State space models for binary responses with generalized extreme value inverse link: An approximate Bayesian approach using hidden Markov models

Carlos. A. Abanto-Valle[†], Hernán B. Garrafa-Aragón[§1] and Victor H. Lachos[‡]

[†]Department of Statistics, Federal University of Rio de Janeiro, Brazil

[§] Escuela de Ingenieria Estadistica de la Universidad Nacional de Ingenieria, Lima-Peru

[‡] Department of Statistics, University of Connecticut, USA.

Abstract

In this article binary state space mixed models (BSSMM) using a flexible skewed inverse link function based on the generalized extreme value (GEV) distribution introduced by (Abanto-Valle et al., 2015) are revisited. Commonly used probit, cloglog and loglog links are prone to link misspecification because of their fixed skewness. The GEV inverse link is flexible to fit the skewness in the response curve with a extra free shape parameter. Bayesian estimation of the parameters of BSSMM in general, and BSSMM with GEV inverse link (BSSMM-GEV) in particular, is usually regarded as challenging, since the likelihood function is a high-dimensional multiple integral. We apply a novel approach to make Bayesian inference in BSSMM-GEV feasible. First, we approximate the likelihood function by integrating out the latent states by using hidden Markov model (HMM) machinery to evaluate an arbitrarily accurate approximation of the likelihood function. Second, we get the posterior mode by using a numerical optimization routine, and third, we use importance sampling to sample from the posterior distribution of the parameters using a multivariate normal distribution with mean and variance given by the posterior mode and the inverse of the Hessian matrix evaluated at the posterior mode. However, the HMM approximation leads to a simple formula for decoding, i.e., estimating the latent process. The proposed methods are illustrated with a real dataset and the results showed that the BSSMM-GEV fits better than the traditional probit, cloglog and loglog inverse links.

Keywords: Binary time series, cloglog link, GEV link, loglog link, probit link, HMM, state space models.

[1]Correspondence to: Hernán B. Garrafa-Aragón, Escuela de Ingeniería Estadística, Facultad de Ingeniería Económica y Ciencias Sociales, Universidad Nacional de Ingeniería, Av Túpac Amaru 210-Rimac, Apartado 1301, Lima-Perú. e-mail: hgarrafa@gmail.com

1 Introduction

Binary response data with two possible outcomes are often encountered in statistical modeling. Time series of binary responses may adequately be described by generalized linear models (McCullagh and Nelder, 1989). However, these might not be adequate if the observations are correlated over time. To address the serial correlation that might be present, West et al. (1985) used generalized linear state space models in a conjugate Bayesian setup. Further studies of this topic have been conducted by Fahrmeir (1992), Song (2000), Carlin and Polson (1992), Czado and Song (2008), Abanto-Valle and Dey (2014) and Abanto-Valle et al. (2015), among others.

Consider a binary time series $\{Y_t, t = 1, \ldots, T\}$, taking the values 0 or 1 with probability of success given by π_t and which is related with a time-varying covariate vector $\mathbf{x}_t = (x_{t1}, \ldots, x_{tk})^\top$ and a $q-$ dimensional latent state variable θ_t. We consider a generalized linear state space model framework for binary responses in the following way

$$Y_t \sim \mathscr{B}er(\pi_t) \qquad t = 1, \ldots, T, \tag{1}$$

$$\pi_t = F(\mathbf{x}_t'\beta + \mathbf{S}_t'\theta_t), \tag{2}$$

$$\theta_t = \mathbf{G}_t\theta_t + \eta_t \qquad \eta_t \sim \mathscr{N}_q(\mathbf{0}, \mathbf{W}_t). \tag{3}$$

In the above setup the observed process $\{Y_t\}$ is described by equations (1)-(2), where $\pi_t = P(Y_t = 1 \mid \theta_t, \mathbf{x}_t, \mathbf{S}_t)$ is the conditional probability of success, $\mathbf{S}_t$ is a $q-$ dimensional vector, β is a $k-$ dimensional vector of regression coefficients, and $\mathbf{x}_t = (x_{t1}, \ldots, x_{tk})'$ is a $k \times 1$ vector of covariates. The system process is defined as a first-order Markov process in Equation (3), where $\mathbf{G}_t$ is the $q \times q$ transition matrix, $\mathbf{W}_t$ is the covariance matrix of error terms η_t, $\mathscr{B}er(.)$ and $\mathscr{N}_q(.,.)$ indicate the Bernoulli and $q-$dimensional normal distributions respectively. In the terminology of generalized linear models (McCullagh and Nelder, 1989), F is the inverse link function. For ease of exposition, we refer to F as the link function in this article.

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 (see, for instance, Albert and Chib, 1993; Basu and Mukhopadhyay, 2000*a*,*b*). However, as Chen and Shao (1999) argued, when the latent probability of a given binary response approaches 0 at a different rate than it approaches 1, symmetric link functions may not be adequate to fit binary data, potentially causing substantial bias in the mean response estimates (Czado and Santner, 1992). To deal with this problem some asymmetric links have been proposed in the literature. Two of the commonly adopted asymmetric link functions are the complementary loglog (cloglog) and the loglog. However, these two links have fixed skewness and lack flexibility, not allowing knowing how much skewness is incorporated.

Considerable research have been conducted trying to introduce flexibility of skewness as well as tail behavior into the link functions. For example, Stukel (1988) proposed a two-parameter class of generalized logistic models, Kim et al. (2007) used the skewed generalized t-link, and Bazán et al. (2010) adopted the skewed probit links and some variants with different parameterizations. Wang and Dey (2010), Wang and Dey (2011) and Jiang et al. (2013) introduced the flexible class of link functions as an appropriate model for binary cross sectional data. Among them, the GEV link is very flexible because of a free shape parameter, providing great flexibility in fitting a wide range of skewness values in the response curve.

On the other hand, state-space models for binary responses were used by Carlin and Polson (1992) and Song (2000) without including covariates. Czado and Song (2008) introduced covariates for binary state-space models with probit link and called the resulting class the binary state-space mixed models (BSSMM). Abanto-Valle and Dey (2014) extended the BSSMM to scale mixture of normal links and developed an ad-hoc MCMC method to sample from the posterior distribution of parameter and latent states. More recently, Abanto-Valle et al. (2015) introduced the BSSMM-GEV by using the Just Another Gibbs Sampler (JAGS) in the R package for the estimation procedure. However, the resulting MCMC algorithm has some undesirable features. In particular, the procedure is quite involved, requiring a large number of computer-intensive simulations. In addition, the computational cost increases rapidly with the sample size.

In this paper, we compare the BSSMM by assuming three standard link functions and the GEV link. We consider the logit, cloglog, loglog and and gev links. We call the corresponding binary state space mixed model as BSSMM-PROBIT, BSSMM-CLOGLOG, BSSMM-LOGLOG and BSSMM–GEV. We also, apply an alternative Bayesian estimation method to the BSSMM class of models. First, we approximate the likelihood function by integrating out the latent states, as suggested by Langrock (2011) and Langrock et al. (2012). Second, we get the maximum a posteriori by using a numerical optimization routine, and third, we use importance sampling to sample from the posterior distribution of the parameters using a multivariate normal distribution with mean and variance given by the maximum a posteriori and the inverse of the Hessian matrix, evaluated at the maximum a posteriori, respectively.

The remainder of this paper is organized as follows. Section 2 gives a brief review about the GEV distribution. Section 3 outlines the setup of the BSSMM models for the three flexible link functions as well as the likelihood approximation and evaluation procedure using HMM methods and the Bayesian approach for parameter estimation. Section 4 shows two criteria for model comparison. Section 5 is devoted to the application and model comparison of all the six models using a real dataset. Finally, some concluding remarks and suggestions for future developments are given in Section 6.

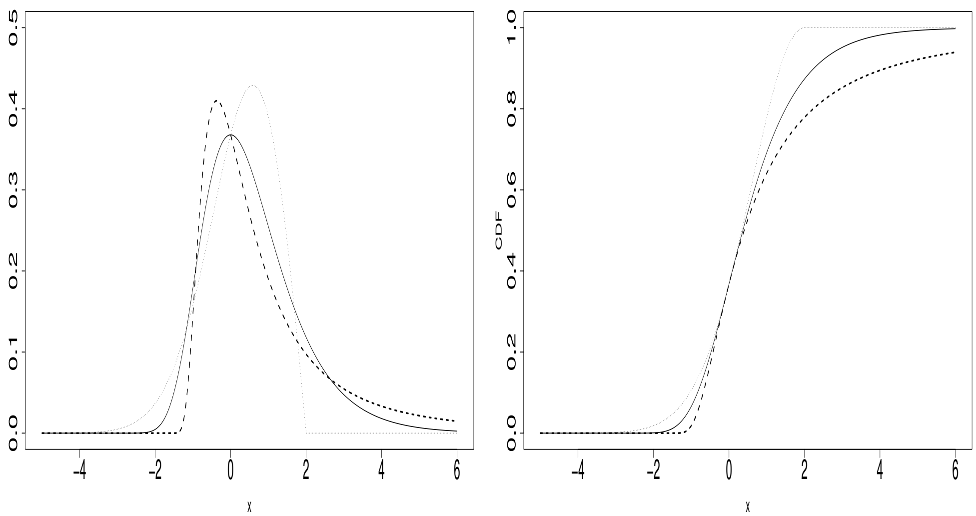

Figure 1: Left: pdf plot of GEV distribution. Right: cdf plot of GEV distribution. Solid line ($\xi = 0$), dashed line ($\xi = 0.6$), and dotted line ($\xi = -0.6$).

2 Generalized extreme value link

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

$$G(x) \quad = \quad \exp\left[-\left\{1+\xi\frac{x-\mu}{\sigma}\right\}_{+}^{-\frac{1}{\xi}}\right], \tag{4}$$

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 (4) 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 (Wang and Dey, 2010, 2011). 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 do the approach rates 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 in the GEV model, Wang and Dey (2010) and Wang and Dey (2011) extended the skewness measure of Arnold and Groeneveld (1995) to the GEV distribution in terms of its mode. Wang and Dey (2010) and Wang and Dey (2011) showed that, based on this skewness definition, the GEV distribution is negatively skewed for $X < -0.307$ and positively skewed for $X > -0.307$.

3 Binary response state-space mixed models with GEV link

3.1 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

$$Y_t \sim \mathscr{B}er(\pi_t) \qquad t = 1,\ldots,T \tag{5}$$

$$\pi_t = P(Y_t = 1 \mid \theta_t, \mathbf{x}_t, \beta) = F(\mathbf{x}_t'\beta + \theta_t) \tag{6}$$

$$\theta_t = \phi\theta_{t-1} + \tau\eta_t. \tag{7}$$

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 ofthe GEV distribution described in Section 2, so 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. We assume that η_t are independent and normally distributed with mean zero and unit variance, $\mid\phi\mid< 1$, i.e., the latent state process is stationary and $\theta_1 \sim \mathscr{N}(0, \frac{\tau^2}{1-\delta^2})$. Clearly θ_t represents a time-specific effect on the observed process.

3.2 Likelihood evaluation by iterated numerical integration

To formulate the likelihood in the BSSMM-GEV, we require the conditional distributions of the random variables y_t, given θ_t $(t = 1,\ldots,T)$, and of the random variables θ_t, given θ_{t-1} $(t = 2,\ldots,T)$. We denote these by

$p(y_t \mid \theta_t)$ and $p(\theta_t \mid \theta_{t-1})$, respectively. The likelihood of the model defined by equations (5) , (6) and (7) can then be derived as

$$\begin{aligned}
\mathscr{L} &= \int \ldots \int p(y_1, \ldots, y_T, \theta_1, \ldots, \theta_T) d\theta_T \ldots d\theta_1 \\
&= \int \ldots \int p(y_1, \ldots, y_T \mid \theta_1, \ldots, \theta_T) \\
&\times \; p(\theta_1, \ldots, \theta_T) d\theta_T \ldots d\theta_1 \\
&= \int \ldots \int p(\theta_1) p(y_1 \mid \theta_1) \\
&\times \prod_{t=2}^{T} \left[p(y_t \mid \theta_t) p(\theta_t \mid \theta_{t-1}) \right] d\theta_T \ldots d\theta_1. \qquad (8)
\end{aligned}$$

Hence, the likelihood is a higher-order multiple integral that cannot be evaluated analytically. Through numerical integration, using a simple rectangular rule based on m equidistant intervals, $B_i = (b_{i-1}, b_i)$, $i = 1, \ldots, m$, with midpoints b_i^* and length b, the likelihood can be approximated as follows:

$$\begin{aligned}
\mathscr{L} &\approx b^T \sum_{i_1=1}^{m} \ldots \sum_{i_T=1}^{m} p(\theta_1 = b_{i_1}^*) p(y_1 \mid \theta_1 = b_{i_1}^*) \\
&\times \prod_{t=2}^{T} p(\theta_t = b_{i_t}^* \mid \theta_{t-1} = b_{i_{t-1}}^*) p(y_t \mid \theta_t = b_{i_t}^*) \\
&= \mathscr{L}_{\text{approx}} . \qquad (9)
\end{aligned}$$

This approximation can be made arbitrarily accurate by increasing m, provided that the interval (b_0, b_m) covers the essential range of the latent process. We note that this simple midpoint quadrature is by no means the only way in which the integral can be approximated (cf. Langrock, 2011).

3.3 Fast evaluation of the approximate likelihood using HMM techniques

The approximate likelihood, in the form given in (9), can be evaluated numerically, but the evaluation will usually be computationally intractable since it involves m^T summands. However, instead of the brute force summation in (9), an efficient recursive scheme can be used to evaluate the approximate likelihood. To see this, we note that the numerical integration essentially corresponds to a discretization of the state space, i.e., the support of the latent process θ_t. Therefore, the approximate likelihood given in (9) can be evaluated using the well-established tools for HMMs, which are models that have exactly the same dependence structure, but with a finite and hence discrete state space (cf. Langrock, 2011; Langrock et al., 2012). In the given scenario, the discrete states correspond to the intervals B_i, $i = 1, \ldots, m$, in which the state space has been partitioned. A key property of HMM, which we exploit here, is that the likelihood can be evaluated efficiently using the so-called forward algorithm, a recursive scheme which iteratively moves forward along the time series, updating the likelihood and the state probabilities in each step (Zucchini et al., 2016). For an HMM, applying the forward algorithm results in a convenient closed-form matrix product expression for the likelihood, and this is exactly what is obtained also for the BSSMM with GEV link:

$$\mathscr{L}_{\text{approx}} = \delta \mathbf{P}(y_1)\Gamma \mathbf{P}(y_2)\Gamma \mathbf{P}(y_3)\cdots \Gamma \mathbf{P}(y_{T-1})\Gamma \mathbf{P}(y_T)\mathbf{1}^\top . \tag{10}$$

Here, the $m \times m$-matrix $\Gamma = \left(\gamma_{ij}\right)$ is the analogue to the transition probability matrix in case of an HMM, defined by $\gamma_{ij} = p(\theta_t = b_j^* \mid \theta_{t-1} = b_i^*) \cdot b$, which is an approximation of the probability of the latent process changing from some other value in the interval B_i to some value in the interval B_j, this conditional probability is determined by Eq. (7). The vector δ is the analogue to the Markov chain's initial distribution in case of an HMM, here defined such that δ_i is the density of the $\mathscr{N}(0, \frac{\tau^2}{1-\phi^2})$-distribution — the stationary distribution of the latent process — multiplied by b. Furthermore, $\mathbf{P}(y_t)$ is an $m \times m$ diagonal matrix with the ith diagonal entry $p(y_t \mid \theta_t = b_i^*)$, so it is the analogue to the matrix comprising the state-dependent probabilities in case of an HMM. This conditional probability is determined by Eq. (5). Finally, $\mathbf{1}^\top$ is a column vector of ones. Using the matrix product expression given in (10), the computational effort required to evaluate the approximate likelihood is linear in the number of observations, T, and quadratic in the number of intervals used in the discretization, m.

In practice, this means that the likelihood can typically be calculated in a fraction of a second, even for T in the thousands and say $m = 100$, a value which renders the approximation virtually exact. Furthermore, the approximation can be made arbitrarily accurate by increasing m (and potentially widening the interval $[b_0, b_m]$).

It should be noted here that, although we are using the HMM forward algorithm to evaluate the (approximate) likelihood, the specifications of δ, Γ and $\mathbf{P}(y_t)$ given above do not exactly define an HMM, since in general the row sums of Γ will only approximately equal one, and the components of the vector δ will only approximately sum to one. If desired, this can be remedied by scaling each row of Γ and the vector δ to total 1.

3.4 Bayesian inference for BSSMM with GEV link

Because we have some constraint in the original parametric space ($\beta \in \mathbb{R}^p$,$\{|\phi| < 1, \tau > 0, \mid \xi \mid < 0.6\}$) of the BSSMM-GEV, we consider transformations for the common parameters, as follows: $\psi = \log\left(\frac{1+\phi}{1-\phi}\right)$, $\omega = \log(\tau)$, and $\kappa = \log\left(\frac{0.6+\xi}{0.6-\xi}\right)$. Let $\varphi = (\beta^\top, \psi, \omega, \kappa)^\top$ and $p(\varphi)$ be the prior distribution of φ. From equation

(10), we then obtain the posterior distribution up to a normalization constant

$$p(\varphi \mid \mathbf{y}_T) \propto p(\varphi)\mathscr{L}_{\text{approx}}(\varphi). \tag{11}$$

Suposse we wish to calculate an expectation $E_{p(\varphi|\mathbf{y}_T)}[h(\varphi)]$, which can be calculated as

$$\begin{aligned} E_{p(\theta|\mathbf{y}_T)}[h(\varphi)] &= \frac{\int h(\varphi)p(\varphi \mid \mathbf{y}_T)d\varphi}{\int p(\varphi \mid \mathbf{y}_T)d\varphi} \\ &= \frac{\int \frac{h(\varphi)p(\varphi|\mathbf{y}_T)}{q(\varphi)}q(\varphi)d\varphi}{\int \frac{p(\varphi|\mathbf{y}_T)}{q(\varphi)}d\varphi} \\ &= \frac{E_{q(\varphi)}\left[h(\varphi)\omega(\varphi)\right]}{E_{q(\varphi)}\left[\omega(\varphi)\right]}, \end{aligned} \tag{12}$$

where $\varphi(\varphi) = \frac{p(\varphi|\mathbf{y}_T)}{q(\varphi)}$ and now $E_q[.]$ denotes an expected value with respect to $q(\theta)$. Therefore a sample of independent draws $\varphi_1, \ldots, \varphi_m$ from $q(\theta)$ can be used to estimate $E_{p(\varphi|\mathbf{y}_T)}[h(\varphi)]$ by

$$\bar{h} = \frac{\sum_{i=1}^{m} \omega(\varphi_i)h(\varphi_i)}{\sum_{i=1}^{m} \omega(\varphi_i)}. \tag{13}$$

It have been shown that using one sample $\varphi_i's$ in estimating the ratio in (12) is more efficient than using two samples (one for the numerator and another for denominator) (Chen et al., 2008). It follows from the strong law of large numbers that $\bar{h} \to E_{p(\theta|\mathbf{y}_T)}[h(\theta)]$ as $m \to \infty$ almost surely (Geweke, 1989). A variance of $\bar{h}(\theta)$ can be consistently estimated by $\sum_{i=1}^{m} \omega(\theta_i)^2[h(\theta_i) - \bar{h}]^2/[\sum_{i=1}^{m} \omega(\theta_i)]^2$.

4 Model comparison criteria

Given the wide range of candidate models, it has become increasingly important to be able to discriminate between competing models for a given application. Another popular metric of summary statistics for Bayesian model comparison is the deviance information criterion (DIC) proposed by Spiegelhalter et al. (2002). This criterion is based on the posterior mean of the deviance. It can be approximated by $\overline{D} = \sum_{q=1}^{Q} D(\theta_q)/Q$, where $D(\theta) = -2\log f(\mathbf{y}_T \mid \theta) = -2\log \mathscr{L}(\theta)$. The DIC can be estimated using the Monte Carlo output by $\widehat{\text{DIC}} = \overline{D} + \widehat{p_D} = 2\overline{D} - D(\bar{\theta})$, where $\widehat{p_D}$ is the effective number of parameters, and can be evaluated as $\widehat{p_D} = \bar{D} - D(\bar{\theta})$. Given the comparison of two alternative models, the model that best fits a dataset is the model with the smallest

DIC value. It is important to integrate out all latent variables in the deviance calculation, since this yields a more appropriate penalty term $\widehat{p_D}$. For all these criteria, the evaluation of the likelihood function $\mathscr{L}(\theta)$ is a key aspect. However, for the BSSMM-GEV it can be evaluated using results given in Subsection 3.2 and 3.3.

Finally, we use the Log-Predictive Score (LPS, Delatola and Griffin, 2011), which can be estimated as: $\widehat{LPS} = -\frac{1}{T}\sum_{t=1}^{T} \log p(y_t \mid \mathbf{y}_{t-1}, \bar{\theta})$.

5 Case study: Deep brain stimulation on attention reaction time

To illustrate the method developed in Section 3 applied to binary responses, we consider responses from a monkey performing the attention paradigm described in Smith et al. (2009). 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 Smith et al., 2009, for a more detailed description of the experiment). Thus our behavioral dataset for this experiment is composed of a time series of binary observations with a 1 corresponding to the reward being delivered and a 0 corresponding to the 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 2 and 3. Dividing the results into periods when stimulation was 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[2]. For this dataset we fit the binary state-space model with three standard link functions (PROBIT, CLOGLOG and LOGLOG), as well as the GEV link function defined in previous sections, where π_t is modeled by

$$\pi_t = P(Y_t = 1 \mid \theta_t) = F(\beta_0 + \beta_1 t + \theta_t).$$

As before, $F(.)$ represents the cdf associated with the corresponding standard link functions in PROBIT, CLOGLOG and LOGLOG models. In the GEV case, let $G(x)$ be the cumulative distribution function at x for the

[2] We thank Anne C. Smith for making the dataset available at her website: http://www.annecsmith.net/deepbrainstimulation.html

GEV distribution with $\mu = 0$ and $\sigma = 1$. Then, $F(\beta_0 + \beta_1 t + \theta_t) = 1 - G(-\beta_0 - \beta_1 t - \theta_t)$, where θ_t is the arousal state of the macaque monkey at time t, for $t = 1, \ldots, 1250$. Let $\varphi = (\beta^\top, \psi, \omega, \kappa)^\top$ the parameter vector of the BSSMM-GEV. We assume that the parameters are prior independent, such that $p(\varphi) = p(\beta)p(\psi)p(\omega)p(\kappa)$, and set the densities as follows: $\beta \sim \mathscr{N}_2(\mathbf{0}, 100\mathbf{I_2})$, $\psi \sim \mathscr{N}(4.5, 100)$, $\omega \sim \mathscr{N}(-1.5, 100)$, where $\mathscr{N}_2(.,.)$ and $\mathscr{N}(.,.)$ denote the bivariate normal and univariate normal distributions, respectively. All these priors are slightly flat. We impose a prior in the scale of $\xi \sim \mathscr{U}(-0.6, 0.6)$ and given the transformation $\kappa = \log\left(\frac{0.6+\xi}{0.6-\xi}\right)$, we obtain the prior of κ using the inverse transformation method. We set the common parameters of the BSSMM-PROBIT, BSSMM-CLOGLOG and BSSMM-LOGLOG as in the BSSMM-GEV case.

Now, we apply the method described in Section 3 to fit the BSSMM-PROBIT, BSSMM-CLOGLOG, BSSMM-LOGLOG and BSSMM-GEV models. All the calculations were performed using stand-alone code developed by us using the Rcpp interface in R. First, we approximated the likelihood using $b_m = -b_0 = 3$ and $m = 50, 100, 150, 200, 400$. Second, we obtained numerically the posterior mode using the *optim* routine in the R package. Table 1 reports the results for each model fitted here. It is important to observe that for m above 150 the results are almost the same. Finally, we apply the importance sampling algorithm to draw a random sample from the posterior distribution of the parameters using a multivariate normal distribution with mean and variance given by the posterior mode the inverse of the Hessian matrix evaluated at the posterior mode, using $b_m = -b_0 = 3$ and $m = 400$, respectively. For each model, we draw a sample of size 500. We introduce an extra resample step in the procedure in order to get a final sample of size 1000, which is used to calculate the moments and posterior 95% credibility intervals given in Table 2.

Table 2 shows, that for all the models considered here, the posterior means of ϕ 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 0.0068 and 0.0094. For GEV model we found that the posterior mean and 95% credibility interval for ξ are -0.5573 and (-0.5923, -0.4376). Notice that according to Section 2, this result indicates the data favor positively skewed link functions, which more closely correspond to the LOGLOG among the standard link functions we consider here.

To assess the goodness of the estimated models, we calculate the DIC and the LPS criteria described in Section 4 to compare models using different link functions. The minimum value of the DIC and LPS give the best fit. Table 3 summarizes the DIC and LPS for our four models. Both criteria select the BSSMM-GEV as the best model for DBS dataset. This confirms our observation that the data support positively skewed link functions, namely the BSSM-LOGLOG standard link, as well as BSSM-GEV with negative skewed parameter.

Table 1: Posterior mode of the parameters obtained when fitting the BSSMM-PROBIT, BSSMM-CLOGOLOG, BSSMM-LOGLOG and BSSMM-GEV respectively, for the DBS dataset (using $m = 50, 100, 150, 200, 400$ and $b_{max} = -b_{min} = 3$) and time in minutes to get the posterior mode.

BSSMM-PROBIT						
m	β_0	β_1	ψ	ω	κ	time
50	0.8764	-0.0015	5.5013	-2.4521	–	0.08
100	0.8775	-0.0015	5.4982	-2.4551	–	0.23
150	0.8775	-0.0015	5.4982	-2.4551	–	0.43
200	0.8775	-0.0015	5.4982	-2.4551	–	0.65
400	0.8775	-0.0015	5.4982	-2.4551	–	2.45
BSSMM-CLOGLOG						
m	β_0	β_1	ψ	ω	κ	time
50	2.0374	-0.0034	5.9804	-2.3497	–	0.10
100	1.1571	-0.0025	6.2162	-2.4319	–	0.34
150	0.6355	-0.0017	4.9109	-2.2532	–	0.63
200	0.6355	-0.0017	4.9109	-2.2532	–	0.94
400	0.6355	-0.0017	4.9109	-2.2532	–	3.58
BSSMM-LOGLOG						
m	β_0	β_1	ψ	ω	κ	time
50	1.8118	-0.0021	5.5862	-2.3484	–	0.11
100	1.8110	-0.0021	5.5844	-2.3499	–	0.34
150	1.8110	-0.0021	5.5844	-2.3499	–	0.65
200	1.8110	-0.0021	5.5844	-2.3499	–	0.99
400	1.8110	-0.0021	5.5844	-2.3499	–	3.59
BSSMM-GEV						
m	β_0	β_1	ψ	ω	κ	time
50	0.6141	-0.0012	4.6957	-2.3492	-2.8248	0.16
100	0.6176	-0.0012	4.6958	-2.3382	-2.8126	0.45
150	0.6182	-0.0012	4.6958	-2.3388	2.8113	0.73
200	0.6182	-0.0012	4.6958	-2.3388	2.8113	1.46
400	0.6182	-0.0012	4.6958	-2.3388	2.8113	4.17

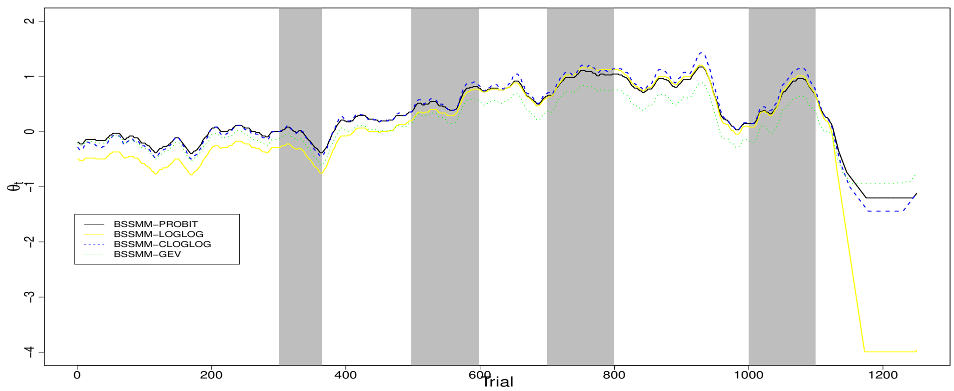

Figure 2: Estimation results for the DBS dataset. Decoded θ_t using the Viterbi algorithm. BSSMM-PROBIT:solid black line, BSSMM-LOGLOG: solid yellow line, BSSMM-GEV: dotted green line, BSSMM-CLOGLOG: doted blue line.

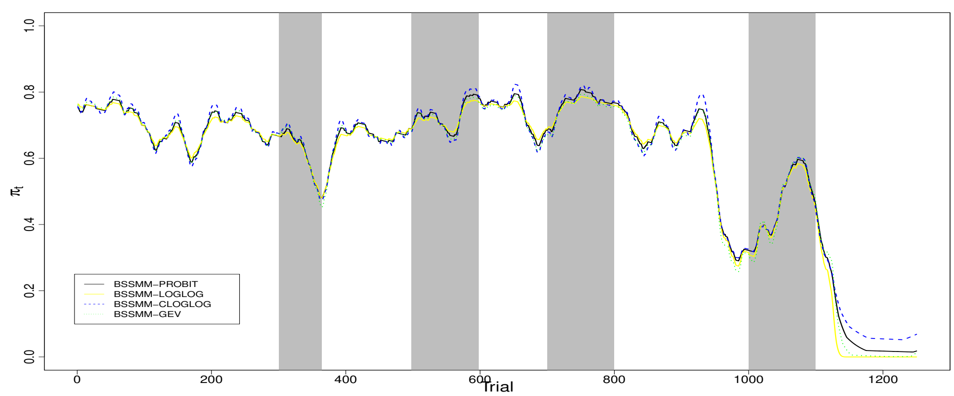

Figure 3: Estimation results for the DBS dataset. π_t obtained the decoded state θ_t using the Viterbi algorithm. BSSMM-PROBIT:solid black line, BSSMM-LOGLOG: solid yellow line, BSSMM-GEV: dotted green line, BSSMM-CLOGLOG: doted blue line

Figure 2 shows the decoded states using the Viterbi algorithm for each of the fitted models. Different line types and colors indicate the decoded states for each of the four fitted models respectively. All the estimates follow a similar pattern, but there are large differences between the estimates, especially in the last OFF period.

In Figure 3, we plot the probability of a correct response computed using the four fitted models. In this case the estimated probability is less constrained and tracks the data independently of the ON/OFF information of

Table 2: Estimation results for the DBS data set. First row: Posterior mean. Second row: Posterior 95% credible interval in parentheses.

Parameter	BSSMM-PROBIT	BSSMM-CLOGLOG	BSSMM-LOGLOG	BSSMM-GEV
β_0	0.9452	0.7948	1.8268	0.8098
	(-0.0279,1.7241)	(-0.0542,1.2372)	(0.5522,3.0517)	(-0.0529,1.1369)
β_1	-0.0017	-0.0021	-0.0022	-0.0016
	(-0.0028,-0.0003)	(-0.0028,-0.0010)	(-0.0035,-0.0008)	(-0.0020,-0.0004)
ϕ	0.9952	0.9924	0.9946	0.9898
	(0.9871,0.9981)	(0.9864,0.9952)	(0.9838,0.9988)	(0.9835,0.9954)
τ^2	0.0068	0.0084	0.0094	0.0077
	(0.0032,0.0134)	(0.0048,0.0185)	(0.0029,0.0200)	(0.0022,0.0104)
ξ	–	–	–	-0.5573
	–	–	–	(-0.5923,-0.4376)

Table 3: DBS dataset. DIC: deviance information criterion. LPS: Log-Predictive Score

Model	DIC	Rank	LPS	Rank
BSSM-PROBIT	1441.23	4	0.5737	4
BSSM-CLOGLOG	1454.04	3	0.5792	3
BSSM-LOGLOG	1433.60	2	0.5695	2
BSSM-GEV	1429.89	1	0.5694	1

the stimulation. In all the cases, on average the response curve lies around the 0.75th 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 also show that the performance does not improve during the first stimulation period. Overall, however, all the models results indicate 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).

6 Conclusions

In this article, we present an easy-to-implement Bayesian estimation approach for the BSSMM-GEV. While we focus on practical and computational aspects of fitting these models to real data, may well exist interest in deriving theoretical properties of the estimators.

We illustrated our methods through an empirical application, which showed that BSSMM-GEV model provides better model fit than the BSSMM-PROBIT, BSSMM-CLOGLOG and BSSMM-LOGLOG in terms of parameter estimates, interpretation and robustness aspects.

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 link with other skewed links, such as the skew normal or the skew Student-t. Finally, we can use the distribution function of the asymmetric Laplace distribution as a link for quantile modeling of BSSMM. Nevertheless, a deeper investigation of these modifications in the context of BSSMM models is beyond the scope of the present paper, but our results provide stimulating topics for further research.

Acknowledgements

The research of Carlos A. Abanto-Valle was supported by the CNPq. V.H. Lachos acknowledges support from Grant FAPESP grant 2018/05013-7.

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.

Abanto-Valle, C. A., Dey, D. K., and Jian, X. (2015), "Binary state space mixed models with flexible link functions: a case study on deep brain stimulation on attention reaction times," *Statistics and its Interface*, 8, 187–194.

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

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

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.

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., Huan, L., Ibrahim, J., and Kim, S. (2008), "Bayesian Variable Selection and Computation for Generalized Linear Models with Conjugate Priors," *Bayesian Analysis*, 3, 585–614.

Chen, M.-H., 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. K. (2008), "State space mixed models for longitudinal observations with binary and binomial responses," *Statistical Papers*, 49, 691–714.

Delatola, E.-I., and Griffin, J. E. (2011), "Bayesian Nonparametric Modelling of the Return Distribution with Stochastic Volatility," *Bayesian Analysis*, 6, 901–926.

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.

Geweke, J. (1989), "Bayesian Inference in Econometric Models Using Monte Carlo Integration," *Econometrica*, 57, 1317–1339.

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*, 72(4), 2180–2204.

Kim, S., Chen, M.-H., and Dey, D. K. (2007), "Flexible generalized t-link models for binary response data," *Biometrika*, 95(1), 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.

Langrock, R., MacDonald, I. L., and Zucchini, W. (2012), "Some nonstandard stochastic volatility models and their estimation usingstructured hidden Markov models," *Journal of Empirical Finance*, 19, 147–161.

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

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.

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.

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

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.

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

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.

Zucchini, W., MacDonald, I. L., and Langrock, R. (2016), *Hidden Markov Models for Time Series: An Introduction Using R, 2nd Edition*, Boca Raton, FL: Chapman & Hall.