

A dynamic Bayesian approach for probability of default and stress test

Taeyoung Kim^a, Yousung Park^{1,b}

^aKorea Enterprise Data Co. Ltd., Korea; ^bDepartment of Statistics, Korea University, Korea

Abstract

Obligor defaults are cross-sectionally correlated as obligors share common economic conditions; in addition obligors are longitudinally correlated so that an economic shock like the IMF crisis in 1998 lasts for a period of time. A longitudinal correlation should be used to construct statistical scenarios of stress test with which we replace a type of artificial scenario that the banks have used. We propose a Bayesian model to accommodate such correlation structures. Using 402 obligors to a domestic bank in Korea, our model with a dynamic correlation is compared to a Bayesian model with a stationary longitudinal correlation and the classical logistic regression model. Our model generates statistical financial statement under a stress situation on individual obligor basis so that the generated financial statement produces a similar distribution of credit grades to when the IMF crisis occurred and complies with Basel IV (Basel Committee on Banking Supervision, 2017) requirement that the credit grades under a stress situation are not sensitive to the business cycle

Keywords: Bayesian model, dynamic longitudinal correlation, probability of default, stress test

1. Introduction

Credit risk is the risk that an obligor or group of obligors fail to pay the timely payment of interest or repay the amount borrowed. The most important one in measuring the credit risk is how to estimate the probability that an obligor does not pay back in the future as the probability of default plays a key role in pricing, portfolio management, and capital allocation. Two main issues have been raised in estimating the probability of default. The first issue is how to include default correlations among obligors and over different time points in estimating the probability of default. Obligor maintain close relationship with sellers, buyers, and commercial transactions. A company may face an increased risk if one of its major customers goes bankrupt, implying that cross-sectional correlations exist among obligors due to common economic conditions (Davis and Lo, 2001; Egloff *et al.*, 2004; Li, 2016). Additionally, longitudinal dependence in default rates also has been investigated (e.g., Wendin, 2006; Bade *et al.*, 2011). Longitudinal dependency is caused by the cyclical behavior of the default intensity varying over time (Hu *et al.*, 2002; Wendin, 2006; Duffie *et al.*, 2007; Heynderickx *et al.*, 2016) so that a particular obligor's possibility of default is influenced by the past probability of default.

Wendin (2006) used a generalized linear mixed model to reflect the cross-sectional and longitudinal correlations in calculating the probability of default. However, this model belongs to an area level model due to the high dimensional problem that arises when an obligor-specific unit level model is

¹ Corresponding author: Department of Statistics, Korea University, 145 Anam-ro, Seongbuk-gu, Seoul 02841, Korea.
E-mail: yspark@korea.ac.kr

used. To resolve the dimensional problem, we use the method of CreditMetrics proposed by J.P Morgan (Xiao, 2002) and use a new dynamic time series model without a presumed trend for longitudinal dependence of defaults. Dynamic state-space Bayesian models have been proposed for longitudinal dependence of multivariate time series (Cargnoni *et al.*, 1997; Reilly *et al.*, 2001; Gosh *et al.*, 2014). Among them, Cargnoni *et al.* (1997) and Reilly *et al.* (2001) used a pure random walk process to reflect the dynamic longitudinal dependence of a multinomial time series; therefore, they assumed that the multinomial time series contains a stochastic trend without any deterministic trend. However, the trend should be tested to examine if it is deterministic or stochastic rather than assumed as the trend types have totally different implications in the longitudinal behavior of defaults (Greene, 2017).

The second issue is to adequately assess the ability that an obligor contractually perform under adverse economic conditions or an occurrence of unexpected events referred to as stress circumstances. Stress situations dramatically change financial statements and make the possibility of an obligor's default higher. Banks want to know each obligor's ability under a stress situation at a future time; therefore, banks estimate extreme values of financial variables related to the probability of default, implying that the distributions of the relevant financial variables are assumed to know. However, this is not possible in the real loan world mainly because of too small number of default obligors. To overcome this difficulty, we suggest an estimating function of the probability of default that is symmetric about financial variables and the corresponding coefficients. Owing to this symmetry, we estimate percentiles of the financial variables from the percentiles of the corresponding coefficients so that we can predict individual obligor's probability of default under a certain stress situation.

We include the latent variables of Albert and Chib (1993) in our hierarchical Bayesian model to accommodate cross-sectional correlations among obligors and longitudinal correlations between different time points. The cross-sectional correlations are embodied by a regression model of latent variables on financial variables, whereas longitudinal correlations are reflected by non-stationary regression coefficients depending on past values of the coefficients.

The remainder of this paper is divided into four sections. In Section 2, we present a dynamic hierarchical Bayesian model and an additional Bayesian model with a stationary correlation structure. This stationary Bayesian model is used to discuss the advantage of our dynamic correlation structure. Full conditional distributions are also presented to implement the MCMC to calculate the probability of default. In Section 3, we apply our model to 402 obligors to a domestic bank in Korea. Our model is compared with the stationary Bayesian model and the classical regression model. We show how to estimate the distribution of financial variables through the corresponding regression coefficients to construct stress situations and to predict the probability of default under a stress situation. The credit ratings from the probability of default under a stress situation are compared to the observed crediting ratings at IMF crisis in 1998. We also examine whether or not our dynamic correlation structure preserves the lasting impact of a stress situation by using the transition matrix of credit rates. Section 4 includes the concluding remarks.

2. A dynamic hierarchical Bayesian model

Let y_{it} be independent Bernoulli random variable with $y_{it} = 1$ when the i^{th} obligor is default at year t and $y_{it} = 0$ otherwise. Here, because the observed number of obligors may not be the same at each time point, i ranges from 1 to I_t for $t = 1, 2, \dots, \mathcal{T}$. Let $\Pr[y_{it} = 1] = \pi_{it}$ be the probability of default. This π_{it} links to p -dimensional covariate \mathbf{x}_{it} by $\pi_{it} = H(\mathbf{x}_{it}^T \mathbf{b})$ where \mathbf{b} is $p \times 1$ parameter to be estimated and H is a link function such as a logistic or probit function. One difficulty arising from this specification of π_{it} is that the posterior of \mathbf{b} cannot be expressed by an explicit form because of

discreteness of y_{it} (Albert and Chibs, 1993). Thus, we use a continuous latent variable z_{it} defined as $y_{it} = 1$ if $z_{it} > 0$ and $y_{it} = 0$ otherwise (Albert and Chibs, 1993; Oh *et al.*, 2003). We then assume that the latent variable z_{it} follows a t -distribution with location parameter $\mathbf{x}_{it}^T \mathbf{b}$ and scale parameter $1/0.634^2$:

$$z_{it} \sim t_8 \left(\mathbf{x}_{it}^T \mathbf{b}, \frac{1}{0.634^2} \right). \quad (2.1)$$

As shown by Albert and Chibs (1993), we have

$$\pi_{it} = P[y_{it} = 1] = P[z_{it} > 0 \mid \mathbf{b}] \approx \frac{e^{\mathbf{x}_{it}^T \mathbf{b}}}{1 + e^{\mathbf{x}_{it}^T \mathbf{b}}}.$$

This indicates that latent variable z_{it} can be used as an approximation of the logistic function. It was also shown by Albert and Chibs (1993) that the t -distribution given by (2.1) can be rewritten as a scale mixture normal distribution :

$$z_{it} \mid \lambda, \mathbf{b} \sim N \left(\mathbf{x}_{it}^T \mathbf{b}, \frac{1}{0.634^2 \lambda} \right), \quad (2.2)$$

where λ is a random variable with the Gamma distribution denoted by $G(8/2, 8/2)$.

We now modify (2.2) for the distribution of z_{it} to reflect cross-sectional and longitudinal correlations as given by

$$z_{it} = \mathbf{x}_{it}^T \mathbf{b}_t + \xi_{it}, \quad (2.3)$$

where ξ_{it} is normally distributed with mean 0, variance $1/(0.634^2 \lambda_t)$ where λ_t is independently distributed with $G(8/2, 8/2)$, and $\text{Corr}(\xi_{it}, \xi_{it'}) = \rho_{ii'}$ when $t = t'$ and 0 otherwise. We also assume that \mathbf{b}_t in model (2.3) varies over time t with the following prior of dynamics:

$$\Delta \mathbf{b}_t = \mathbf{\Gamma} \Delta \mathbf{b}_{t-1} + \boldsymbol{\epsilon}_t \quad \text{and} \quad \boldsymbol{\epsilon}_t \sim N(0, \mathbf{\Psi}) \quad \text{for } t = 1, 2, 3, \dots, \mathcal{T}, \quad (2.4)$$

where $\Delta \mathbf{b}_t = \mathbf{b}_t - \mathbf{b}_{t-1}$, and $\mathbf{\Gamma} = \text{Diag}(\gamma_1, \gamma_2, \dots, \gamma_p)$ and $\mathbf{\Psi} = \text{Diag}(\psi_1, \psi_2, \dots, \psi_p)$ with dimension p .

The dynamics of the latent variable z_{it} in model (2.3) is regulated by model (2.4) and behaves like a process with a stochastic trend or deterministic trend according to estimated $\mathbf{\Gamma}$ in which an unexpected impact such as a stress circumstance is permanent on b_{kt} (the k^{th} element of \mathbf{b}_t) when $\hat{\gamma}_k = 0$ for $k = 1, \dots, p$ whereas it is temporary when $\hat{\gamma}_k < 0$. The cross-sectional correlation $\rho_{ii'}$ between obligors i and i' in model (2.3) can be suffered from a curse of dimensionality due to large number of obligors. We approximate $\rho_{ii'}$ by the correlation between obligor's returns in equity market such as the CreditMetrics currently used in J.P Morgan. The CreditMetrics is defined by $\rho_{ii'} = w_i \times w_{i'}$ where w_i is the square root of the determinant coefficient (i.e., R^2) obtained from a simple regression with the return of obligor i as a regress and the index return (e.g., S & P 500) as a regressor. To avoid a high dimensionality in estimating $\rho_{ii'}$, we use these CreditMetrics as a proxy of the cross-correlation $\rho_{ii'}$ and treat them as covariates.

To complete our model specification, independent non-informative uniform priors of $(-1, 1)$ are imposed on γ_j and an Inverse Gamma distribution $\text{IG}(a_\psi, b_\psi)$ on ψ_j where a_ψ and b_ψ are known.

All of these model specifications produce the following joint posterior distribution.

$$\begin{aligned}
z_{ii}, \mathbf{b}_t, \Gamma, \Psi, \lambda_t \mid y_{ii} \sim & \prod_{t=1}^{\mathcal{T}} \prod_{i=1}^{I_t} [I(z_{ii} > 0, y_{ii} = 1) + I(z_{ii} \leq 0, y_{ii} = 0)] \\
& \times \prod_{t=1}^{\mathcal{T}} \left[0.634 \lambda_t^{\frac{1}{2}} \mid \mathbf{R}_t \mid^{-\frac{1}{2}} \exp \left(-\frac{\lambda_t}{2} (\mathbf{z}_t - \mathbf{X}_t \mathbf{b}_t)^T \mathbf{R}_t^{-1} (\mathbf{z}_t - \mathbf{X}_t \mathbf{b}_t) \right) \right] \\
& \times \prod_{t=1}^{\mathcal{T}} \lambda_t^{\frac{8}{2}-1} \exp \left(-\frac{8}{2} \lambda_t \right) \times \prod_{j=1}^p \left(\frac{1}{\psi_j} \right)^{a_\psi-1} \exp \left(-b_\psi \left(\frac{1}{\psi_j} \right) \right) \\
& \times \prod_{t=1}^{\mathcal{T}} \mid \Psi \mid^{-\frac{1}{2}} \exp \left(-\frac{1}{2} (\mathbf{b}_t - \mathbf{b}_{t-1} - \Gamma \Delta \mathbf{b}_{t-1})^T \Psi^{-1} (\mathbf{b}_t - \mathbf{b}_{t-1} - \Gamma \Delta \mathbf{b}_{t-1}) \right),
\end{aligned}$$

where $\mathbf{z}_t = (z_{t1}, z_{t2}, \dots, z_{tI_t})^T$, \mathbf{R}_t is the correlation matrix of \mathbf{z}_t , and \mathbf{X}_t is the corresponding covariate matrix. This joint posterior distribution is implemented from the following full conditional distributions from (2.5) to (2.9):

$$\mathbf{z}_t \mid \mathbf{b}_t, \lambda_t \sim N \left[\mathbf{X}_t \mathbf{b}_t, \frac{1}{0.634^2 \lambda_t} \mathbf{R}_t \right] \quad (2.5)$$

with z_{it} truncated above (below) if $y_{it} = 1$ ($y_{it} = 0$),

$$\mathbf{b}_t \mid \mathbf{z}_t, \lambda_t, \Psi, \Gamma \sim N \left[\boldsymbol{\mu}_{\mathbf{b}_t}, \Sigma_{\mathbf{b}_t} \right], \quad (2.6)$$

where

$$\begin{aligned}
\Sigma_{\mathbf{b}_t} &= \left[\Psi^{-1} + \mathbf{X}_t^T \left(\frac{1}{0.634^2 \lambda_t} \mathbf{R}_t \right)^{-1} \mathbf{X}_t \right]^{-1}, \\
\boldsymbol{\mu}_{\mathbf{b}_t} &= \Sigma_{\mathbf{b}_t} \left[\Psi^{-1} (\mathbf{b}_{t-1} + \Gamma \nabla \mathbf{b}_{t-1}) + \mathbf{X}_t^T \left(\frac{1}{0.634^2 \lambda_t} \mathbf{R}_t \right)^{-1} \mathbf{z}_t \right],
\end{aligned}$$

the conditional posterior distribution of ψ_j is

$$\psi_j \mid \mathbf{b}_t, \gamma_j, a_\psi, b_\psi \sim \text{IG} \left[\frac{\mathcal{T}}{2} + a_\psi, \frac{\sum_{t=1}^{\mathcal{T}} (b_{tj} - b_{t-1,j} - \gamma_j (b_{t-1,j} - b_{(t-2)j}))^2}{2} + b_\psi \right], \quad (2.7)$$

the conditional posterior distribution of γ_j is

$$\gamma_j \mid \mathbf{b}_t, \psi_j \sim N \left[\frac{\sum_{t=1}^{\mathcal{T}} \nabla b_{tj} b_{t-1,j}}{\sum_{t=1}^{\mathcal{T}} (\nabla b_{t-1,j})^2}, \frac{\psi_j}{\sum_{t=1}^{\mathcal{T}} (\nabla b_{t-1,j})^2} \right], \quad (2.8)$$

and, finally, the conditional distribution of λ_t is

$$\lambda_t \mid \mathbf{z}_t, \mathbf{b}_t \sim \Gamma \left(\frac{1}{2} + 4, \frac{0.634^2 (\mathbf{z}_t - \mathbf{X}_t \mathbf{b}_t)^T \mathbf{R}_t^{-1} (\mathbf{z}_t - \mathbf{X}_t \mathbf{b}_t)}{2} + 4 \right). \quad (2.9)$$

Table 1: Default observations in the service industry

Year	Observations in data	Number of defaults	Annual default rate
2000	78	1	1.28%
2001	102	2	1.96%
2002	111	4	3.60%
2003	111	6	5.41%
Total	402	13	3.23%

Table 2: Financial ratios

Category	Variable name
Profitability	Net Borrowings Dependency Ratio (x_{1t})
Stability	Borrowing Structure (x_{2t})
Activity	Inventory Turnover Ratio (x_{3t})
Activity	Trade Receivables Turnover Ratio (x_{4t})
Cash Flow	Cash Flow/Total Borrowings (x_{5t})
Stability	Long Term Debt Repayment Capacity (x_{6t})

The conditional distributions from (2.5) to (2.9) are iteratively implemented by the MCMC procedures until a convergence criterion is met.

In order to examine the dynamic longitudinal correlation structure defined by (2.4), we consider an additional Bayesian model with a stationary longitudinal correlation structure as given by

$$z_{it} = \mathbf{x}_{it}^T \mathbf{b} + \xi_{it}$$

where ξ_{it} is normally distributed with mean 0, variance $1/(0.634^2\lambda)$, and $\text{Corr}(\xi_{it}, \xi_{i't'}) = \phi^{|t-t'|} \rho_{ii'}$ for $i \neq i'$ in which $\pi(\mathbf{b}) \propto 1$ and $\phi \sim \text{Uniform}(-1, 1)$ are assumed. This stationary model therefore replaces the dynamic longitudinal correlation of (2.4) with a stationary autoregressive process with order 1, and \mathbf{b}_t with \mathbf{b} not depending on time t . We do not present the posterior distribution and full conditional distributions because the derivations are the same as those for the dynamic hierarchical Bayesian model (the derivations can be provided upon request).

3. Empirical study

A domestic bank in Korea collected a data to build its internal credit risk model from 402 companies in total during four years from 2000 to 2003. All these 402 companies belonged to the service industry. We apply this data to the three models, the dynamic hierarchical Bayesian model (dynamic model), the Bayesian model with a stationary correlation structure(stationary model), and the classical logistic model. The time period is taken to see how the stress scenario generated from our dynamic model is close to IMF situations in terms of credit ratings and how long the influence of the generated stress lasts. The time period seems to be a point in time that deviated to some extent from the IMF shock, but economic conditions were not significantly different. As summarized in Table 1, only 13 default companies were observed among them. While developing a credit risk model, six financial ratios are used as covariates and classified into four categories: profitability, stability, activity, and cash flow (Table 2).

We diagnosed the convergence of the estimates in the two Bayesian models based on the MCMC output. This could be done by calculating the potential scale reduction factor (Gelman and Rubin, 1992). We simulate 5 parallel independence sequences with five different starting points: the posterior mean estimate of a parameter, mean \pm its standard deviation, and mean $\pm 2 \times$ its standard deviation.

Table 3: Accuracy ratios of the three models

Dynamic model	Stationary model	Logistic model
0.894	0.766	0.754

Table 4: Accuracy ratios for training data and testing data

Data	Dynamic model	Stationary model	Logistic model
Training data	0.896	0.764	0.756
Test data	0.809	0.652	0.416

Table 5: Posterior means and 95% credibility interval of \mathbf{b}_t

Coefficient	2000	2001	2002	2003
Borrowings dependency (b_{1t})	1.5823 (0.0849, 3.0834)	-2.0383 (-2.7879, -0.9307)	-0.0640 (-0.1911, 0.0449)	0.1296 (0.0212, 0.2866)
Borrowing structure (b_{2t})	0.0990 (-0.0172, 0.1979)	-0.3024 (-0.4520, -0.1265)	-0.0291 (-0.0797, -0.0051)	-0.0003 (-0.0089, 0.0076)
Inventory turnover ratio (b_{3t})	0.1615 (0.0056, 0.3315)	-0.0978 (-0.1428, -0.0427)	0.0097 (0.0028, 0.0167)	0.0227 (0.0064, 0.0422)
Trade receivable ratio (b_{4t})	-1.5504 (-2.2203, -0.2928)	0.2228 (0.0596, 0.3981)	0.0155 (-0.0106, 0.0423)	0.0093 (-0.0065, 0.0285)
Cash flow/total borrowings (b_{5t})	0.0038 (-0.0898, 0.1031)	0.0289 (-0.0309, 0.0837)	-0.0079 (-0.0274, 0.0060)	0.0008 (-0.0037, 0.0061)
Debt repayment capacity (b_{6t})	-6.0417 (-19.9681, 5.3674)	-3.9798 (-9.0627, 0.5776)	-0.9978 (-2.2263, 0.0507)	0.2334 (-0.4738, 1.1359)

Each of the five sequence is simulated with the lengths, 1000, 2000, 5,000, 10,000, 20,000, and 40,000 (after discarding the first half of the simulation). All potential scale reduction factors were close to 1 after 20,000 iteration, indicating that 40,000 simulation runs in our data analysis are acceptable. The running time is an important factor when systemizing the model for a credit risk management system. We calculated the running times of our personal computer for both dynamic and stationary models. The running time of the dynamic model was 30 seconds, while the stationary model was about 18 minutes, showing a clear difference in the running time.

3.1. Model comparisons

In order to compare dynamic, stationary, and logistic models, we use Accuracy Ratio (AR) (Sobehart and Keenan, 2001) to measure model discriminatory ability to differentiate default obligors from non-default obligors. AR ranges from 0 to 1 and the closer AR is to 1, the better. Table 3 shows that the dynamic model is better than the two other models. To check model generalization, we divide the data into two parts: the first part consisting of the obligors from 2000 to 2002 is the training data for model fit, and the second consisting of the 2003 year data is the testing data for model generalization. We fit the two Bayesian models and the logistic regression model to the training data and then apply the fitted models to the testing data to calculate the ARs of the three models (Table 4).

Table 4 shows that the AR of the dynamic model is higher than those of the stationary model and the logistic model. This implies that the dynamic model is more robust than the two other models, and the dynamic correlation structure is more suitable to the real data than the stationary correlation.

From now on, we only discuss the dynamic model. The estimates of \mathbf{b}_t for $t = 2000, 2001, 2002,$ and 2003 and their 95% credibility intervals are presented in Table 5, showing that the cash flow represented by cash flow/total borrowings and the stability represented by debt repayment capacity are

not significant. We also observe that \mathbf{b}_t changes the sign along with time t by its dynamic correlation structure.

To check if obligors have higher credit risk due to close economical relationship, we calculate the Bayes factors with and without cross-sectional correlations. They are -208.2 for with model and -226.9 for without model, indicating that the model with cross-sectional correlations is significantly better than without model by the criterion of Congdon (2014).

3.2. Constructing statistical stress test

Under adverse economic conditions or the occurrence of unexpected events referred to as the stress circumstance, the credit quality of obligors could be changed according to how much the corresponding financial statement is changed. The change of credit quality includes the change in value caused not only by possible default events, but also by upgrades and downgrades in credit quality. Most domestic banks have predicted the obligor's grade under the stress circumstance by making a scenario for the obligor's financial statement expected at the stress situation. However, this scenario might be incorrect and often is not possible for each obligor.

We propose an alternative method for making a statistical stress situation for individual obligor's basis and calculating each obligor's probability of default (PD) under the stress situation. Note, from model (2.3), that each obligor's PD is obtained by the probit model:

$$\pi_{ii} = \Phi(0.634 \sqrt{\lambda_t} \mathbf{x}_{ii}^T \mathbf{b}_t), \quad (3.1)$$

where λ_t and \mathbf{b}_t are estimated under the dynamic model.

This probability π_{ii} is symmetric about the financial ratio \mathbf{x}_{ii} and the regression coefficient \mathbf{b}_t . In other words, an increase of $\mathbf{x}_{ii}^T \mathbf{b}_t$ is expected under a stress situation, and the increase comes from increasing \mathbf{x}_{ii} or increasing \mathbf{b}_t . In a Bayesian perspective, \mathbf{x}_{ii} is a constant but \mathbf{b}_t is a random variable with a completely known posterior distribution. Therefore, instead of predicting \mathbf{x}_{ii} under a stress situation, we choose an appropriate value of \mathbf{b}_t using its posterior distribution so that the chosen value of \mathbf{b}_t yields a high value of the PD which is expected in the stress situation. We then generate each obligor's financial statement under a stress situation by equating these two quantities:

$$\mathbf{x}_{ii}^T \mathbf{b}_t^* = \mathbf{x}_{ii}^{*T} \mathbf{b}_t, \quad (3.2)$$

where \mathbf{x}_{ii} and \mathbf{x}_{ii}^* are financial statements observed in a normal situation and generated in a stress situation for the i^{th} obligor, respectively, and \mathbf{b}_t and \mathbf{b}_t^* are the posterior mean in a normal situation and the presumed value in a stress situation, respectively. The $\mathbf{x}_{ii}^T \mathbf{b}_t^*$ in the left of (3.2) is known because \mathbf{x}_{ii} is observed and \mathbf{b}_t^* is determined by its posterior distribution in right hand side. Therefore, the value of \mathbf{x}_{ii}^* is easily generated by the identity equation of (3.2).

For example, Table 6 shows that for year 2003, the coefficient of 'net borrowings dependency ratio' has mean 0.1296 and its 75th percentiles is 0.1694. This 75th percentile is supposed to the value of 'net borrowings dependency ratio' presumed under a stress situation. An observed value of the 'net borrowings dependency ratio' from an obligor at $t = 2003$ was 1.34. Therefore, since we have $0.2270 = 1.34 \times 0.1694$ under the stress situation, the value of 'net borrowings dependency ratio' under the stress situation becomes 1.75 ($= 0.2270/0.1296$).

The left-hand side of Table 6 shows the posterior means of the six coefficients and the 75th percentiles. The right-hand side of Table 6 provides the mean values of the six financial ratios and those under the stress situation corresponding to the 75th percentiles of the coefficients given in Table 6.

Table 6: Generation of financial ratios under a stress situation (95% credibility intervals in parentheses)

Variable name	Coefficients		Financial ratios	
	Under normal	Under stress	Under normal	Under stress
Borrowings dependency (b_{t1})	0.1296 (0.0212, 0.2866)	0.1694	1.34	1.75
Borrowings structure (b_{t2})	-0.0003 (-0.0089, 0.0076)	0.0021	68.61	0.00
Inventory turnover ratio (b_{t3})	0.0227 (0.0064, 0.0422)	0.0287	49.24	62.36
Trade receivables ratio (b_{t4})	0.0093 (-0.0065, 0.0285)	0.0147	12.24	19.35
Cash flow/total borrowings (b_{t5})	0.0008 (-0.0037, 0.0061)	0.0022	36.67	107.19
Debt repayment capacity (b_{t6})	0.2334 (-0.4738, 1.1359)	0.4620	0.65	1.30

Table 7: The proportion of grades of corporate bonds issued from 2000 to 2003

Grade	AAA	AA	A	BBB	BB	B	CCC ~ C
Portion	6.86%	10.29%	22.19%	32.29%	22.57%	3.80%	2.00%

Source: National Information & Credit Evaluation, Inc

Table 8: The distribution of grades under the stress circumstance

Grade	Normal*	IMF (1998)*	Stress
AAA	6.86%	0.50%	1.71%
AA	10.29%	2.00%	6.75%
A	22.19%	25.50%	21.66%
BBB	32.29%	36.50%	30.13%
BB	22.57%	23.50%	26.85%
B	3.80%	3.50%	7.43%
CCC ~ C	2.00%	1.74%	5.47%
Investment Grade	71.63%	64.50%	60.25%
Junk Grade	28.37%	35.50%	39.75%

Source: National Information & Credit Evaluation, Inc

In order to assign a credit grade to individual obligor's PD, we use the distribution of the grades of corporate bonds issued for four years from 2000 to 2003 as summarized in Table 7. Using (3.1) to calculate PD, we determine six cut-off points of the PDs to identify the seven credit grades matching to the proportion of grades in Table 7. We then calculate the PDs under the stress situations generated from the identity equation (3.2) with \mathbf{b}_t^* being the 75th percentiles of \mathbf{b}_t and apply the PDs to the six cut-off points to obtain the corresponding credit grades under the stress situation. The resulting grades are provided in the last column of Table 8. We compare these predicted grades with the observed grades of the bonds issued in 1988 when IMF occurred, which is listed in the second column of Table 8. The proportions of both grades are close to each other. The difference of the junk grade (BB~C) is as small as 4% point. We also observe that the proportion of the junk grade is predicted to increase by 11.4% point in the stress situations (Table 8).

It is known that the credit grades under a stress situation are not sensitive to the business cycle. Namely, the grades under a stress situation tend to remain in a steady way but not in a dynamic way. As a shock on \mathbf{b}_t spreads to the coming years thanks to the dynamic structure of \mathbf{b}_t , our model fulfills this desirable property as shown in the transition matrix provided in Table 9. For constructing the transition matrix, we obtain the coefficient \mathbf{b}_t^* for $t = 2003$ in the same way as described in Table 6.

Table 9: Transition matrix for grades under a stress circumstance from 2003 to 2004

		T = 2004						SUM	
		AAA	AA	A	BBB	BB	B		CCC ~ C
T = 2003	AAA	1 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (100%)
	AA	0 (0%)	6 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	6 (100%)
	A	0 (0%)	0 (0%)	16 (84.2%)	3 (15.8%)	0 (0%)	0 (0%)	0 (0%)	19 (100%)
	BBB	0 (0%)	0 (0%)	4 (10.5%)	31 (81.6%)	3 (7.9%)	0 (0%)	0 (0%)	38 (100%)
	BB	0 (0%)	0 (0%)	0 (0%)	2 (6.3%)	29 (90.7%)	1 (3.0%)	0 (0%)	32 (100%)
	B	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	6 (85.7%)	1 (14.3%)	7 (100%)
	CCC~C	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	2 (25%)	6 (75%)	8 (100%)

We replace \mathbf{b}_{t-1} in the dynamic of (2.4) with this \mathbf{b}_t^* to forecast \mathbf{b}_t for $t = 2004$, and then apply (3.1) to forecast the PDs at $t = 2004$ for all obligors in 2003. Finally, we use again the same six cut-off points as in Table 8 to predict the 2004 grade of each obligor in 2003. Table 9 is the transition matrix for grades under the stress circumstance from 2003 to 2004. This table shows that the grades between the two consecutive years remain the same in most cases and move only one-step upward or downward, satisfying the desirable property in a stress circumstance.

4. Concluding remarks

We proposed a dynamic hierarchical Bayesian model to accommodate cross-sectional correlations among obligor's defaults in the same industry without a curse of dimensionality, and longitudinal correlations of defaults over successive time points with a steady change of credit grades under stress situation. We showed that the dynamic model outperformed another Bayesian model with a stationary correlation structure as well as the classical logistic regression model in terms of Accurate Ratio and model generalization. We proposed how to statistically construct individual obligor's financial statement under a stress situation. We demonstrated that the generated financial statement produced a similar distribution of credit grades to when the IMF crisis occurred in 1998, and our dynamic structure of regression coefficients complied with Basel IV (2017) requirement that the credit grades under a stress situation are not supposed to be sensitive to the business cycle. There was no choice for comparison with IMF crisis, but the use of old data and no application to the financial crisis of 2007–2008 are the weakness of this study. The applications of our model to credit risk data after the financial crisis is needed if the data is available.

Acknowledgement

I dedicate this paper to the late Taeyoung Kim, the first author. This paper was based on his doctoral dissertation although it was completely rewritten and updated.

References

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

- Bade B, Roesch D, and Scheule H (2011). Default and recovery risk dependencies in simple credit risk, *European Financial Management*, **71**, 120–144.
- Basel Committee on Banking Supervision (2017). Basel IV.
- Cargnoni C, Muller P, and West M (1997). Bayesian forecasting of multinomial time series through conditional Gaussian dynamic models, *Journal of the American Statistical Association*, **92**, 640–647.
- Condon P (2014). *Applied Bayesian Modeling* (2nd Ed.), John Wiley & Sons, New York.
- Davis M and Lo V (2001). Infectious defaults, *Quantitative Finance*, **1**, 382–387.
- Duffie D, Saita L, and Wang K (2007). Multi-period corporate default prediction with stochastic covariates, *Journal of Financial Economics*, **83**, 635–665.
- Egloff D, Leippold M, and Vanini P (2004). *A Simple Model of Credit Contagion*, Preprint, Swiss Banking Institute, University of Zurich.
- Gelman A and Rubin DB (1992). Inference from iterative simulation using multiple sequences, *Statistical Science*, **7**, 457–472.
- Ghosh A, Mukhopadhyay S, Roy S, and Bhattacharya S (2014). Bayesian inference in nonparametric dynamic state-space models, *Statistical Methodology*, **21**, 35–48.
- Greene WH (2017). *Econometric Analysis* (8th ed), Pearson Education, New York.
- Heynderickx W, Cariboni J, Schoutens W, and Smits B (2016). The relationship between risk-neutral and actual default probabilities: the credit risk premium, *Applied Econometrics*, **84**, 4066–4081
- Hu YT, Kiesel R, and Perraudin W (2002). The estimation of transition matrices for sovereign credit ratings, *Journal of Banking & Finance*, **26**, 1383–1406.
- Li W (2016). Probability of default and default correlations, *Journal of Risk and Financial Management*, **9**, 1–19.
- Oh M, Choi JW, and Kim D (2003). Bayesian inference and model selection in latent class logit models with parameter constraints: an application to market segmentation, *Journal of Applied Statistics*, **30**, 191–204.
- Reilly C, Gelman A, and Katz J (2001). Poststratification without population level information on the poststratifying variable, with application to political polling, *Journal of the American Statistical Association*, **96**, 640–647.
- Sobelhart J and Keenan S (2001). Measuring default accurately, *Credit Risk Special Report*, *Risk*, **14**, 31–33
- Wendin JEP (2006). *Bayesian Method in Portfolio Credit Risk Management* (Doctoral Thesis), Swiss Federal Institute of Technology.
- Xiao JY (2002). *Obligor in CreditMetrics*, Research Technical Note, RiskMetrics Group.