THE UNIVERSITY OF MELBOURNE

Melbourne Institute Working Paper Series

Working Paper No. 16/08

A Bayesian Simulation Approach to Inference on a Multi-State Latent Factor Intensity Model

Chew Lian Chua, G. C. Lim and Penelope Smith

MELBOURNE INSTITUTE
of Applied Economic and Social Research

A Bayesian Simulation Approach to Inference on a Multi-State Latent Factor Intensity Model*

Chew Lian Chua[†], G. C. Lim[†] and Penelope Smith[‡]

[†] Melbourne Institute of Applied Economic and Social Research
The University of Melbourne
[‡] Westpac Banking Corporation, Sydney

Melbourne Institute Working Paper No. 16/08

ISSN 1328-4991 (Print)

ISSN 1447-5863 (Online)

ISBN 978-0-7340-3284-3

August 2008

* This research was funded by a grant from the Melbourne Centre for Financial Studies, a consortium of Melbourne, Monash and RMIT Universities, and the Financial Services Institute of Australasia.

Melbourne Institute of Applied Economic and Social Research
The University of Melbourne
Victoria 3010 Australia
***Telephone* (03) 8344 2100**
***Fax* (03) 8344 2111**
***Email* melb-inst@unimelb.edu.au**
***WWW Address* http://www.melbourneinstitute.com**

Abstract

This paper provides a Bayesian approach to inference on a multi-state latent factor intensity model to manage the problem of highly analytically intractable pdfs. The sampling algorithm used to obtain posterior distributions of the model parameters includes a particle filter step and a Metropolis-Hastings step within a Gibbs sampler. A simulated example is conducted to show the feasibility and accuracy of this sampling algorithm. The approach is applied to the case of credit ratings transition matrices.

Keywords: Non-linear non-Gaussian state space model; Auxiliary particle filter.

JEL classification: C11; C15

1 Introduction

Economic and financial analysis which involves estimating latent factor models and/or estimating multi-state models have grown with the availability of computer simulation algorithms (see for examples, McNeil and Wendin (2007), Fiorentini et al. (2004), and Bauwens and Veredas(2004)). Recently, Koopman, Lucas and Monteiro (2008) proposed a multi-state latent factor intensity model for credit ratings to allow for multiple origins and destination states. Their approach is set in a continuous time framework, and their application is quite specific because they note that the "econometric issues related to the generalization are intricate and the computational consequences are severe (p.422)". Part of the difficulty lies with a likelihood which is a high dimensional integral with analytically intractable pdfs.

Bayesian econometrics have grown in importance as a way to manage complicated likelihood functions. This paper proposes a version of the Gibbs sampler which includes a particle filter step and a Metropolis-Hastings (MH) step to estimate a multi-state latent factor intensity model[1]. As noted by McNeil and Wendin (2007) MCMC algorithms such as the Gibbs sampler can be used to estimate models with complex latent structures such as serially correlated random effects and/or multivariate random effects capturing heterogeneity across industry sectors. Since the multi-state latent factor intensity model is a highly non-linear state space model, the particle filter of Pitt and Shepard (1999) within a Gibbs sampler provides an appropriate way of simulating the time-dependent latent factor from the intractable conditional filtered density.

The paper is organised as follows. Section 2 presents a general multi-state latent factor intensity model and sets out the likelihood. The Bayesian approach and the sampling algorithms are set out in Section 3. A simulation study to validate the algorithm is described in Section 4 and an empirical case study is contained in Section 5. Concluding remarks are in Section 6.

2 Multi-state latent factor intensity model

In this section, the multi-state latent factor intensity model is presented. Let s denote the s^{th} transition from one state to another out of a total of S

[1] For an analysis of credit ratings in a Bayesian framework without allowing for latent variables, see Das, Fan and Geng (2002).

possible transitions and let k be the k^{th} unit in the sample of K transiting from one state to another. The instantaneous probability $\lambda_{sk}(t_i)$ of unit k engaging in transition s at time t_i is:

$$\lambda_{sk}(t_i) = R_{sk}(t_i) \exp\left[\eta_s + \gamma_s w_k(t_i) + \alpha_s \psi(t_i)\right]. \tag{1}$$

Here $R_{sk}(t_i)$ is a dummy variable that takes the value of one if unit k is 'at risk' of transition. The unknown parameters of the model are η_s, γ_s and α_s. The vector n_s represents the constant reference-level log-intensity of transition type s, γ_s measures the sensitivity of unit k's association to changes in observable explanatory variables $w_k(t_i)$ and α_s measures the sensitivity of unit k's association to changes in an unobservable common dynamic latent factor $\psi(t_i)$. A deterministic hazard function $H_{sk}(t_i)$ can also be attached to equation (1); in this paper we let $H_{sk}(t_i) = 1$.

The latent variable $\psi(t_i)$ accounts for unobserved dependence between the transition histories in a parsimonious way. We make the assumption that $\psi(t_i)$ follows an AR(1) process:

$$\psi(t_i) = \rho\psi(t_{i-1}) + \varepsilon(t_i), \; i = 1, ..., N \tag{2}$$

$$\varepsilon(t_i) \sim N(0, \sigma^2) \tag{3}$$

where all of the roots of ρ lie outside the unit circle. However we note that this assumption is not central to the method we propose and that the model could easily be generalized to accommodate alternative functional forms and higher order AR processes. As α_s and σ are not simultaneously identified, we normalize the parameter space so that $\sigma = 1$.

2.1 Likelihood function

Let $Y_{sk}(t_i)$ be a function that is equal to 1 when unit k experiences a transition event of type s at time t_i and zero otherwise and let $\tau_i = t_i - t_{i-1}$ be the duration between events. If we define

$$z_i = \{\tau, R_{11}(t_i), ..., R_{SK}(t_i), Y_{11,}(t_i), ..., Y_{SK}(t_i)\} \tag{4}$$

then the likelihood function of θ conditional on ψ_i, for the i^{th} event time is:

$$p(z_i|\theta,\psi(t_i),\rho)=\prod_{k=1}^{K}\prod_{s=1}^{S}\exp\left(Y_{sk}(t_i)\left(\eta_s+\gamma_s' w_{kt}+\alpha_s\psi(t_i)\right)-R_{sk}(t_i)\int_{t_i-1}^{t_i}\lambda_{sk}(t_i)dt\right) \tag{5}$$

where $\theta=\{\eta_1,...,\eta_S,\gamma_1',...,\gamma_S',\alpha_1,...,\alpha_S\}$ and $\rho=\{\rho_1,...,\rho_r\}$. Then, the likelihood function of θ for the whole sample period is then

$$\begin{aligned} p(z|\theta,\rho) &= \prod_{i=1}^{N} p(z_i|\theta,\mathcal{F}_{t_{i-1}},\rho) \qquad (6)\\ &= \prod_{i=1}^{N}\int p(z_i|\theta,\psi(t_i),\rho)p(\psi(t_i)|\mathcal{F}_{t_{i-1}},\rho)d\psi(t_i) \end{aligned}$$

where $\mathcal{F}_{t_{i-1}}$ denotes the history of all observations up to time t_{i-1}, z is the collection of $\{z_1,\, z_2,\, ...,\, z_N\}$ and $p(\psi(t_i)|\mathcal{F}_{t_{i-1}},\rho)$ is a predictive density which is defined as:

$$p(\psi(t_i)|\mathcal{F}_{t_{i-1}},\rho)=\int p(\psi(t_i)|\psi(t_{i-1}),\rho)p(\psi(t_{i-1})|\mathcal{F}_{t_{i-1}},\theta,\rho)d\psi(t_{i-1}) \tag{7}$$

3 Bayesian Inference

Before Bayes' theorem can be applied to the likelihood, the first step is to specify a prior distribution for (θ,ρ). We assume that each parameter is *a priori* independent, and to cater for model comparison the elicited priors have to be proper. Thus, the joint prior pdf is given as

$$p(\theta,\rho)=I(\rho)\prod_{s=1}^{S}p(\eta_s)p(\gamma_s)p(\alpha_s), \tag{8}$$

where the prior pdfs are: $\eta_s\sim N(\underline{\eta_s},\sigma^2_{\eta_s})$, $\gamma_s\sim N(\underline{\gamma_s},\sigma^2_{\gamma_s})$ and $\alpha_s\sim N(\underline{\alpha_s},\sigma^2_{\alpha_s})$. Elements in ρ are bounded within an uniform distribution, $U(-1,1)$, to ensure stationary in $\psi(t_i)$. The indicator function $I(\rho)$ is such that $I(\rho)=1$ if the roots of ρ are within the range of the uniform distributions.

The second step is to combine $p(\theta,\rho)$ with the likelihood function in (6) to give the joint posterior pdf for (θ, ρ):

$$p(\theta, \rho|z) \propto p(\theta, \rho)p(z|\theta, \rho) \tag{9}$$

Given that inferences on the parameters are made from their marginal posterior pdfs which are highly intractable, we propose a Gibbs sampler that encompasses an Auxiliary Particle Filter (APF) (see Pitt and Shepard (1999)) for simulating $\psi(t_i)$ and a Metropolis-Hastings (MH) algorithm for θ.

It would be appropriate at this juncture, to note briefly the rationale for the APF. Producing draws of $\psi(t_i)$ via a filtering method involves the repetition of two basic steps: predicting and updating. These steps recursively produces sequences of draws of $\psi(t_i)$ from the filtered density $p(\psi(t_i)|\mathcal{F}_{t_i}, \theta, \rho)$. In the prediction step, $p(\psi(t_{i+1})|\mathcal{F}_{t_i}, \theta, \rho)$ is obtained by projecting $p(\psi(t_i)|\mathcal{F}_{t_i}, \theta, \rho)$ one-step ahead from $p(\psi(t_{i+1})|\psi(t_i), \rho)$

$$p(\psi(t_{i+1})|\mathcal{F}_{t_i}\, \theta, \rho) = \int p(\psi(t_{i+1})|\psi(t_i), \rho)p(\psi(t_i)|\mathcal{F}_{t_i}, \theta, \rho)d\psi(t_i) \tag{10}$$

and as new information arrive at t_{i+1}, the filtered density is updated via Bayes theorem:

$$p(\psi(t_{i+1})|\mathcal{F}_{t_{i+1}}, \theta, \rho) \propto p(z_i|\theta, \mathcal{F}_{t_{i-1}}, \rho)p(\psi(t_{i+1})|\mathcal{F}_{t_i}, \theta, \rho). \tag{11}$$

Although the filter approach is conceptually straightforward, because the multi-factor latent intensity model is a non-linear non-gaussian state space model, the filtered density is intractable. We thus adopt the APF to recursively simulate $p(\psi(t_i)|\mathcal{F}_{t_i}, \theta, \rho)$ from period $i = 1$ to N. As noted in Aguilar and West (2000), the APF is a proven technique for sequential updating of simulation-based summaries for filtered[2] distribution for time-evolving states; for example Chib et al. (2002) used the APF to estimate a stochastic volatility model with jumps.

3.1 Gibbs sampler

This subsection implement the main steps of our proposed sampling algorithm while the subsequent two subsections provide the steps involved in

[2] In the filtering literature, various methods have been proposed as well, for example sequential importance sampling and sampling importance resampling approaches. Doucet et al. (2000) offer an excellent survey on other approaches.

the APF and MH algorith respectively.

1. Choose an arbitrary starting point for $(\theta^{(j)}, \rho^{(j)})$ and set $j = 0$.
2. Given $(\theta^{(j)}, \rho^{(j)})$, sample $\widetilde{\psi}(t_i)^{(j+1)}$, $i = 1, ..., N$, using the APF algorithm set out below.
3. Sample $\rho^{(j+1)}$ from

$$\rho^{(j+1)}|\widetilde{\psi}(t_1)^{(j+1)}, ..., \widetilde{\psi}(t_N)^{(j+1)}, \theta^{(j)} \sim N(\overline{\rho}, \overline{\sigma}_p^2)_{I(\rho)} \tag{12}$$

where $\overline{\rho} = \left(\sum_{i=2}^{N} \widetilde{\psi}^2(t_{i-1})\right)^{-1} \left(\sum_{i=2}^{N} \widetilde{\psi}(t_{i-1})\psi(t_i)\right), \overline{\sigma}_p^2 = \left(\sum_{i=2}^{N} \widetilde{\psi}^2(t_{i-1})\right)^{-1}$.
Ensure that the stationarity condition $-1 < \rho^{(j+1)} < 1$ holds via rejection sampling.
4. Given $\rho^{(j+1)}$, sample $\theta^{(j+1)}$ using a random walk MH algorithm.
5. Set $j = j + 1$ and return to step 2.[3].

3.1.1 Auxiliary Particle Filter algorithm

1. Given $\theta^{(j)}$ and $\rho^{(j)}$ obtain G draws of $\psi(t_0)$ from $N(0, 1)$ and set $i = 1$.[4]
2. Given $\psi(t_{i-1})^{(g)}$ from $p(\psi(t_{i-1})|\mathcal{F}_{t_{i-1}}, \theta^{(j)}, \rho^{(j)})$, compute

$$\widehat{\psi}(t_i)^{*(g)} = \rho^{(j)}\psi(t_{i-1})^{(g)} \; g = 1, ..., G, \tag{13}$$

and

$$w_g = \frac{p(z_i|\theta^{(j)}, \widehat{\psi}(t_i)^{*(g)}, \rho^{(j)})}{\sum_{l=1}^{G} p(z_i|\theta^{(j)}, \widehat{\psi}(t_i)^{*(l)}, \rho^{(j)})} \; g = 1, ..., G. \tag{14}$$

3. Construct the following CDF

$$c_g = c_{g-1} + w_g \; g = 1, ..., G. \tag{15}$$

[3]Note that we have experimented with different MCMC setups, including a version of the MH algorithm with an APF step for Ψ. We found that the Gibbs sampler version reported here is more stable and that it ensures convergence of the parameters as the chain progresses.

[4]We assume that $\psi(t_{-1})^{(g)}$ are zeros, thus $\psi(t_0)|\mathcal{F}_0, \theta^{(j)} \sim N(0, 1)$.

and starting from c_1 find the first c_g which is greater than u, where u is drawn from $U(0,1)$.

4. Select the associated $\widehat{\psi}(t_i)^{*(g)}$ and $\psi(t_{i-1})^{(g)}$ and set $\widehat{\psi}(t_i)^{*(k_1)} = \widehat{\psi}(t_i)^{*(g)}$ and $\psi(t_{i-1})^{(k_1)} = \psi(t_{i-1})^{(g)}$.

5. Repeat steps 3 to 4 R times to obtain $\{\widehat{\psi}(t_i)^{(k_1)}, ..., \widehat{\psi}(t_i)^{(k_R)}\}$ and $\{\psi(t_{i-1})^{(k_1)}, ..., \psi(t_{i-1})^{(k_R)}\}$.

6. For each k_l, simulate

$$\psi(t_i)^{*(l)} \sim N(\rho^{(j)}\psi(t_{i-1})^{(k_l)}, 1) \ l = 1, ..., R \tag{16}$$

and compute

$$w_l^* = \frac{p(z_i|\theta^{(j)}, \psi(t_i)^{*(l)}, \rho^{(j)})}{p(z_i|\theta^{(j)}, \widehat{\psi}(t_i)^{*(k_l)}, \rho^{(j)})}. \tag{17}$$

7. Construct the CDF for π_l^*

$$c_l^* = c_{l-1}^* + \frac{w_l^*}{\sum_{r=1}^{R} w_r^*} \ l = 1, ..., R \tag{18}$$

and starting from c_1^* find the first c_l^* which is greater than u^*, where u^* is drawn from $U(0,1)$.

8. Select the associated $\psi(t_i)^{*(l)}$ and set $\psi(t_i)^{(1)} = \psi(t_i)^{*(l)}$.

9. Repeat step 7 G times to obtain $\{\psi(t_i)^{(1)}, ..., \psi(t_i)^{(G)}\}$. Note that these draws are deemed to be from $p(\psi(t_i)|\mathcal{F}_{t_i}, \theta^{(j)}, \rho^{(j)})$.

10. Compute the likelihood[5] at i

$$\widehat{p}(z_i|\theta^{(j)}, \rho^{(j)}) = \left(G^{-1}\sum_{g=1}^{G} w_g\right)\left(R^{-1}\sum_{l=1}^{R} w_l^*\right). \tag{19}$$

11. Take expectation of

$$\widetilde{\psi}(t_i)^{(j+1)} = G^{-1}\sum_{g=1}^{G} \psi(t_i)^{(g)}. \tag{20}$$

[5] See Pitt (2002).

12. Set $i = i + 1$ and return to step 2 until $i = N$.

13. Compute the likelihood for the whole sample

$$\widehat{p}(z|\theta^{(j)}, \rho^{(j)}) = \prod_{i=1}^{N} \widehat{p}(z_i|\theta^{(j)}, \rho^{(j)}). \tag{21}$$

3.1.2 Metropolis-Hastings algorithm

1. Given $\theta^{(j)}$, generate a candidate $\theta^{*(j)}$ from a random walk transition density $q(\theta^{*(j)}, \theta^{(j)})$.

2. Calculate the acceptance probability

$$\alpha(\theta^{(j)}, \theta^{*(j)}) = \min\left[\frac{p(\theta^{*(j)})\widehat{p}(z|\theta^{*(j)}, \rho^{(j+1)})}{p(\theta^{(j)})\widehat{p}(z|\theta^{(j)}, \rho^{(j+1)})}, 1\right] \tag{22}$$

   where $\widehat{p}(z|\theta^{*(j)}, \rho^{(j+1)})$ and $\widehat{p}(z|\theta^{(j)}, \rho^{(j+1)})$ are computed from the APF algorithm.

3. Generate an independent random variable u from $U(0, 1)$.

4. Set $\theta^{(j)} = \theta^{*(j)}$ if $u < \alpha(\theta^{(j)}, \theta^{*(j)})$ or else $\theta^{(j)} = \theta^{(j)}$.

5. Repeat steps 1 to 4 J times. Note that this is to allow for burn-in.

6. Set $\theta^{(j+1)} = \theta^{(j)}$.

3.2 Marginal likelihood and likelihood function estimation

Marginal likelihood

An important part of the analysis is to compare competing models. In the Bayesian context, one approach to model comparison is to evaluate the marginal likelihoods of different models. These marginal likelihoods measure how well the competing models predict the observed data. The marginal likelihood is defined as the integral of the likelihood function with respect to the prior density:

$$p(\mathcal{F}_T) = \int p(\theta, \rho)p(z|\theta, \rho)d\theta d\rho. \tag{23}$$

where $p(z|\theta, \rho) = \prod_{i=1}^{N} \int p(z_i|\theta, \rho, \mathcal{F}_{t_{i-1}})p(\psi(t_{i+1})|\mathcal{F}_{t_i}, \theta, \rho)d\psi(t_{i+1})$.

However, in our paper, the marginal likelihoods are analytically impossible to compute. We have adopted the Modified Harmonic Mean (MHM) method of Gelfand and Dey (1994) to obtain the marginal likelihood. The advantage of this approach is that it can be employed along with almost any sampling technique and it uses the posterior parameters' draws to obtain the marginal likelihood:

$$\widehat{p}(\mathcal{F}_T)^{-1} = \frac{1}{J}\sum_{j=1}^{J} \frac{f(\theta^{(j)}, \rho^{(j)})}{p(\theta^{(j)}, \rho^{(j)})\widehat{p}(z|\theta^{(j)}, \rho^{(j)})} \tag{24}$$

where $f(\theta, \rho)$ is a density function with supported constraint within the posterior support of (θ, ρ) and which ideally approximates the posterior pdf. Geweke (1999) suggests a truncated multivariate normal distribution with different sets of truncation values, $\delta \in (0, 1)$, for $f(\theta, \rho)$:

$$\theta \sim N\left(\begin{array}{c} \widehat{\theta} \\ \widehat{\rho} \end{array}, \widehat{\Sigma}_{\theta,\rho}\right)_{I(\Gamma)} \tag{25}$$

where $\widehat{\theta} = J^{-1}\sum\limits_{j=1}^{J}\theta^{(j)}, \widehat{\rho} = J^{-1}\sum\limits_{j=1}^{J}\rho^{(j)}$,and $\widehat{\Sigma}_{\theta} = J^{-1}\sum\limits_{j=1}^{J}\left[\begin{array}{c} \theta^{(j)} - \widehat{\theta} \\ \rho^{(j)} - \widehat{\rho} \end{array}\right]\left[\begin{array}{c} \theta^{(j)} - \widehat{\theta} \\ \rho^{(j)} - \widehat{\rho} \end{array}\right]'$. $I(\Gamma)$ is an indictor function such that $I(\Gamma) = 1$ if $\left(\left[\begin{array}{c} \theta^{(j)} - \widehat{\theta} \\ \rho^{(j)} - \widehat{\rho} \end{array}\right]'\widehat{\Sigma}_{\theta}^{-1}\left[\begin{array}{c} \theta^{(j)} - \widehat{\theta} \\ \rho^{(j)} - \widehat{\rho} \end{array}\right]\right) \leq$ q where q is such that $P(\chi^2_a < q) = \delta$ and a is the dimension of (θ, ρ). Note that in computing $f(\theta, \rho)$ an additional normalising constant δ is added to ensure $f(\theta, \rho)$ integrates to unity.

It is worth noting that the marginal likelihood will allow one to test which alternative specifications is important for predicting the probabilities of transitions. However, if there is no clear 'winning' model, one may wish to account for this uncertainty by combining the dynamics of several specifications. This can be achieved by averaging the dynamics of each candidate model with the weights being the posterior model probabilities.

4 Simulation study

To check that our technique for obtaining the posterior distributions of the model parameters works in practice, we apply the algorithm to a simulated data set with $K = 300$ units. The DGP allows for three possible states with the third state being an absorbing state (hence $S = 4$). We further assume that the starting process of $\psi(t_i)$ is zero, and the signs of α_s are restricted such that $\alpha_s : \alpha_{down} < 0$ for transitions to a lower state and $\alpha_s : \alpha_{up} > 0$ for transitions to a higher state. The parameter values that were used to simulate the data are shown in the first column of table 1. The interval between transition events is determined by an exponential distribution with $\sum_{k=1}^{K} \sum_{s=1}^{S} \lambda_{sk}(t_i)$. An univariate multinomial distribution is used to determine which k^{th} unit is experiencing a transition at t_i with probabilities of transition given as

$$\pi_k(t_i) = \frac{\sum_{s=1}^{S} \lambda_{sk}(t_i)}{\sum_{l=1}^{K} \sum_{s=1}^{S} \lambda_{sl}(t_i)}, \quad k = 1, ..., K. \tag{26}$$

Next, we determine the k^{th} unit state from a multinomial distribution with the probability of the type of transition given by

$$\pi_{sk}(t_i) = \frac{\lambda_{sk}(t_i)}{\sum_{l=1}^{S} \lambda_{lk}(t_i)}, \quad s = 1, ..., S. \tag{27}$$

The DGP will end when all the units have entered the absorbing default state.

The priors for the parameters are assumed to be uninformative with the hyper-parameters outlined in the third column of table 1. The sampling scheme[6] was run for 12000 draws and the first 2000 draws were discarded.

From table 1, we note that the true values are close to the posterior means of (θ, ρ) and within the $\pm$ one standard deviation range. In other words, as shown in figure 1, the true values are well within their respective marginal posterior pdfs. In addition, we also perform the convergence diagnostic test of Heidelberger and Welch to check for convergence in draws of (θ, ρ). This test suggest that the draws of (θ, ρ) have converged. Finally, figure 2 shows that the posterior mean of $\psi(t_i)$ and the actual latent path are almost identical with a correlation which is close to 1. These results indicate that our proposed sampling scheme has successfully estimated the MLFI model.

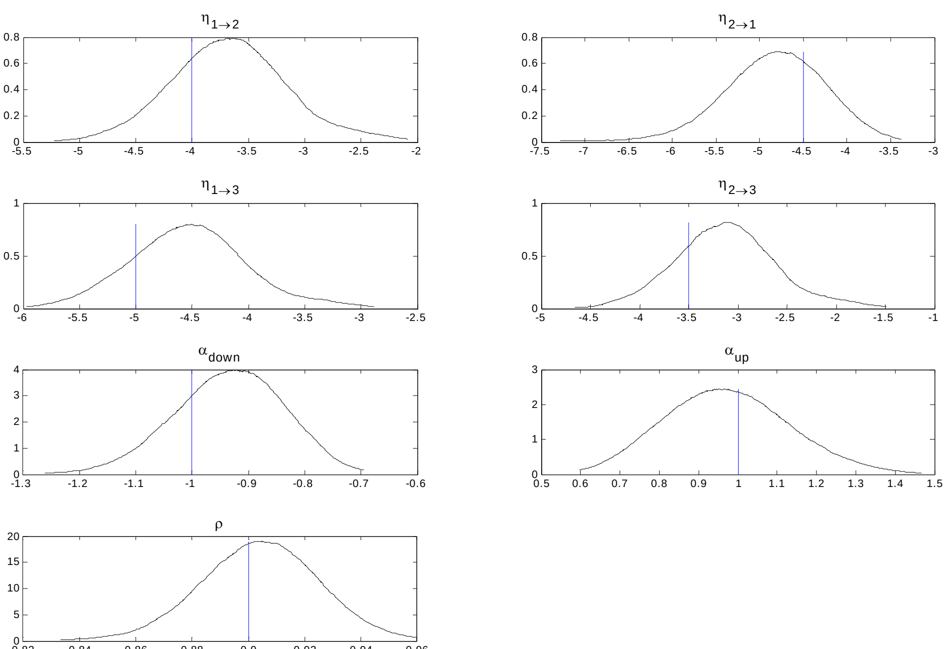
[6] The program used for the estimation is written in Matlab code. In the algorithm, we also set $G = 100, R = 200$, and $J = 200$.

Table 1: Simulated data from the multi-state latent factor intensity model

			Posterior	Heidelberger &
Parameter	True Value	Prior	Mean	Welch test
$\eta_{1\to2}$	-4.0	$N(0,100)$	-3.7305	passed
			(0.4765)	
$\eta_{2\to1}$	-4.5	$N(0,100)$	-4.7658	passed
			(0.5380)	
$\eta_{1\to3}$	-5.0	$N(0,100)$	-4.5896	passed
			(0.4848)	
$\eta_{2\to3}$	-3.5	$N(0,100)$	-3.1850	passed
			(0.4752)	
α_{up}	1.0	$N(0,100)_{\alpha_{up}>0}$	0.9765	passed
			(0.0946)	
α_{down}	-1.0	$N(0,100)_{\alpha_{down}<0}$	-0.9356	passed
			(0.1516)	
ρ	0.9	$U(-1,1)$	0.9038	passed
			(0.0202)	

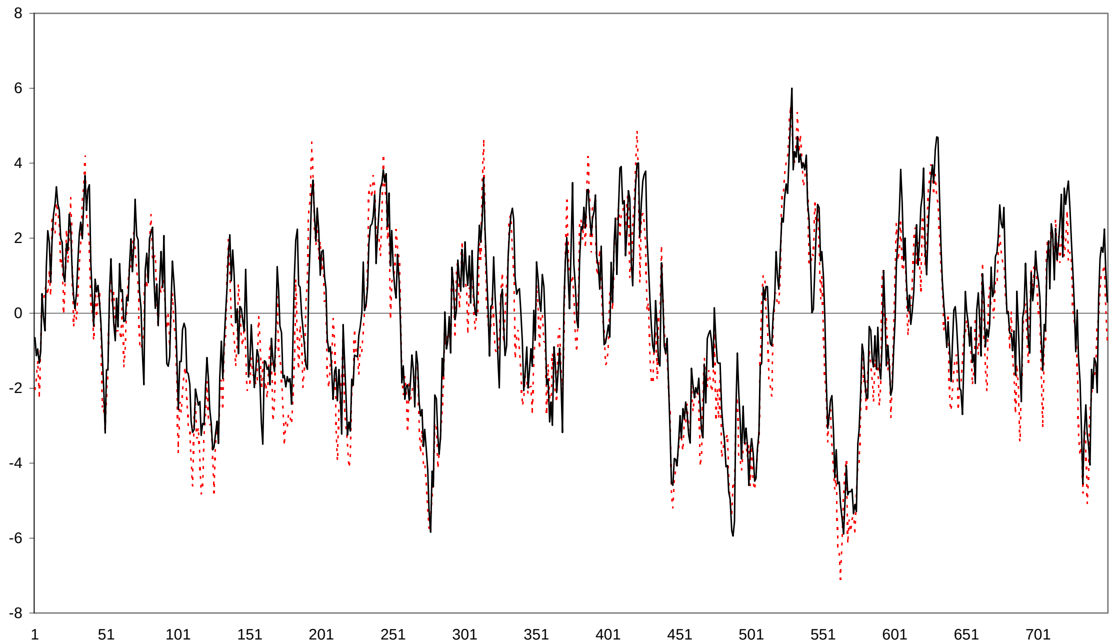
The standard deviations are given in the parentheses.

Figure 1. Estimated marginal posterior pdfs and actual values



The vertical lines are the actual values.

Figure 2. Posterior estimates and actual[7] latent values of $\psi(t)$



The thin line is the posterior mean and the dotted line is the path of of $\psi(t)$.

5 Empirical Study of S&P Credit Ratings

Migration ratings transition matrices play an important role in credit decisions and they have become even more important under the Basel II Capital accord because ratings can be used to determine the size of a bank's capital buffer. A good understanding of the dynamic behavior of migration transition matrices and its co-variation with economic conditions is thus important from the perspective of both the finance industry and the regulatory authorities. In this empirical section, we consider the monthly migration credit ratings of 1049 firms for the sample period between 1995:02 to 2005:04. This data is sourced from the Standard & Poor's CompuStats database.

We classify the S&P ratings into 4 rating groups (defined below) and

[7]For a neater presentation, the 95% confidence interval is not plotted in figure 2, however, the actual path is well within the confidence interval.

consider transitions between these broad groups.

$$\begin{aligned} A &= \{AAA, AA+, AA, AA-, A+, A, A-\} \\ B &= \{BBB+, BBB, BBB-, BB+, BB, BB-, B+, B, B-\} \\ C &= \{CCC+, CCC, CCC-, CC+, CC, CC-, C+, C, C-\} \\ D &= Default \end{aligned}$$

As shown in Table 2, rather than focusing on 12 possible transitions, because there are negligible transitions for the cases: $A \rightarrow C$, $A \rightarrow D$, $C \rightarrow A$, $D \rightarrow A$ and $D \rightarrow C$ in the sample period, we are left with 7 possible transitions. These are $A \rightarrow B$, $B \rightarrow A$, $B \rightarrow C$, $B \rightarrow D$, $C \rightarrow B$, $C \rightarrow D$ and $D \rightarrow B$.

Table 2: Number of Transitions in the sample period

	To			
From	A	B	C	D
A	–	369	1	4
B	222	–	459	98
C	3	99	–	36
D	1	57	2	–

In addition to restricting $\alpha_s : \alpha_{down} < 0$ for transitions to a lower state and $\alpha_s : \alpha_{up} > 0$ for transitions to a higher state, we restrict $\gamma_s : \gamma_{down} < 0$ for transitions to a lower state and $\gamma_s : \gamma_{up} > 0$ for transitions to a higher state. We assume that the priors for the parameters are fairly uninformative but proper whereby the prior hyper-parameters are $\underline{\eta_s} = \ln \sum_{i=1}^{N} \sum_{k=1}^{K} Y_{sk}(t_i) - \ln \sum_{i=1}^{N} \sum_{k=1}^{K} R_{sk}(t_i)$, $\underline{\alpha}_s = \underline{\gamma}_s = 0$, and $\sigma^2_{\eta_s} = \sigma^2_{\alpha_s} = \sigma^2_{\gamma_s} = 100^2$. Note that $\underline{\eta_s}$ in this case is the ML estimate. The duration between rating events is monthly and this example may be viewed as a discrete time version of the multi-state latent factor intensity model with more than one firm transiting each month. We allow for sampling scheme of 12000 draws, and discard the first 2000 draws as burns-in. Table 1 reports the estimates of the parameters and the marginal likelihoods for five models.

Columns 2 and 3 contain the estimates generated from the simpliest model: $\lambda_{sk}(t_i) = R_{sk}(t_i)\exp(\eta_s)$. Column 2 contains the maximum likelihood estimator of η_s while column 3 contains their Bayesian counterparts[8].

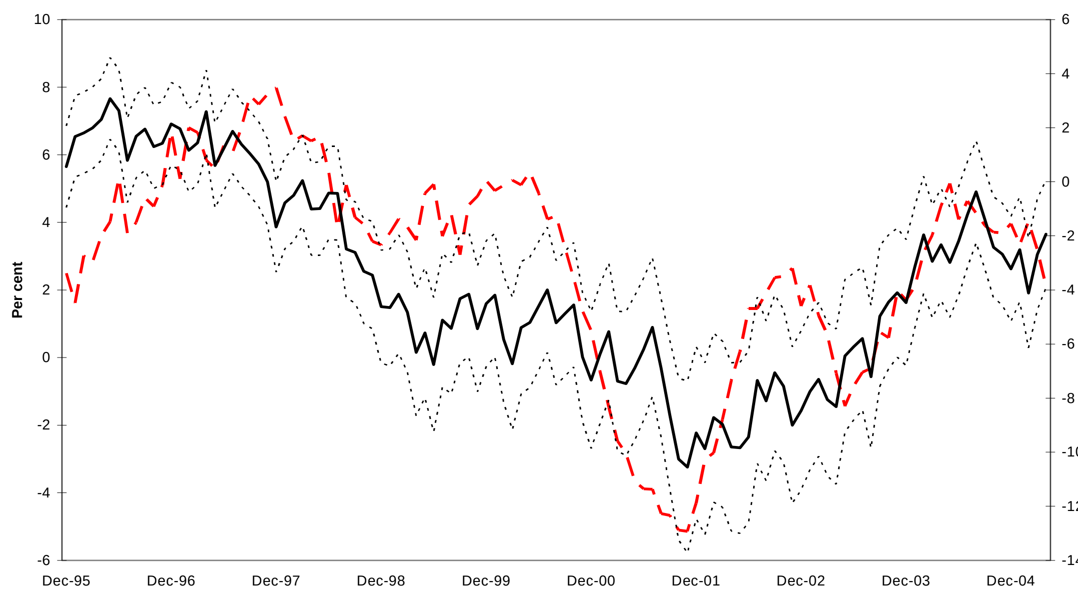
[8]The Metropolis-Hasting algorithm is used to estimate this model.

As expected with fairly uninformative priors, the Bayesian estimates are almost identical to those of the MLE.

Columns 4 and 5 report estimates for models which include a macroeconomic variable, i.e. $\lambda_{sk}(t_i) = R_{sk}(t_i)\exp(\eta_s + \gamma_s w_k(t_i))$ where $w_k(t_i)$ is a macroeconomic variable. M_2 is the model which includes the growth of the industrial production index (IPI), while M_3 is the model which includes the growth of employment (Emp). A comparison of the log marginal likelihood estimates show that M_2 and M_3 have informational gain over M_1.

Next, we consider a model with an unobserved common component and denote this model as M_4. The log marginal likelihood show that M_4 has considerable information gain over the preceding models. In the last model M_5, we relax the assumption of a common latent variable, instead we allow for state-dependent latent variables such that each $\psi^s(t_i)$ follows a random walk process. However, the log marginal likelihood shows that M_5 is not an improvement over the model with a common latent component. This may not be too surprising as 60% of total migration ratings are from A to B and B to C.

Figure 3. Paths of the estimated unobserved common component and the industrial production index



The dashed line is the year-ended growth of industrial production index. The dotted lines are the 95 % confidence interval.

Table 3: Posterior means, posterior standard deviations and Marginal likelihoods

Parameter	MLE	M_1	M_2	M_3	M_4	M_5
log $p(z\|M_i)$		-8313	-8191	-8236	-8134	-8156
ρ	–	–	–	–	0.9713	1
					(0.0239)	
$\eta_{A\to B}$	-4.4917	-4.4923	-4.1218	-4.0211	-5.2960	-5.7446
		(0.0515)	(0.0536)	(0.0595)	(0.3321)	(0.9050)
$\eta_{B\to A}$	-5.7531	-5.7560	-5.8749	-5.8343	-5.6279	-5.7245
		(0.0663)	(0.0857)	(0.0842)	(0.1195)	(0.6363)
$\eta_{B\to C}$	-5.0267	-5.0264	-4.6545	-4.5577	-5.8182	-6.6572
		(0.0464)	(0.0500)	(0.0548)	(0.3290)	(0.8709)
$\eta_{B\to D}$	-6.5708	-6.5866	-6.1984	-6.1104	-7.3680	-8.2878
		(0.1004)	(0.1048)	(0.1067)	(0.3445)	(1.3502)
$\eta_{C\to B}$	-5.3140	-5.3186	-5.3860	-5.3751	-5.0671	-4.7074
		(0.1007)	(0.1021)	(0.1073)	(0.1493)	(0.9567)
$\eta_{C\to D}$	-6.3256	-6.3333	-6.1796	-6.0438	-7.4239	-5.1199
		(0.1650)	(0.1637)	(0.1611)	(0.3703)	(1.2388)
$\eta_{D\to B}$	-4.6576	-4.6679	-4.7401	-4.7183	-4.3857	-5.3805
		(0.1339)	(0.1343)	(0.1388)	(0.1855)	(1.4483)
α_{down}	–	–	–	–	-0.1952	-0.2650
					(0.0438)	(0.0556)
α_{up}	–	–	–	–	0.0585	0.2088
					(0.0208)	(0.0780)
γ_{down}	–	–	-13.4419	-0.4114	–	–
			(0.8655)	(0.0319)		
γ_{up}	–	–	3.2337	0.0584	–	–
			(1.4934)	(0.0373)		

The plot in Figure 3 suggests that the unobserved common component in M_4 co-move with various phases of the US economy. The rise and fall of $\widehat{\Psi}$ between the 1995 to 2001 are in tandem with the boom and bust of the dot com episode, and the trough of $\widehat{\psi}_{t_i}$ correspond to the US recession of 2001. In addition, the correlations between IPI and $\widehat{\Psi}$, and Emp and $\widehat{\Psi}$ are about 0.69 and 0.55, respectively, and they provide further evidence of co-movement. Indeed, this finding is in line with several empirical studies (Nickell et al. (2000), Bangia et al. 2002, Koopman and Lucas (2005), Hu et al. (2002)) that systematic credit risk factor correlate with macro-economic conditions. The extracted latent variable is acting as a broad-based general proxy for macroeconomic conditions.

6 Conclusions

This paper has proposed a Bayesian approach to estimate a multi-state latent factor intensity model to manage the problem of highly analytically intractable pdfs. Our proposed sampling algorithm to estimate the model included an APF filter step and a MH step within a Gibbs sampler. In order to access the feasiblity and accuracy of the sampling algorithm, we conducted a simulated example. Estimation of the parameters and the latent variable were inline with the simulated values. We applied the method to the case of credit ratings transistion matrices from S&P 500. The results revealed that the latent variable provided a good proxy for broad-based macroeconomic conditions. In fact, the latent variable was a better proxy for macroeconomic conditions than either the industrial production index or the employment variable.

References

[1] Aguilar, O., and West, M. (2000), 'Bayesian Dynamic Factor Models and Portfolio Allocation', *Journal of Business and Economic Statisitcs*, 18, 338-357.

[2] Bangia, A., Diebold, F.X., Kronimus, A,. Schagen,C. and Schuermann,T. (2002), 'Ratings migration and the Business Cycle with appli-

cations to credit portfolio stress testing', *Journal of Banking & Finance*, 26, 445-474.

[3] Bauwens, L. and Veredas, D. (2004), 'The Stochastic Conditional Duration Model: A Latent Variable Model for the Analysis of Financial Durations', *Journal of Econometrics*, 119, 381-412.

[4] Chib, S., Nardari, F., and Shephard, N. (2002), 'Markov Chain Monte Carlo Methods for Stochastic Volatility Models', *Journal of Econometrics*, 108, 281-316.

[5] Das, S.R., Fan, R., and Geng, G. (2002) 'Bayesian Migration in Credit Ratings Based on Probabilities of Default', *Journal of Fixed Income*, 12, 17–23.

[6] Doucet, A. Godsill, S. and Andrieu C. (2000), 'On Sequential Monte Carlo Sampling Methods for Bayesian Filtering', *Statistics and Computing*, 10, 197-208.

[7] Fiorentini G., Sentana, E., and Shephard. N. (2004), 'Likelihood-based Estimation of Latent Generalised ARCH Structures', *Econometrica*, 72, 1481-1517.

[8] Gelfand, A.E. and Dey, D.K. (1994), 'Bayesian Model Choice: Asymptotics and Exact Calculations', *Journal of the Royal Statistical Society Series* B, 56, 501-514.

[9] Geweke, J. (1999), 'Using Simulation Methods for Bayesian Econometric Models: Inference, Development, and Communciation', *Econometric Reviews* 18, 1-73.

[10] Hu,Y.T., Kiesel R., and Perraudin W. (2002), 'The Estimation of Transition Matrices for Sovereign Credit Ratings', *Journal of Banking and Finance*, 26, 1383-1406.

[11] Koopman, S.J., Lucas, A. and Monteiro, A.A. (2008), 'The Multi-State Latent Factor Intensity Model for Credit Rating Transitions', *Journal of Econometrics, 142,*399-424.

[12] Koopman, S.J., and Lucas, A. (2005), 'Business and Default Cycles for Credit Risk', *Journal of Applied Econometrics*, 20, 311-323.

[13] McNeil, A.J., and Wendin, J. (2007), 'Bayesian Inference for Generalized Linear Mixed Models of Portfolio Credit Risk', *Journal of Empirical Finance,* 14, 131-149.

[14] Nickell, P., Perraudin, W. and Varotto, S. (2000), 'Stability of Rating Transitions', *Journal of Banking & Finance*, 10, 423-444.

[15] Pitt, M. and Shephard, N. (1999), 'Filtering via simulation: auxiliary particle filter," *Journal of the American Statistical Association* 94, 590-599.