State space models for binary responses with Generalized Extreme Value inverse link

C. A. Abanto-Valle[†1], Dipak K. Dey[§] and Xun Jiang[§]

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

[§] Department of Statistics, University of Connecticut, USA

Abstract

State space models (SSM) for binary time series using a flexible skewed inverse link function based on the generalized extreme value (GEV) distribution are introduced. Commonly used probit and logit links are prone to link misspecification because of their fixed skewness. The GEV inverse link is flexible in fitting the skewness in the response curve with a free shape parameter. Markov chain Monte Carlo (MCMC) methods for Bayesian analysis of SSM with GEV inverse link are implemented using the WinBUGS package, a freely available software. Model comparison relies on the deviance information criterion (DIC). The flexibility of the propose model 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 GEV inverse link fit better over the usual probit and logit inverse links.

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

1 Introduction

In many areas of application of statistical modeling one encounters observations that take one of two possible forms. Such binary data are often measured with covariates or explanatory variables that either continuous or discrete or categorical. Time series of binary responses may adequately be described by Generalized linear models (McCullagh and Nelder, 1989). However, if serial correlation is present or if the observations are overdispersed, these models may not be adequate, and several approaches can be taken. Generalized linear state space models also address those problems and are treated in a paper

[1]Correspondence to: Carlos A. Abanto-Valle, Department of Statistics, Federal University of Rio de Janeiro, Caixa Postal 68530, CEP: 21945-970, RJ, Brazil. e-mail: cabantovalle@im.ufrj.br

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) and Czado and Song (2008) 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 covariates vector $\mathbf{x}_t = (x_{t1}, \ldots, x_{tk})'$ 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

$$\begin{aligned} Y_t &\sim \mathscr{B}er(\pi_t) \qquad\qquad t = 1, \ldots, T & (1)\\ \pi_t &= F(\mathbf{x}_t'\beta + \mathbf{S}_t'\theta_t) & (2)\\ \theta_t &= \mathbf{H}_t\theta_{t-1} + \eta_t \qquad \eta_t \sim \mathscr{N}_q(\mathbf{0}, \mathbf{W}_t). & (3)\end{aligned}$$

In the above setup the observed process $\{Y_t\}$ is described by equations (5)-(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 $\mathbf{H}_t$ is the $q \times q$ transition matrix, $\mathbf{W}_t$ is the covariance matrix of error term η_t, $\mathscr{B}er(.)$ and $\mathscr{N}_q(.,.)$ indicate the Bernoulli and the $q-$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 problems symmetric links are widely used in the literature (Albert and Chib, 1993; Basu and Mukhopadhyay, 2000a,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 be not useful to fit binary data and asymmetric link functions must be considered. In this case 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 some asymmetric links are considered in the literature. For example, Kim et al. (2008) used the skewed generalized t-link, Bazán et al. (2010) the skewed probit links and some variants with different parameterizations and more recently Wang and Dey (2010) and Wang and Dey (2011) introduced the GEV link, as an appropriate and flexible model for the binary data and to overcome the constraint for the skewed generalized t-link models. With a free shape parameter, the GEV distribution provides great flexibility in fitting a wide range of skewness in the response curve.

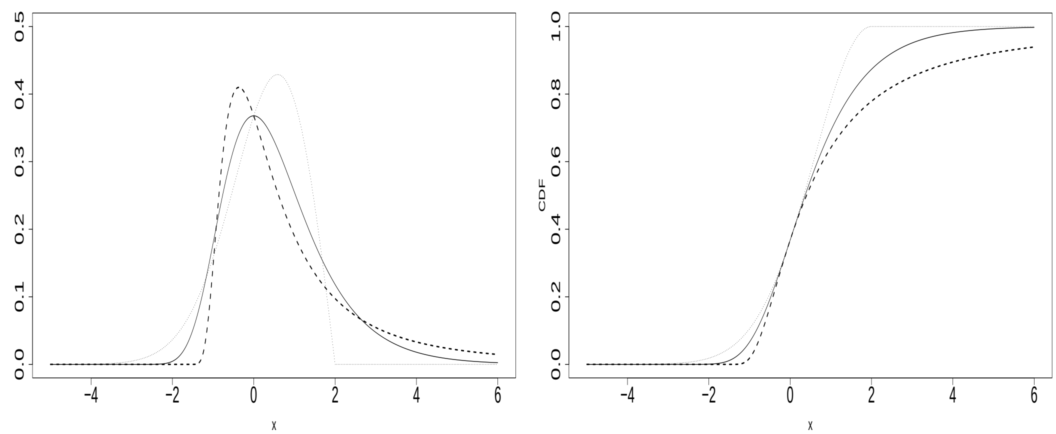

Figure 1: Left: Probability density function plots. Right: cumulative distribution function (CDF) plots of a Weibull distribution ($\mu = 0, \sigma = 1, \xi = -0.5, dashed$), Gumbel distribution ($\mu = 0, \sigma = 1, \xi = 0.0, solid$), and Fréchet distribution ($\mu = 0, \sigma = 1, \xi = 0.5, dotted$).

State space model for binary responses with probit link have been 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 as binary state space mixed models (BSSMM). More recently, Abanto-Valle and Dey (2012) extended it to normal scale mixture links.

In this paper, we extend the BSSMM by assuming the GEV distribution as a link. Inference in the BSSMM–GEV model is performed under a Bayesian paradigm via MCMC methods, which permits to obtain the posterior distribution of parameters by simulation starting from reasonable prior assumptions on the parameters. 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 WinBUGS software to implement the BSSMM–GEV model although WinBUGS uses a single-move sampler. Compared with the multiple-move sampler, the single-move sampler produces higher correlated posterior samples. However, such dependency can be compensated by running a longer Markov chain. On the other hand, the gain in efficiency in using complex sampling schemes is "largely outweighed by the ease of implementation" in WinBUGS.

The remainder of this paper is organized as follows. Section 2 gives a brief review about the GEV distribution. Section 3 outlines the BSSMM-GEV model as well as the Bayesian estimation procedure using MCMC methods. Section 4 is devoted to the application and model comparison among the logit and probit links using a real data set. Finally, some concluding remarks and suggestions for future developments are given in Section 5.

2 Generalized extreme value models

The GEV link models are based on the Generalized Extreme Value (GEV) distribution. Given a sequence of independent and identically distributed random variables $Y_1, Y_2, \ldots, Y_n$, the extreme value theory considers parametric models for the maximum $M_n = \max\{Y_1, \ldots, Y_n\}$. The exact distribution of M_n is known given a specified distribution of the Y_i's. On the other hand, in the absence of such specification, extreme value theory considers the existence of $lim_{n\to\infty} P[\{(M_n - b_n)/a_n\} \leq y] \equiv F(y)$ with two sequences of real numbers $a_n > 0$ and b_n. If $F(y)$ is non-degenerated, it belongs to either the Gumbel, the Fréchet or the Weibull class of distributions. All these three families of distributions can be expressed under a common distribution function as follows:

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

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. A more detailed discussion on the extreme value distributions can be found in Coles (2001) and Smith (2003). Extreme value analysis finds wide application in many areas, including climatology (Coles et al., 2003; Huerta and Sansó, 2007; Sang and Gelfand, 2009), environmental science (Smith, 1989; Thompson et al., 2001), financial strategy of risk management (Dahan and Mendelson, 2001; Morales, 2005), to stock returns data (Kunihama et al., 2011; Nakajima et al., 2011), and biomedical data processing (Roberts, 2000).

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). The Gumbel distribution is least positively skewed distribution in the GEV class when ξ is non-negative. Figure 1 (left) provides a plot of the probability distribution of the GEV class which shows the flexibility of such distribution. Figure 1 (right) shows the response curves with ξ equal to -0.5, 0 and 0.5. 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 (2010) and Wang and Dey (2011) extended the skewness measure of Arnold and Groeneveld (1995) for 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 $X < \log 2 - 1$ and positively skewed for $X > \log 2 - 1$.

3 Binary responses 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. As before, $\mathbf{x}_t$ is a $k \times 1$ vector of covariates. We assume that

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

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

$$\theta_t = \delta\theta_{t-1} + \tau\eta_t, \tag{7}$$

where, $G(x)$ represents the cumulative distribution function at x for the GEV distribution with $\mu = 0$ and $\sigma = 1$ and unknown shape parameter ξ. 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 $\theta_0 \sim \mathscr{N}(0, \frac{\tau^2}{1-\delta^2})$. Clearly θ_t represents a time-specific effect on the observed process. Under a Bayesian paradigm, we use MCMC methods to conduct the posterior analysis in the next subsection.

3.2 Inference procedure

A Bayesian approach to parameter estimation of the model defined by equations (5)-(7), techniques using Monte Carlo simulation via Markov Chain (MCMC) is adopted. Suppose that the model depends on a parameter vector $\Psi = (\beta', \delta, \tau^2, \xi)'$ and let $\theta_{0:T} = (\theta_0, \theta_1, \ldots, \theta_T)'$ be the latent states. Then the likelihood function $L(\Psi)$ is not easy to calculate. The Bayesian approach for estimating the parameters in the model uses the data augmentation principle, which considers $\theta_{0:T}$ as latent parameters. The joint posterior density of parameters and latent variables can be written as

$$p(\theta_{0:T}, \Psi \mid \mathbf{y}_{1:T}) \quad \propto \quad p(\mathbf{Y}_{1:T} \mid \theta_{0:T}, \Psi, \mathbf{y}_{1:T}) p(\theta_{0:T} \mid \Psi) p(\Psi), \tag{8}$$

where

$$p(\mathbf{Y}_{1:T} \mid \theta_{0:T}, \Psi) \quad = \quad \prod_{t=1}^{T} \{\pi_t^{Y_t}(1-\pi_t)^{1-Y_t}\} \tag{9}$$

$$p(\theta_{0:T} \mid \Psi) \quad = \quad \phi(\theta_0 \mid 0, \frac{\tau^2}{1-\delta^2}) \prod_{t=1}^{T} \phi(\theta_t \mid \delta\theta_{t-1}, \tau^2), \tag{10}$$

where π_t is given by equation (6) 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. We assume the prior distribution as

$$p(\Psi) \quad = \quad p(\beta)p(\delta)p(\tau^2)p(\nu).$$

The prior distributions are set as: $\beta \sim \mathcal{N}_k(\beta_0, \Sigma_0)$, $\delta \sim \mathcal{N}_{(-1,1)}(\delta_0, \sigma_\delta^2)$, $\xi \sim \mathcal{U}(-0.5, 0.5)$ and $\tau^2 \sim \mathcal{IG}(\frac{n_0}{2}, \frac{T_0}{2})$, where $\mathcal{N}_k(.,.)$, $\mathcal{N}_{(a,b)}(.,.)$,$\mathcal{U}(a,b)$ and $\mathcal{IG}(.,.)$ denote the $k-$variate normal, the truncated normal on interval (a,b), the uniform distribution on interval (a,b) and the inverse gamma distributions respectively.

We can evaluate Equation (8) using standard Monte Carlo Markov Chain methods in WinBUGS (Lunn et al., 2000). Implementation in this software merely requires specifying the model setup in equations (5)-(7), as well as priors for the unknown parameters $p(\Psi)$.

4 Application

To illustrate the technique 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 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 estimulation os 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 data set we fit the Binary state space model with GEV link (BSSM-GEV) defined by equations (5) and (7), where π_t is modeled by

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

As before, $G(x)$ represents the cumulative distribution function at x for the GEV distribution with $\mu = 0$ and $\sigma = 1$ and θ_t is the arousal state of the macaque monkey at time t. We set the priors as $\delta \sim \mathscr{N}_{(-1,1)}(0.96, 1000)$, $\tau^2 \sim \mathscr{IG}(0.1, 0.01)$ and $\xi \sim \mathscr{U}(-0.5, 0.5)$. With the same data set, we fit the probit and logit model defined by equations (5) and (7), with $\pi_t = \Phi(\theta_t)$ and $\log(\frac{\pi_t}{1-\pi_t}) = \theta_t$, respectively. We denote them by BSSM-N and BSSM-L. We implemented the three models, BSSM-N, BSSM-L and BSSM-GEV, using the software package WinBUGS, because of its user-friendly model declaration language[3]. WinBUGS 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 450000 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 50th values

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

[3] The WinBUGS code for the BSSM-GEV model is available upon request to the first author.

of the chain were stored. With the resulting 8000 values, we calculated the posterior means, the 95% credible intervals. The MCMC output of all the parameters passed the convergence test of 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 above 0.99, showing higher persistence of the autoregressive parameter for states variables and thus in binary time series. The posterior means of τ^2 are between 0.0081 and 0.0021, being the BSSM-N and BSSM-GEV less variables than the BSSM-L. We found that the posterior mean and 95% credibility interval for the shape parameter ξ are -0.4556 and (-0.4998,-0.3515), respectively. Figure 2, shows the density (right) and CDF function (left) evaluated in -0.4998 (dashed line), -0.4556 (solid line) and -0.3515 (dotted line). From Figure 2 (left), we can see differences, between the rates as them approaches to 1.

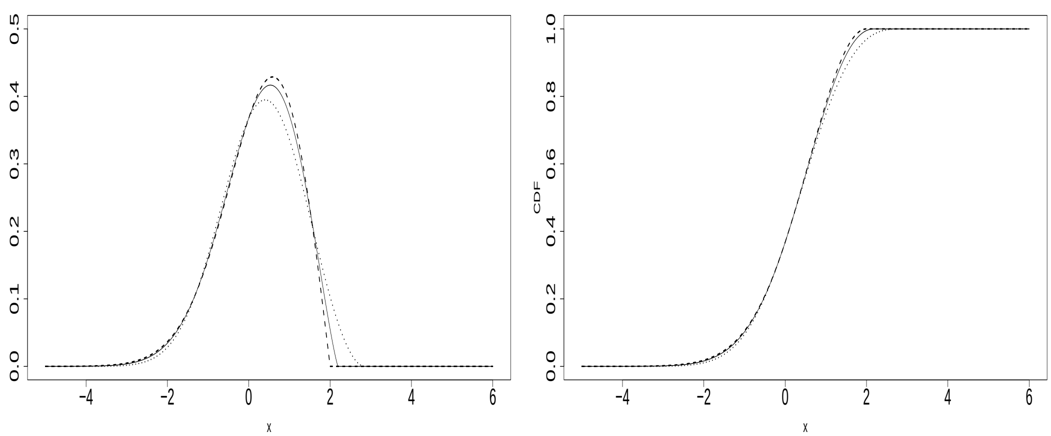

Figure 2: Left: Probability density function plots. Right: cumulative distribution function (CDF) dashed line ($\mu = 0, \sigma = 1, \xi = -0.4998$), solid line ($\mu = 0, \sigma = 1, \xi = 0.4556$), and dotted line ($\mu = 0, \sigma = 1, \xi = -0.3515$).

Figure 3 shows the posterior smoothed mean for the states θ_t for each one of the models fitted. The solid, dotted and dashed lines indicate the posterior smoothed mean for the BSSM-GEV, BSSM-N and BSSM-L, respectively. All the estimates follow a similar pattern, but there are expressive differences between the estimates, specially in the last OFF period.

In Figure 4,we plot the posterior smoothed mean for the probability of a correct response computed using the BSSM-N (dotted line), BSSM-L (dashed line) and BSSM-GEV (solid line) 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, specially with the BSSM-GEV. 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. (2009).

To assess the goodness of the estimated models, we calculate the deviance information criterion, DIC (Spiegelhalter et al., 2002) to compare models using different link functions. The DIC is easily calculated using WinBUGS. The minimum value of the DIC gives the best fit. In this context, p_D is a measure of model complexity. We compare the BSSM-N, BSSMM-L and BSSMM-GEV models. From Table 2, the DIC selects the BSSM-GEV as the best model for the monkey performance data set.

5 Conclusions

In this paper we have proposed a class of state space mixed models for longitudinal binary data using a GEV distribution as an extension of Czado and Song (2008) and Abanto-Valle and Dey (2012). In this setup, the shape parameter is estimated along with model fitting. The flexibility in links is important to avoid link misspecification. An attractive aspect of the model is that can be easily implemented, under a Bayesian perspective, via MCMC by using the WinBUGS package. We illustrated the methods

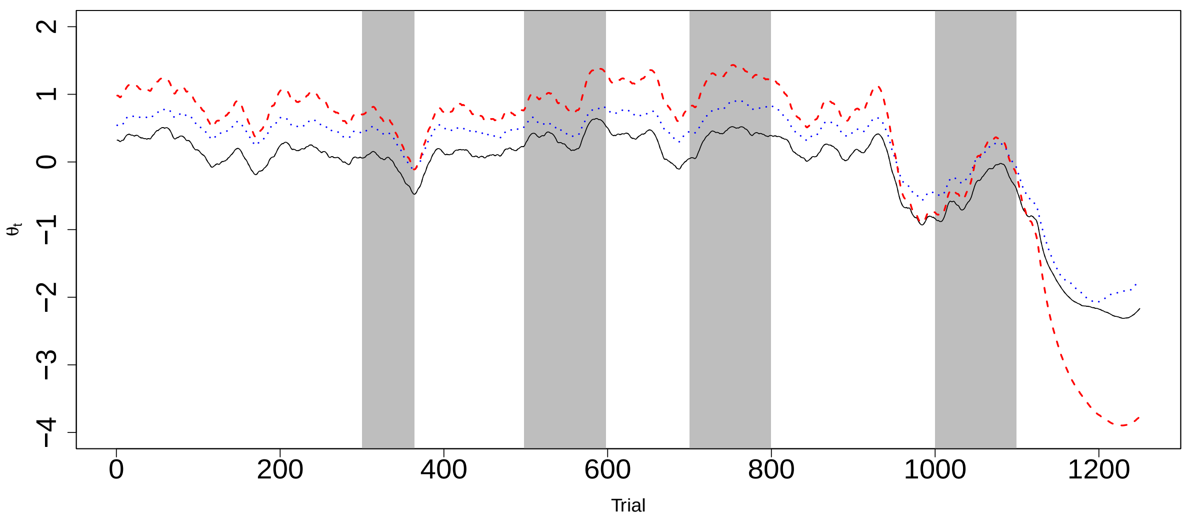

Figure 3: Estimation results for the monkey performance data set. Posterior smoothed mean of θ_t. BSSM-N: dotted line, BSSM-L: dashed line, BSSM-GEV: solid line

Table 1: Estimation results for monkey performance data set. First row: Posterior mean. Second row: Posterior 95% credible interval in parentheses.

Parameter	Model		
	BSM-N	BSSM-L	BSSM-GEV
	0.9945	0.9958	0.9936
δ	(0.9870,0.9992)	(0.9886,0.996)	(0.9832,0.993)
	0.0081	0.0211	0.0096
τ^2	(0.0043,0.0142)	(0.0091,0.0396)	(0.0047,0.0213)
–	–		-0.4558
ξ	–	–	(-0.4988,-0.3515)

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

Model	DIC	p_D	Rank
BSSM-N	1434.0	39.7	2
BSSM-L	1435.0	39.6	3
BSSM-GEV	1427.0	42.3	1

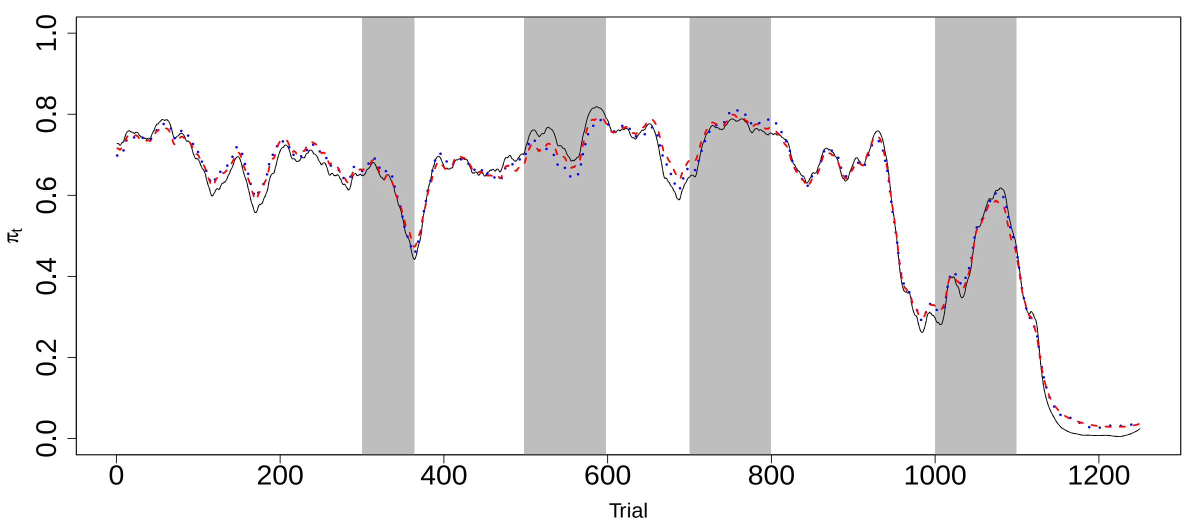

Figure 4: Estimation results for the monkey performance data set. Posterior smoothed mean of π_t. BSSM-N: dotted line, BSSM-L: dashed line, BSSM-GEV: solid line

through an empirical application with the monkey performance data set. Since the WinBUGS software automatically computes the DIC, we used it for model comparison. Empirical findings show that the BSSM-GEV model provides better model fitting than the BSSM-N and BSSM-L models.

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 note the same, we can compare the performance of the GEV link with 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. Nevertheless, a deeper investigation of those modifications is beyond the scope of the present paper, but provides stimulating topics for further research.

Acknowledgements

The research of Carlos A. Abanto-Valle was supported by the CNPq grant 202052/2011-7

References

Abanto-Valle C, Dey D. 2012. State space mixed models for binary responses with scale mixture of normal distributions links. Technical Report 22, Department of Statistics, University of Connecticut.

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

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

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

Basu S, Mukhopadhyay S. 2000b. Binary response regression with normal scale mixture links. In Dey DK, Ghosh SK, Mallick BK (eds.) *Generalized Linear Models: A Bayesian perspective.* 231–239.

Bazán JL, Bolfarine H, Branco MD. 2010. A framework for skew-probit links in binary regression. *Communications in Statistics - Theory and Methods* **39**: 678–697.

Carlin BP, Polson NG. 1992. Monte carlo bayesian methods for discrete regression models and categorical time series. In Bernardo JM, Berger JO, Dawid AP, Smith AFM (eds.) *Bayesian Statistics. Vol. 4.* 577–586.

Chen MH, Dey DK, Shao QM. 1999. A new skewed link model for dichotomus quantal response data. *Journal of the Amercian Statistical Association* **94**: 1172–1186.

Coles S, Pericchi LR, Sisson S. 2003. A fully probabilistic approach to extreme rainfall modeling. *Journal of Hydrology* **273**: 35–50.

Coles SG. 2001. *An introduction to statistical modeling of extreme values*. New York: Springer.

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

Czado C, Song PXK. 2008. State space mixed models for longitudinal observations with binary and binomial responses. *Statistical Papers* **49**: 691–714.

Dahan E, Mendelson H. 2001. An extreme-value model of concept testing. *Management Science* **47**: 102–116.

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.

Heidelberger P, Welch PD. 1983. Simulation run length control in the presence of an initial transient. *Operations Research* **31**: 1109–1144.

Huerta G, Sansó B. 2007. Time-varying models fro extreme values. *Environmental and Ecological Statistics* **14**: 285–299.

Kim S, Chen MH, Dey DK. 2008. Flexible generalized t-link models for binary response data. *Biometrika* **95**: 93–106.

Kunihama T, Omori Y, Zhang Z. 2011. Efficient estimation and particle filter for max-stable processes. *Journal of Time Series Analysis* .
URL http://dx.doi.org/10.1111/j.1467-9892.2011.00740.x

Lunn DJ, Thomas A, Best N, Spiegelhalter D. 2000. WinBUGS- a bayesian modelling frameworks: concepts, structure and extensibility. *Statistics and Computing* **10**: 325–337.

McCullagh P, Nelder JA. 1989. *Generalized linear models*. London: Chapman and Hall, 2nd ed edition.

Morales CF. 2005. *Estimation of max-stable processes using monte carlo methods with applications to financial risk assessment*. Ph.D. thesis, Department of Statistics and Operations Research, University of North Carolina, Chapel Hill.

Nakajima J, Kunihama T, Omori Y, Frühwirth-Schnatter S. 2011. Generalized extreme value distribution with time-dependence using the AR and MA models in state space form. *Journal of Computational Statistics and Data Analysis* .
URL http://dx.doi.org/10.1016/j.csda.2011.04.017

Roberts S. 2000. Extreme value statistics for novelty detection in biomedical data processing. *Science, Measurement & Technology IEE Procedures* **147**: 363–367.

Sang H, Gelfand A. 2009. Hierarchical modeling for extreme values observed over space and time. *Enviromental and Ecological Statistics* **16**: 407–426.

Smith AC, Shah SA, Hudson AE, Purpura KP, Victor JD. 2009. A bayesian statistical analysis of behavioural facilation associated deep brain stimulation. *Journal of Neurosicence Methods* **183**: 263–476.

Smith LR. 1989. Extreme value of enviromental time series: an application to trend detection in ground-level ozone. *Statistical Science* **4**: 367–393. With discussion.

Smith RL. 2003. Statistics of extremes, with applications in environment, insurance and finance. In Finkenstadt B, Rootzen H (eds.) *Extreme values in finance, telecommunications and the environment*. 1–78.

Song PK. 2000. Monte carlo kalman filter and smoothing for multivariate discrete state space models. *The Canadian Journal of Statistics* **28**: 641–652.

Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. 2002. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B* **64**: 583–640.

Thompson ML, Reynolds J, Cox LH, Guttorp P, Sampson PD. 2001. A review of statistical methods for the meteorological adjustment of tropospheric ozone. *Atmospheric Enviroment* **35**: 617–630.

Wang X, Dey DK. 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, Dey DK. 2011. Generalized extreme value regression for ordinal response data. *Environmental and Ecological Statistics* **18**: 619–634.

West M, Harrison PJ, Migon HS. 1985. Dynamic generalized linear models and bayesian forecasting. *Journal of the American Statistical Association* **136**: 209–220. With discussion.