State space mixed models for binary responses with scale mixture of normal distributions links

C. A. Abanto-Valle[†][1] **and Dipak K. Dey**[§]

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

[§] Department of Statistics, University of Connecticut

Abstract

We propose a state space mixed models for binary time series where the inverse link function is modeled to be a cumulative distribution function of the scale mixture of normal (SMN) distributions. Specific inverse links examined include the normal, Student-t, slash and the variance gamma links. We use the threshold latent approach (Albert and Chib, 1993) to represent the binary system as a linear state space model. Using a Bayesian paradigm, an efficient Markov chain Monte Carlo (MCMC) algorithm is introduced for parameter estimation. We illustrate the proposed methods with real data set. Empirical results showed that the slash inverse link fit better over the usual inverse probit link.

Keywords: Binary time series, longitudinal data, Markov chain Monte Carlo, probit, scale mixture of normal links, 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 by West et al. (1985) in a conjugate Bayesian setting. They have been subject to further research by

[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

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

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

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

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

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 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, the probit link is widely used in the literature. Albert and Chib (1993) using the data augmentation principle introduced the threshold latent approach to deal with the symmetric probit and Student-t links in a elegant way. Other symmetric links using normal scale mixture links in a nonparametric setup are described in Basu and Mukhopadhyay (2000a) and Basu and Mukhopadhyay (2000b). The binary state space model with probit link using the threshold approach (Albert and Chib, 1993) 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). They justified that including regression variables is appealing as it would enable us to quantify the relationship between the probability of success and covariates.

In this paper, we extend the BSSMM with probit link (Czado and Song, 2008) by assuming the flexible class of scale mixtures of normal (SMN) links (Lange and Sinsheimer 1993; Chow and Chan 2008) and the univariate latent states follow a first order autoregressive process. Interestingly, this rich class contains as proper elements the normal (BSSMM–N), Student-t (BSSMM–T), slash (BSSMM-S) and variance gamma (BSSMM–VG) links. All these distributions have heavier tails than the normal one, and thus can be used for robust inference in these types of models. We refer to this generalization as BSSMM–SMN. Inference in the class of BSSMM–SMN 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. Using the threshold latent approach (Albert and Chib, 1993), we simulate the latent states in an efficient way by using the simulation smoother of de Jong and Shephard (1995).

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

2 Scale mixture of normal distributions

A random variable Y belongs to the SMN family if it can be expressed as

$$Y = \mu + \kappa(\lambda)^{1/2} X, \tag{4}$$

where μ is a location parameter, $X \sim \mathcal{N}(0, \sigma^2)$, λ is a positive mixing random variable with *cdf* $H(. \mid \nu)$ and *pdf* $h(.|\nu)$, ν is a scalar or parameter vector indexing the distribution of λ and $\kappa(.)$ is a positive weight function. As in Lange and Sinsheimer (1993) and Chow and Chan (2008), we restrict our attention to the case in that $\kappa(\lambda) = 1/\lambda$. Given λ, we have $Y|\lambda \sim \mathcal{N}(\mu, \lambda^{-1}\sigma^2)$ and the pdf of Y is given by

$$f_{SMN}(y|\mu, \sigma^2, \nu) = \int_{-\infty}^{\infty} \phi(y|\mu, \lambda^{-1}\sigma^2) dH(\lambda|\nu), \tag{5}$$

where $\phi(. \mid \mu, \sigma^2)$ denotes the density of the univariate $\mathscr{N}(\mu, \sigma^2)$ distribution. From equation (5), we have that the *cdf* of the SMN distributions is given by

$$\begin{aligned} F_{SMN}(y|\mu, \sigma^2, \nu) &= \int_{-\infty}^{y} \int_{-\infty}^{\infty} \phi(u|\mu, \lambda^{-1}\sigma^2) dH(\lambda|\nu) du \\ &= \int_{-\infty}^{\infty} \Phi\left(\frac{\lambda^{1/2}[y-\mu]}{\sigma}\right) dH(\lambda|\nu), \end{aligned} \tag{6}$$

where $\Phi(.)$ is the *cdf* of the standard normal distribution. The notation $Y \sim SMN(\mu, \sigma^2, \nu)$ will be used when Y has *pdf* (5) and *cdf* (6). As was mentioned above, the SMN family constitutes a class of thick-tailed distributions including the normal, the Student-t, the Slash and variance gamma distributions, which are obtained respectively by choosing the mixing variables as: $\lambda = 1$, $\lambda \sim \mathscr{G}(\frac{\nu}{2}, \frac{\nu}{2})$, $\lambda \sim \mathscr{B}e(\nu, 1)$ and $\lambda \sim \mathscr{IG}(\frac{\nu}{2}, \frac{\nu}{2})$, where $\mathscr{G}(.,.)$, $\mathscr{B}e(.,.)$ and $\mathscr{IG}(.,.)$ denote the gamma, beta and inverse gamma distributions respectively.

3 Binary responses state space mixed models with normal scale mixture links

In this section we introduce the BSSM with SMN links using a latent variable representation in order to develop an efficient MCMC algorithm for parameter estimation.

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. According to Albert and Chib (1993), we introduce T latent variables $Z_1, \ldots, Z_T$, such that

$$Y_t = \begin{cases} 1 & Z_t > 0 \\ 0 & Z_t \leq 0 \end{cases}. \tag{7}$$

We assume that

$$\begin{aligned} \pi_t &= P(Y_t = 1 \mid \theta_t, \mathbf{x}_t, \beta) \\ &= P(Z_t > 0 \mid \theta_t, \mathbf{x}_t, \beta) = F_{SMN}(\mathbf{x}_t'\beta + \theta_t) = \int_{-\infty}^{\infty} \Phi(\lambda_t^{\frac{1}{2}}[\mathbf{x}_t'\beta + \theta_t]) dH(\lambda_t|\nu) \end{aligned} \tag{8}$$

which is the *cdf* in (6) with $\mu = 0$ and $\sigma^2 = 1$. Using the latent threshold vector $\mathbf{Z}_{1:T} = (Z_1, \ldots, Z_T)'$, we have the linear state space model with

$$
\begin{aligned}
Z_t &= \mathbf{x}_t'\beta + \theta_t + \lambda_t^{-1/2}\varepsilon_t \qquad (9)\\
\theta_t &= \delta\theta_{t-1} + \tau\eta_t \qquad (10)\\
\lambda_t &\sim p(\lambda_t|\nu), \qquad (11)
\end{aligned}
$$

where, the innovations ε_t and η_t are assumed to be mutually independent and normally distributed with mean zero and unit variance and $p(\lambda_t \mid \nu)$ is the mixing density. We assume that $\mid \delta \mid < 1$, i.e., the latent state process is stationary and $\theta_0 \sim \mathcal{N}(0, \frac{\tau^2}{1-\delta^2})$. The equations (9) and (10), conditioned on δ, the vector β and the mixing variable λ_t, represent jointly a linear state space model. 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. Conditionally to λ_t , some derivations are common to all members of the BSSMM-SMN family (see Appendix for details).

3.2 Inference procedure

A Bayesian approach to parameter estimation of the model defined by equations (9)-(11), techniques using Monte Carlo simulation via Markov Chain (MCMC) is adopted. Suppose that the model depends on a parameter vector $\Psi = (\beta', \delta, \tau^2, \nu)'$. 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 $Z_{1:T}$, $\theta_{0:T}$ and $\lambda_{1:T}$ as latent parameters. The joint posterior density of parameters and latent variables can be written as

$$
p(\mathbf{Z}_{1:T}, \theta_{0:T}, \lambda_{1:T}, \Psi \mid \mathbf{y}_{1:T}) \propto p(\mathbf{Z}_{1:T} \mid \theta_{0:T}, \lambda_{1:T}, \Psi, \mathbf{y}_{1:T}) p(\theta_{0:T} \mid \Psi) p(\lambda_{1:T} \mid \Psi) p(\Psi), \qquad (12)
$$

where

$$p(\mathbf{Z}_{1:T} \mid \theta_{0:T}, \lambda_{1:T}, \Psi, \mathbf{y}_{1:T}) = \prod_{t=1}^{T} \left[\{1(Z_t \geq 0)1(y_t = 1) + 1(Z_t < 0)1(y_t = 0)\} \phi(Z_t \mid \mathbf{x}_t'\alpha + \theta_t, \lambda_t) \right], \tag{13}$$

$$p(\theta_{0:T} \mid \Psi) = \phi(\theta_0 \mid 0, \frac{\tau^2}{1-\delta^2}) \prod_{t=1}^{T} \phi(\theta_t \mid \delta\theta_{t-1}, \tau^2), \tag{14}$$

$$p(\lambda_{1:T} \mid \Psi) = \prod_{t=1}^{T} p(\lambda_t \mid \nu), \tag{15}$$

where $1(X \in A)$ denotes an indicator function that is equal to 1 if the random variable X is contained in the set A and zero otherwise, and $p(\Psi)$ indicates the prior distribution. We assume the prior distribution as

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

For the common parameters of the BSSMM-SMN class, the prior distributions are set as: $\beta \sim \mathcal{N}_k(\beta_0, \Sigma_0)$, $\delta \sim \mathcal{N}_{(-1,1)}(\delta_0, \sigma_\delta^2)$ and $\tau^2 \sim \mathcal{IG}(\frac{n_0}{2}, \frac{T_0}{2})$, where $\mathcal{N}_k(.,.)$, $\mathcal{N}_{(a,b)}(.,.)$, $\mathcal{IG}(.,.)$ denote the $k-$variate normal, the truncated normal on interval (a,b) and the inverse gamma distributions respectively. The $p(\nu)$ is specified for each member of the BSSMM-SMN class.

As the posterior distribution in (12) is intractable analytically, we draw random samples of Ψ, $\mathbf{Z}_{1:T}$, $\lambda_{1:T}$ and $\theta_{0:T}$ from their full conditional distributions using the Gibbs sampling. The sampling scheme is described by the following algorithm:

Algorithm 1

1. Set $i = 0$ and get starting values for the parameters $\Psi^{(i)}$ and the latent variables $\theta_{0:T}^{(i)}$, $\lambda_{1:T}^{(i)}$;
2. Draw $\mathbf{Z}_{1:T}^{(i+1)} \sim p(\mathbf{Z}_{1:T} | \theta_{0:T}^{(i)}, \lambda_{1:T}^{(i)}, \Psi^{(i)}, \mathbf{y}_{1:T})$;
3. Draw $\theta_{0:T}^{(i+1)} \sim p(\theta_{0:T} | \lambda_{1:T}^{(i)}, \Psi^{(i)}, \mathbf{Z}_{1:T}^{(i+1)}, \mathbf{y}_{1:T})$;
4. Draw $\lambda_{1:T}^{(i+1)} \sim p(\lambda_{1:T} | \theta_{1:T}^{(i+1)}, \Psi^{(i)}, \mathbf{Z}_{1:T}^{(i+1)}, \mathbf{y}_{1:T})$;
5. Draw $\Psi^{(i+1)} \sim p(\Psi | \theta_{1:T}^{(i+1)}, \lambda_{1:T}^{(i+1)}, \mathbf{Z}_{1:T}^{(i+1)}, \mathbf{y}_{1:T})$;
6. Set $i = i + 1$ and return to step 2 until achieving convergence.

Cycling through 2 to 5 is a complete sweep of this sampler. The MCMC sampler will require us to perform many thousands of sweeps to generate samples from the posterior distribution $p(\mathbf{Z}_{1:T}, \theta_{0:T}, \lambda_{1:T}, \Psi \mid \mathbf{y}_{1:T})$. Details on the full conditionals of Ψ, the mixing variables $\lambda_{1:T}$, the threshold variables $\mathbf{Z}_{1:T}$ are given in the Appendix. Conditional on $\lambda_{1:T}$ equations (9)-(11) define a linear state space model, we sample the latent states $\theta_{0:T}$ in step 3 of Algorithm 1 using the simulation smother of de Jong and Shephard (1995).

3.3 Out-of-sample forecasting

We have that $K-$step ahead prediction density can be calculated using the composition method through the following recursive procedure:

$$
\begin{aligned}
p(Z_{T+K} \mid \mathbf{y}_{1:T}) &= \int \Big[p(Z_{T+K} \mid \lambda_{T+K}, \theta_{T+K}, \Psi) \\
&\times \; p(\theta_{T+K} \mid \Psi, \mathbf{y}_{1:T}) p(\lambda_{T+K} \mid \Psi) p(\Psi \mid \mathbf{y}_{1:T}) \Big] d\theta_{T+K} d\lambda_{T+K} d\Psi, \\
p(\theta_{T+K} \mid \Psi, \mathbf{y}_{1:T}) &= \int \Big[p(\theta_{T+K} \mid \Psi, \theta_{T+K-1}) p(\theta_{T+K-1} \mid \Psi, \mathbf{y}_{1:T}) \Big] d\theta_{T+K-1}.
\end{aligned}
$$

Evaluation of these integrals is straightforward, by using Monte Carlo approximation. To initialize a recursion, we use N draws $\{\theta_T^{(i)}, \lambda_T^{(i)}, \Psi^{(i)}\}_{i=1}^N$ from the MCMC sample. Then given these N draws, sample N draws $\theta_{T+k}^{(i)}$ from $p(\theta_{T+k} \mid \Psi^{(i)}, \theta_{T+k-1}^{(i)})$ and $\lambda_{T+k}^{(i)}$ from $p(\lambda_{T+k} \mid \Psi^{(i)})$ for $i = 1, \ldots, N$ and $k = 1, \ldots, K$, by using equations (10) and (11), respectively. Finally, sample N draws $\{\mathbf{y}_{T+k}^{(i)}\}_{i=1}^N$ from $p(Z_{T+k} \mid \lambda_{T+k}, \theta_{T+k}^{(i)}, \Psi^{(i)})$. With draws from Z_{T+k}, θ_{T+k} and λ_{T+k} the conditional probability π_{T+k} can be calculated easily.

4 Application

A binary time series of infant sleep status were recorded in a 120 min electroencephalographic (EEG) sleep pattern study (Sttoffer et al., 1998). Careful consideration should be given to the lability of state and the disruption of the expected rapid eye movement (REM) and non-REM components of the neonatal or infant sleep cycle. So, here it is considered that $Y_t = 1$ if during minute t the infant was judged to be in REM sleep cycle and otherwise $Y_t = 0$. Two time-varying covariates are considered. Let x_{t1} be the number of body movements during the minute t and x_{t2} the number of body movements due not to

Table 1: Estimation results for the infant sleep data set. First row: Posterior mean. Second row: Posterior 95% credible interval in parentheses. Third row: MC error. Fourth row: CD statistics. Fifth row: Inefficiency factors.

Parameter	BSSMM-N	BSSMM-T	BSSMM-S	BSSMM-VG
	-0.0479	-0.0281	-0.0550	0.3185
	(-1.9425,1.6770)	(-2.3300,2.5256)	(-2.2316,2.0157)	(-1.7192,4.4206)
β_0	0.0483	0.0944	0.0842	0.1764
	1.39	1.43	0.68	1.25
	6.04	15.08	14.03	31.43
	0.2706	0.4336	0.3134	0.2775
	(-0.0833,0.6245)	(-0.0837,1.1298)	(-0.0942,0.7338)	(-0.0791,0.6670)
β_1	0.0044	0.0115	0.0056	0.0054
	0.00	-0.32	0.37	-1.27
	1.12	2.97	1.33	1.29
	-0.4679	-0.6755	-0.5196	-0.4647
	(-0.9525,-0.0257)	(-1.5048,-0.0866)	(-1.0588)-0.0372)	(-0.9336,-0.0207)
β_2	0.0056	0.0131	0.0068	0.0060
	0.11	0.28	-0.72	1.29
	1.12	2.62	1.33	1.30
	0.9336	0.9402	0.9349	0.9424
	(0.8333,0.9868)	(0.8528,0.9890)	(0.8341,0.9887)	(0.8512,0.9922)
δ	0.0011	0.0011	0.0012	0.0019
	-0.72	0.21	-0.46	-0.18
	1.67	2.25	1.72	5.47
	0.3060	03423	0.3566	0.2789
	(0.0663,1.0169)	0.0628,1.1807)	(0.0667,1.2465)	(0.0512,0.9980)
τ^2	0.0149	0.0206	0.0226	0.0205
	0.53	-1.27	0.14	1.51
	7.04	6.82	6.85	11.66
	–	7.5621	6.4002	6.2396
	–	(2.0410,30.6290)	(2.0310,13.2640)	(1.7800,12.8040)
ν	–	0.4931	0.1265	0.1118
	–	-0.56	-1.74	1.65
	–	9.19	3.94	3.15

sucking during minute t. As in Czado and Song (2008) our objective is to investigate wether or not the probability of being in the REM sleep status is significantly related to the two types of body movements x_{t1} or x_{t2}. Our analysis differs from them in the sense that we compare the fit of the probit (BSSMM-N), the Student-t (BSSMM-T), slash (BSSMM-S) and variance gamma (BSSMM-VG) links. So, from equation (8), we have that the conditional probability of succes is given by

$$\pi_t = P(Y_t = 1 \mid \beta, \theta_t, \lambda_t) = F_{MSN}(\beta_0 + \beta_1 x_{t1} + \beta_1 x_{t2} + \theta_t),$$

and the state equation is an AR(1) process as in equation (10). We set the prior as: $\delta \sim \mathscr{N}_{-1,1}(0.95, 100)$, $\tau^2 \sim \mathscr{IG}(2.5, 0.125)$ and $\beta \sim \mathscr{N}_3(\boldsymbol{\beta}_0, \Sigma_0)$, where $\boldsymbol{\beta}_0 = \mathbf{0}$ and $\Sigma_0 = 500^2\mathbf{I}_3$, $\mathbf{0}$ indicates a 3×1 vector os zeros and $\mathbf{I}_3$ the identity matrix of order 3. The prior distributions on the shape parameters were chosen as: $\nu \sim \mathscr{G}(5.0, 0.8)$ for the BSSMM-S and BSSMM-VG models, respectively. In BSSMM-T, we assume a non-informative prior as in Fonseca et al. (2008). The priors' mean and variance for δ are respectively, 0.0032 and 0.3328. So, this prior is equivalent to the uniform distribution on interval $(-1, 1)$, which gives zero mean and variance of 0.3333. All the calculations were performed running stand-alone code developed by us using an open source C++ library for statistical computation, the Scythe statistical library (Pemstein et al., 2011), which is available for free download at `http://scythe.wustl.edu`.

We fit the BSSMM-N, BSSMM-T, BSMM-S and BSSMM-VG models. For each case, we conducted the MCMC simulation for 50000 iterations. In all the cases, the first 10000 draws were discarded as a burn-in period. In order to reduce the autocorrelation between successive values of the simulated chain, only every $20th$ values of the chain were stored. With the resulting 2000 values, we calculated the posterior means, the 95% credible intervals, Monte Carlo errors and the convergence diagnostic (CD) statistics (Geweke, 1992). If the sequence of the recorded MCMC output is stationary, it converges in distribution to the standard normal. According to the CD the null hypothesis that the sequence of 2000 draws is stationary was accepted at the 5% level, CD $\in (-1.96, 1.96)$, for all the parameters in all the models considered here. Table 1 summarizes the results. The inefficiency factor is defined by $1 + \sum_{s=1}^{\infty} \rho_s$ where ρ_s is the sample autocorrelation at lag s. It measures how well the MCMC chain mixes (see, e.g, Kim et al., 1998). It is the estimated ratio of the numerical variance of the posterior sample mean to the variance of the sample mean from uncorrelated draws. When the inefficiency factor is equal to m, we need to draw MCMC samples m times as many as the number of uncorrelated samples. From Table 1 examining the inefficiency coeficients, we found that our algorithm produces a good mixing of the MCMC chain. As expected, Figures 1 and 2 show a rapid decay of autocorrelations for all the parameters for the BSSMM-S on the infant sleep data sets.

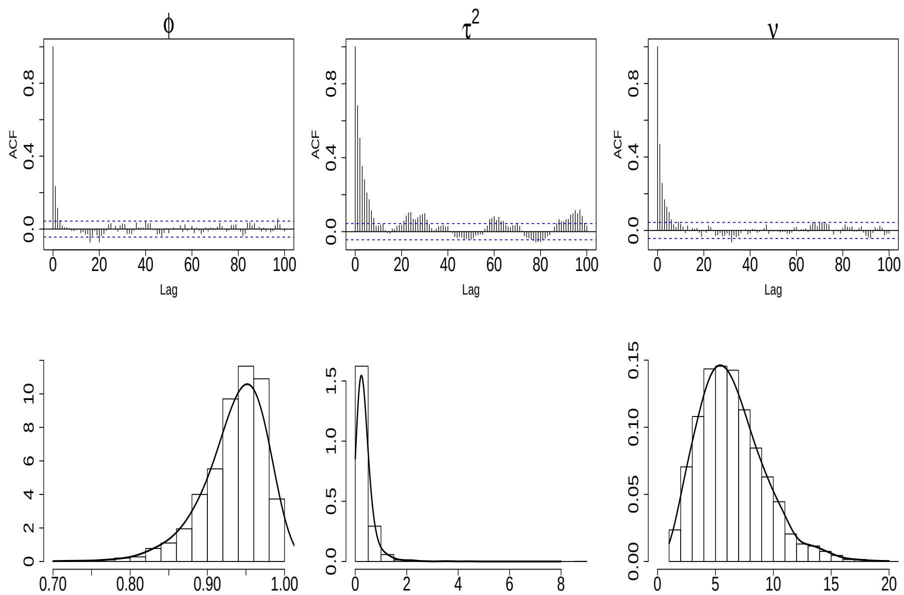

Figure 1: Estimation results for the infant sleep status data set (BSSMM-S). The top row shows plots of sample autocorrelations and the bottom row shows posterior histograms and overlayed density estimates, left: δ, middle: τ^2 and right:ν.

For all the models, the posterior means of δ are above 0.93, showing higher persistence of the autoregressive parameter for states variables and thus in binary time series. The posterior means τ^2 are between 0.27 and 0.34. This values are consistent with the results in Czado and Song (2008). The magnitude of the tail fatness is measured by the shape parameter ν in the BSSMM-T, BSSMM-S and BSSMM-VG models. In each case, the posterior means of ν are almost 7.5, 6.4 and 6.3, respectively. These results seem to indicate that the measurement errors of the Z_t threshold variables are better explained by heavy-tailed distributions, as a consequence the corresponding links could be more convenient than the normal probit link.

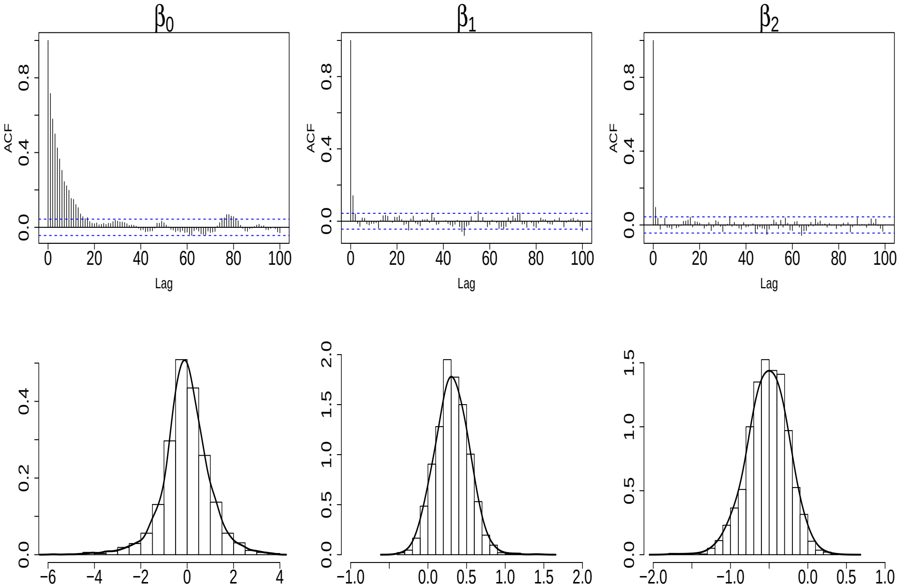

Figure 2: Estimation results for the infant sleep status data set (BSSMM-S). The top row shows plots of sample autocorrelations and the bottom row shows posterior histograms and overlayed density estimates, left: β_0, middle: β_1 and right:β_2

From Table 1, we found empirically that the influence of the number of body movements (x_1) is marginal, since the corresponding 95% credible interval for β_1 contains the zero value. On the other hand, the influence of the number of body movements not due to sucking (x_2) is detected to be statistically significant. The negative value of the posterior mean for β_2 shows that a higher number of body movements not due to sucking will reduce the probability of the infant being in REM sleep.

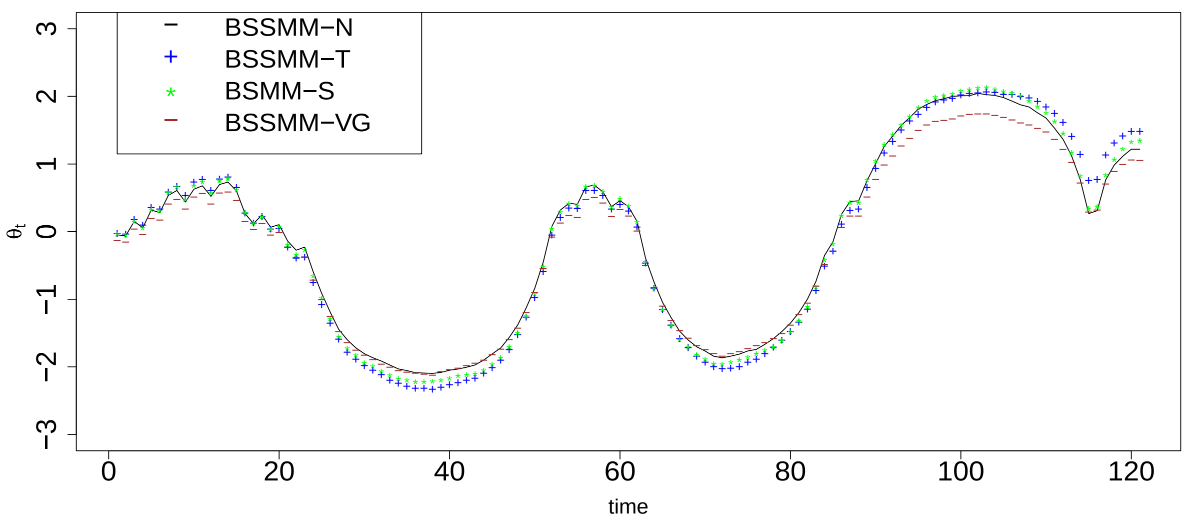

Figure 3: Estimation results for the infant sleep status data set. Posterior smoothed mean of θ_t. BSSMM-N: solid line, BSSM-T: $+$, BSSMM-S:$*$ and BSSM-VG: $-$

Figure 3 shows the posterior smoothed mean of states variables, θ_t. The states could be considered as an underlying continuous "sleep state". We found some differences between the fit of the different models, but in general the results are according with Czado and Song (2008).

To assess the goodness of the estimated models, we calculate the deviance information criterion, DIC

(Spiegelhalter et al., 2002; Celeux et al., 2006). The minimum value of the DIC gives the best fit. In this context, p_D is a measure of model complexity. We compare the BSSMM-N, BSSMM-T, BSSMM-S and BSSMM-VG models. From Table 2, the DIC selects the BSSMM-S model as the best model.

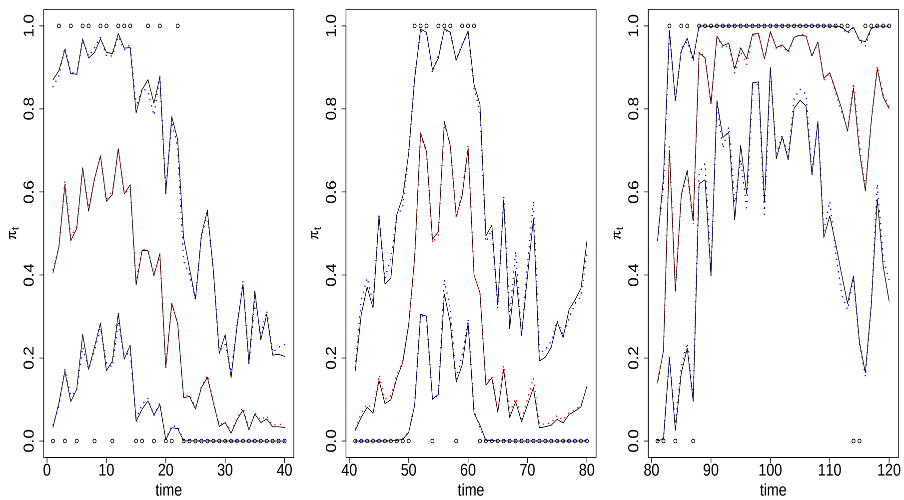

Figure 4: Posterior smoothed mean of the conditional probability and 95% credibility limits for π_t for BSSMM-N (solid line) and BSSMM-S (doted line). The points are the true observations

Figure 4 shows the posterior smoothed mean and 95% credibility intervals of π_t based on the MCMC output for the BSSMM-N (solid line) and BSSMM-S (dotted line) links. We divide the entire data in three panels with the objective to better visualization of the results. Both models show similar smoothed probabilities.

We investigate the predictive ability of the BSSMM-N and BSSMM-S models. For this purpose, we divide our data set in an initial group with 110 observations and in the second group, we left the next

Table 2: Data infat sleep: Model Comparison.

	DIC	p_D	Rank
BSSMM-N	100.17	18.52	4
BSSMM-T	99.18	18.47	2
BSSMM-S	94.76	16.86	1
BSSMM-VG	99.35	18.57	3

10 observations for out-of-sample forecast. We fit the BSSMM-N and BSSMM-S models, parameter estimation for this data set is not reported here. We simulate out-of-sample states using the methods described in section 3.3. As by product of the MCMC simulation, we can found the conditional probability of success π_{110+k} for the next 10 observations for each model considered. We report the posterior mean an 95% credibility intervals results in Figure 5. The solid line corresponds to BSSMM-N and the dotted line to the BSSMM-S. Both of models show a similar behavior. It is also important to emphasize that in general, we do not advocate the use of the BSSMM-SMN models in all situations but recommend using the models discussed here to assess the robustness of the conclusions, replacing the normal assumption with a more flexible model if this provides a more appropriate analysis.

5 Conclusions

In this paper we proposed a class of state space mixed models for longitudinal binary data using mixture of scale normal links as an extension of Czado and Song (2008). The models include both deterministic and random predictors. We studied three specific sub-classes, viz.the Student-t, slash and the variance gamma links. Under a Bayesian perspective, we constructed an algorithm based on Markov chain Monte Carlo(MCMC) simulation methods to estimate all the parameters and latent quantities of our proposed BSSMM-SMN links. The latent states are efficiently simulated using the simulation smoother of de Jong and Shephard (1995). Accordingly, with the DIC criteria all the BSSMM-SMN links fit better than the normal probit link and the BSSMM-S, the best fit.

This article makes certain contributions, but several extensions are still possible. First, we focus on

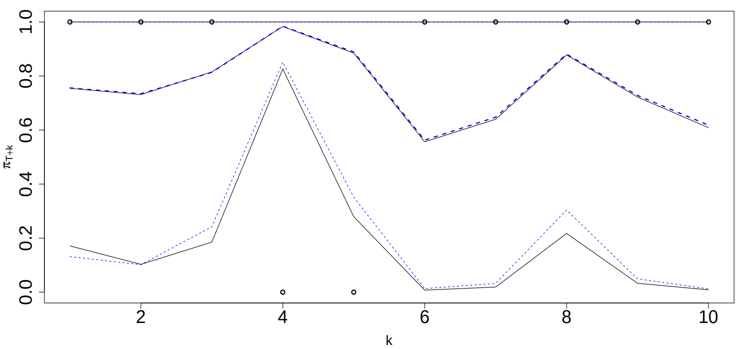

Figure 5: Posterior mean of the k step forecast probability and the 95% credibility limits for π_{T+k} for BSSMM-N (solid line) and BSSMM-S (doted line). The points are the true observations

symmetrical links, but if the rate of zeros or ones are note the same, skewed links as the skew normal or the skew Student-t are good alternatives. Nevertheless, a deeper investigation of those modifications is beyond the scope of the present paper, but provides stimulating topics for further research.

Appendix : The Full conditionals

In this appendix, we describe the full conditional distributions for the parameters, the threshold variables $\mathbf{Z}_{1:T}$ and the mixing variables $\lambda_{1:T}$ of the BSSMM-SMN class.

Full conditional distribution of β, δ and τ^2

For the common parameters of the BSSMM-SMN class, the prior distributions are set as: $\beta \sim \mathscr{N}_k(\beta_0, \Sigma_0)$, $\delta \sim \mathscr{N}_{(-1,1)}(\delta_0, \sigma_\delta^2)$ and $\tau^2 \sim \mathscr{IG}(\frac{T_0}{2}, \frac{S_0}{2})$. So, the full conditional of β is given by

$$\beta \mid \mathbf{Z}_{1:T}, \theta_{0:T}, \lambda_{1:T} \quad \sim \quad \mathscr{N}_k(\beta_1, \Sigma_1), \tag{16}$$

where $\Sigma_1 = [\Sigma_0^{-1} + \sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t' \lambda_t]^{-1}$ and $\beta_1 = \Sigma_1 [\Sigma_0^{-1} \beta_0 + \sum_{t=1}^{T} \mathbf{x}_t (Z_t - \theta_t) \lambda_t]$.

We have the following full conditional for δ is given by

$$p(\delta \mid \theta_{\mathbf{0:T}}, \tau^{\mathbf{2}}) \quad \propto Q(\delta) \exp\{-\tfrac{a_\delta}{2\tau^2}(\delta - \tfrac{b_\delta}{a_\delta})^2\} \mathbb{I}_{|\delta|<1}, \tag{17}$$

where $Q(\delta) = \sqrt{1-\delta^2} \exp\{-\frac{1}{2\tau^2}[(1-\delta^2)\theta_0^2\}$, $a_\delta = \sum_{t=1}^{T} \theta_{t-1}^2 + \frac{\tau^2}{\sigma_\delta^2}$, $b_\delta = \sum_{t=1}^{T} \theta_{t-1}\theta_t + \delta_0 \frac{\tau^2}{\sigma_\delta^2}$ and $\mathbb{I}_{|\delta|<1}$ is an indicator variable. As $p(\delta \mid \theta_{0:T}, \tau^2)$ in (17) does not have closed form, we sample it by using the Metropolis-Hastings algorithm with truncated $\mathcal{N}_{(-1,1)}(\frac{b_\delta}{a_\delta}, \frac{\tau^2}{a_\delta})$ as the proposal density.

Finally, the full conditional of τ^2

$$\tau^2 \mid \theta_{0:T}, \delta \quad \sim \quad \mathcal{IG}(\frac{T_1}{2}, \frac{S_1}{2}), \tag{18}$$

where $T_1 = T_0 + T + 1$ and $S_1 = S_0 + [(1-\delta^2)\theta_0^2] + \sum_{t=1}^{T} (\theta_t - \delta\theta_{t-1})^2$.

Full conditional of $\mathbf{Z}_{1:T}$

Since the latent variable $\mathbf{Z}_{1:T}$ are conditional independent given $\theta_{0:T}$ and $\lambda_{1:T}$, we have that the full conditional distribution of $\mathbf{Z}_{1:T}$ is given by

$$\begin{aligned} p(\mathbf{Z}_{1:T} \mid \mathbf{Y}_{1:T}, \theta_{0:T}, \lambda_{1:T}, \beta) &= \prod_{t=1}^{T} p(Z_t \mid \theta_t, \lambda_t, \beta) \\ &= \prod_{t=1}^{T} \left[\{1(Z_t \geq 0)1(y_t = 1) + 1(Z_t < 0)1(y_t = 0)\} \phi(Z_t \mid \mathbf{x}_t'\alpha + \theta_t, \lambda_t) \right], \end{aligned} \tag{19}$$

where $\phi(x \mid \mu, \sigma^2)$ indicates the density of the normal density with mean μ and variance σ^2. So, A density $p(Z_t \mid \theta_t, \lambda_t, \beta)$ is a truncate normal density according with the value of Y_t.

Full conditional of λ_t and ν

• **BSSMM-T case** As $\lambda_t \sim \mathcal{G}(\frac{\nu}{2}, \frac{\nu}{2})$, then $\lambda_t \mid Z_t, \theta_t, \beta, \nu \sim \mathcal{G}(\frac{\nu+1}{2}, \frac{[Z_t - \mathbf{x}_t'\beta - \theta_t]^2 + \nu}{2})$. An important modeling assumption is the regularization penalty $p(\nu)$ on the tail thickness. A default Jeffreys' prior was developed by Fonseca et al. (2008), with a number of desirable properties particularly when learning a fat-tail from a finite data set. The default Jeffreys's prior for ν takes the form

$$p(\nu) \quad \propto \quad \left(\frac{\nu}{\nu+3}\right)^{\frac{1}{2}} \left\{ \psi'\left(\frac{\nu}{2}\right) - \psi'\left(\frac{\nu+1}{2}\right) - \frac{2(\nu+3)}{\nu(\nu+1)^2} \right\}^{\frac{1}{2}}, \tag{20}$$

where $\psi'(a) = \frac{d\{\psi(a)\}}{da}$ and $\psi(a) = \frac{d\{\log\Gamma(a)\}}{da}$ are the trigamma and digamma functions, respectively. The interesting feature of this prior is its behavior as ν goes to infinity and it has polynomial tails of the form $p(\nu) \propto \nu^{-4}$. In this case, the tail of the prior decays rather fast for large values of ν and assessing the degree of tail thickness can require prohibitively large samples. Then, the full conditional of ν is

$$p(\nu \mid \lambda_{1:T}) \quad \propto \quad p(\nu) \frac{\left[\frac{\nu}{2}\right]^{\frac{T\nu}{2}} \mathrm{e}^{-\frac{\nu}{2}[\sum_{t=1}^{T}(\lambda_t - \log\lambda_t)]}}{[\Gamma(\frac{\nu}{2})]^T} \mathbb{I}_{2<\nu\leq 40}. \tag{21}$$

As (21) does not have closed form, we sample ν by using the Metropolis-Hastings algorithm. The proposal density used are $\mathscr{N}_{(2<\nu<40)}(\mu_\nu, \tau_\nu^2)$, with $\mu_\nu = x - \frac{q'(x)}{q''(x)}$ and $\tau_\nu^2 = (-q''(x))^{-1}$, where x is the value of the previous iteration, $q(.)$ is the logarithm of the conditional posterior density, and $q'(.)$ and $q''(.)$ are the first and second derivatives respectively.

• **BSSMM-S case**

Using the fact that $\lambda_t \sim \mathscr{B}e(\nu, 1)$, then $\lambda_t \mid Z_t, h_t, \theta \sim \mathscr{G}_{(0<\lambda_t<1)}(\nu + \frac{1}{2}, \frac{1}{2}[Z_t - \mathbf{x}_t'\beta - \theta_t]^2)$, the right truncated gamma distribution. Assuming that a prior distribution of $\nu \sim \mathscr{G}(a_\nu, b_\nu)$, the full conditional distribution of ν is $\mathscr{G}_{\nu>1}(T + a_\nu, b_\nu - \sum_{t=1}^{T}\log\lambda_t)$, i.e., the left truncated gamma distribution.

• **BSSMM-VG case**

As $\lambda_t \sim \mathscr{I}\mathscr{G}(\frac{\nu}{2}, \frac{\nu}{2})$, the full conditional of λ_t is given by

$$p(\lambda_t \mid Z_t, h_t, \nu) \quad \propto \lambda_t^{-\frac{\nu}{2}+\frac{1}{2}-1} \mathrm{e}^{-\frac{1}{2}(\lambda_t[Z_t - \mathbf{x}_t'\beta - \theta_t]^2 + \frac{\nu}{\lambda_t})}, \tag{22}$$

which is the generalized inverse gaussian distribution, $\mathscr{G}\mathscr{I}\mathscr{G}(-\frac{\nu}{2} + \frac{1}{2}, [Z_t - \mathbf{x}_t'\beta - \theta_t]^2, \nu)$.

We assume the prior distribution of ν as $\mathscr{G}(a_\nu, b_\nu)\mathbb{I}_{0<\nu\leq 40}$. Then, the full conditional of ν is

$$p(\nu \mid \mathbf{y}_{1:T}, \mathbf{h}_{0:T}, \lambda_{1:T}) \quad \propto \quad \frac{\left[\frac{\nu}{2}\right]^{\frac{T\nu}{2}} \nu^{a_\nu - 1} \mathrm{e}^{-\frac{\nu}{2}\sum_{t=1}^{T}[(\frac{1}{\lambda_t} + \log\lambda_t) + 2b_\nu]}}{[\Gamma(\frac{\nu}{2})]^T} \mathbb{I}_{0<\nu\leq 40}. \tag{23}$$

Thus, we sample ν by using the Metropolis-Hastings algorithm as in the case of the BSSMM-T model with proposal density $\mathscr{N}_{(0,40)}(\mu_\nu, \tau_\nu^2)$.

Acknowledgements

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

References

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

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, K GS, K MB (eds.) *Generalized Linear Models: A Bayesian perspective.* 231–239.

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.

Celeux G, Forbes F, Robert CP, Titterington DM. 2006. Deviance information criteria for missing data models (with discussion). *Bayesian Analysis* **4**: 651–705.

Chow STB, Chan JSK. 2008. Scale mixtures distributions in statistical modelling. *Australian & New Zeland Journal of Statistics* **50**: 135–146.

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

de Jong P, Shephard N. 1995. The simulation smoother for time series models. *Biometrika* **82**: 339–350.

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.

Fonseca TCO, Ferreira MAR, Migon HS. 2008. Objective Bayesian analysis for the student-t regression model. *Biometrika* **95**: 325–333.

Geweke J. 1992. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In Bernardo JM, Berger JO, Dawid AP, Smith AFM (eds.) *Bayesian Statistics. Vol. 4.* 577–586.

Kim S, Shepard N, Chib S. 1998. Stochastic volatility: likelihood inference and comparison with ARCH models. *Review of Economic Studies* **65**: 361–393.

Lange KL, Sinsheimer JS. 1993. Normal/independent distributions and their applications in robust regression. *J. Comput. Graph. Stat* **2**: 175–198.

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

Pemstein D, Quinn KV, Martin AD. 2011. The scythe statistical library: An open source c++ library for statistical computation. *Journal of Statistical Software* **42**: 1–26.

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.

Sttoffer DS, Schert MS, Richardson GA, Day NL, Coble PA. 1998. A Walsh-Fourier analysis of the effects of moderate maternal alcohol consumption on neonatal sleep-state cycling. *Journal of the American Statistical Association* **83**: 954–963.

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.