be substantial bias in the mean response estimates (Czado and Santner, 1992). To deal with this problem, asymmetric links have been considered in the literature.

There are lots of research done to include skewness into the link function. For example, Stukel (1988) proposed a two-parameter class of generalized logistic models, Kim et al. (2008) used the skewed generalized t-link, and Bazán et al. (2010) adopted the skewed probit links and some variants with different parameterizations. More recently Bazán et al. (2014) introduced the power normal (PN) link. By introducing a power parameter, the PN link achieves great flexibility in both positive and negative directions in a symmetric fashion.

It is also worth noting that binary state space mixed models (BSSMM) with probit link have been considered by Czado and Song (2008) who carried out an MCMC estimation. Recently, Abanto-Valle and Dey (2014) extended the model of Czado and Song (2008) using certain scale mixtures of Gaussians for the inverse link function. The estimation of these models is not straightforward because the latent states are included in the distribution function. Efficient MCMC algorithms were developed using the threshold approach of Albert and Chib (1993).

In this paper, we propose a new class of links functions to binary regression models based in the inverse of skewed distributions obtained by exponentiation. An extension of usual links as logit, probit, cloglog and loglog to a class of power links which include these links as particular cases. Thus, correspondents BSSMM-power-logit, BSSMM-power-probit, BSSMM-power-cloglog and BSSMM-power-loglog are proposed. We then fit the models under a Bayesian paradigm using MCMC methods, which allows estimation of tthe posterior distribution of parameters based on reasonable prior assumptions.

While there are a growing number of increasingly sophisticated and complex sampling schemes being developed that allow efficient convergence, we make use here of a free off-the-shelf software package, JAGS (Plummer, 2003) running inside R, using the rjags package. The rjags package provides an interface from R to the JAGS library for Bayesian data analysis usign Markov Chain Monte Carlo (MCMC) to generate a sequence of dependent samples from the posterior distribution of the parameters. JAGS makes use of a single-move sampler, which, compared with a multi-move sampler, is likelt to generate posterior samples which are more highly correlated . This dependency can be reduced by running a longer Markov chain and thinning the samples. The gain in efficiency by using complex sampling schemes is to some extent is outweighed by the ease of implementation in JAGS.

The remainder of this paper is organized as follows. Section 2 introduces the new class of links considering Power distributions. Section 3 outlines the setup of the BSSMM models for the flexible link functions proposed as well as the corresponding Bayesian estimation procedure using MCMC methods. Section 4 is devoted to the application and model comparison of all the models proposed with alternativer models in the literature using a real data set. Finally, some concluding remarks and suggestions for future developments are given in Section 5.

2 The power distributions and alternative links functions

We first introduce some notation that will be used throughout the paper, then propose the power distribution which include the power normal (PN) distribution (Gupta and Gupta, 2008) and describe some related properties of this distribution.

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

$$g_p(x \mid \mu, \sigma^2, \lambda) = \frac{\lambda}{\sigma} h\left(\frac{x-\mu}{\sigma}\right)\left[H\left(\frac{x-\mu}{\sigma}\right)\right]^{\lambda-1}, \quad (1)$$

where $h(.)$ and $H(.)$ are, respectively, the probability density function (pdf) and the cumulative distribution function (cdf) of of a continuous univariate distribution supported on the whole real line $< -\infty, \infty >$ named *baseline distribution*.

If $\lambda = 1$, the density of X in (1) reduces to the density of the $h(\mu, \sigma^2)$. The *standard* Power distribution is obtained from equation (1) with $\mu = 0$ and $\sigma^2 = 1$. It is denoted by $Z \sim \mathcal{P}(\lambda)$. Its corresponding pdf and cdf are given by

$$f_p(z \mid \lambda) = \lambda h(z)[H(z)]^{\lambda-1}, \quad (2)$$

$$F_p(z) = [H(z)]^{\lambda}. \quad (3)$$

2.1 Some power distributions

2.1.1 The power normal distributions

A univariate random variable X is said to follow a Power Normal distribution, $X \sim \mathcal{PN}(\mu, \sigma^2, \lambda)$ if the density of this distribution has the form

$$g_{PN}(x \mid \mu, \sigma^2, \lambda) = \frac{\lambda}{\sigma} \phi\left(\frac{x-\mu}{\sigma}\right)\left[\Phi\left(\frac{x-\mu}{\sigma}\right)\right]^{\lambda-1}, \quad (4)$$

where $\phi(.)$ and $\Phi(.)$ are, respectively, the probability density function (pdf) and the cumulative distribution function (cdf) of the standard normal distribution.

If $\lambda = 1$, the density of X in (4) reduces to the density of the $N(\mu, \sigma^2)$. Specifically the cdf and pdf of the standard power normal distribution are given respectively by

$$F_{PN}(z; \lambda) = (\Phi(z))^{\lambda}, \quad f_{PN}(z; \lambda) = \lambda \frac{e^{-\frac{1}{2}z^2}}{\sqrt{2\pi}} (\Phi(z))^{\lambda-1}.$$

This distribution was studied by (Gupta and Gupta, 2008).

2.1.2 The power logistic distributions

A univariate random variable X is said to follow a Power Logistic distribution, $X \sim \mathcal{PL}(\mu, \sigma^2, \lambda)$ if the density of this distribution has the form

$$g_{PL}(x \mid \mu, \sigma^2, \lambda) \quad = \quad \frac{\lambda}{\sigma} l\left(\frac{x-\mu}{\sigma}\right)\left[L\left(\frac{x-\mu}{\sigma}\right)\right]^{\lambda-1}, \tag{5}$$

where $l(.)$ and $L(.)$ are, respectively, the probability density function (pdf) and the cumulative distribution function (cdf) of the standard logistic distribution. If $\lambda = 1$, the density of X in (5) reduces to the density of the $L(\mu, \sigma^2)$. Specifically the cdf and pdf of the standard power logistic distribution are given respectively by

$$F_{PL}(z;\lambda) = \frac{1}{(1+\exp(-z))^{\lambda}} \equiv (1+\exp(-z))^{-\lambda}, \quad f_{PL}(z;\lambda) = \frac{\lambda \exp(-x)}{(1+\exp(-x))^{\lambda+1}}.$$

This distribution is named generalized logistic of type I (Johnson et al., 1995) and has also been called the skew-logistic distribution.

2.1.3 The power cloglog distributions

A univariate random variable X is said to follow a Power Cloglog distribution, $X \sim \mathcal{PCLL}(\mu, \sigma^2, \lambda)$ if the density of this distribution has the form

$$g_{CLL}(x \mid \mu, \sigma^2, \lambda) \quad = \quad \frac{\lambda}{\sigma} cll\left(\frac{x-\mu}{\sigma}\right)\left[CLL\left(\frac{x-\mu}{\sigma}\right)\right]^{\lambda-1}, \tag{6}$$

where $cll(.)$ and $CLL(.)$ are, respectively, the probability density function (pdf) and the cumulative distribution function (cdf) of the standard complement of Gumbel or extreme (minimum) value distribution named here as cloglog distribution and defined as $cll(.) = \exp(. - \exp(.))$ and $CLL(.) = 1 - \exp(-\exp(.))$

If $\lambda = 1$, the density of X in (6) reduces to the density of the $CLL(\mu, \sigma^2)$. Specifically the cdf and pdf of the standard power cloglog distribution are given respectively by

$$F_{PCLL}(z;\lambda) = [1-\exp(-\exp(z)]^{\lambda}, \quad f_{PCLL}(z;\lambda) = \lambda \exp(z - \exp(z))\,[1-\exp(-\exp(z))]^{\lambda-1}.$$

2.1.4 The power loglog distributions

A univariate random variable X is said to follow a Power loglog distribution, $X \sim \mathcal{PLL}(\mu, \sigma^2, \lambda)$ if the density of this distribution has the form

$$g_{LL}(x \mid \mu, \sigma^2, \lambda) \;=\; \frac{\lambda}{\sigma} ll\left(\frac{x-\mu}{\sigma}\right)\left[LL\left(\frac{x-\mu}{\sigma}\right)\right]^{\lambda-1}, \tag{7}$$

where $ll(.)$ and $LL(.)$ are, respectively, the probability density function (pdf) and the cumulative distribution function (cdf) of the standard Gumbel or extreme (maximum) value distribution named here as loglog distribution and defined as $ll(.) = \exp(-. - \exp(-.))$ and $LL(.) = \exp(-\exp(-.))$.

If $\lambda = 1$, the density of X in (7) reduces to the density of the $LL(\mu, \sigma^2)$. Specifically the cdf and pdf of the standard power loglog distribution are given respectively by

$$F_{PLL}(z; \lambda) = [\exp(-\exp(-z))]^{\lambda}, \quad f_{PLL}(z; \lambda) = \lambda \exp(-z - \exp(-z))\,[\exp(-\exp(-z))]^{\lambda-1}.$$

2.2 Alternative links functions

Usual links functions in binary regression models are probit, logit, cloglog and loglog, which are based in the cdf of known distributions. As displayed in Figure 1, logit and probit are symmetric links but cloglog and loglog are asymmetric links.

Considering the cdf of Power distributions defined above, alternative links functions to known link function in the literature can be proposed. As example, the Figure 2 displays the cdf fuctions of the Power distributions for various values of λ. In this case when $\lambda = 1$ usual links probit, logit, cloglog and loglog are retrieved but when $\lambda \neq 1$ news and flexible links are obtained. In this case, λ is a structural parameter associated with the choice of the link function, controling the skewness and will be estimated since the data.

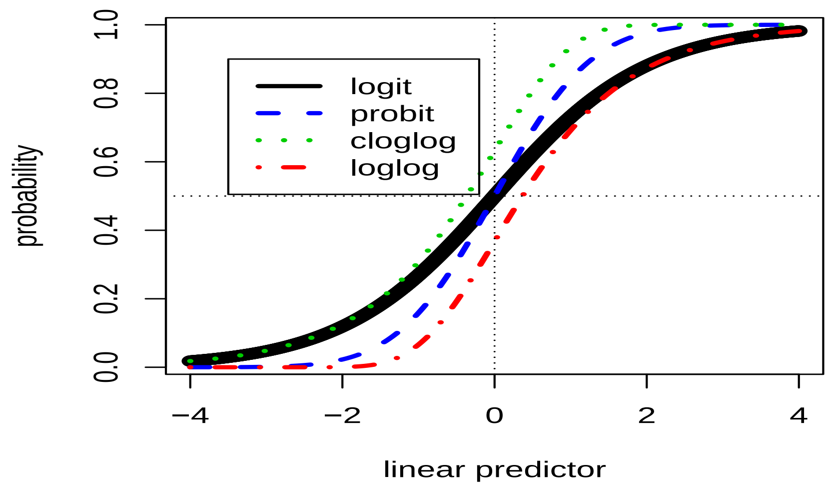

Figure 1: Usual links functions based in the cdf of known distributions.

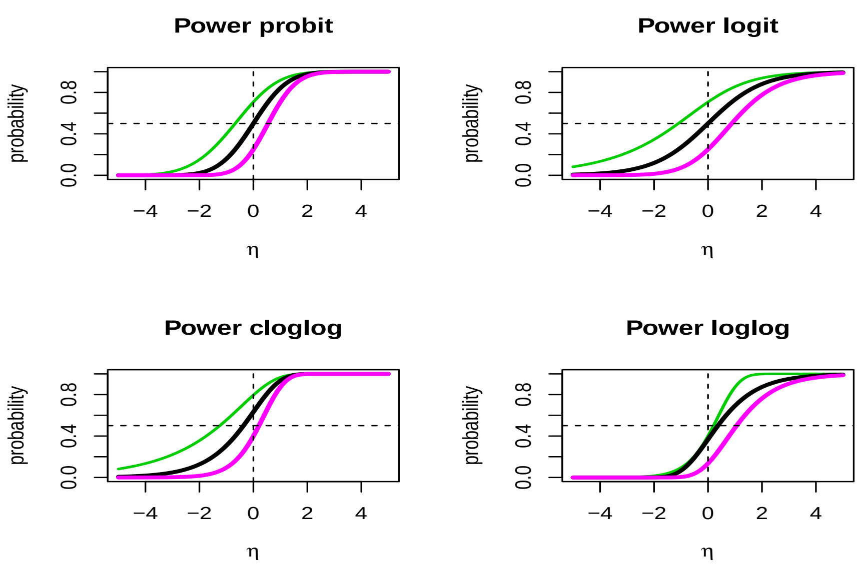

Figure 2: plot of inverse of links functions based in the cdf of Power distributions considering different values of λ. green line ($\lambda = -0.5$), black line ($\lambda = 1$), and blue line ($\lambda = 2$).

3 Binary response state space mixed models with skewed links

3.1 Model setup

Let $\mathbf{Y}_{1:T} = (Y_1, \ldots, Y_T)'$ denote T independent binary observations and let $\mathbf{x}_t$ denote a $k \times 1$ vector of covariates. We assume that

$$
\begin{aligned}
Y_t &\sim \mathcal{B}er(\pi_t) \qquad\qquad t = 1, \ldots, T && (8)\\
\pi_t &= P(Y_t = 1 \mid \theta_t, \mathbf{x}_t, \boldsymbol{\beta}) = F_p(\mathbf{x}_t'\beta + \theta_t) && (9)\\
\theta_t &= \delta\theta_{t-1} + \tau\eta_t, && (10)
\end{aligned}
$$

where, $F(x) = F_p(x)$, represents the cdf at x for the standard power distribution with unknown shape parameter λ defined in (3). Considering the distributions defined above, we name the eigth possible models as BSSMM-P-PROBIT, BSSMM-P-LOGIT, BSSMM-P-LOGLOG and BSSMM-P-CLOGLOG, BSSMM-PROBIT, BSSMM-LOGIT and BSSMM-LOGLOG and BSMM-CLOGLOG, where "P" indicates the power version of the model. We assume that η_t are independent and normally distributed with mean zero and unit variance, $\mid \delta \mid < 1$, i.e., the latent state process is stationary and $\theta_0 \sim \mathcal{N}(0, \frac{\tau^2}{1-\delta^2})$. The latent state, θ_t, represents time-specific effects in the observed process. Using the Bayesian paradigm, MCMC methods are employed to compute posteriors in the next subsection.

3.2 Inference procedure

We estimate the model defined by equations (8)-(10) using Monte Carlo Markov Chain (MCMC). The number of parameters required varies between choice of link model. The model depends on a parameter vector Ψ, where $\Psi = (\boldsymbol{\beta}', \delta, \tau^2, \omega)'$ to power distributions with $\omega = \log\lambda$ and on $\Psi = (\boldsymbol{\beta}', \delta, \tau^2)'$ in the PROBIT, LOGIT, LOGLOG and CLOGLOG model cases, respectively. Let $\boldsymbol{\theta}_{0:T} = (\theta_0, \theta_1, \ldots, \theta_T)'$ constitute the latent states. The Bayesian approach for estimating the model treats $\theta_{0:T}$ as latent parameters themselves and updates them in each step of MCMC. The joint posterior density of parameters and latent variables can be written as

$$
p(\boldsymbol{\theta}_{0:T}, \Psi \mid \mathbf{y}_{1:T}) \;\propto\; p(\mathbf{Y}_{1:T} \mid \boldsymbol{\theta}_{0:T}, \Psi, \mathbf{y}_{1:T}) p(\boldsymbol{\theta}_{0:T} \mid \Psi) p(\Psi), \qquad (11)
$$

where

$$p(\mathbf{Y}_{1:T} \mid \boldsymbol{\theta}_{0:T}, \Psi) = \prod_{t=1}^{T} \{\pi_t^{Y_t}(1-\pi_t)^{1-Y_t}\} \tag{12}$$

$$p(\boldsymbol{\theta}_{0:T} \mid \Psi) = \phi(\theta_0 \mid 0, \frac{\tau^2}{1-\delta^2}) \prod_{t=1}^{T} \phi(\theta_t \mid \delta\theta_{t-1}, \tau^2), \tag{13}$$

where π_t is given by equation (9) and $\phi(x \mid \mu, \sigma^2)$ denotes the normal density with mean μ and variance σ^2 evaluated at x and $p(\Psi)$ indicates the prior distribution. The prior distributions of Ψ can be written as

$$p_P(\Psi) = p(\boldsymbol{\beta})p(\delta)p(\tau^2)p(\omega).$$

The prior distributions for individual parameters are set as: $\boldsymbol{\beta} \sim \mathcal{N}_k(\boldsymbol{\beta}_0, \boldsymbol{\Sigma}_0)$, $\delta \sim \mathcal{N}_{(-1,1)}(\delta_0, \sigma_\delta^2)$, $\tau^2 \sim \mathcal{IG}(\frac{n_0}{2}, \frac{T_0}{2})$, $\xi \sim \mathcal{U}(-u_0, u_0)$, $\omega \sim \mathcal{N}(\omega_0, W_0)$, where $\mathcal{N}_k(.,.)$, $\mathcal{N}_{(a,b)}(.,.)$,$\mathcal{U}(a,b)$, $\mathcal{IG}(.,.)$, $\mathcal{G}(.,.)$ denote the $k-$variate normal, the truncated normal on interval (a,b), the uniform distribution on interval (a,b), the inverse gamma distribution and the gamma distribution respectively. When PROBIT, LOGIT, LOGLOG and CLOGLOG are considered, prior specification to ω is not required.

We can evaluate Equation (11) using standard Markov Chain Monte Carlo methods in JAGS (Plummer, 2003). Implementation in this software merely requires specifying the model setup in equations (8)-(10), as well as priors for the unknown parameters, $p(\Psi)$.

4 Case study: Deep brain stimulation on attention reaction time

To illustrate the technique applied to binary responses, we consider binary responses from a monkey performing the attention paradigm described in Shah et al. (2009) and Smith et al. (2009). The task consists of making a saccade to a visual target, a (variable length) period of fixation on the target, detection of a change in target color and then a bar release. Sustained attention is required becauseto receive a reward, the animal must release the bar within a brief time window cued by the change in target color (see Smith et al., 2009, for a more detailed description of the experiment). The behavioral data set for this experiment is composed of a time series of binary observations with a 1 (0) corresponding to reward being delivered (not delivered) at each trial, respectively. The goal of the experiment is to determine whether declining performance as a result of fatigue can be corrected by application of deep brain stimulation (DBS) resulting in previous levels of performance. In this specific data set, the monkey performed 1250 trials. Stimulation was applied during 4 periods across trials 300-364, 498-598, 700-799 and 1000-1099, indicated by shaded gray regions in Figures 3 and 4. Dividing the results into periods when stimulation is applied ("ON") and not applied ("OFF"), there are 240 correct responses out of 367 trials during the ON periods and 501 correct responses from 883 trials during the off periods. Out of 1250 observations, 741 (or 59.28%) are correct responses. For this dataset we fit the BSSMM-PROBIT, BSSMM-LOGIT, BSMM-LOGLOG, BSSMM-CLOGLOG and their correspondents power versions: BSSMM-P-PROBIT, BSSMM-P-LOGIT, BSSMM-P-LOGLOG, BSSMM-P-CLOGLOG including BSSMM-GEV, where π_t is modeled by

$$\pi_t \quad = \quad P(Y_t = 1 \mid \theta_t) = F(\theta_t).$$

As before, $F(.)$ represents the cdf associated with the corresponding link functions in non power and power models, where θ_t is the arousal state of the macaque monkey at time t. We set the priors as $\delta \sim \mathcal{N}_{(-1,1)}(0.96, 10^3)$, $\tau^2 \sim \mathcal{IG}(0.01, 0.01)$. For the power models, we assume $\omega \sim \mathcal{N}(0, 10^6)$. For each case, we conducted the MCMC simulation for 900,000 iterations. In all the cases, the first 100000 draws were discarded as a burn-in period. In order to reduce the autocorrelation between successive values of the simulated chain, thinning was performed i.e. only every 100th value of the chain was stored. With the resulting 8000 values, we calculated the posterior means and the 95% credibility intervals (CI). The MCMC output of all the parameters passed the convergence test of Geweke (1992) and Heidelberger and Welch (1983), available for free with the CODA package with the R software.

From Table 1, we found that for all the models considered here, the posterior means of δ are close to 1, showing higher persistence of the autoregressive parameter for states variables and thus in binary time series. The posterior means of τ^2 are between 0.0064 and 0.0246. Some comments about the λ parameter.

Table 1: Parameter estimation results for monkey performance data set. First row: Posterior mean. Second row: Posterior 95% credibility intervals interval in parentheses.

	Model			
	PROBIT	LOGIT	LOGLOG	CLOGLOG
	0.9949	0.9953	0.9958	0.9959
δ	(0.9865,0.9994)	(0.9877,0.9995)	(0.9893,0.9995)	(0.9886,0.9997)
	0.0097	0.0246	0.0124	0.0115
τ^2	(0.0043,0.0190)	(0.0108,0.0487)	(0.0052,0.0252)	(0.0044,0.0239)
	Model			
	P-PROBIT	P-LOGIT	P-LOGLOG	P-CLOGLOG
	0.9959	0.9978	0.9963	0.9965
δ	(0.9889,0.9996)	(0.9938,0.9998)	(0.9877,0.9997)	(0.9900,0.9997)
	0.0092	0.0150	0.0125	0.0064
τ^2	(0.0035,0.0196)	(0.0061,0.0309)	(0.0053,0.0260)	(0.0022,0.0156)
	1.1773	3.6086	1.6498	2.5262
λ	(0.4754,2.9163)	(1.4030,7.2880)	(0.2090,6.4051)	(0.9524,6.2985)

To compare the goodness of the estimated models, we calculate the Watanabe-Akaike information criterion, WAIC (Watanabe, 2010; 2013; Gelman et al., 2014), to compare models using different link functions. The WAIC is defined as

$$WAIC \quad = \quad -2(\text{lppd} - p_{WAIC}), \tag{14}$$

where lppd means the log pointwise predictive density defined by

$$\text{lppd} \quad = \quad \sum_{t=1}^{T} \log \int p(y_t \mid \Psi, \boldsymbol{\theta}_{0:T}) p(\boldsymbol{\theta}_{0:T} \mid \mathbf{y}_{1:T}) d\Psi d\boldsymbol{\theta}_{0:T} \tag{15}$$

and

$$p_{WAIC} \quad = \quad 2\sum_{t=1}^{T} \log E_{\Psi,\boldsymbol{\theta}_{0:T}|\mathbf{y}_{1:T}}[p(y_t \mid \Psi, \boldsymbol{\theta}_{0:t})] - E_{\Psi,\boldsymbol{\theta}_{0:T}|\mathbf{y}_{1:T}}[\log p(y_t \mid \Psi, \boldsymbol{\theta}_{0:t})]. \tag{16}$$

To compute both measures, the lppd and the p_{WAIC}, in practice we can evaluate it using the MCMC output. We label it as Ψ^s and $\boldsymbol{\theta}^s_{0:T}$, $s = 1, \ldots, S$. So approximated versions obtained for computation are given by

$$\begin{aligned}\hat{\text{lppd}} &= \sum_{t=1}^{T} \log\left(\frac{1}{S}\sum_{s=1}^{S} p(y_t \mid \boldsymbol{\Psi}, \boldsymbol{\theta}_t)\right)\\ p_{WAIC} &= 2\sum_{t=1}^{T}\left(\log\left(\frac{1}{S}\sum_{s=1}^{S} p(y_t \mid \boldsymbol{\Psi}, \boldsymbol{\theta}_t)\right) - \frac{1}{S}\sum_{s=1}^{S} \log p(y_t \mid \Psi, \theta_t)\right).\end{aligned}$$

The minimum value of the WAIC gives the best fit. In this context, p_{WAIC} is a measure of model complexity. Table 2 summarizes the WAIC and p_{WAIC} for our eigth models. The WAIC selects the BSSMM-P-LOGLOG as the best model for the monkey performance data set, although BSSMM-LOGLOG is close as well. This confirms our observation that the data supports a positively skewed link function.

Table 2: Monkey performance data set. DIC: deviance information criterion, p_D: effective number of parameters.

Model	WAIC	p_{WAIC}	Rank
BSSMM-PROBIT	1418.3	39.13	5
BSSMM-LOGIT	1421.2	38.01	6
BSSMM-LOGLOG	1412.3	34.79	2
BSSMM-CLOGLOG	1426.5	42.56	8
BSSMM-P-PROBIT	1417.8	38.94	4
BSSMM-P-LOGIT	1414.1	34.89	3
BSSMM-P-LOGLOG	1411.8	34.88	1
BSSMM-P-CLOGLOG	1421.9	44.16	7

Figure 3 shows the posterior smoothed mean for the states θ_t for each one of the four models classified as the better models by the WAIC criterion. Different line types and colors indicate the posterior smoothed mean for the four fitted models respectively. All the estimates follow a similar pattern, but there are expressive differences between the estimates.

As before, in Figure 4, we plot the posterior smoothed mean for the probability of a correct response computed using the four fitted models, the better models according to the WAIC crite-

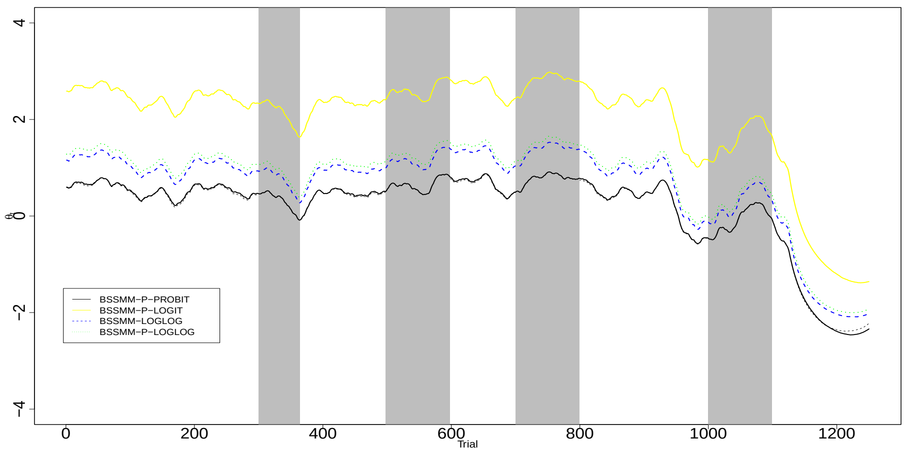

Figure 3: Estimation results for the monkey performance data set. Posterior smoothed mean of θ_t.

rion. In this case the estimated probability is less constrained and tracks the data independently of the stimulation-ON/OFF information. In all the cases, on average the response curve lies around the 0.75 level but decreases are observed at the end of the first stimulation-ON period around trial 375, at the end of the 4th OFF period around trial 950 and for the remainder of the experiment from trial 1100 onwards, with some slight differences starting around 1150. All the models are able to account for stimulation effect. The results indicate that stimulation has a positive influence on the performance. However, they show that the performance does not improve during the first stimulation period. Overall, however, all the models results highlight an abrupt step-like decline in performance towards the end of the experiment, around trial 950, which undergoes a significant increase during the final stimulation period before a final significant drop to zero. All the results are consistent with Smith et al. (2009).

5 Conclusions

In this paper we have proposed three flexible classes of state space mixed models for longitudinal binary data using power distributions as extensions of Czado and Song (2008) and Abanto-Valle and Dey (2014). In this setup, the parameters controling the skewness are estimated along with

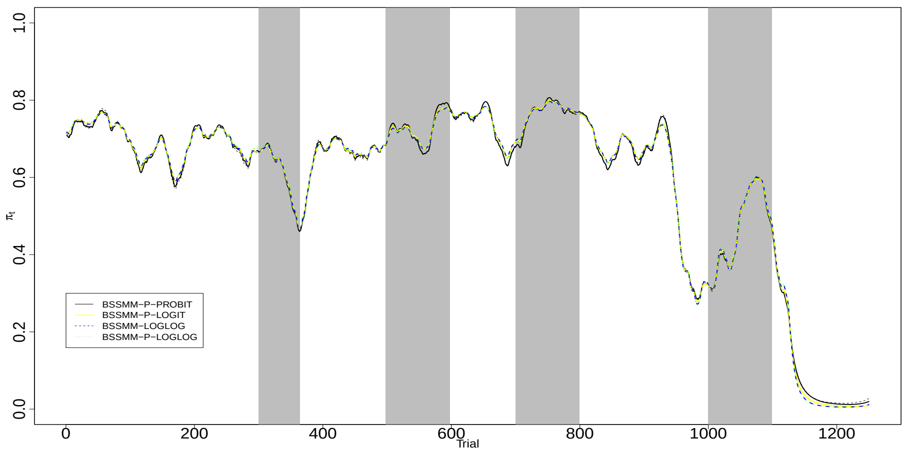

Figure 4: Estimation results for the monkey performance data set. Posterior smoothed mean of π_t.

model fitting. The flexibility in links is important to avoid link misspecification. An attractive aspect of the model is that it can be easily implemented, under a Bayesian perspective, via MCMC by using JAGS. We illustrated the methods through an empirical application with the monkey performance data set. WAIC measure is used it for model comparison. Empirical findings show that the BSSMM with power links are extremely robust in model fitting no matter the data favors left skewed, symmetric or right skewed links.

This article makes certain contributions, but several extensions are still possible. First, we focus on binary observations, but the setup can be extended to binomial and ordinal data. Langrock (2011) has shown that methods which are well-known for hidden Markov models (HMMs) can be applied in order to perform a fast and accurate numerical integration for the likelihood function in general state space models in order to get maximum likelihood-based estimators. Nevertheless, a deeper investigation of those modifications in the context of BSSM models is beyond the scope of the present paper, but provides stimulating topics for future research.

References

Abanto-Valle, C. A., and Dey, D. K. (2014), "State space mixed models for binary responses with scale mixture of normal distributions links," *Computational Statistics and Data Analysis*, 71, 274–287.

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

Basu, S., and Mukhopadhyay, S. (2000*a*), "Bayesian Analysis of Binary Regression Using Symmetric and Asymmetric links," *Sankhyã: The Indian Journal of Statistics, Series B*, 62, 372–379.

Basu, S., and Mukhopadhyay, S. (2000*b*), "Binary Response Regression with Normal Scale Mixture Links," in *Generalized Linear Models: A Bayesian perspective.*, eds. D. K. Dey, S. K. Ghosh, and B. K. Mallick, New York: Marcell Decker, pp. 231–239.

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.

Bazán, J. L., Romeo, J. S., and Rodrigues, J. (2014), "Bayesian skew-probit regression for binary response data," *Brazilian Journal of Probability and Statistics*, forthcoming.

Carlin, B. P., and Polson, N. G. (1992), "Monte Carlo Bayesian methods for discrete regression models and categorical time series," in *Bayesian Statistics. Vol. 4.*, eds. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, Oxford, U.K.: Clarendon Press, pp. 577–586.

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

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

Czado, C., and Song, P. X.-K. (2008), "State space mixed models for longitudinal observations with binary and binomial responses," *Statistical Papers*, 49, 691–714.

Fahrmeir, L. (1992), "Posterior mode estimation by extended Kalman filtering for multivariate dynamic generalized linear models," *Journal of the American Statistical Association*, 87, 501–509.

Gelman, A., Hwang, J., and Vehtari, A. (2014), "Understanding predictive information criteria for Bayesian models," *Statistics and Computing*, 24, 997–1016.

Geweke, J. (1992), "Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments," in *Bayesian Statistics. Vol. 4.*, eds. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, Oxford, U.K.: Clarendon Press, pp. 577–586.

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

Heidelberger, P., and Welch, P. D. (1983), "Simulation run length control in the presence of an initial transient," *Operations Research*, 31, 1109–1144.

Johnson, N. L., Kotz, S., and Balakrishnan, N. (1995), *Continuous Univariate Distributions*, Vol. 2, New York: Wiley.

Kim, S., Chen, M. H., and Dey, D. K. (2008), "Flexible generalized t-link models for binary response data," *Biometrika*, 95, 93–106.

Langrock, R. (2011), "Some applications of nonlinear and non-Gaussian state-space modelling by means of hidden Markov models," *Journal of Applied Statistics*, 38, 2955–2970.

McCullagh, P., and Nelder, J. A. (1989), *Generalized linear models*, 2nd ed edn, London: Chapman and Hall.

Plummer, M. (2003), JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling,, in *Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003). March*, pp. 20–22.

Shah, S. A., Baker, J. L., Ryou, J. W., Purpura, K. P., and Schiff, N. D. (2009), "Modulation of arousal regulation with central thalamic deep brain stimulation," *Conf Proc IEEE Eng Med Biol Soc*, 1, 3314–3317.

Smith, A. C., Shah, S. A., Hudson, A. E., Purpura, K. P., and Victor, J. D. (2009), "A Bayesian statistical analysis of behavioral facilitation associated with deep brain stimulation," *Journal of Neuroscience Methods*, 183, 267–276.

Song, P. K. (2000), "Monte Carlo Kalman filter and smoothing for multivariate discrete state space models," *The Canadian Journal of Statistics*, 28, 641–652.

Stukel, T. A. (1988), "Generalized Logistic Models," *Journal of the American Statistical Association*, 83, 426–431.

Watanabe, S. (2010), "Asymtotic equivalence of Bayes crross validation and widely applicable information criterion in singular learning theory," *Journal of Machine Learning Research*, 14, 867–897.

Watanabe, S. (2013), "A widely applicable Bayesian information criterion," *Journal of Machine Learning Research*, 14, 867–897.

West, M., Harrison, P. J., and Migon, H. S. (1985), "Dynamic Generalized Linear Models and Bayesian Forecasting," *Journal of the American Statistical Association*, 136, 209–220. With discussion.