Binary response state space models with Scale Mixture of Normal Links

Renata. S. Bueno[†] and C. A. Abanto-Valle[†1]

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

Abstract

Observation-driven state space models with scale mixture of normal links are presented for binary time series as a robust alternative to the usual normal setup which is commonly used in the literature. We develop an efficient Markov chain Monte Carlo (MCMC) estimation procedure for the proposed state space models. An application using the (aggregated) Tokyo rainfall data set (Knorr-Held, 1999) is analised.

Keywords: Markov chain Monte Carlo, Probit, Scale Mixture of Normal distributions, State Space Models.

1 Introduction

Among models for time series of dichotomous observations, the class of state space models or parameter-driven models (Cox, 1981) seems to have gained a great deal of popularity. See for example Fahrmeir (1992), Carlin and Polson (1992) and Czado and Song (2008), among others. A binary state space model consists of two processes: In the first observed process $\{y_t\}$, the conditional distribution of y_t is given by a $q-$dimensional state variable $\boldsymbol{\theta}_t$ is Bernoulli, namely $y_t \mid \boldsymbol{\theta} \sim \mathcal{B}er(\mu_t)$, where $\mathcal{B}er(.)$ indicates a Bernoulli distribution and the conditional probability of success $\mu_t = P(y_t = 1 \mid \boldsymbol{\theta}_t)$, follows the observation equation,

$$\mu_t = g(\mathbf{F}_t'\boldsymbol{\theta}_t) \tag{1}$$

with a given link function $g^{-1}(.)$ as in generalized linear models (e.g. McCullagh and Nelder, 1989) and a known $q-$dimensional vector F_t comprised of the time-varying covariates. In the second process, the state variables $\{\boldsymbol{\theta}\}$ are assumed to follow a $q-$dimensional Markov process, governed by the state equation,

$$\boldsymbol{\theta}_t = \mathbf{G}_t\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t \tag{2}$$

where $\mathbf{G}_t$ is a $q \times q$-dimensional transition matrix and the error vector $\boldsymbol{\omega}_t$ has zero mean.

[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

2 Scale Mixture of Normal distributions

Scale mixtures of normal distributions, which play a very important role in statistical modeling, are derived by mixing a normally distributed random variable (Z) with a non-negative scale random variable (λ), as follows

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

where μ is a location parameter, λ is a positive mixing random variable with probability density function (pdf) $h(\lambda|\boldsymbol{\nu})$, independent of $Z \sim \mathcal{N}(0, \sigma^2)$, where $\boldsymbol{\nu}$ 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 $\kappa(\lambda) = 1/\lambda$ in this paper. Thus, given λ, $Y|\lambda \sim \mathcal{N}(\mu, \lambda^{-1}\sigma^2)$ and the pdf of Y is given by

$$f(y|\mu, \sigma^2, \nu) = \int_0^\infty \mathcal{N}(y|\mu, \lambda^{-1}\sigma^2)h(\lambda|\boldsymbol{\nu})d\lambda. \tag{3}$$

From a suitable choice of the mixing density $h(.|\boldsymbol{\nu})$, a rich class of continuous symmetric and unimodal distribution can be described by the density given in (3) that can readily accommodate a thicker-than-normal process. Note that when $\lambda = 1$ (a degenerate random variable), we retrieve the normal distribution. Apart from the normal model, we explore 3 different types of heavy-tailed densities based on the choice of the mixing density $h(.|\boldsymbol{\nu})$. These are as follows.

- *The student-t distribution*, $Y \sim \mathcal{T}(\mu, \sigma^2, \nu)$

  The use of the student-t distribution as an alternative robust model to the normal distribution has frequently been suggested in the literature (Little, 1988; Lange et al., 1989). For the student-t distribution with location μ, scale σ and degrees of freedom ν, the pdf can be expressed in the following hierarchical form:

  $$Y|\mu, \sigma^2, \nu, \lambda \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{\lambda}\right), \qquad \lambda|\nu \sim \mathcal{G}(\nu/2, \nu/2). \tag{4}$$

  where $\mathcal{G}(a, b)$ denotes the Gamma distribution.

- *The slash distribution*, $Y \sim \mathcal{S}(\mu, \sigma^2, \nu)$, $\nu > 0$.

  This distribution presents heavier tails than those of the normal distribution and it includes the normal case when $\nu \uparrow \infty$. Its can be expressed as

  $$Y|\mu, \sigma^2, \nu, \lambda \sim N\left(\mu, \frac{\sigma^2}{\lambda}\right), \qquad \lambda|\nu \sim \mathcal{B}e(\nu, 1), \tag{5}$$

  where $\mathcal{B}e(., .)$ denotes the beta distribution. The slash distribution has been mainly used in simulation studies because it represents some extreme situations depending on the value of ν, see for example Andrews et al. (1972), Morgenthaler and Tukey (1991) and Wang and Genton (2006).

- *The variance gamma distribution,* $Y \sim \mathcal{VG}(\mu, \sigma^2, \nu)$, $\nu > 0$.

  The symmetric variance gamma (VG) distribution was first proposed by Madan and Seneta (1990) to model share market returns. The VG distribution is controlled by the shape parameter $\nu > 0$, presents heavier tails than those of the normal distribution and has a similar SMN density representation to the Student-t distribution. It can be shown that the VG density can be expressed as

$$f(y|\mu, \sigma, \nu) = \int_0^\infty N\left(y|\mu, \frac{\sigma^2}{\lambda}\right) \mathcal{IG}(\lambda|\frac{\nu}{2}, \frac{\nu}{2}) d\lambda. \tag{6}$$

  Thus, the VG distribution is equivalent to the following hierarchical form:

$$Y|\mu, \sigma^2, \nu, \lambda \sim N\left(\mu, \frac{\sigma^2}{\lambda}\right), \qquad \lambda|\nu \sim \mathcal{IG}(\frac{\nu}{2}, \frac{\nu}{2}), \tag{7}$$

  where $\mathcal{IG}(a, b)$ is the inverse gamma distribution.

3 Binary State Space Models with Scale Mixture of normal links

We start with state space models for binary response variables, in which a latent variable representation is utilized to develop MCMC algorithms for parameter estimation.

3.1 Model formulation

Consider a binary data y_t where the binary response vector is denoted by $\mathbf{y}_{1:T} = (y_1, \ldots, y_T)'$. We adopt the so-called threshold approach (e.g. Albert and Chib, 1993) where y_t is generated through dichotomization of an underlying continuous process z_t, given by the one-to-one correspondence

$$Y_t = 1 \Leftrightarrow Z_t \geq 0 \qquad t = 1, \ldots, T. \tag{8}$$

With the unobservable or latent threshold variable vector $\mathbf{z}_{1:T} = (z_1, \ldots, z_T)'$, a binary state space model can be rewritten as follows,

$$z_t = F_t'\boldsymbol{\theta}_t + \lambda_t^{-\frac{1}{2}}\epsilon_t \tag{9}$$

$$\boldsymbol{\theta}_t = \mathbf{G}_t\boldsymbol{\theta}_t + \boldsymbol{\omega}_t \tag{10}$$

$$\lambda_t \sim h(\lambda_t \mid \nu) \tag{11}$$

where $\epsilon_t \overset{\text{iid}}{\sim} \mathcal{N}(0, 1)$ and $\boldsymbol{\omega} \overset{\text{iid}}{\sim} \mathcal{N}(0, \mathbf{W}_t)$. Equations (9) and (10) conditional on the mixture variable λ_t together represent a linear Gaussian state space model.

From the one-to-one relationship (9) and $\lambda_t = 1$ for every $t = 1, \ldots, T$, that the marginal distribution of y_t given both state variable $\boldsymbol{\theta}_t$ follows a probit model of the form

$$\mu_t = P(y_t = 1 \mid \boldsymbol{\theta}_t) = \Phi(\mathbf{F}'\boldsymbol{\theta}_t) \tag{12}$$

where $\Phi(.)$ denotes the cumulative distribution function of $\mathcal{N}(0,1)$.
Other symmetric link functions are obtained from equation (9). If $\lambda_t \sim \mathcal{G}(\nu/2, \nu/2)$, $\lambda_t \sim \mathcal{B}e(\nu, 1)$ and $\lambda_t \mid \mathcal{IG}(\nu/2, \nu/2)$ the symmetric Student-t-, Slash and variance-gamma links are obtained.

3.2 Markov Chain Monte Carlo methods

A Bayesian approach to parameter estimation in the BSSM-SMN links (Binary state space models with Scale Mixture of Normal links) defined by equations (9) and (10) and (11) relies on MCMC techniques.
Suposse that the BSSM-SMN links depends on a parameter vector $\mathbf{\Psi}$. Then the likelihood function $L(\mathbf{\Psi})$ is not easy to calculate. The Bayesian approach for estimating the parameters in the BSSM-SMN links uses the data augmentation principle, which considers $\mathbf{z}_{1:T}$, $\boldsymbol{\theta}_{0:T}$ and $\boldsymbol{\lambda}_{1:T}$ as latent parameters.The joint posterior density of parameters and latent variables can be written as

$$p(\mathbf{z}_{1:T}, \boldsymbol{\theta}_{0:T}, \boldsymbol{\lambda}_{1:T}, \mathbf{\Psi} \mid \mathbf{y}_{1:T}) \quad \propto \quad p(\mathbf{z}_{1:T} \mid \boldsymbol{\lambda}_{1:T}, \mathbf{h}_{0:T}, \mathbf{y}_{1:T}) p(\boldsymbol{\theta}_{0:T} \mid \mathbf{\Psi}) p(\boldsymbol{\lambda}_{1:T} \mid \mathbf{\Psi}) p(\mathbf{\Psi}), \tag{13}$$

We sample from (13) using the Gibbs sampling algorithm. In particular we will update the state variables jointly using the simulation smoother of de Jong and Shephard (1995).

References

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

Andrews DF, Bickel PJ, Hampel FR, Huber PJ, Rogers WH, Tukey J. 1972. *Robust Estimates of Location: Survey and Advances.* Princeton, NJ: Princeton University Press.

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.

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

Cox DR. 1981. Statistical analysis of time series, some recent developments. *Scandinavian Journal of Statistics* **8**: 93–115.

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 extendedkalman filtering for multivariate dynamic generalized linear models. *Journal of the American Statistical Association* **87**: 501–509.

Knorr-Held L. 1999. Conditional prior proposals in dynamic models. *Scandinavian Journal of Statistics* **26**: 129–144.

Lange KL, Little R, Taylor J. 1989. Robust statistical modeling using t distribution. *Journal of the American Statistical Association* **84**: 881–896.

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

Little R. 1988. Robust estimation of the mean and covariance matrix from data with missing values. *Applied Statistics* **37**: 23–38.

Madan D, Seneta E. 1990. The variance gamma (v.g) model for share market return. *Journal Business* **63**: 511–524.

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

Morgenthaler S, Tukey J. 1991. *Configural Polysampling: A Route to Practical Robustness*. New York: Wiley.

Wang J, Genton MG. 2006. The multivariate skew-slash distribution. *Journal of Statistical Planning and Inference* **136**: 209–220.