$\mathcal{I}^2\mathcal{SDS}$

The Institute for Integrating Statistics in Decision Sciences

Technical Report TR-2008-2

February 20, 2008

A Short Note on Bayesian Analysis of Dynamic Probit Models

Refik Soyer
Department of Decision Sciences
The George Washington University

Minje Sung
School of Business Administration
Ajou University, Suwon, Korea

A SHORT NOTE ON BAYESIAN ANALYSIS OF DYNAMIC PROBIT MODELS

Refik Soyer
Department of Decision Sciences
The George Washington University, Washington, DC, USA

Minje Sung
School of Business Administration, Ajou University, Suwon, Korea

Abstract

In this note we consider a dynamic probit model where the coefficients follow a first-order Markov process. We present an exact Gibbs sampler for Bayesian analysis of the model using the data augmentation approach of Albert and Chib (1993) and the forward filtering backward sampling algorithm of Fruhwirth-Schnatter (1994) for dynamic linear models. We discuss how our approach can be used for probit based Markov regression models and discuss Markov order selection in these dynamic models.

Key words: Markov models, Bayesian inference, longitudinal data, dynamic linear models, model selection.

1. Introduction: Dynamic Probit Model

Time varying coefficient models for categorical longitudinal data has been considered by authors such as Shephard and Pitt (1997), Gamerman (1998), Kauermann (2000), and more recently by Fruhwirth-Schnatter and Fruhwirth (2006). Most of the previous work have considered logit-type state-space models. As noted by Fruhwirth-Schnatter and Fruhwirth (2006), Markov chain Monte Carlo (MCMC) approaches proposed by Shephard and Pitt (1997) and Gamerman (1998) for the analysis of these models are based on Metropolis-Hastings algorithm which requires specification of a proposal density in high dimensions. To alleviate this the authors proposed a data augmentation based MCMC method for analysis of dynamic logit models. A simpe version of a dynamic probit model has been considered by Andrieu and Daucet (2002) where the authors used particle filtering for Bayesian analysis.

In what follows, we consider probit-type state-space models and develop an exact Gibbs sampler for Bayesian analysis of this class of models. Our approach is an extension of the data augmentation approach of Albert and Chib (1993) to dynamic probit models where we implement the *forward filtering backward sampling* algorithm of Fruhwirth-Schnatter (1994).

We consider a binary time series X_t and we define a dynamic probit model similar to considered by Andrieu and Daucet (2002) as

$$Pr\{X_t = 1|\pi_t\} = \pi_t \text{ with } \pi_t = \Phi(\boldsymbol{F}_t\boldsymbol{\theta}_t) \tag{1.1}$$

where $\boldsymbol{F}_t$ is a $1 \times K$ covariate vector and $\boldsymbol{\theta}_t$ is a $K \times 1$ vector of regression parameters. We define the dynamic nature of the model via a state equation for $\boldsymbol{\theta}_t$

$$\boldsymbol{\theta}_t = \boldsymbol{G}\boldsymbol{\theta}_{t-1} + \boldsymbol{w}_t \tag{1.2}$$

with $\boldsymbol{w}_t$'s are uncorrelated multivariate normal error vectors with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{W}_\theta$ and $\boldsymbol{G}$ is the speficified transition matrix of the model. It is most common to assume that $\boldsymbol{G}$ is an identity matrix implying a *steady model* in the sense of West and Harrison (1997). Thus, in our development we consider

$$\boldsymbol{\theta}_t = \boldsymbol{\theta}_{t-1} + \boldsymbol{w}_t. \tag{1.3}$$

The procedure can be easily extended for a general known transition $\boldsymbol{G}$ as well as for certain cases where $\boldsymbol{G}$ is unknown. These are discussed in Soyer and Sung (2008).

We can extend this for longitudinal data for individuals $i = 1, \ldots, N$. In this case we write the above as

$$Pr\{X_{it} = 1|\pi_{it}\} = \pi_{it} \text{ with } \pi_{it} = \Phi(\boldsymbol{F}_{it}\boldsymbol{\theta}_t) \tag{1.4}$$

and assume the same state equation (1.3) for all individuals.

2. Bayesian Inference

We first consider the case for the ith individual Following Albert and Chib (1993), we can define the above model by using independent latent variables Z_{it} such that

$$X_{it} = \begin{cases} 1 & \text{if } Z_{it} > 0 \\ 0 & \text{otherwise.} \end{cases} \tag{2.1}$$

If we assume that Z_{it}'s are normally distributed with mean $\boldsymbol{F}_{it}\boldsymbol{\theta}_t$ and variance 1, that is, $Z_{it} \sim N(\boldsymbol{F}_{it}\boldsymbol{\theta}_t, 1)$, then we have the probit model

$$\pi_{it} = \Phi(\boldsymbol{F}_{it}\boldsymbol{\theta}_t). \tag{2.2}$$

Given the above setup we can develop a Gibbs sampler for the inference using the data augmentation algorithm of Albert and Chib (1993) with the algorithm proposed by Fruhwirth-Schnatter (1994) for dynamic linear models.

Given observed data $D = \{X_{it};\, t = 1, \ldots, T\}$, we can design a Gibbs sampler using full posterior conditional distributions $p(\Theta|D, \boldsymbol{Z_i^T})$ and $p(\boldsymbol{Z_i^T}|D, \Theta)$ with vectors $\Theta = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_T)$ and $\boldsymbol{Z_i^T} = (Z_{i1}\, Z_{i2} \cdots Z_{iT})$. In obtaining $p(\boldsymbol{Z_i^T}|D, \Theta)$, we note that Z_{it}'s are independent random variables and use

$$(Z_{it}|\boldsymbol{\theta}_t, X_{it} = 1) \sim N(\boldsymbol{F}_{it}\boldsymbol{\theta}_t, 1)\, I(Z_{it} > 0)$$

$$(Z_{it}|\boldsymbol{\theta}_t, X_{it} = 0) \sim N(\boldsymbol{F}_{it}\boldsymbol{\theta}_t, 1)\, I(Z_{it} < 0).$$

In implementation of the Gibbs sampler, we can directly draw from the joint posterior distribution of $p(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_T|\boldsymbol{Z_i^T})$ using the using the *forward filtering* *backward sampling* algorithm of Fruhwirth-Schnatter (1994) which is given in West and Harrison (1997) for Kalman filter type models. It is possible to adopt the algorithm for our case as will be discussed next.

We define $\boldsymbol{Z_i^t} = (\boldsymbol{Z_i^{t-1}}, Z_{it}),\ t = 1,\ \ldots, T$ and note that, similar to the Bayesian dynamic linear models of West and Harrison (1997), using the Markov structure of our model we can write $p(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_T | \boldsymbol{Z_i^T})$ as

$$p(\boldsymbol{\theta}_T|\boldsymbol{Z_i^T})\, p(\boldsymbol{\theta}_{T-1}|\boldsymbol{\theta}_T, \boldsymbol{Z_i^{T-1}}) \cdot \cdots \cdot p(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2, \boldsymbol{Z_i^1}), \tag{2.3}$$

where the first term $p(\boldsymbol{\theta}_T|\boldsymbol{Z_i^T})$ is available from standard DLM updating. We can start the sampling from $\boldsymbol{\theta}_T$ and then sequentially sample $\boldsymbol{\theta}_{T-1}, \ldots, \boldsymbol{\theta}_1$ using densities $p(\boldsymbol{\theta}_{t-1}|\boldsymbol{\theta}_t, \boldsymbol{Z_i^{t-1}})$ for $t = T-1, \ldots, 2$. The required distributions can be obtained using the state equation of the DLM. We can write

$$p(\boldsymbol{\theta}_{t-1}|\boldsymbol{\theta}_t, \boldsymbol{Z_i^{t-1}}) \propto p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1}, \boldsymbol{Z_i^{t-1}}) p(\boldsymbol{\theta}_{t-1}|\boldsymbol{Z_i^{t-1}}), \tag{2.4}$$

where

$$(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1}, \boldsymbol{Z_i^{t-1}}) \sim Normal(\boldsymbol{\theta}_{t-1}, \boldsymbol{W}_\theta). \tag{2.5}$$

Standard DLM setup yields

$$(\boldsymbol{\theta}_{t-1}|\boldsymbol{Z_i^{t-1}}) \sim Normal(\boldsymbol{m_{t-1}}, \boldsymbol{C_{t-1}}) \tag{2.6}$$

where

$$\boldsymbol{m_t} = \boldsymbol{m_{t-1}} + \boldsymbol{R_t}\boldsymbol{F}'_{it}(1 + \boldsymbol{F}_{it}\boldsymbol{R_t}\boldsymbol{F}'_{it})^{-1} e_t \tag{2.7}$$

with $e_t = Z_{it} - \boldsymbol{F}_{it}\boldsymbol{m_{t-1}}$ is a scalar, $\boldsymbol{R_t} = \boldsymbol{C_{t-1}} + \boldsymbol{W}_\theta$ and

$$\boldsymbol{C_t} = \boldsymbol{R_t} - \boldsymbol{R_t}\boldsymbol{F}'_{it}(1 + \boldsymbol{F}_{it}\boldsymbol{R_t}\boldsymbol{F}'_{it})\boldsymbol{F}_{it}\boldsymbol{R_t}. \tag{2.8}$$

It follows from the above that

$$(\boldsymbol{\theta}_{t-1}|\boldsymbol{\theta}_t, \boldsymbol{Z_i^{t-1}}) \sim Normal(\boldsymbol{h_{t-1}}, \boldsymbol{H_{t-1}}) \tag{2.9}$$

where

$$\boldsymbol{h}_{t-1} = \boldsymbol{m_{t-1}} + \boldsymbol{C_{t-1}}\boldsymbol{R_t}^{-1}(\boldsymbol{\theta}_t - \boldsymbol{m_{t-1}}) \tag{2.10}$$

and

$$\boldsymbol{H}_{t-1} = \boldsymbol{C}_{t-1} - \boldsymbol{C}_{t-1}\boldsymbol{R}_t^{-1}\boldsymbol{C}_{t-1}. \tag{2.11}$$

In the case where we have a prior on W_θ which is an inverse Wishart form given by

$$\boldsymbol{W}_\theta^{-1}|\boldsymbol{R}, r \sim Wish(\boldsymbol{R}, r), \tag{2.12}$$

where the scale matrix $\boldsymbol{R}$ and degrees of freedom $r > K$ are known quantities, the above results will be all conditional on W_θ. The full conditional of W_θ can be obtained as proportional to

$$|\mathbf{W}_\theta^{-1}|^{(r+T-K-1)/2} exp\Big[-\frac{1}{2}tr\Big\{\Big(\boldsymbol{R} + \sum_{t=1}^{T}(\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-1})(\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-1})'\Big)\boldsymbol{W}_\theta^{-1}\Big\}\Big], \tag{2.13}$$

which is again a Wishart density with degrees of freedom, $(r+T)$, and scale matrix $\Big(\boldsymbol{R} + \sum_{t=1}^{T}(\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-1})(\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-1})'\Big)$.

When we consider data from M individuals for T time periods given by $D = \{X_{it};\, t = 1, \ldots, T;\, i = 1, \ldots, M\}$ where X_{it}'s are conditionally independent cross the individuals as well as time, we define $\boldsymbol{Z}_t$ as a $M \times 1$ vector of latent variables Z_{it}. We now assume that $\boldsymbol{Z}_t$ has a multivariate normal distribution with independent components denoted as $\boldsymbol{Z}_t \sim N(\boldsymbol{F}_t\boldsymbol{\theta}_t, \boldsymbol{I}_M)$ where $\boldsymbol{F}_t$ is a $M \times K$ matrix of covariates and $\boldsymbol{I}_M$ is an $M \times M$ identity matrix. Since Z_{it}'s are conditionally independent random quantities, their full conditionals will stay the same. In updating $\boldsymbol{\theta}_t$'s we will have

$$\boldsymbol{m}_t = \boldsymbol{m}_{t-1} + \boldsymbol{R}_t\boldsymbol{F}_t'(\boldsymbol{I}_M + \boldsymbol{F}_t\boldsymbol{R}_t\boldsymbol{F}_t')^{-1}\boldsymbol{e}_t \tag{2.14}$$

with $\boldsymbol{e}_t = \boldsymbol{Z}_t - \boldsymbol{F}_t\boldsymbol{m}_{t-1},\ \boldsymbol{R}_t = \boldsymbol{C}_{t-1} + \boldsymbol{W}_\theta$ and

$$\boldsymbol{C}_t = \boldsymbol{R}_t - \boldsymbol{R}_t\boldsymbol{F}_t'(\boldsymbol{I}_M + \boldsymbol{F}_t\boldsymbol{R}_t\boldsymbol{F}_t')\boldsymbol{F}_t\boldsymbol{R}_t. \tag{2.15}$$

where $\boldsymbol{e}_t$ is a $M \times 1$ vector. Updating of $\boldsymbol{W}_\theta$ will stay the same.

Thus, the proposed approach provides an exact Gibbs sampler for Bayesian inference in the dynamic probit model (1.1) with evolution equation (1.2). Extension of the approach to multinomial dynamic probit models is considered in Soyer and Sung (2008).

3. Markov Regression Model Representation

We consider a $q-th$ order nonhomogeneous Markov chain model defined as

$$Pr\{X_{it}=1|\pi_{it}, X_{i,t-1},\cdots,X_{i,t-q}\}=\pi_{it} \text{ with } \pi_{it}=\Phi(\boldsymbol{F}_{it}\boldsymbol{\theta}_t) \quad (3.1)$$

where

$$\boldsymbol{F}_{it}=(1\, X_{i,t-1}\cdots X_{i,t-q}) \text{ and } \boldsymbol{\theta}_t=(\theta_{0t}\,\theta_{1t}\cdots\theta_{qt})'.$$

Thus, the transition probability π_{it} is given by

$$\pi_{it}=\Phi(\theta_{0t}+\sum_{j=1}^{q}\theta_{j,t}\,X_{i,t-j}). \quad (3.2)$$

For $q=1$ we obtain the first-order chain model

$$\pi_{it}=\Phi(\theta_{0t}+\theta_{1,t}\,X_{i,t-1})$$

implying a transition matrix

$$\boldsymbol{P}_t=\begin{bmatrix}\Phi(\theta_{0t}+\theta_{1,t}\,X_{i,t-1}) & 1-\Phi(\theta_{0t}+\theta_{1,t}\,X_{i,t-1})\\ \Phi(\theta_{0t}) & 1-\Phi(\theta_{0t})\end{bmatrix}. \quad (3.3)$$

The above model provides an nonhomogeneous extension of Bayesian Markov regression models of Erkanli et al. (2001) who used logistic link functions. It also gives an alternative class of Bayesian nonhomogeneous Markov chain models considered in Sung et al. (2007).

We note that the transition probabilities can also be dependent on covariates and such dependence can be easily incorporated by defining components of $\boldsymbol{F}_{it}\boldsymbol{\theta}_t$

accordingly. The Bayesian inference results presented in the previous section can be easily used for analysis of the model.

4. Markov Order and Variable Selection in Dynamic Probit Models

As pointed by Kass and Raftery (1995), Bayesian model comparison/selection is made using Bayes factors which are obtained as the ratio of marginal likelihoods $p(D|i)$ under two competing models $i = 1, 2$. In many problems $p(D|i)$ is not available in an analytical form and its evaluation using posterior Monte Carlo samples is not a trivial task. Thus, various alternatives to marginal likelihoods have been suggested for model selection using Monte Carlo samples; see for example Gelfand (1996).

However, in certain problems where a Gibbs sampler is used and all the full conditional distributions are known, it is possible to approximate the marginal likelihoods from the posterior samples using a method introduced by Chib (1995). For example, Chib (1995) showed that how marginal likelihood $p(D|i)$ can be obtained for the static probit regression model i with a given set of independent variables. In so doing, Chib (1995) used the data augmentation approach of Albert and Chib (1993) and discussed computation of $p(D|i)$ from the Gibbs sampler output. In what follows we will illustrate how the approach by Chib (1995) can be extended for the dynamic probit model and discuss how the approach can be used for Markov regression model order selection. Thus, we present a Bayesian approach for order selecton in nonhomogeneous Markov chains which have not been considered by Sung et al. (2007).

Note that suppresing dependence on model i the marginal likelihood for a particular model is expressed as

$$p(D) = \frac{p(D|\Theta)p(\Theta)}{p(\Theta|D)}, \qquad (4.1)$$

where Θ is a vector of parameters. As pointed out by Chib (1995) the above holds for any value of Θ, say Θ^*, and the value of posterior density $p(\Theta^*|D)$ can be estimated by

$\widehat{p}(\Theta^*|D)$ using Monte Carlo samples. Since $p(D|\Theta^*)$ and $p(\Theta^*)$ can be evaluated at Θ^*, the log marginal likelihood can be estimated as

$$ln\,\widehat{p}(D) = ln\,p(D|\Theta^*)\; + ln\,p(\Theta^*) - \; ln\,\widehat{p}(\Theta^*|D). \tag{4.2}$$

In evaluating (4.2), the only term which is not readily available is $\widehat{p}(\Theta^*|D)$, but as shown in Chib (1995) this can be obtained using the outputs from the Gibbs sampler.

In our case, we also have the the latent variables $\boldsymbol{Z^T} = (\boldsymbol{Z_j^T};\, j = 1, \ldots, M)$ and the parameter vector $\Theta = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_T, \boldsymbol{W}_\theta)$. To estimate $p(\Theta^*|D)$ we need the full conditionals $p(\boldsymbol{Z^T}|\Theta,\, D)$ and $p(\Theta|\boldsymbol{Z^T}, D)$ that are available to us. We can write

$$p(\Theta^*|D) = \int p(\Theta^*|\boldsymbol{Z^T}, D)p(\boldsymbol{Z^T}|D)\, d\boldsymbol{Z^T} \tag{4.3}$$

and note that we have samples available from $p(\boldsymbol{Z}^{\boldsymbol{T}}|D)$ via the Gibbs sampler. thus, (4.3) can be approximated as

$$p(\Theta^*|D) \approx \frac{1}{G}\sum_{g=1}^{G} p(\Theta^*|(\boldsymbol{Z^T})^{(g)}, D), \tag{4.4}$$

where $(\boldsymbol{Z^T})^{(g)}$ are samples from the posterior distribution $p(\boldsymbol{Z}^{\boldsymbol{T}}|D)$. Note that $p(\Theta^*|\boldsymbol{Z^T}, D)$ can be written as

$$p(\boldsymbol{\theta}_T^*|\boldsymbol{Z^T}, \boldsymbol{W}_\theta^*)\, p(\boldsymbol{\theta}_{T-1}^*|\boldsymbol{\theta}_T^*, \boldsymbol{Z^{T-1}}, \boldsymbol{W}_\theta^*) \cdots p(\boldsymbol{\theta}_1^*|\boldsymbol{\theta}_2^*, \boldsymbol{Z^1}, \boldsymbol{W}_\theta^*)\, p(\boldsymbol{W}_\theta^*|\boldsymbol{Z^T}). \tag{4.5}$$

In the above all terms are immediately available except the last one, that is, $p(\boldsymbol{W}_\theta^*|\boldsymbol{Z^T})$. We can obtain this term as

$$p(\boldsymbol{W}_\theta^*|\boldsymbol{Z^T}) = \int p(\boldsymbol{W}_\theta^*|\boldsymbol{\theta^T})p(\boldsymbol{\theta^T}|\boldsymbol{Z^T})d\boldsymbol{Z^T} \tag{4.6}$$

where $\boldsymbol{\theta^T} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_T)$. The above can be approximated as

$$p(\boldsymbol{W}_\theta^*|\boldsymbol{Z^T}) \approx \frac{1}{G}\sum_{g=1}^{G} p(\boldsymbol{W}_\theta^*|\boldsymbol{\theta}_1^{(g)}, \ldots, \boldsymbol{\theta}_T^{(g)}) \tag{4.7}$$

where $\boldsymbol{\theta}_1^{(g)}, \ldots, \boldsymbol{\theta}^{(g)}$ are samples from the posterior $p(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_T | \boldsymbol{Z^T})$. Note that $p(\boldsymbol{W}_\theta^* | \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_T)$ is the full conditional which is a Wishart density.

Thus, all components of (4.2) are now available and we can evaluate $ln\, \widehat{p}(D)$ for a given model as reflected by the order of the Markov process and/or by the variables included in (3.1).

Implementation of the proposed exact Gibbs sampler for the dynamic probit models and the order selection approach are illustrated using real life longitudinal data in Soyer and Sung (2008).

References

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

Andrieu, C. and Daucet, A. (2002). Particle filtering for partially observed Gaussian state space models. *Journal of the Royal Statistical Society, Series B*, 64:827-836.

Cargnoni C., Mueller P., and West M. (1997). Bayesian forecasting of multinomial time series through conditionally Gaussian dynamic models. *Journal of the American Statistical Association*, 92:640-647.

Chib, S. (1995). Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association*, 90:1313-1321.

Erkanli A., Soyer R., and Angold A. (2001). Bayesian analyses of longitudinal binary data using Markov regression models of unknown order. *Statistics in Medicine*, 20:755-770.

Fruhwirth-Schnatter, S. (1994). Data augmentation and dynamic linear models, *Journal of Time Series Analysis*, 15:183-202.

Fruhwirth-Schnatter, S. and Fruhwirth, R. (2006). Auxiliary mixture sampling with applications to logistic models. Technical Report 2006-14, Department of Applied Statistics, Johannes Kepler Univerity Linz.

Gamerman, D. (1998). Markov chain Monte Carlo for dynamic generalized linear models. *Biometrika*, 85:215-227.

Gelfand, A. E. (1996). Model determination using sampling-based methods. In *Markov Chain Monte Carlo in Practice*. (Eds. W. R. Gilks, S. Richardson and D. J. Spiegelhalter), 145-161, Chapman & Hall.

Kass, R. E. and Raftery, A. E. (1995). Bayes factors, *Journal of the American Statistical Association*, 90:773-795.

Kauermann, G. (2000). Modeling longitudinal data with ordinal response by varying coefficients, *Biometrics*, 56:692-698.

Shephard, N. and Pitt, M. K. (1997). Likelihood analysis of non-Gaussian measurement time-series. *Biometrika*, 84:653-667.

Soyer, R. and Sung, M. (2008). Bayesian analysis of dynamic probit models: Application to longitudinal data. Manuscript in preparation.

Sung, M., Soyer, R. and Nhan, N. (2007). Bayesian analysis of nonhomogeneous Markov chains: Applications to mental health data. *Statistics in Medicine*, 26: 3000-3017.

West, M. and Harrison, J. (1997). *Bayesian Forecasting and Dynamic Models*. 2nd Edition, Springer-Verlag.