Statistics and Its Interface Volume 8 (2015) 187–194

Binary state space mixed models with flexible link functions: a case study on deep brain stimulation on attention reaction time*

Carlos A. Abanto-Valle, Dipak K. Dey†, and Xun Jiang

State space models (SSM) for binary time series data using a flexible skewed link functions are introduced in this paper. Commonly used logit, cloglog and loglog links are prone to link misspecification because of their fixed skewness. Here we introduce two flexible links as alternatives, they are the generalized extreme value (GEV) and the symmetric power logit (SPLOGIT) links. Markov chain Monte Carlo (MCMC) methods for Bayesian analysis of SSM with these links are implemented using the JAGS package, a freely available software. Model comparison relies on the deviance information criterion (DIC). The flexibility of the proposed model is illustrated to measure effects of deep brain stimulation (DBS) on attention of a macaque monkey performing a reaction-time task [19]. Empirical results showed that the flexible links fit better over the usual logit and cloglog links.

Keywords and phrases: Binary time series, GEV link, Logit link, Markov chain Monte Carlo, Probit link, State space models.

1. INTRODUCTION

Binary response data with two possible outcomes are often encountered in statistical modeling. Generalized linear models [17] can be used to model time series of binary response. However, it might not be adequate if the observations are correlated over time. To address the serial correlation that might be presenting, West et al. [25] used Generalized linear state space models in a conjugate Bayesian setup. Further researches on this topic have been followed by Fahrmeir [12], Song [20], Carlin and Polson [7] and Czado and Song [11] among others.

A critical issue in modeling binary response data is the choice of the links. In the context of binary regression, logit and probit links are two widely used symmetric link functions in the literature [2, 4, 5]. However, as Chen et al. [8] have argued, when the latent probability of a given binary response approaches 0 with different rate as it approaches 1, symmetric link functions may not be adequate to fit binary data and result in substantial bias in the mean response estimates [10]. To deal with this problem some asymmetric links are considered in the literature. Two of the commonly adopted asymmetric link functions are complementary loglog (cloglog) and loglog. However, these two links have fixed skewness and lack the flexibility to let the data tell how much skewness should be incorporated.

There are lots of research done to introduce flexibility of skewness as well as tail behavior into the link functions. For example, Stukel [22] proposed a two-parameter class of generalized logistic models, Kim et al. [15] used the skewed generalized t-link, and Bazán et al. [6] adopted the skewed probit links and some variants with different parameterizations. Wang and Dey [23], Wang and Dey [24] and Jiang et al. [14] introduced the flexible class of link functions as an appropriate model for the binary cross sectional data. Among them the GEV and SPLOGIT class of links are not only very flexible but they also include many standard symmetric links as special cases. Specifically, with a free shape parameter, the GEV distribution provides great flexibility in fitting a wide range of skewness in the response curve. Alternatively by introducing a power parameter on a baseline logit link and its mirror reflection, the SPLOGIT link achieves great flexibility in both positive and negative directions in a symmetric fashion. In terms of tail behavior, scale mixture of normal link (Choy and Chan 9) is a rich class of symmetric link functions that contain many standard links (e.g., probit, Student-t) as special cases.

State space model for binary responses have been used by Carlin and Polson [7] and Song [20] without including covariates. Czado and Song [11] introduced covariates for binary state space models with probit link and called the resulting class as binary state space mixed models (BSSM). More recently, Abanto-Valle and Dey [1] extended it to scale mixture of normal (SMN) links, which produces a general class of symmetric links.

In this paper, we compare the BSSM by assuming three standard link functions and two flexible link functions. The three standard links we consider here are logit, cloglog and loglog and we call the corresponding state space model BSSM-LOGIT, BSSM-CLOGLOG and BSSM-LOGLOG.

*The research of Carlos A. Abanto-Valle was supported by the CNPq-Brazil grants 303421/2012-6 and 481159/2013-4, and FAPERJ grant E-26/110.359/2014.
†Corresponding author.

The two flexible links are the SPLOGIT link and the GEV link with corresponding models BSSM–SPLOGIT and BSSM–GEV. We then fit the models and the inferences are done under a Bayesian paradigm via MCMC methods, which permits to obtain the posterior distribution of parameters based on reasonable prior assumptions. Despite the growing number of advanced sampling schemes developed with various degree of sophistication and complexity, the idea to trade off the easy-to-use techniques with more efficient but complicated techniques may be unattractive to general practitioners. Therefore, we adopt the JAGS software [18] running inside the R package to implement the three models although the JAGS software uses a single-move algorithm to draw a sample from the joint posterior distribution. Compared with a multi-move sampler, the single-move sampler produces higher correlated posterior samples. Such dependency can be compensated by running a longer Markov chain. On the other hand, the gain in efficiency in using complex sampling schemes to some extent is outweighed by the ease of implementation in JAGS.

The remainder of this paper is organized as follows. Section 2 gives a brief review about the GEV distribution. Section 3 introduces the symmetric power logit models. Section 4 outlines the setup of the BSSM models for the three flexible link functions as well as the corresponding Bayesian estimation procedure using MCMC methods. Section 5 conducts a simulation study about the robustness of the flexible link functions. Section 6 is devoted to the application and model comparison of all the six models using a real data set. Finally, some concluding remarks and suggestions for future developments are given in Section 7.

2. GENERALIZED EXTREME VALUE LINK

The GEV link models are based on the Generalized Extreme Value (GEV) distribution, which is given by

$$(1) \qquad G(x) = \exp\left[-\left\{1+\xi\frac{x-\mu}{\sigma}\right\}_{+}^{-\frac{1}{\xi}}\right],$$

where $\mu \in R$ is the location parameter, $\sigma \in R^+$ is the scale parameter, $\xi \in R$ is the shape parameter and $x_+ = \max(x, 0)$. The distribution in Model (1) is called the GEV distribution. Its importance as a link function arises from the fact that the shape parameter ξ purely controls the tail behavior of the distribution [23, 24]. Figure 1 provides a comparison of pdf and cdf plots of the GEV class with different ξ to show the flexibility of such distributions. By looking at the cdf plot it is obvious that as the values of the shape parameter change, so does the approaching rate to 1 and 0.

Since the usual definition of skewness $\mu_3 = \{E(X-\mu)^3\}\{E(X-\mu)\}^{-\frac{3}{2}}$ does not exist for large positive values of X's for the GEV model, Wang and Dey [23] and Wang and Dey [24] extended the skewness measure of Arnold and Groeneveld [3] for the GEV distribution in terms of its mode. Wang and Dey [23] and Wang and Dey [24] showed

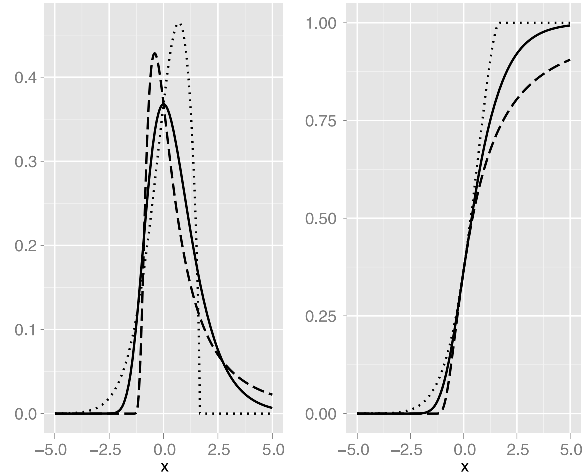

Figure 1. Left: pdf plot of GEV distribution. Right: cdf plot of GEV distribution. Solid line ($\xi = 0$), dashed line ($\xi = 0.6$), and dotted line ($\xi = -0.6$).

that, based on this skewness definition, the GEV distribution is negatively skewed for $\xi < -0.307$ and positively skewed for $\xi > -0.307$.

3. SYMMETRIC POWER LOGIT MODELS

Jiang et al. [14] propose a general class of flexible link functions based on a symmetric baseline link function and its mirror reflection. Suppose F_0 is a baseline link function for which the pdf is symmetric about zero, the symmetric power distribution (as link function as well) based on F_0 is defined as

$$(2) \quad F(x,r) = F_0^r(\frac{x}{r})\mathbf{I}_{(0,1]}(r) + \left(1 - F_0^{\frac{1}{r}}(-rx)\right)\mathbf{I}_{(1,+\infty)}(r),$$

where $\mathbf{I}$ is the indicator function. Considering F_0 as the cdf of a logistic distribution which will lead us to SPLOGIT link adopted in this paper. By combining the power of the baseline link and its reflection in one single family great flexibility of skewness can be achieved in both positive and negative directions. Also, scaling x by the same parameter r in the formulation can prevent the mode of the pdf to be too far away from zero. Clearly, by introducing an additional parameter r in logit baseline in the form of (2), the skewness of SPLOGIT distribution can be adjusted from its baseline to achieve more flexibility in modeling asymmetric data.. However, the construction of (2) ensures that the flexibility is achieved symmetrically with respect to $r = 1$ and thus can accomondate greater skewnewss for both directions by choosing appropriate tail.

Under Arnold and Groeneveld [3]'s measure, the skewness of SPLOGIT distribution can be found analytically as $\gamma_M = 1 - 2(\frac{r}{r+1})^r$ for $0 < r < 1$, and $\gamma_M = 2(\frac{1}{r+1})^{1/r} - 1$ for $r > 1$. As a result, it is negatively skewed when $0 < r < 1$, positively skewed when $r > 1$, and reduces to the symmetric

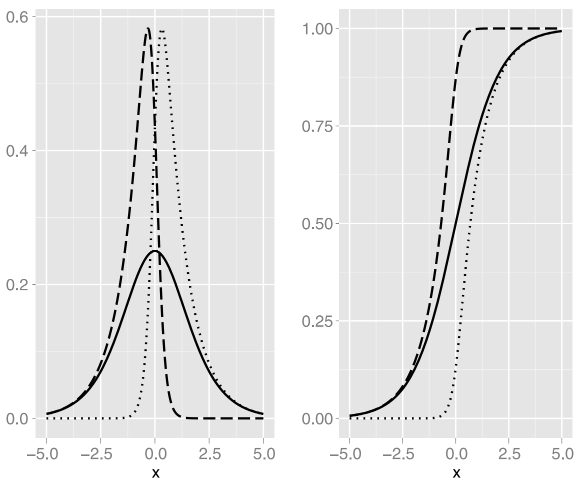

Figure 2. Left: pdf plot of SPLOGIT distribution. Right: cdf plot of SPLOGIT distribution. Solid line ($r = 1$), dashed line ($r = 0.2$), and dotted line ($r = 5$).

logit link when $r = 1$. The range of skewness provided by splogit family is unlimited, reaching -1 and 1 respectively, as r tends to 0 and $+\infty$. Figure 2 compares the pdf and cdf plots of SPLOGIT distribution with different r values. It is clear that the pdf associated with $r = 0$ is just the standard logistic distribution, and the pdf asscociated with $r = 0.2$ is the mirror reflection of the pdf asscociated with $r = 5$ since 0.2 is the reciprocal of 5.

4. BINARY RESPONSES STATE SPACE MIXED MODELS WITH GEV AND SPLOGIT LINKS

4.1 Model setup

Let $\mathbf{Y}_{1:T} = (Y_1, \ldots, Y_T)'$, where $Y_t, t = 1, \ldots, T$, denote T independent binary random variables. Suppose $\mathbf{x}_t$ is a $k \times 1$ vector of covariates. We assume that

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

In the GEV case, $F(x) = 1 - G(-x)$, where $G(x)$ represents the cdf at x for the GEV distribution with $\mu = 0$ and $\sigma = 1$ and unknown shape parameter ξ. Notice that the GEV link specified here is the mirror reflection of GEV distribution discribed in Section 2, thus is positively skewed for $\xi < -0.307$ and negatively skewed for $\xi > -0.307$. Also, when $\xi = 0$, the GEV model reduces to CLOGLOG model. In the case of SPLOGIT, $F(x) = F(x, r)$ defined in (2). It is negatively skewed when $r < 1$ and positively skewed when $r > 0$. As stated before, we name the two odels BSSM-GEV and BSSM-SPLOGIT, respectively. 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 $\theta_0 \sim \mathcal{N}(0, \frac{\tau^2}{1-\delta^2})$. Clearly θ_t represents a time-specific effect on the observed process. Only for the objective of comparison, we include the BSSM with slash link [see 1, for details about the BSSM with slash link] and denote it as BSSM-SLASH. Under a Bayesian paradigm, we use MCMC methods to conduct the posterior analysis in the next subsection.

4.2 Inference procedure

Here we develop a Markov Chain Monte Carlo (MCMC) procedure to make inference about the model defined by (3)–(5) under the Bayesian paradigm. It is obvious that the model depends on a parameter vector Ψ, where $\Psi = (\boldsymbol{\beta}', \delta, \tau^2, \xi)'$ in GEV case, $\Psi = (\boldsymbol{\beta}', \delta, \tau^2, r)'$ in SPLOGIT case and $\Psi = (\boldsymbol{\beta}', \delta, \tau^2, \nu)'$ in SLASH model. Let $\boldsymbol{\theta}_{0:T} = (\theta_0, \theta_1, \ldots, \theta_T)'$ be the latent states. The Bayesian approach for estimating model parameters 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

$$
\begin{aligned}
p(\boldsymbol{\theta}_{0:T}, \Psi \mid \mathbf{y}_{1:T}) &\propto p(\mathbf{Y}_{1:T} \mid \boldsymbol{\theta}_{0:T}, \Psi, \mathbf{y}_{1:T}) \\
(6) \qquad &\times p(\boldsymbol{\theta}_{0:T} \mid \Psi)p(\Psi),
\end{aligned}
$$

where

$$
\begin{aligned}
(7) \qquad p(\mathbf{Y}_{1:T} \mid \boldsymbol{\theta}_{0:T}, \Psi) &= \prod_{t=1}^{T} \{\pi_t^{Y_t}(1-\pi_t)^{1-Y_t}\} \\
p(\boldsymbol{\theta}_{0:T} \mid \Psi) &= \phi(\theta_0 \mid 0, \frac{\tau^2}{1-\delta^2}) \\
(8) \qquad &\times \prod_{t=1}^{T} \phi(\theta_t \mid \delta\theta_{t-1}, \tau^2),
\end{aligned}
$$

and π_t is given by equation (4) 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. In GEV, SPLOGIT and SLASH models the three prior distributions of Ψ can be written as

$$
\begin{aligned}
p_{GEV}(\Psi) &= p(\boldsymbol{\beta})p(\delta)p(\tau^2)p(\xi), \\
p_{SPLOGIT}(\Psi) &= p(\boldsymbol{\beta})p(\delta)p(\tau^2)p(r), \\
p_{SLASH}(\Psi) &= p(\boldsymbol{\beta})p(\delta)p(\tau^2)\prod_{t=1}^{T} p(\lambda_t|\nu)p(\nu).
\end{aligned}
$$

The prior distributions for individual parameters are set as: $\boldsymbol{\beta} \sim \mathcal{N}_k(\boldsymbol{\beta}_0, \boldsymbol{\Sigma}_0)$, $\delta \sim \mathcal{N}_{(-1,1)}(\delta_0, \sigma_\delta^2)$, $\tau^2 \sim \mathcal{IG}(\frac{n_0}{2}, \frac{T_0}{2})$, $\xi \sim \mathcal{U}(-u_0, u_0)$, $r \sim \mathcal{G}(r_0, r_0)$, $\lambda_t \sim \mathcal{B}e(\nu, 1), t = 1, \ldots, t$ and $\nu \sim \mathcal{G}(u_0, v_0)$, 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.

Table 1. Simulation results based on 200 replicates. Mean is the average of posterior mean. SD is the average of posterior standard deviation. Bold numbers indicate the fit under the true model

True model: BSSMM-LOGIT												
	Fitted model											
	LOGIT		CLOGLOG		LOGLOG		GEV		SPLOGIT		SLASH	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
β_0	**0.036**	**0.174**	-0.475	0.122	0.432	0.120	-0.415	0.144	-0.040	0.406	0.024	0.131
β_1	**1.036**	**0.109**	0.703	0.076	0.703	0.080	0.672	0.091	0.822	0.155	0.733	0.115
δ	**0.902**	**0.078**	0.878	0.088	0.874	0.099	0.878	0.089	0.889	0.092	0.891	0.083
τ^2	**0.132**	**0.159**	0.090	0.121	0.107	0.157	0.079	0.098	0.074	0.075	0.101	0.083
$\xi/r/\nu$							-0.141	0.204	1.045	0.511	4.313	2.044
True model: BSSMM-CLOGLOG												
	Fitted model											
	LOGIT		CLOGLOG		LOGLOG		GEV		SPLOGIT		SLASH	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
β_0	0.819	0.262	**0.046**	**0.168**	1.043	0.191	-0.014	0.180	0.023	0.402	0.516	0.183
β_1	1.626	0.149	**1.041**	**0.099**	1.157	0.136	1.068	0.124	1.108	0.243	1.109	0.133
δ	0.929	0.036	**0.929**	**0.035**	0.894	0.060	0.926	0.036	0.926	0.036	0.927	0.039
τ^2	0.194	0.140	**0.089**	**0.074**	0.219	0.248	0.072	0.074	0.095	0.077	0.093	0.078
$\xi/r/\nu$							0.073	0.193	0.560	0.291	4.815	2.108
True model: BSSMM-LOGLOG												
	Fitted model											
	LOGIT		CLOGLOG		LOGLOG		GEV		SPLOGIT		SLASH	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
β_0	-0.664	0.258	-1.082	0.189	**-0.052**	**0.167**	-0.750	0.178	-0.107	0.414	-0.470	0.184
β_1	1.590	0.145	1.126	0.129	**1.023**	**0.094**	0.872	0.099	1.160	0.230	1.072	0.125
δ	0.931	0.037	0.902	0.058	**0.931**	**0.036**	0.933	0.034	0.931	0.036	0.931	0.037
τ^2	0.171	0.124	0.210	0.247	**0.071**	**0.052**	0.052	0.038	0.097	0.076	0.079	0.056
$\xi/r/\nu$							-0.413	0.107	1.956	0.874	4.680	2.049

We can evaluate Equation (6) using standard MCMC methods in JAGS [18]. Implementation in this software merely requires specifying the model setup in equations (3)–(5), as well as priors for the unknown parameters $p(\Psi)$.

5. SIMULATION STUDY

Here we conduct a simulation study to investigate the robustness of the BSSM-GEV, BSSM-SPLOGIT and BSSM-SLASH models against link misspecification when the data are generated from different standard models. We generate our data from (3)–(5) with F set to be the cdf corresponding to LOGIT, CLOGLOG and LOGLOG links. Under Arnold and Groeneveld [3]'s measure, the skewness associated with the three links are 0, -0.264 and 0.264. For each of the true model we independently generate 200 datasets of sample sizes $T = 800$. For each dateset, we generate one covariates $\mathbf{x}_t, t = 1, \ldots, T$ from independent standard normal distributions. The true values of regression coefficients are set to be $\boldsymbol{\beta} = (\beta_0, \beta_1)' = (0, 1)'$. We set other parameters $\delta = 0.95$ and $\tau = 0.2$. Then we fit the BSSM-LOGIT, BSSM-CLOGLOG, BSSM-LOGLOG, BSSM-GEV, BSSM-SPLOGIT and BSSM-SLASH models to each set of the generated data and compare the outcomes. The prior distributions of parameters are set as: $\delta \sim \mathcal{N}_{-1,1}(0.95, 100)$, $\tau^2 \sim \mathcal{IG}(0.25, 0.01)$ and $\boldsymbol{\beta} = (\beta_0, \beta_1)' \sim \mathcal{N}_2(\boldsymbol{\beta}_0, \Sigma_0)$, where $\boldsymbol{\beta}_0 = \mathbf{0}$ and $\Sigma_0 = 500^2\mathbf{I}_2$, $\mathbf{0}$ indicates a 2×1 vector os zeros and $\mathbf{I}_2$ the identity matrix of order 2. The prior for shape parameter ξ in BSSM-GEV model is set to follow $\mathcal{U}(-0.6, 0.6)$. The prior for power parameter r in BSSM-SPLOGIT model is set to follow $\mathcal{G}(1, 1)$. The prior for shape parameter ν in BSSM-SLASH model is set to follow $\mathcal{G}(5.0, 0.8)$.

Table 1 summarizes the average posterior mean and posterior standard deviation of the parameters under different combination of true and fitted models. Notice the true value of $\beta_0 = 0$ and $\beta_1 = 1$, it is clear that the fit under the true model (bold numbers) recovers the original value of the parameters very nicely in all three cases. Since BSSM-GEV and BSSM-SPLOGIT models are the two flexible models account for skewed data, we pay special attention to shape parameter ξ in BSSM-GEV and power parameter r in BSSM-SPLOGIT. When the true models are BSSM-LOGIT, BSSM-CLOGLOG and BSSM-LOGLOG respectively, $\xi = -0.141, 0.073, -0.413$, and $r = 1.045, 0.560, 1.956$. It is clear from Section 2 and 3 that the values of ξ and r reflect the skewness of BSSM-LOGIT (symmetric), BSSM-CLOGLOG (left skewed) and BSSM-LOGLOG (right-skewed). Also, from Section 4 we see that BSSM-CLOGLOG is a special case of BSSM-GEV when $\xi = 0$, this is clearly confirmed in the BSSM-GEV fitting by an average fitted value of $\xi = 0.073$, $\beta_0 = -0.014$ and $\beta_1 = 1.068$, all extremely close to the parameter setup of

Table 2. Percentage of best performance and average DIC under different model fitting among 200 replicates. The best performance is determined as the lowest DIC

	True model: BSSMM-LOGIT				
	Fitted model				
	CLOGLOG	LOGLOG	GEV	SPLOGIT	SLASH
% best	16.0%	20.5%	24.0%	1.5%	38.0%
avg DIC	934.9	935.0	933.4	937.1	933.5
	True model: BSSMM-CLOGLOG				
	Fitted model				
	LOGIT	LOGLOG	GEV	SPLOGIT	SLASH
% best	0.0%	14.5%	68.5%	3.5%	13.5%
avg DIC	770.8	768.9	755.5	762.7	765.5
	True model: BSSMM-LOGLOG				
	Fitted model				
	LOGIT	CLOGLOG	GEV	SPLOGIT	SLASH
% best	0.5%	14.5%	70.5%	1.0%	13.5%
agv DIC	778.6	778.1	768.5	774.4	773.8

the true model. Similarly, it is also clear from Section 3 that BSSM-LOGIT is a special case of BSSM-SPLOGIT when $r = 1$, which is also confirmed by the fact that $r = 1.045$, $\beta_0 = -0.040$ and $\beta_1 = 0.822$ under the BSSM-SPLOGIT fit. Since the BSSM-SLASH model does not accommodate skewed data, it behaves similarly as BSSM-LOGIT model when the true model is symmetric (BSSM-LOGIT).

Model comparison results have been summarized in Table 2. We exclude fitting the true model so that only the misspecified models are compared. By looking at the percentage of best performers, we can clearly see that the BSSM-GEV model is extremely robust against link misspecifications with 24.0% best performance in BSSM-LOGIT case, 68.5% in BSSM-CLOGLOG case and 70.5% in BSSM-LOGLOG case. SLASH model performs the best in BSSM-LOGIT case with 38.0% best performance, and 13.5% each in other two cases. For BSSM-SPLOGIT case, if we only look at the percentage of best performance, it would be tempting to conclude that it performs even worse than some standard link function. However, closer examination of the average DIC value shows that in BSSM-CLOGLOG and BSSM-LOGLOG case, BSSM-SPLOGIT model performs much better than other standard links. It is clear that BSSM-GEV model always outperforms BSSM-SPLOGIT a little, therefore making the percentage of best performance of BSSM-SPLOGIT look bad. In conclusion, our three flexible link models performs well against link misspecification with BSSM-GEV model definitely stands out as the best model.

6. CASE STUDY: DEEP BRAIN STIMULATION ON ATTENTION REACTION TIME

To illustrate the technique applied to binary responses, we consider responses from a monkey performing the attention paradigm described in Smith et al. [19]. The task consisted of making a saccade to a visual target followed by a variable period of fixation on the target and detection of a change in target color followed by a bar release. This standard task requires sustained attention because in order to receive a reward, the animal must release the bar within a brief time window cued by the change in target color [see 19, for a more detailed description of the experiment]. Thus our behavioral data set for this experiment are composed of a time series of binary observations with a 1 corresponding to reward being delivered and a 0 corresponding to reward not being delivered at each trial, respectively. The goal of the experiment is to determine whether, once performance has diminished as a result of spontaneous fatigue, deep brain stimulation (DBS) allows the animal to recover its pre-fatigue level of performance. In this experiment, 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[1]. For this data set we fit the Binary state space model with three standard link functions (LOGIT, CLOGLOG and LOGLOG), as well as three flexible link functions (SLASH, SPLOGIT and GEV) defined in previous sections, 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 standard link functions in LOGIT, CLOGLOG and LOGLOG models. For SLASH and SPLOGIT case, $F(x)$ represents the cumulative distribution function at

[1]We thank Anne C. Smith for making the data set available on her website: http://www.ucdmc.ucdavis.edu/anesthesiology/research/smith_Bayesian.html.

Table 3. Estimation results for monkey performance data set. First row: Posterior mean. Second row: Posterior 95% HPD interval in parentheses

	Model					
	LOGIT	CLOGLOG	LOGLOG	SLASH	SPLOGIT	GEV
	0.9954	0.9968	0.9959	0.9873	0.9969	0.9939
δ	(0.9895,0.9999)	(0.9921,0.9999)	(0.9909,0.9999)	(0.9889,0.9999)	(0.9931,0.9999)	(0.9864,0.9999)
	0.0233	0.0110	0.0117	0.0118	0.0087	0.0083
τ^2	(0.0085,0.0411)	(0.0046,0.0191)	(0.0041,0.0217)	(0.0043,0.0215)	(0.0029,0.0165)	(0.0029,0.0146)
	–	–	–	5.1536	4.0230	-0.5396
$\nu/r/\xi$	–	–	–	(1.3417, 9.1360)	(1.3568,7.1885)	(-0.5998, -0.4229)

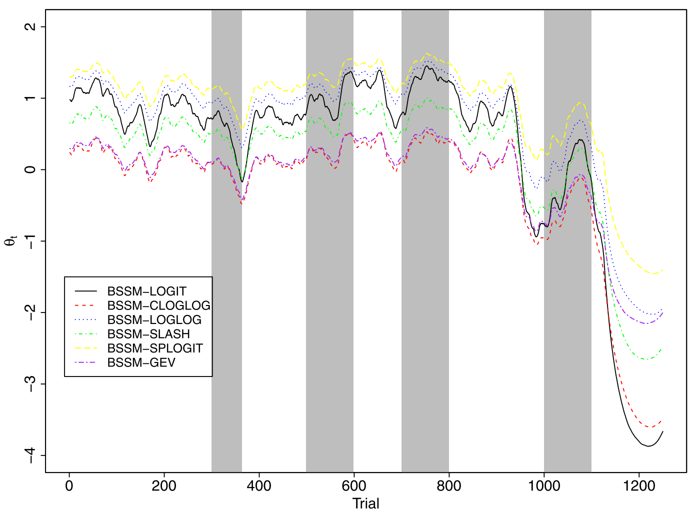

Figure 3. Estimation results for the monkey performance data set. Posterior smoothed mean of θ_t.

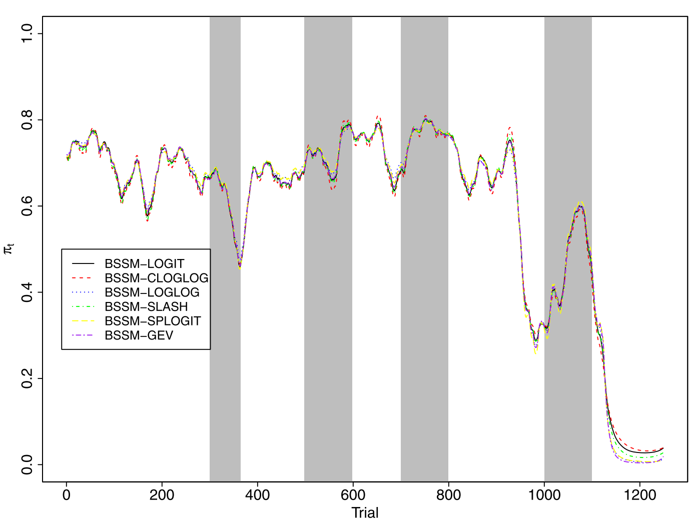

Figure 4. Estimation results for the monkey performance data set. Posterior smoothed mean of π_t.

x for the SLASH and SPLOGIT distributions. In GEV case, let $G(x)$ be the cumulative distribution function at x for the GEV distribution with $\mu = 0$ and $\sigma = 1$, then $F(\theta_t) = 1 - G(-\theta_t)$. θ_t is the arousal state of the macaque monkey at time t. We set the priors as $\delta \sim \mathcal{N}_{(-1,1)}(0.95, 1000)$, $\tau^2 \sim \mathcal{IG}(0.1, 0.01)$, $\nu \sim \mathcal{G}(5,1)$ for SLASH model, $\xi \sim \mathcal{U}(-0.6, 0.6)$ for GEV model and $r \sim \mathcal{G}(1,1)$ for SPLOGIT model. We fit these six models denoted by BSSM-LOGIT, BSSM-CLOGLOG, BSSM-LOGLOG, BSSM-SLASH, BSSM-SPLOGIT and BSSM-GEV, using the software package JAGS, because of its user-friendly model declaration language[2]. JAGS is not designed to handle extremely large models and data sets (e.g., >2000 trials). Other software may be preferable in these situations. For each case, we conducted the MCMC simulation for 550000 iterations. In all the cases, the first 50000 draws were discarded as a burn-in period. In order to reduce the autocorrelation between successive values of the simulated chain, only every 100th values of the chain were stored. With the resulting 5000 values, we calculated the posterior means, the 95% HPD intervals. The MCMC output of all the parameters passed the convergence test of Heidelberger and Welch [13], available for free with the CODA package with the R software.

From Table 3, 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 0.0085 and 0.0233. For GEV model we found that the posterior mean and 95% HPD interval for ξ are -0.5396 and $(-0.5998, -0.4229)$, while for SPLOGIT model the posterior mean and 95% HPD interval for r are 4.0230 and (1.3568, 7.1885). Notice that from Section 4 we see that both values indicate the data favors positively skewed link functions, which corresponds to LOGLOG among the standard link functions we consider here. The posterior mean and 95% HPD interval for ν in SLASH model are 5.1536 and (1.3417, 9.1360).

To assess the goodness of the estimated models, we calculate the deviance information criterion, DIC [21] to compare models using different link functions. The deviance function is defined as:

[2] The JAGS codes for the BSSM-SLASH, BSSM-SPLOGIT and BSSM-GEV models are available upon request to the first author.

Table 4. Monkey performance data set. DIC: deviance information criterion

Model	DIC	Rank
BSSM-LOGIT	1423.9	5
BSSM-CLOGLOG	1433.1	6
BSSM-LOGLOG	1414.7	1
BSSM-SLASH	1420.6	4
BSSM-SPLOGIT	1415.7	2
BSSM-GEV	1417.6	3

$$\begin{aligned} D(\Psi, \boldsymbol{\theta}_{0:T}) &= -2\log[p(\mathbf{y}_{1:T} \mid \Psi, \boldsymbol{\theta}_{0:T})] \\ &= -2\sum_{t=1}^{T}[y_t \log(\pi_t) + (1-y_t)\log(1-\pi_t)]. \end{aligned} \tag{9}$$

The deviance information criterion (DIC) is defined by

$$\begin{aligned} \text{DIC} &= -2E_{\Psi,\boldsymbol{\theta}_{0:T}|\mathbf{y}_{1:T}}[D(\Psi, \boldsymbol{\theta}_{0:T})] + p_D \\ &= \bar{D}(\Psi, \boldsymbol{\theta}_{0:T}) + p_D, \end{aligned} \tag{10}$$

where $E_{\Psi,\boldsymbol{\theta}_{0:T}|\mathbf{y}_{1:T}}$ denotes the expectation taken with respect to the posterior distribution of Ψ and $\boldsymbol{\theta}_{0:T}$ given the data $\mathbf{y}_{1:T}$. The second term p_D in (10) is the effective number of parameters, which measures the complexity of the model. Specifically, p_D is defined as the difference between the deviance evaluated at the posterior mean of the parameters and the posterior mean of the deviance:

$$p_D = \bar{D}(\Psi, \boldsymbol{\theta}_{0:T}) - D(\bar{\Psi}, \bar{\boldsymbol{\theta}}_{0:T}). \tag{11}$$

Computing the DIC is straightforward in an MCMC implementation. Monitoring both $(\Psi, \boldsymbol{\theta}_{0:T})$ and $D(\Psi, \boldsymbol{\theta}_{0:T})$ in MCMC updates, at the end one can estimate the D by the sample mean of the simulated values of D and the $D(\bar{\Psi}, \bar{\boldsymbol{\theta}}_{0:T})$ by plugging in the sample means of the simulated posterior values of Ψ and $\boldsymbol{\theta}_{0:T}$. A lower values of DIC indicates a better-fitting model. The DIC is easily calculated using JAGS. Table 4 summarizes the DIC for our six models. The DIC selects the BSSM-LOGLOG as the best model for the monkey performance data set, although BSSM-SPLOGIT and BSSM-GEV are close as well. This confirms our observation that the data supports positively skewed link function, namely the BSSM-LOGLOG standard link, as well as BSSM-GEV and BSSM-SPLOGIT with positive skewed parameters.

Additionally, we use another measure of global fit, by considering the posterior mean of unstandardized residuals $e_i = y_i - E(Y_i)$, that is the sum of absolute residuals (SAR) defined as $\text{SAR} = \sum_{i=1}^{n} |e_i|$. As before, lower values of SAR indicate a better fit. Table 5 shows the comparing time for the competing models. The BSSM-GEV minimizes the SAR. According to the SAR criterion, skewed link models give better fit than the logit link.

Figure 3 shows the posterior smoothed mean for the states θ_t for each one of the models fitted. Different line types and colors indicate the posterior smoothed mean for the six fitted models respectively. All the estimates follow a similar pattern, but there are expressive differences between the estimates, specially in the last OFF period.

Table 5. Monkey performance data set. Sum of absolute residuals (SAR) and computing time (in minutes) for 50000 swaps for the competing models

Model	SAR	Rank	Computing time
BSSM-LOGIT	470.32	6	3.17
BSSM-CLOGLOG	472.28	5	2.79
BSSM-LOGLOG	469.75	4	3.50
BSSM-SLASH	469.28	3	9.55
BSSM-SPLOGIT	468.84	2	15.63
BSSM-GEV	467.35	1	3.20

In Figure 4, we plot the posterior smoothed mean for the probability of a correct response computed using the six fitted models. In this case the estimated probability is less constrained and tracks the data independent 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 result 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. [19].

7. CONCLUSIONS

In this paper we have proposed three flexible classes of state space mixed models for longitudinal binary data using GEV, SPLOGIT and SMN distributions as extensions of Czado and Song [11] and Abanto-Valle and Dey [1]. In this setup, the parameters controling the skewness and tail behavior are estimated along with 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 a simulation study and an empirical application with the monkey performance data set. DIC measure is used it for model comparison. Empirical findings show that the BSSM-GEV and BSSM-SPLOGIT model 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. Second, if the rate of zeros or ones are not the same, we can compare the performance of our flexible links with other skewed links as the skew normal or the skew Student-t. In such case, it is necessary to develop efficient sampler for the states variables. Langrock [16] 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.

Received 13 October 2013

REFERENCES

[1] 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. MR3131970

[2] Albert, J., and Chib, S. [1993], "Bayesian analysis of binary and polychotomous response data," *Journal of the American Statistical Association*, 88, 669–679. MR1224394

[3] Arnold, B., and Groeneveld, R. [1995], "Measuring skewness with respect to the mode," *The American Statistician*, 49, 34–38. MR1341197

[4] Basu, S., and Mukhopadhyay, S. [2000*a*], "Bayesian analysis of binary regression using symmetric and asymmetric links," *Sankhyā: The Indian Journal of Statistics, Series B*, 62, 372–379. MR1834162

[5] 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. MR1893792

[6] 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

[7] 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.

[8] 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. MR1731481

[9] Choy, S. T. B., and Chan, J. S. K. [2008], "Scale mixtures distributions in statistical modelling," *Australian & New Zeland Journal of Statistics*, 50, 135–146. MR2516871

[10] Czado, C., and Santner, T. [1992], "The effect of link misspecification on binary regression inference," *Journal of Statistical Planning and Inference*, 33, 213–231. MR1190622

[11] Czado, C., and Song, P. K. [2008], "State space mixed models for longitudinal observations with binary and binomial responses," *Statistical Papers*, 49, 691–714. MR2439014

[12] 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. MR1235581

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

[14] Jiang, X., Dey, D., Prunier, R., Wilson, A., and Holsinger, K. [2013], "A new class of flexible link function with application to spatially correlated species co-occurrence in cape floristic region," *The Annals of Applied Statistics*, 7, 2180–2204. MR3161718

[15] Kim, S., Chen, M. H., and Dey, D. K. [2008], "Flexible generalized t-link models for binary response data," *Biometrika*, 95, 93–106. MR2409717

[16] 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. MR2859846

[17] McCullagh, P., and Nelder, J. A. [1989], *Generalized Linear Models*, 2nd edn, London: Chapman and Hall. MR0727836

[18] 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.

[19] 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.

[20] Song, P. K. [2000], "Monte Carlo Kalman filter and smoothing for multivariate discrete state space models," *The Canadian Journal of Statistics*, 28, 641–652. MR1793117

[21] 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, Series B*, 64, 583–640. MR1979380

[22] Stukel, T. A. [1988], "Generalized logistic models," *Journal of the American Statistical Association*, 83, 426–431. MR0971368

[23] Wang, X., and Dey, D. K. [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

[24] Wang, X., and Dey, D. K. [2011], "Generalized extreme value regression for ordinal response data," *Environmental and Ecological Statistics*, 18, 619–634. MR2860108

[25] 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. MR0786598

Carlos A. Abanto-Valle
Departament of Statistics
Federal University of Rio de Janeiro
Caixa Postal 68530, CEP: 21945-970
Rio de Janeiro
Brazil
E-mail address: cabantovalle@im.ufrj.br

Dipak K. Dey
Department of Statistics
University of Connecticut
215 Glenbrook Rd, U-4120
Storrs, CT 06269
USA
E-mail address: dipak.dey@uconn.edu

Xun Jiang
Medical Sciences Biostatistics
Amgen
One Amgen Center Drive
Thousand Oaks, CA 91320
USA
E-mail address: xunj@amgen.com