# Bayesian Inference for Predicting the Default Rate Using the Power Prior

Seong W. Kim[1], Young Sook Son[2] and Sanga Choi[3]

## Abstract

Commercial banks and other related areas have developed internal models to better quantify their financial risks. Since an appropriate credit risk model plays a very important role in the risk management at financial institutions, it needs more accurate model which forecasts the credit losses, and statistical inference on that model is required. In this paper, we propose a new method for estimating a default rate. It is a Bayesian approach using the power prior which allows for incorporating of historical data to estimate the default rate. Inference on current data could be more reliable if there exist similar data based on previous studies. Ibrahim and Chen (2000) utilize these data to characterize the power prior. It allows for incorporating of historical data to estimate the parameters in the models. We demonstrate our methodologies with a real data set regarding SOHO data and also perform a simulation study.

*Keywords* : Default rate; Bayesian approach; power prior; AR(1) model; historical data; Gibbs sampling.

## 1. Introduction

Since early 1990s there has been a considerable amount of work to develop sophisticated systems measuring credit risks, which are arisen from diverse areas such as banking, finance, insurance, and other related fields. These systems or models are intended to quantify and aggregate resulting credit risks. In particular, much efforts have been well recognized by bank regulators. The Basel Capital Accord formally encourages banks to have their own internal risk models. The ultimate purpose of credit risk models is to forecast the probability of losses that may occur in banks' credit portfolios. Thus an appropriate model should be built in accordance with being conceptually sound, easily applicable, and empirically feasible.

1) Division of Applied Mathematics, Hanyang University, Ansan, 426-791, Korea.
   Correspondence : seong@hanyang.ac.kr
2) Department of Statistics, Chonnam National University, Kwangju, 500-757, Korea.
3) Department of Mathematics, Ajou University, Suwon, 443-749, Korea.

Credit scoring is a procedure to measure the degree of the credit explicitly based on accumulated customers data. Since it relies upon real data through statistical methodologies, it is usually more reliable than subjective or judgmental methods. The data initially classifies into two groups, which are the *good* and the *bad* groups. In retail lending, most of banks have their own credit scoring system (CSS) to evaluate the individual credit, and thus, loan processes are appropriately executed. In commercial lending, a similar procedure should be performed to evaluate corporate credit risks. This system is often called the credit rating system (CRS). These systems can accommodate banks' competitiveness resulting in the increase of performing customers with high profitability.

When you build credit evaluation models, the first step is 'data gathering'. We quite often see that good samples are a lot larger than bad samples. However, in practice, we take both good and  bad samples with equal sizes. If all the good samples are used, the resulting entire samples depend too much upon the characteristics of the good samples (cf. Thomas et *al.*, 2002; Caouette et *al.*, 1998). On the contrary, the equal sample size should increase predictability for both the goods and bads simultaneously. Once the data are collected, several statistical methodologies can be applied to compute the default rate of each individual. These methodologies include the logistic regression, Chaid analysis, neural networks, and etc. It is not a major issue in this article to explore modelling techniques, and thus we do not mention them in detail. The next step is to create the risk band and calculate the default probability in each risk band. Because equal sample sizes are used, we need to adjust the band probabilities by estimating the 'mean default rate' for the underlying population.

In recent years, default rate models in portfolio contexts have been studied and a decent amount of literature regarding this problem has been published till now. An initial attempt was done by Fons (1991) who formulates a statistical model to forecast aggregated issuer-based default rates. In this study, Fons found that about 52% of the variation in historical default rates could be explained using only two factors, credit quality and the state of the economy. Carey (1998) examined two characteristics of private debt portfolio credit risk loss rate distributions empirically. Gordy (2000) compared risk measures calculated for the same portfolio but using different models. Nickel, Perraudin and Varotto (2001) predicted the credit risk of a large portfolio of dollar denominated eurobonds using rating-based and equity-based approaches. Emmer and Tasche (2003) developed approximate formula to calculate credit risks.

It is well known that default rates are dependent upon time. In this article, we consider a first order autoregressive model (AR(1)),

$$r_t - r = \phi(r_{t-1} - r) + \epsilon_t, \tag{1}$$

where $|\phi| < 1$, $\epsilon_t$ follows a Gaussian white noise with mean 0 and variance $\tau$, and $t$ is an observed time. To estimate the mean default rate $r$ we propose a Bayesian approach using the power prior and it is compared with the conditional least square (CLS) estimation from a frequentist perspective.

In the Bayesian approach it is a major issue to specify the priors for the parameters in the model. For Bayesian analysis, a specified prior distribution is essentially required. However, it is not easy to specify a meaningful prior distribution for the parameters in each model since it is requiring contextual interpretations of a large number of parameters. Thus it needs to look for some useful and automated specifications. Reference priors can be used to address it. However, they may result in improper posterior probabilities. Recently, Berger and Pericchi (1996) have proposed the intrinsic prior which provides a comprehensive solution to ambiguous problems. In Bayesian analysis, Jeffreys's prior (Jeffreys, 1946) is commonly used for reference priors. However, for certain models it is not appropriate because of computation and analytic problems (cf. Kim and Ibrahim, 2000). In addition, reference priors do not reflect any real prior information that one may need for a specific situation. These cases are often encountered when the current study is similar to the previous studies in measuring the response and covariates. The data from  past studies may be used for real prior information for the current study when the current study is similar to the previous studies in measuring the response and covariates.  However, in the case of using data from a previous study as a prior information, it needs a caution that one should not use them blindly or in a semiautomatic fashion when constructing informative priors, because the information contained in them may be inappropriate for the research problem at hand. The data arising from previous studies are referred as 'historical data' (cf. Ibrahim and Chen, 2000).

A useful informative prior on historical data is the power prior of Ibrahim and Chen (2000)  because it inherently automates the informative prior specification for every conceivable models in the model space. The power prior is defined by the likelihood function based on the historical data, raised to a power $a_0$, where $a_0 (0 \le a_0 \le 1)$ is a scalar parameter that controls the influence of the historical data on the current study.

The initial idea of the power prior can be traced to Diaconis and Ylvisker (1979) and Morris (1983), where they studied conjugate priors for exponential families. However, these two authors only considered the situation in which the power $a_0$ is a fixed constant. When $a_0$ is random, the formulation becomes quite complicated and theoretical properties of the powers remain largely unknown. Chen, Ibrahim

and Shao (1999) have provided the theoretical properties of power priors for the class of various models.

This paper organized as follows. In section 2 the overview of the power prior is presented. In section 3 we propose the AR(1) model with the power prior. Section 4 contains two numerical examples, involving simulated and real datasets. We finish this paper with a brief discussion in section 5.

## 2. The power prior

Let $D$ denote the data from the current study, and let $L(\theta \mid D)$ be the likelihood function of the current study, where $\theta$ is a vector of interesting parameters. Suppose that historical data $D_0$ from a similar experiment are available. Further, let $\pi_0(\theta \mid \cdot)$ denote the prior distribution for $\theta$ before the historical data $D_0$ are observed. We often call $\pi_0(\theta \mid \cdot)$ the initial prior distribution for $\theta$. From Ibrahim and Chen (2000) the joint power prior distribution is characterized as

$$\pi(\theta \mid D_0, a_0) \propto [L(\theta \mid D_0)]^{a_0} \pi_0(\theta \mid c_0),\tag{2}$$

where $c_0$ is a specified hyperparameter for the initial prior, and $a_0$ is a prior parameter that brings up the weight for the historical data relative to the likelihood of the current study. The parameter $a_0$ can be interpreted as a precision parameter. It is reasonable that the range of $a_0$ is restricted to be between 0 and 1. The parameter $a_0$ controls the heaviness of the tails of the prior for $\theta$. As $a_0$ becomes smaller, the tails of (2) gets heavier. When we fix $a_0 = 1$, (2) can be the posterior distribution of $\theta$ from the previous study. When $a_0 = 0$, the prior distribution does not depend on the historical data and thus (2) can be a usual prior. An important role of $a_0$ is controlling the influence of the historical data on the current data. Such control may be important when the sample sizes of two studies are quite different or there is heterogeneity between the previous and the current study. For example, it is the case that a financial crisis has occurred in the past year.

The hierarchical power prior is completed by specifying a prior distribution for $a_0$. A natural and common choice of prior distribution for $a_0$ is a beta distribution. Then the joint power prior becomes

$$\pi(\theta, a_0 \mid D_0) \propto [L(\theta \mid D_0)]^{a_0} \pi_0(\theta \mid c_0) \pi(a_0 \mid \gamma_0),\tag{3}$$

where $\pi(a_0 \mid \gamma_0) \propto a_0^{\delta_0 - 1}(1 - a_0)^{\lambda_0 - 1}$ and $\gamma_0 = (\delta_0, \lambda_0)$ are the specified hyperparameter vector. A beta prior for $a_0$ appears to be the most natural to use and leads to the

most natural elicitation scheme. The prior in (3) does not have a closed form in general. However, a desirable feature of (3) is that it creates heavier tails for the marginal prior of $\theta$ than (2). Thus (3) is more flexible in weighting the historical data. If a small prior weight is desired, we choose the prior mean of $a_0$, $\mu_{a_0} = \delta_0/(\delta_0 + \lambda_0)$ to be small. Otherwise, $\mu_{a_0} \geq 0.5$ may be suitable. The joint power prior in (3) can be generalized when multiple historical datasets are available. Suppose that there are $M$ historical datasets, and let $D_{0k} = (n_{0k}, y_{0k}, X_{0k})$ be the historical data based on the $k$th study, and let $D_0 = (D_{01}, \cdots, D_{0M})$. In this case, it is desirable to define a precision parameter $a_{0k}$ for each historical study and take the distribution for $a_{0k}$'s to be i.i.d. beta distribution with parameters $(\delta_0, \lambda_0)$, for $k = 1, \cdots, M$. Let $a_0 = (a_{01}, \cdots, a_{0M})$, then the joint power prior in (3) can be generalized as

$$\pi(\theta, a_0 \mid D_0) \propto \left\{ \prod_{k=1}^{M} [L(\theta \mid D_{0k})]^{a_{0k}} \pi_0(a_{0k} \mid \gamma_0) \right\} \pi_0(\theta \mid c_0).$$

## 3. The AR(1) model with the power prior

Suppose that we have actual default rates observed in equally spaced intervals. It could be the case that the high default interval is more likely to be followed by another high default interval. So we propose an AR(1) model for estimating 'mean default rate'. In (1) let $r$ denote the underlying mean default rate (unit: %). Suppose we have a set of default rates calculated by $n$ intervals, denoted by $r = (r_1, \cdots, r_n)$. Let $\theta = (r, \phi, \tau)$ be the model parameters. Then the likelihood function is

$$L(\theta \mid r) = (2\pi\tau)^{-n/2} \exp\left\{ -\frac{1}{2\tau} \sum_{t=2}^{n} [r_t - r - \phi(r_{t-1} - r)]^2 \right\}. \tag{4}$$

Suppose that we have historical data from a past year, calculated by $n_0$ intervals. We denote this by $r^* = (r_1^*, \cdots, r_{n_0}^*)$. Then the hierarchical power prior is specified as

$$\pi(\theta, a_0 \mid r^*) \propto [L(\theta \mid r^*)]^{a_0} \pi_0(\theta) \pi(a_0)$$

We specify a beta prior for $\pi(a_0)$ with hyperparameters $(\delta_0, \lambda_0)$, a truncated normal for $r$, denoted by $TN(\mu_0, \sigma_0^2)$ with $0 < r < \infty$, and uniform for $\phi$, denoted by $U(-1, 1)$. Meanwhile, we give an inverse gamma for $\tau$, denoted by $IG(\gamma_0, \eta_0)$. Under the assumption of independent *priori*, the joint power prior for $(\theta, a_0)$ is expressed as

$$\pi(r,\phi,\tau,a_0 \mid r^*) \propto [L(\theta \mid r^*)]^{a_0} a_0^{\delta_0-1} (1-a_0)^{\lambda_0-1} \exp\left\{-\frac{(r-\mu_0)^2}{2\sigma_0^2}\right\} \frac{1}{\tau^{\gamma_0+1}} e^{-\eta_0/\tau}, \qquad (5)$$

where $L(\theta \mid r^*)$ is (4) with $(r^*, n_0)$ in place of $(r, n)$ and $(\delta_0, \lambda_0, \mu_0, \sigma_0, \gamma_0, \eta_0)$ are known hyperparameters. Thus the posterior density is given by

$$p(\theta, a_0 \mid r, r^*) = L[\theta \mid r] \cdot \pi(\theta, a_0 \mid r^*)$$

$$\propto \tau^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\tau}\sum_{t=2}^{n}[r_t - r - \phi(r_{t-1}-r)]^2\right\}$$

$$\times \tau^{-\frac{n_0 a_0}{2}} \exp\left\{-\frac{a_0}{2\tau}\sum_{t=2}^{n_0}[r_t^* - r - \phi(r_{t-1}^*-r)]^2\right\}$$

$$\times a_0^{\delta_0-1}(1-a_0)^{\lambda_0-1}\exp\left\{-\frac{(r-\mu_0)^2}{2\sigma_0^2}\right\}\frac{1}{\tau^{\gamma_0-1}}e^{-\eta_0/\tau}. \qquad (6)$$

If the joint power prior (5) is proper, then the posterior in (6) will be proper.

Theorem 1   Suppose that

$$\pi_0(\theta) \propto \exp\left\{-\frac{(r-\mu_0)^2}{2\sigma_0^2}\right\}\frac{1}{\tau^{\gamma_0+1}}e^{-\eta_0/\tau},$$

where $(\mu_0, \sigma, \gamma_0, \eta_0)$ are specified hyperparameters, and assume that $\pi(a_0) \propto a_0^{\delta_0-1}(1-a_0)^{\lambda_0-1}$ with specified hyperparameters $(\delta_0, \lambda_0)$. Then (5) is proper.

proof: To show (5) to be finite, set

$$R = \sum_{t=2}^{n_0}[r_t^* - r - \phi \cdot (r_{t-1}^* - r)]^2.$$

Then

$$\int \pi(r,\phi,\tau,a_0 \mid r^*)$$

$$\propto \int_0^1 \int_{-1}^1 \int_a^b \int_0^\infty \tau^{-\frac{n_0 a_0}{2}} \exp\left\{-\frac{a_0 R}{2\tau}\right\} a_0^{\delta_0-1}(1-a_0)^{\lambda_0-1}\exp\left\{-\frac{(r-\mu_0)^2}{2\sigma_0^2}\right\}\tau^{-(\gamma_0+1)}$$

$$\times \exp\left(-\frac{\eta_0}{\tau}\right) d\tau \, dr \, d\phi \, da_0$$

$$= \int_0^1 \int_{-1}^1 \int_a^b \exp\left\{-\frac{(r-\mu_0)^2}{2\sigma_0^2}\right\} a_0^{\delta_0-1}(1-a_0)^{\lambda_0-1}\left(\frac{a_0}{2}R+\eta_0\right)^{-(\frac{n_0 a_0}{2}+\gamma_0)}$$

$$\times \Gamma\left(\frac{n_0 a_0}{2}+\gamma_0\right) dr \, d\phi \, da_0 . \qquad (7)$$

Since

$$(\frac{a_0}{2}R+\eta_0)^{-(\frac{n_0a_0}{2}+\gamma_0)} \leq \eta_0^{-(\frac{n_0a_0}{2}+\gamma_0)},$$

(7) is less than or equal to

$$\int_0^1 \int_{-1}^1 \int_a^b \exp\left\{-\frac{(r-\mu_0)^2}{2\sigma_0^2}\right\} a_0^{\delta_0-1} (1-a_0)^{\lambda_0-1} \eta_0^{-(\frac{n_0a_0}{2}+\gamma_0)} \Gamma(\frac{n_0a_0}{2}+\gamma_0) dr\, d\phi\, da_0. \quad (8)$$

Finally, (8) is less than or equal to

$$\begin{cases} K \cdot \Gamma(\gamma_0)\eta_0^{-(n_0+\gamma_0)} & \text{if } 0 < \eta_0 < 1,\ 0 < \gamma_0 < 1, \\ K \cdot \Gamma(n_0+\gamma_0)\eta_0^{-(n_0+\gamma_0)} & \text{if } 0 < \eta_0 < 1,\ \gamma_0 > 1, \\ K \cdot \Gamma(\gamma_0) & \text{if } \eta_0 > 1, 0 < \gamma_0 < 1, \\ K \cdot \Gamma(n_0+\gamma_0) & \text{if if } \eta_0 > 1,\ \gamma_0 > 1, \end{cases} \quad (9)$$

where

$$K = \int_0^1 a_0^{\delta_0-1} (1-a_0)^{\lambda_0-1} da_0 \cdot \int_{-1}^1 d\phi \cdot \int_a^b \exp\left\{-\frac{(r-\mu_0)^2}{2\sigma_0^2}\right\} dr.$$

Since (9) is finite we complete the proof.

# 4. Numerical results

First, we consider the full conditional distributions in order to implement Gibbs sampling of (Gelfand and Smith (1990)). From (6), the full conditional posterior densities of parameters are expressed as the follows

- $h(r \mid \phi, \tau, a_0, r, r^*) \propto TN(\mu_r, \sigma_r^2)$ *with* $0 < r < \infty$,

  where

  $$\mu_r = \sigma_r^2 \left\{ a_0(1-\phi)\sigma_0^2 \sum_{t=2}^{n_0} (r_t^* - \phi r_{t-1}^*) + (1-\phi)\sigma_0^2 \sum_{t=2}^n (r_t - \phi r_{t-1})\mu_0\tau \right\} / (\tau\sigma_0^2)$$

  and

  $$\sigma_r^2 = \tau\sigma_0^2 \left\{ (n_0-1)a_0(1-\phi)^2\sigma_0^2 + (n-1)(1-\phi)^2\sigma_0^2 + \tau \right\}^{-1}.$$

- $h(\phi \mid r, \tau, a_0, r, r^*) \propto TN(\mu_\phi, \sigma_\phi^2)$ with $-1 < \phi < 1$,

  where

  $$\mu_\phi = \sigma_\phi^2 \left\{ a_0 \sum_{t=2}^{n_0} (r_{t-1}^* - r)(r_t^* - r) + \sum_{t=2}^n (r_{t-1} - r)(r_t - r) \right\} / \tau$$

  and

  $$\sigma_\phi^2 = \tau \left\{ a_0 \sum_{t=2}^{n_0} (r_{t-1}^* - r)^2 + \sum_{t=2}^n (r_{t-1} - r)^2 \right\}^{-1}.$$

- $h(\tau \mid r, \phi, a_0, r, r^*) \propto IG(\gamma, \eta)$,

where

$$\gamma = \frac{a_0 n_0 + n}{2} + \gamma_0$$

and

$$\eta = \frac{1}{2} \left\{ a_0 \sum_{t=2}^{n_0} [r_t^* - r - \phi(r_{t-1}^* - r)]^2 + \sum_{t=2}^{n} [r_t - r - \phi(r_{t-1} - r)]^2 \right\} + \eta_0. \bullet$$
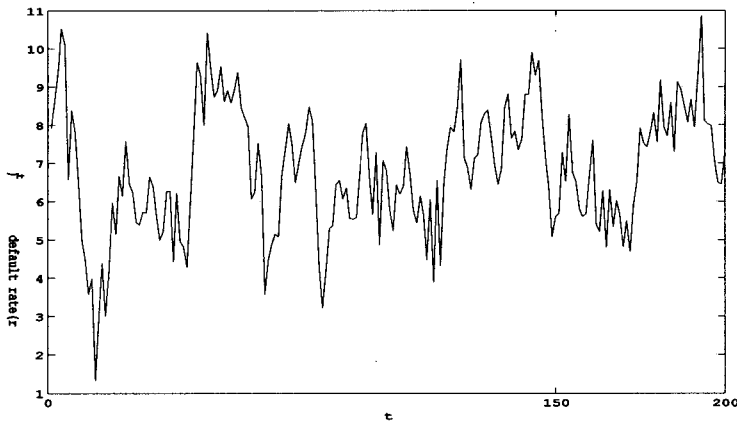
$$h(a_0 \mid r, \phi, \tau, r, r^*) \propto a_0^{\delta_0 - 1} (1 - a_0)^{\lambda_0 - 1} (2\pi\tau)^{-a_0 n_0/2} \exp\left\{ \frac{-a_0}{2\tau} \sum_{t=2}^{n_0} [r_t^* - r - \phi(r_{t-1}^* - r)]^2 \right\}.$$

Remark 1 All of the full conditional densities are in standard forms except that of $a_0$. However, the conditional posterior density of $a_0$ is log-concave provided $\delta_0 > 1$ and $\lambda_0 > 1$. Thus, we use the Adaptive Rejection Sampling (ARS) of Gilks and Wild (1992) to generate random variates from $h(a_0 \mid \cdot)$.

Example 1: We perform a simulation study and the results are obtained using the Gibbs sampling along with the ARS. We use 10,000 Gibbs iterations with burn-in sample of size 1000. For our illustration, we simulate 150 observations for historical data and 50 observations for current data from the AR(1) model,
$$r_t - 6.5 = 0.7(r_{t-1} - 6.5) + \epsilon_t, \quad t = 2, ..., 200,$$
where $\epsilon_t \sim N(0,1)$. <Figure 4.1> shows 200 simulated data and <Table 4.1> describes the descriptive statistics for the simulated data. <Table 4.3> gives the posterior estimates of the model parameters for several values of $(\delta_0, \lambda_0)$ in the case of truncated normal prior $TN(7,1)$ for $r$. From <Table 4.3>, we see that as the weight of $a_0$ increases so does the posterior mean of $a_0$, $E(a_0 \mid r^*, r)$. <Table 4.3> also indicates the standard deviations of parameters decrease and highest posterior density (HPD) intervals get narrower as the weight of $a_0$ increases. Further, the HPD intervals are not severe to modest changes in $(\mu_{a_0}, \sigma_{a_0})$. From the <Table 4.3>, it is easy to see that when the historical data are incorporated the default rate $r$ is closer to the true one than when they are not. <Figure 4.2> shows the marginal posterior densities of the default rate $r$ for two different priors, uniform and truncated normal with three choices of $(\mu_{a_0}, \sigma_{a_0})$. They are (0.05,0.0217), (0.5,0.0498), and (0.95,0.0217). From <Figure 4.2> we see that both marginal density curves are a little flatter as the prior mean of $a_0$ decreases, but they are almost the same for all three choices of $(\mu_{a_0}, \sigma_{a_0})$ . We also see that the standard deviation of marginal posterior distribution of $r$ for uniform prior is somewhat larger than for the truncated normal. Although we do not present here, we obtained the same numerical results of standard deviations between two priors.

<Figure 4.1> Plot of simulated AR(1) data:

$$r_t - 6.5 = 0.7(r_{t-1} - 6.5) + \epsilon_t, \quad \{\epsilon_t\} \sim N(0,1)$$

<Table 4.1>  Summary of AR(1) time series data.

| data | size | mean | SD | Min. | Max. |
|------|------|------|-----|------|------|
| historical data | 150 | 6.726 | 1.7130 | 1.3220 | 10.5309 |
| current data | 50 | 7.0979 | 1.3690 | 4.6844 | 10.8624 |
| total | 200 | 6.8195 | 1.6385 | 1.3220 | 10.8624 |

<Table 4.2> AR(1) time series data: Results of conditional least square estimation.

| data | $r$ | $\phi$ | $\tau$ | p-value[†] |
|------|-----|--------|--------|----------|
| historical data | $6.8142^{***}$ | $0.7772^{***}$ | 1.1788 | 0.3256 |
| current data | $6.9285^{***}$ | $0.6962^{***}$ | 1.0028 | 0.2837 |
| total data | $6.8971^{***}$ | $0.7655^{***}$ | 1.1200 | 0.1322 |

$^{***}$:significant at p-value<0.0001.

[†] : p-value at lag 12 in the Portmanteau test for testing $H_0$:the residual series is a white noise.

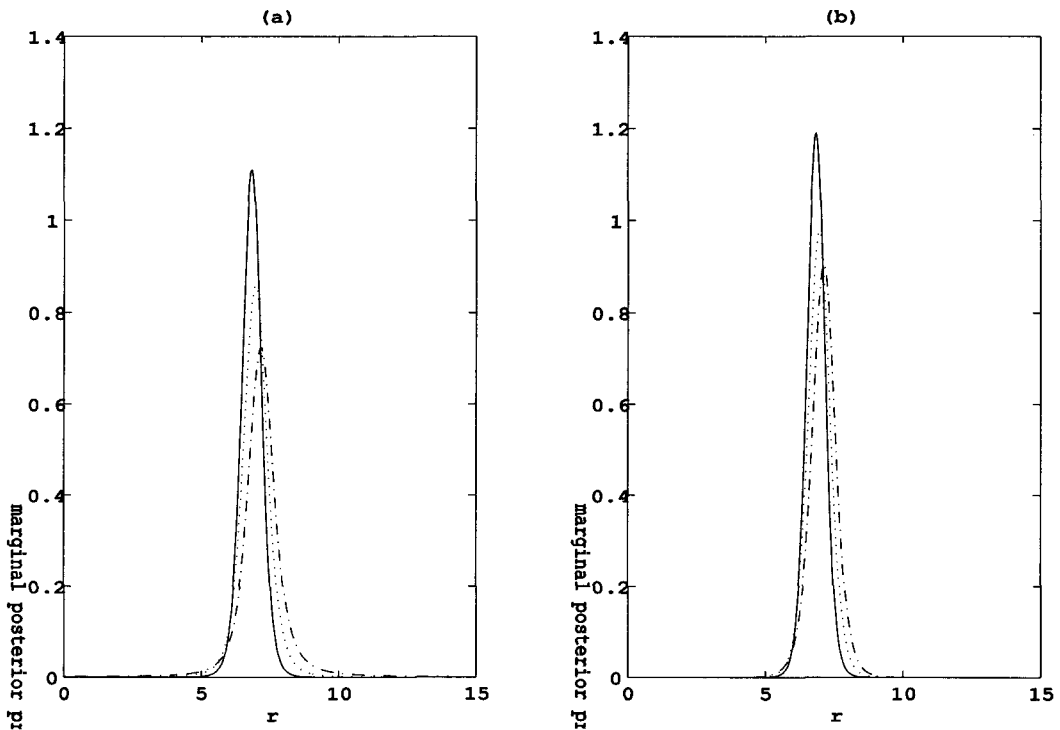<Table 4.3> AR(1) time series data: posterior estimates of model parameters with $\pi(r) \sim TN(7,1)$.

| $E(a_0 \mid r^*, r)$ | $\mu_{a0}$ | $\sigma_{a0}$ | Parameter | Mean | SD | Median | 95% HPD Interval |
|---|---|---|---|---|---|---|---|
| 0 (with probability 1) | | | $r$ | 7.1417 | 0.5065 | 7.1491 | (6.1092, 8.1503) |
| | | | $\phi$ | 0.7158 | 0.1057 | 0.7161 | (0.5094, 0.9269) |
| | | | $\tau$ | 1.0004 | 0.2120 | 0.9720 | (0.6255, 1.4247) |
| 0.0220 | 0.0500 | 0.0217 | $r$ | 7.1102 | 0.5118 | 7.1116 | (6.0583, 8.1165) |
| | | | $\phi$ | 0.7277 | 0.1008 | 0.7296 | (0.5285, 0.9201) |
| | | | $\tau$ | 1.0145 | 0.2079 | 0.9890 | (0.6310, 1.4175) |
| 0.1022 | 0.2000 | 0.0398 | $r$ | 7.0482 | 0.4848 | 7.0555 | (6.0625, 7.9878) |
| | | | $\phi$ | 0.7425 | 0.0883 | 0.7433 | (0.5634, 0.9095) |
| | | | $\tau$ | 1.0465 | 0.1904 | 1.0281 | (0.6986, 1.4202) |
| 0.2971 | 0.5000 | 0.0498 | $r$ | 6.9570 | 0.4518 | 6.9568 | (6.0133, 7.8109) |
| | | | $\phi$ | 0.7621 | 0.0704 | 0.7623 | (0.6219, 0.8964) |
| | | | $\tau$ | 1.0867 | 0.1633 | 1.0702 | (0.7888, 1.4113) |
| 0.5695 | 0.8000 | 0.0398 | $r$ | 6.8809 | 0.3943 | 6.8822 | (6.1058, 7.6644) |
| | | | $\phi$ | 0.7698 | 0.0575 | 0.7691 | (0.6568, 0.8831) |
| | | | $\tau$ | 1.1082 | 0.1368 | 1.0976 | (0.8496, 1.3746) |
| 0.8014 | 0.9500 | 0.0217 | $r$ | 6.8371 | 0.3608 | 6.8377 | (6.1117, 7.5198) |
| | | | $\phi$ | 0.7714 | 0.0514 | 0.7718 | (0.6711, 0.8716) |
| | | | $\tau$ | 1.1209 | 0.1252 | 1.1126 | (0.8813, 1.3637) |
| 1 (with probability 1) | | | $r$ | 6.8192 | 0.3337 | 6.8211 | (6.1984, 7.5305) |
| | | | $\phi$ | 0.7730 | 0.0469 | 0.7733 | (0.6810, 0.8665) |
| | | | $\tau$ | 1.1270 | 0.1139 | 1.1198 | (0.9160, 1.3572) |

Example 2: This example involves the 104-week default rates of SOHO (small office home office) customers in Kook-Min bank from January 2001 to December 2002. For illustration of our methodologies, we use the data from 2001 as the historical data and the data from 2002 as the current data. <Figure 4.3> exhibits the plot of data and <Table 4.4> shows the summary of weekly default rates for the secured loan. The mean and the standard deviation of the data from 2002 are slightly larger than those of the data from 2001. We compute the posterior means of parameters with the joint power prior (2). All procedures for computation is identical with those for simulation. The CLS estimates of the parameters and the residual analysis in <Table 4.5> show that our real data is fitted well for AR(1) model. The CLS estimates of the parameters for the current data are slightly larger than that of the historical data. For example, the default rate for the current data is 1.9211, whereas for the historical data it is 1.6034. In <Table 4.6> the posterior estimates of the parameters are reported for various $(\mu_{a_0}, \sigma_{a_0})$ when we give $TN(2,1)$ for the prior of $r$. From <Table 4.6> we see that the HPD intervals are not sensitive to modest changes in weight of $a_0$. This implies that

the HPD intervals are quite robust in terms of change of $(\mu_{a_0}, \sigma_{a_0})$. <Table 4.6> also shows that as the posterior mean of $a_0$ increases the standard deviation and the length of HPD intervals for all parameters decrease. This is very desirable feature since it illustrates that more precise estimates can be yielded by incorporating historical data. For example, when $a_0$ (with probability 1) the standard deviation and HPD interval for the default rate $r$ are given by $2.2339, (1.5531, 2.2859)$, respectively, whereas, for incorporating historical data (i.e, $a_0 = 1$ with probability 1) these are $0.1387, (1.5338, 2.0731)$. These differences are quite large.

<Table 4.4> Summary of weekly default rate data for extended security.

| data | size | Mean | Std. dev. | Minimum | Maximum |
|---|---|---|---|---|---|
| historical data | 52 | 1.5991 | 0.7347 | 0.6110 | 4.0512 |
| current data | 52 | 1.9360 | 1.0132 | 0.0000 | 4.8780 |
| total | 104 | 1.7676 | 0.8968 | 0.0000 | 4.8780 |



<Figure 4.2> AR(1) time series data: plots of marginal posterior densities for $r$ with (a) $\pi(r) \sim U(0,100)$ and (b) $\pi(r) \sim TN(7,1)$; $(dashed\ curve)(\mu_{a_0}, \sigma_{a_0}) = (0.05, 0.0217)$; $(dotted\ curve)(\mu_{a_0}, \sigma_{a_0}) = (0.5, 0.0498)$; $(solid\ curve)(\mu_{a_0}, \sigma_{a_0}) = (0.95, 0.0217)$.

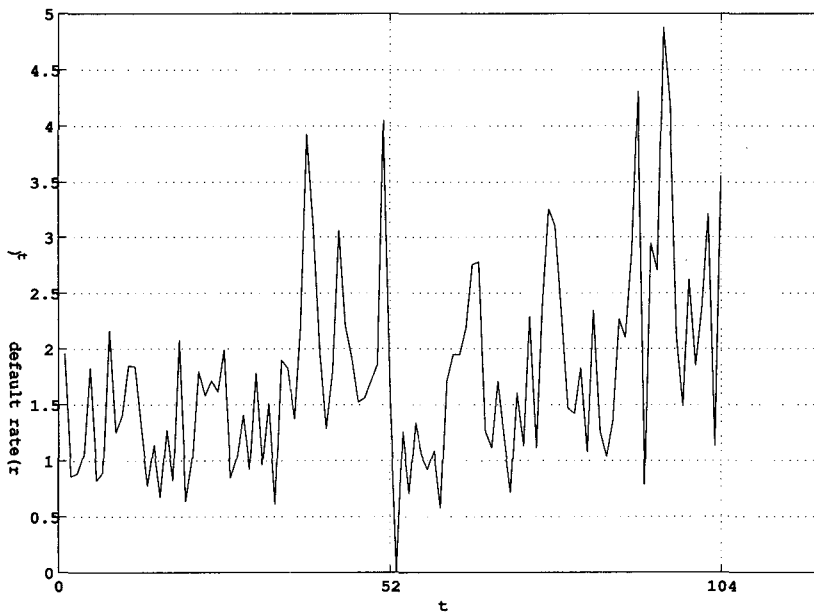<Table 4.5> Weekly default rate data for extended security: Results of conditional least square estimation.

| data | $r$ | $\phi$ | $\tau$ | p-value |
|------|-----|--------|--------|---------|
| historical data | $1.6034^{***}$ | $0.3122^{*}$ | 0.4969 | 0.6057 |
| current data | $1.9211^{***}$ | $0.3616^{**}$ | 0.9173 | 0.6468 |
| total | $1.7795^{***}$ | $0.3756^{***}$ | 0.7020 | 0.4133 |

*: significant at significance level 5%,

**: significant at significance level 1%,

***: significant at p-value <0.0001,

† : p-value at lag 12 in the Portmanteau test for testing $H_0$: the residual series is a white noise.



<Figure 4.3> Plot of weekly default rate data for extended security.

<Table 4.6>   Weekly default rate data for extended security: posterior estimates of model parameters with $\pi(r) \sim TN(2,1)$.

| $E(a_0 \mid r^*, r)$ | $\mu_{a0}$ | $\sigma_{a0}$ | Parameter | Mean | SD | Median | 95% HPD Interval | |
|---|---|---|---|---|---|---|---|---|
| 0 (with probability 1) | | | $r$ | 2.0105 | 0.2339 | 2.0089 | (1.5531, | 2.4859) |
| | | | $\phi$ | 0.3915 | 0.1364 | 0.3909 | (0.1354, | 0.6732) |
| | | | $\tau$ | 0.8760 | 0.1805 | 0.8523 | (0.5543, | 1.2356) |
| 0.0370 | 0.0500 | 0.0224 | $r$ | 1.9993 | 0.2229 | 1.9950 | (1.5450, | 2.4253) |
| | | | $\phi$ | 0.3914 | 0.1322 | 0.3910 | (0.1315, | 0.6571) |
| | | | $\tau$ | 0.8599 | 0.1763 | 0.8389 | (0.5550, | 1.2150) |
| 0.2054 | 0.2500 | 0.0500 | $r$ | 1.9432 | 0.1998 | 1.9396 | (1.5696, | 2.3607) |
| | | | $\phi$ | 0.3942 | 0.1227 | 0.3940 | (0.1616, | 0.6392) |
| | | | $\tau$ | 0.8124 | 0.1533 | 0.7941 | (0.5352, | 1.1168) |
| 0.4859 | 0.5000 | 0.0224 | $r$ | 1.8806 | 0.1739 | 1.8767 | (1.5534, | 2.2419) |
| | | | $\phi$ | 0.3924 | 0.1093 | 0.392 | (0.1849, | 0.6141) |
| | | | $\tau$ | 0.7541 | 0.1237 | 0.7405 | (0.5295, | 1.0004) |
| 0.7013 | 0.7500 | 0.0500 | $r$ | 1.8423 | 0.1568 | 1.8403 | (1.5438, | 2.1608) |
| | | | $\phi$ | 0.3908 | 0.1020 | 0.3907 | (0.1976, | 0.5915) |
| | | | $\tau$ | 0.7211 | 0.1132 | 0.7096 | (0.5152, | 0.9419) |
| 0.9314 | 0.9500 | 0.0224 | $r$ | 1.8094 | 0.1402 | 1.8071 | (1.5347, | 2.0891) |
| | | | $\phi$ | 0.3859 | 0.0957 | 0.3857 | (0.1972, | 0.5713) |
| | | | $\tau$ | 0.6938 | 0.1009 | 0.6833 | (0.5063, | 0.8913) |
| 1 (with probability 1) | | | $r$ | 1.8017 | 0.1387 | 1.7998 | (1.5338, | 2.0731) |
| | | | $\phi$ | 0.3873 | 0.0939 | 0.3876 | (0.2035, | 0.5705) |
| | | | $\tau$ | 0.6857 | 0.0981 | 0.6761 | (0.4988, | 0.8789) |

# 5. Discussion

In this paper we have proposed a new method to estimate the default rate with the AR(1) model. A feature of our method allows for incorporating of historical data to estimate the parameters in the model. When the historical data are incorporated, more precise estimates of parameters can be yielded, especially in case that historical and current data are similar to the previous studies in measuring the response and covariates. Further the power prior seems to be useful in a wide variety of applications, including carcinogenicity studies or clinical trials. They are also quite useful in model selection contexts since they automate the prior elicitation procedure for the prior on the model space, as well as the model parameters arising from the different models. Moreover, when the current data exhibit quite heterogeneity comparing to previous data, our method may be more reliable for forecasting credit losses. Further extensions to our proposed method can be considered. For instance, it can be extended to AR(p) model and

several historical datasets may be incorporated in current data for predicting a default rate.

# Reference

[1] Berger, J.O. and Pericchi, L.R. (1996). The Intrinsic Bayes Factor for Model Selection and Prediction. *Journal of American Statistical Association*, Vol. 91, 109-122.

[2] Caouette, J.B., Altman, E.I., and Narayanan, P. (1998). *Managing Credit Risk*, John Wiley & Sons, Inc.

[3] Carey, M. (1998). Credit Risk in Private Debt Portfolios, J. Finance, LIII, 1363-1387.

[4] Chen, M.-H., Ibrahim, J.G., and Shao, Q.-M. (1999). Prior Elicitation, Variable Selection and Bayesian Computation for Logistic Regression Models. *Journal of Royal Statistical Society* B, Vol. 61, Part 1,223-242.

[5] Dianconis, P. and Ylvisaker, D. (1979). Conjugate Priors for Exponential Families. *Annals of Statistics*, Vol. 7, 269-281.

[6] Emmer, S. and Tasche, D. (2003). Calculating Credit Risk Capital Charges with the One-factor Model, *The European Investment Review Third Annual Conference Geneva*, September 25.

[7] Fons, J.S. (1991). An Approach to Forecasting Default Rates, A Moody's Special Report.

[8] Gelfand, A.E. and Smith, A.F.M. (1990). Sampling Based Approaches to Calculating Marginal Densities. *Journal of the American Statistical Association*, Vol. 87, 523-532.

[9] Gilks, W.R. and Wild, P. (1992). Adaptive Rejection Sampling for Gibbs Sampling. *Applied Statistics*, Vol. 41, No. 2, 337-348.

[10] Gordy, M. (2000). A Comparative Anatomy of Credit Risk Models. *Journal of Banking & Finance*, Vol. 24, 119-149.

[11] Ibrahim, J.G. and Chen, M.-H. (2000). Power Prior Distributions for Regression Models. *Statistical Science*, Vol. 15, 46-60.

[12] Jeffreys, H. (1946). An Invariant Form for the Prior Probability in Estimation Problems, *Proceedings of the Royal Society London*. Ser. A, Vol. 186, 453-461.

[13] Kim, S.W. and Ibrahim, J.G. (2000). On Bayesian Inference for Proportional Hazards Models Using Noninformative Priors. *Life Time Data Analysis*, Vol. 6, 331-341.

[14] Morris, C.N. (1983). Natural Exponential Families with Quadratic Variance Functions: statistical theory. *Annals of Statistics*, Vol. 11, 515-529.

[15] Nickell, P., Perraudin, W., and Varotto, S. (2001). Ratings versus Equity-based

Credit Risk Modelling: an empirical analysis, *Working paper*, Bank of England, Birkbeck College.

[16] Thomas, L.C., Edelman, D.B., and Crook, J.N. (2002). *Credit Scoring and Its Applications*, SIAM, Philadelphia.