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

The Institute for Integrating Statistics in Decision Sciences

Discrete Time Bayesian Mortgage Default Models

Feng Xu
Department of Decision Sciences
The George Washington University, USA

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

Discrete Time Bayesian Mortgage Default Models

Feng Xu
Department of Decision Sciences
The George Washington University, USA

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

Abstract

In this paper we consider Bayesian analysis of mortgage default rate. We present models for the aggregate default data and develop Bayesian inference procedures for analyzing these models. We illustrate implementation of the models by analyzing real mortgage default rates.

1. Introduction and Overview

Increase in the residential mortgage default risk has significant impact on financial markets in the U.S. and across the world, as witnessed during the recent subprime mortgage loan crisis. The residential mortgage market is important to stability of the U.S. economy. The U.S. residential mortgage market has developed tremendously in size over time. As shown in Figure 1, outstanding debt of single-family mortgage loans in the U.S. has grown from around $2.6 trillion in 1990 to above $9.8 trillion in the second quarter of 2006, representing an increase from about 45% to 74.5% of its share in GDP during the period. With the exception of early 1980s, the U.S. national homeownership rate has increased from around 63% to above 68% in the last four decades. Considering the increase of the U.S. population during the same period, the increase in homeownership, measured by absolute numbers, is tremendous. Therefore,

any increase in the market risk, represented by mortgage default rate, will bring significant losses to all entities in the markets.

Reliable estimation of the default risk is of interest to government agencies to set up effective policies for stabilizing the market and is important for mortgage lenders, insurers and guarantors to develop financial instruments to mitigate the negative consequences of an increase in the risk. Besides, valid estimate of mortgage default risk is essential for accurate pricing of the mortgage-backed securities (MBS), which in turn protect the interests of investors in this secondary mortgage market.

U.S. Single-family Mortgage Outstanding & GDP (in $millions)
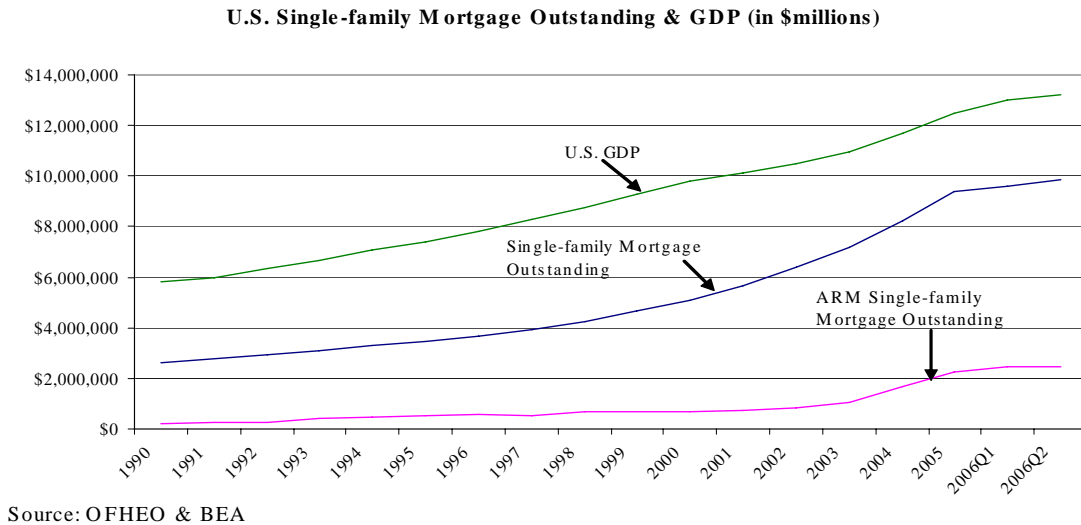

Source: OFHEO & BEA

Figure 1: Single-Family Mortgage Debt Outstanding

Different definitions of mortgage default have been used in the literature by different researchers. For example, Giliberto and Houston (1989) define mortgage default as the "transfer of the legal ownership of the property from the borrower to the lender either through the execution of foreclosure proceedings or the acceptance of a deed in lieu of foreclosure." Others who focus on modeling of the mortgage default risk, simply define the default as being delinquent in mortgage payment for 90 days; see for

example, Capone and Metz (2003). In most default models it is assumed that "default is synonymous with foreclosure" [Ambrose and Capone (1998)].

Research of mortgage default risk has started in late 60s with the aim of predicting default rates, explaining differences in the default rates among mortgages and across mortgage pools, for providing a better understanding of default behavior of borrowers and for assessing the effect of loan, property, and borrower characteristics on default rates; see Capone (2002). A thorough review of the literature can be found in Quercia and Stegman (1992) and Leece (2004).

Initial work by von Furstenberg (1969) studied the default rate at aggregate level using data from Federal Housing Administration (FHA) and Department of Veteran Affairs (VA) mortgage loans. Iimportant factors influencing the default decision of borrowers are identified by works of Herzog and Earley (1970), Williams et al. (1974), Sandor and Sosin (1975). The factors included the loan-to-value (LTV) ratio, interest rate, mortgage terms and payment-to-income (PTI) ratio. The main purpose of the research during this period was to help lenders to assess the default risk of borrowers and to predict default rates [Quercia and Stegman (1992)]. In late 70s focus of the research shifted to studies from borrower's perspective and the default decision was considered as one of the possible choices available to the borrower. The option theoretic basis was recognized under which the default decision being viewed as a put option, and the transaction costs, borrower characteristics and trigger events (such as divorce and loss of job, which impact the ability to pay the mortgage loan) began to gain their importance in residential mortgage default research [(Leece (2004)].

According to Quercia and Stegman (1992), starting in late 80s, researchers became more interested in studying mortgage pools and developing theories for mortgage and MBS pricing. Therefore, the focus of the research was on estimating the probability of default of a large mortgage loan pool. Equity related factors such as housing price and interest rate have been identified as the main factors factors influencing the default risk

by researchers such as Green and Shoven (1986), Schwartz and Torous (1989), Kau *et al.* (1991, 1993) who used an option theoretic approach. On the other hand, some researchers pointed out that non equity related factors, such as trigger events which influence borrower's ability-to-pay for the mortgage loan, were more important in default decisions; see Foster and van Order (1985), Gardner and Mills (1989), Giliberto and Houston (1989), Lekkas *et al.* (1993). More recent work on mortgage default risk includes duration analysis see Lambrecht *et al.* (1997) and competing risks theory where default and prepayment risks are studied together as competing options see Deng (1997), Pavlov (2001), Clapp *et al.* (2001), Calhoun and Deng (2002).

Most of the previous work in modeling mortgage default risk can be classified as the indirect and the direct approaches. The indirect approach is based on the option theoretic framework and studies the default risk by comparing the value of the property and the mortgage loan, while the direct approach uses hazard rate type models to analyze time-to-default probability, or the event of default itself. Implementation of the indirect approach requires individual loan performance data which is difficult to obtain due to privacy issues. In fact any type of individual loan level data may not be readily available for public use. Thus, in many cases the only type of readily accessible default data may be aggregated. Accurate estimation of mortgage default rate at aggregate level is of interest for mortgage lenders, insurers and policy makers. A good estimate of aggregate default rate of the mortgages in the pool underwritten is necessary for accurately pricing the mortgage backed securities.

All of the previous work in mortgage default literature are based on sampling theory methods for statistical estimation and inference. None of these studies considered use of Bayesian methods in modeling the mortgage default rate and thus they have limitations in making probabilistic conclusions with regards to mortgage default risk. In this paper we study residential mortgage default risk by developing Bayesian models for

aggregate default data. In so doing, we use a direct approach for modeling default risk and focus on discrete time models.

In Section 2 we present Bayesian models for analyzing aggregate mortgage default rates over time. In so doing, we introduce logistic beta time series and a random effects types extension of it. As an alternate modeling strategy a Markov modulated beta process is presented in Section 3 for describing discrete time default rates. Implementation of the models to real default data is presented in Section 4.

2. Logistic Beta Time Series Model for Default Rates

In this section we introduce a Bayesian model for a time-series of aggregate mortgage default rates. We consider this time-series as a discrete time beta process and introduce a logistic beta time-series model and a random effects extension of it.

Let the aggregate mortgage default rate at time t denoted by Y_t for $t = 1, 2, ...$ Since Y_t is a proportion of defaulted mortgages at time t, it is measured as a value in the $(0, 1)$ interval. Thus, it is not unreasonable to assume that Y_t follows a Beta distribution at time t. More specifically, we consider a beta distribution denoted as $Y_t \sim Beta(\kappa \alpha_t, \kappa \beta_t)$, with density proportional to

$$p(Y_t | \alpha_t, \beta_t, \kappa) \propto Y_t^{\kappa \alpha_t - 1}(1 - Y_t)^{\kappa \beta_t - 1} \tag{1}$$

where parameters $\alpha_t > 0$, $\beta_t > 0$ such that $\alpha_t + \beta_t = 1$, and $\kappa > 0$ is the precision parameter. The mean and variance of Y_t are given by

$$E[Y_t | \alpha_t, \beta_t, \kappa] = \alpha_t \tag{2}$$

$$Var[Y_t | \alpha_t, \beta_t, \kappa] = \frac{\alpha_t(1 - \alpha_t)}{(\kappa + 1)}. \tag{3}$$

Thus, α_t represents the expected default rate at time t and κ represents our cetainty about it. Time dependence of α_t's can be modeled by using a logit

transformation and incorporating time dependent deterministic covariates $\boldsymbol{X}_t = (1 \, X_{1t}$ $X_{2t}\ldots X_{p-1,t})'$ as

$$logit(\alpha_t) = log(\frac{\alpha_t}{1-\alpha_t}) = \boldsymbol{\theta}'\boldsymbol{X}_t \tag{4}$$

where $\boldsymbol{\theta}' = (\theta_0, \theta_1, ..., \theta_{p-1})$ is a $1 \times p$ vector of unknown regression parameters. We will refer to the above a logistic beta time-series model for the mortgage default rate.

Time series of proportions has been considered in Bayesian literature by several authors such as Quintana and West (1988) who considered logistic-normal distributions of Aitchison and Shen (1980) for the series and Grunwald et al. (1993) who modeled Dirichlet distributed time-series in state-space form. In both papers some numerical methods or approximations have been used for posterior computations. Our proposed logistic beta time-series model is more similar to the setup of Quintana and West (1988), but unlike them because we assume beta distributed time-series.

In our model we assume that given α_t and κ, Y_t's are conditionally independent over time. Therefore, given time-series data $D = (Y_1, Y_2, \ldots, Y_T)$ for T periods, the likelihood function can be expressed as proportional to

$$L(\alpha_1, ..., \alpha_T, \kappa; D) \propto \prod_{t=1}^{T} Y_t^{\kappa\alpha_t - 1}(1 - Y_t)^{\kappa\beta_t - 1} \tag{5}$$

where

$$\alpha_t = \frac{exp(\boldsymbol{\theta}'\boldsymbol{X}_t)}{1 + exp(\boldsymbol{\theta}'\boldsymbol{X}_t)}. \tag{6}$$

In other words, (5) is the likelihood function of κ and $\boldsymbol{\theta}$ given data D. In specifying the prior distribution of the unknown parameters, κ and $\boldsymbol{\theta}$ can be assumed to be independent of each other, that is, $p(\kappa, \boldsymbol{\theta}) = p(\kappa)p(\boldsymbol{\theta})$. Uncertainty about the precision parameter κ, can be described by a gamma prior denoted as

6

$$\kappa \sim Gamma(a_\kappa, b_\kappa) \tag{7}$$

where $a_\kappa, b_\kappa > 0$ are specified prior parameters. For the regression parameter vector $\boldsymbol{\theta}$ we can assume a multivariate normal distribution denoted as

$$\boldsymbol{\theta} \sim MVN(\boldsymbol{\mu}, \boldsymbol{W}) \tag{8}$$

where the mean vector $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{W}$ are known quantities.

The Bayesian analysis of the logistic beta time-series model requires the joint posterior distribution

$$p(\kappa, \boldsymbol{\theta}|D) \propto L(\kappa, \boldsymbol{\theta}; D)p(\kappa)p(\boldsymbol{\theta}) \tag{9}$$

where $L(\kappa, \boldsymbol{\theta}; D)$ is obtained from (5) by substituting α_t's with (6). The posterior distribution (9) can not be obtained in any analytically tractable form, but it can be evaluated using Markov chain Monte Carlo (MCMC) methods, which will be implemented using the winBUGS programming environment of Spiegelhalter, Thomas, Best, and Gilks (1996).

Once the posterior joint density $p(\kappa, \boldsymbol{\theta}|D)$ is obtained via use of the MCMC methods we can develop various posterior inferences using Monte Carlo estimates. Given $[(\kappa^1, \boldsymbol{\theta}^1), \ldots, (\kappa^G, \boldsymbol{\theta}^G)]$, a posterior sample of size G from $p(\kappa, \boldsymbol{\theta}|D)$, the marginal posterior distributions of covariate parameters θ_i's, for $i = 0, 1, \ldots, (p-1)$, can be obtained using histogram or density estimates from the marginal samples. Similarly, posterior samples for α_t's can be generated using

$$\alpha_t^g = \frac{exp((\boldsymbol{\theta}')^g X_t)}{1 + exp((\boldsymbol{\theta}')^g X_t)} \tag{10}$$

for $g = 1, \ldots, G$, and posterior distributions of α_t's can be obtained using density estimates. Any posterior moments of the parameters can be computed by Monte Carlo

averages. For example, we can obtain the posterior mean vector $E(\boldsymbol{\theta}|D)$ via the Monte Carlo integral approximation

$$E(\boldsymbol{\theta}|D) \approx \frac{1}{G}\sum_{g=1}^{G}\boldsymbol{\theta}^g. \tag{11}$$

Posterior probabilities such as $Pr(\alpha_* < \alpha_t < \alpha^* |\, D)$, can be evaluated as

$$Pr(\alpha_* < \alpha_t < \alpha^* |\, D) \approx \frac{1}{G}\sum_{g=1}^{G}\mathbf{1}(\alpha_* < \alpha_t^g < \alpha^*), \tag{12}$$

where $\mathbf{1}(A)$ takes the value 1 if event A occurs and 0 otherwise. Similarly we can obtain a $100(1-\alpha)\%$ *credible interval* for α_t [see for example, Bernardo and Smith (1994), p. 259] which can be considered as Bayesian interval estimator of α_t. Predictive inference about the future default rate Y_{T+1} is made by using ther posterior predictive distribution

$$p(Y_{T+1}|D) = \int\int p(Y_{T+1}|\kappa, \alpha_{T+1})\, p(\kappa, \alpha_{T+1}|D)\, d\kappa\, d\alpha_{T+1} \tag{13}$$

where $p(Y_{T+1}|\kappa, \alpha_{T+1})$ is given by the beta density (1). The integral in (13) can not be evaluated analytically but we can obtain $p(Y_{T+1}|D)$ via the Monte Carlo approximation

$$p(Y_{T+1}|D) \approx \frac{1}{G}\sum_{g=1}^{G}p(Y_{T+1}|\kappa^g, \alpha_{T+1}^g), \tag{14}$$

where α_{T+1}^g is defined in terms of $\boldsymbol{\theta}$ as given by (10).

We note that the logit model given by (4) implies a deterministic model for α_t's given the covariate vectors $\boldsymbol{X}_t$ over time. Hence, the time variation of α_t's is due to the time dependent covariate vectors in the model. Often there may be possible sources of unknown variation that can not be captured by the covariates given by $\boldsymbol{X}_t$. This problem can be solved by including a random effects type term associated with each time period into the model. Thus, the logit model can be modified as

$$logit(\alpha_t) = \boldsymbol{\theta}' \boldsymbol{X}_t + \epsilon_t, \tag{15}$$

where ϵ_t is the random effects term. We assume that ϵ_t's are conditionally independent normal random variables denoted as

$$\epsilon_t | \tau \sim N(0, \tau_\epsilon^{-1}) \tag{16}$$

where the unknown precision is described by the gamma prior, $\tau_\epsilon \sim Gamma(a_\tau, b_\tau)$, with specified parameters a_τ and b_τ. Furthermore, we assume a priori that $\kappa, \boldsymbol{\theta}$ and ϵ_t's are independent of each other.

Under the random effects logit model, the joint posterior distribution of interest is

$$p(\alpha_1, \ldots, \alpha_T, \kappa, \epsilon_1, \ldots, \epsilon_T, \tau_\epsilon | D) \propto \prod_{t=1}^{T} p(Y_t | \alpha_t, \kappa) p(\kappa) p(\boldsymbol{\theta}) p(\epsilon_t | \tau_\epsilon) p(\tau_\epsilon) \tag{17}$$

which can also be written as

$$p(\boldsymbol{\theta}, \kappa, \epsilon_1, \ldots, \epsilon_T, \tau_\epsilon | D) \propto \prod_{t=1}^{T} p(Y_t | \boldsymbol{\theta}, \kappa) p(\kappa) p(\boldsymbol{\theta}) p(\epsilon_t | \tau_\epsilon) p(\tau_\epsilon) \tag{18}$$

using the identity

$$\alpha_t = \frac{exp(\boldsymbol{\theta}' \boldsymbol{X}_t + \epsilon_t)}{1 + exp(\boldsymbol{\theta}' \boldsymbol{X}_t + \epsilon_t)}. \tag{19}$$

As in the original logistic beta time series model, the joint posterior can not be obtained analytically. Thus, MCMC methods will be used for posterior and predictive inferences.

3. A Markov Modulated Beta Process

The logistic beta time series model and its random effects version introduced above assume that the expected default rate α_t depends on covariates $\boldsymbol{X}_t$. Due to the aggregate nature of the data, components of vector $\boldsymbol{X}_t$ will typically include macro socioeconomic variables such as interest rates, housing price index, national unemployment rate, etc. Since such macro factors can not adequately reflect the complex

environment influencing borrowers' default decisions, an alternate modeling strategy is to relate the default rate to the state of an unobservable environmental stochastic process. By assuming the latent process is a Markov process, the class of models are referred to as *Markov modulated processes* [see for example, Ozekici and Soyer (2003)] or *hidden Markov models*. The motivation for this class of models in our case is the fact that factors influencing aggregate default rate are too complex to be captured by few covariates. Thus, we assume that the default rates depend on a random environment which evolves over time according to a Markov process.

Hidden Markov models have been previously considered in the finance literature in modeling commodity returns. For example, Ross (1976) points out that, this class of models can be motivated by the arbitrage pricing theory which states that changes in commodity prices are related to unknown underlying factors. More recently, Ryden et al. (1998) presented empirical evidence supporting hidden Markov models for daily return series and presented likelihood based inference methods.

In modeling the aggregate mortgage default rate, we consider the beta time series model (1)

$$p(Y_t|\alpha_t, \kappa) \propto Y_t^{\kappa\alpha_t - 1}(1 - Y_t)^{\kappa(1-\alpha_t)-1} \tag{20}$$

and define

$$\alpha_t = \alpha(S_t) \tag{21}$$

where S_t is a latent state variable. More specifically, we let S_t denote the state of the environment at time t and given the state of the environment at time t is i, (20) can be written as

$$(Y_t|S_t = i, \kappa) \sim Beta[\kappa\alpha(i), \kappa(1 - \alpha(i))]. \tag{22}$$

We assume that the environmental process $S = \{S_t : t \geq 1\}$ is a Markov chain with time homogeneous transition matrix Π on a discrete state space E. In our development we

10

assume that only the expected default rate, that is, α_t depends on the enviromental process. Given the environmental process and the precision parameter, we assume that Y_t's are conditionally independent random quantities, that is,

$$p(Y_1, Y_2, \ldots, Y_T \mid S, \kappa) = \prod_{t=1}^{T} p(Y_t \mid \alpha(S_t), \kappa). \tag{23}$$

Bayesian analysis of hidden Markov models (HMMs) has been considered by many authors, but most of these considered observation models such as Bernoulli and Gaussian distributions where the implementation of MCMC methods is straightforward. As discussed by Robert et al. (1993), a Gibbs sampler can be easily implemented in these cases. However, Markov modulated beta processes have not been previously considered in the Bayesian HMMs literature.

In the Bayesian setup, in addition to the latent variables S_t, for $t = 1, 2, \ldots, T$, the transition matrix Π, the precision parameter κ and the expected default rates $\alpha(S_t)$, for $t = 1, 2, \ldots, T$ are all unknown quantities. For the components of the transition matrix, we assume that Π_i, the ith row of Π follows a Dirichlet distribution

$$p(\Pi_i) \propto \prod_{j \in E} \pi_{ij}^{\psi_{ij}-1} \tag{24}$$

with specified parameters ψ_{ij}'s and that the Π_i's are independent of each other. If we assign $\psi_{ij} = 1$ for all j in the above then the prior distribution (24) reduces to a joint uniform distribution. We denote the density in (24) as $\Pi_i \sim Dirichlet\{\psi_{ij}; j \in E\}$. Since $0 < \alpha(S_t) < 1$, for a given environment $i \in E$, we assume that

$$\alpha(i) \sim Beta(a(i), b(i)) \tag{25}$$

where $a(i)$ and $b(i)$ are known quantities. Given the environmental process, $\alpha(i)$'s are assumed to be independent random quantities. Furthermore, $\alpha(i)$'s are independent of

Π_i. As in the previous discussion, the precision parameter κ is assumed to be independent of all other quantities and its prior distribution is given by (7).

Given data D, the observed default rates for T periods, we are interested in the joint posterior distribution of all unkown quantities $(\kappa, \boldsymbol{\alpha}, \Pi, S^{(T)})$, where $\boldsymbol{\alpha} = (\alpha(i); i \in E)$ and $S^{(T)} = (S_1, \ldots, S_T)$. The joint posterior $p(\kappa, \boldsymbol{\alpha}, \Pi, S^{(T)}|D)$ can be written as

$$p(\kappa, \boldsymbol{\alpha}, \Pi, S^{(T)}|D) \propto \Big[\prod_{t=1}^{T} \pi(S_t|S_{t-1}) Y_t^{\kappa\alpha(S_t)-1} (1-Y_t)^{\kappa[1-\alpha(S_t)]-1} \Big] \times$$

$$\Big[\prod_{i \in E} p(\Pi_i) \left[\alpha(i)\right]^{a(i)-1} \left[1-\alpha(i)\right]^{b(i)-1} \Big] p(\kappa), \tag{26}$$

where $\pi(S_t|S_{t-1}) = \pi_{S_{t-1},S_t}$ is the transition probability at time t. The joint posterior distribution in (26) can not be evaluated in closed form. We can use MCMC methods to develop posterior and predictive inferences for the Markov modulated beta process model.

4. Empirical Analysis

In this section we implement the discrete time Bayesian models of Sections 2 and 3 using real mortgage default data from the U.S. residential mortgage market. Due to the limitation of access to aggregate default rate data for specific mortgage pools, we use 90-day past due national delinquency rate of FHA insured FRM single-family mortgage loans, from National Delinquency Survey of Mortgage Banker's Association (MBA). The data is recorded on a quarterly base, from quarter 1 of 1992 to quarter 4 of 2005.

In analyzing the aggregate default rates, we consider both the logistic beta time series (LBT) and the Markov modulated beta process (MMBP) models. In the LBT model, we consider equity factors and ability-to-pay factors as the covariates in the model. Two key equity factors, the interest rate and the housing property prices, are included. The interest rate used here is the U.S. Treasury securities at 10-year constant

maturity given by The Federal Reserve Board, and housing prices included are the U.S. Housing Pricing Index (HPI) declared by Office of Federal Housing Enterprise Oversight (OFHEO). Two ability-to-pay factors are also included in the analysis. These are the Homeowner Mortgage Financial Obligations Ratio (FOR Mortgage) from The Federal Reserve Board and the U.S. National Unemployment Rate from Bureau of Labor Statistics.

In the LBT model we use independent normal priors for the logistic regression coefficients with mean 0 and variance 100. The precision parameter κ is assumed to have a Gamma distribution with mean 1 and variance 100. In the random effect extension of the model the precision parameter τ is also assumed to be a gamma with mean 1 and variance 100. Thus, our selection of the parameters imply proper but noninformative priors for all the unknown quantities. A similar strategy is followed for the prior selection in the MMBP model where we consider three environments. In implementation of the MCMC methods the WinBUGS was used as the computing environment. In all cases, simulations were run with 10,000 burn-in iterations followed by 10,000 (thinned) iterations for posterior analyses.

Table 1 below summarizes the posterior results for the LBT model parameters with columns labeled with 2.5% and 97.5% being relative percentiles of posterior samples. Figure 2 below shows the posterior distributions of these parameters.

	mean	sd	2.50%	median	97.50%
θ_0	-5.218	0.9246	-7.05	-5.212	-3.435
$\theta_{interest\ rate}$	-5.253	7.077	-19.5	-5.178	8.428
θ_{HPI}	0.00611	0.001557	0.003052	0.006147	0.009101
$\theta_{unemployment\ rate}$	3.771	6.098	-8.182	3.749	15.77
$\theta_{FOR\ Mortgage}$	-3.562	9.038	-21.14	-3.604	14.06
κ	250.4	48.93	164.1	247.2	356.8

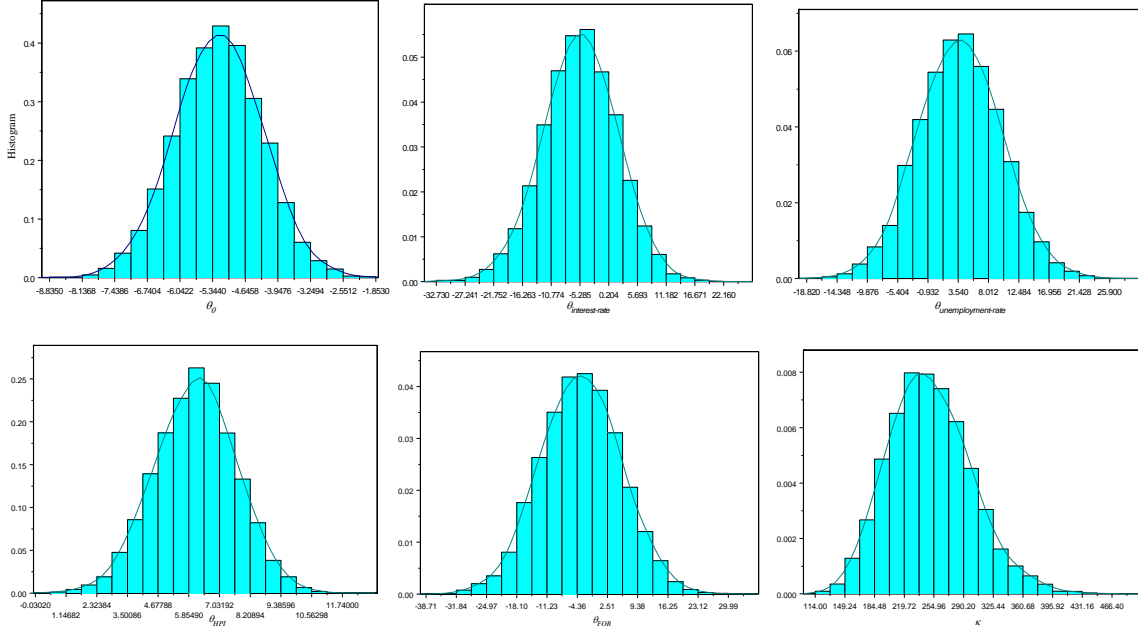Table 1 Posterior Inferences of the LBT Model

Figure 2. Posterior Distributions of LBT Model Parameters

Estimates of aggregate mortgage default rate over all observed quarters are shown in the Figure 3 from which we note that for a large proportion of all 56 quarters, the estimated default rates are larger than the actual observed values from the residential mortgage market. This suggests that besides the four factors included in the model, there may exist other factors influencing the aggregate default risk.

Posterior results from the random effects extension of the LBT model are shown in Table 2. where three random effects parameters (ϵ_{16}, ϵ_{32}, and ϵ_{48}) are also included for illustrative purposes. We note that the posterior results are very similar to the previous case.

Comparison of the actual and the estimated aggregate default rates for random effects LBT model is given in Figure 4. We note that the extension does not provide a significant improvement over the basic LBT model.
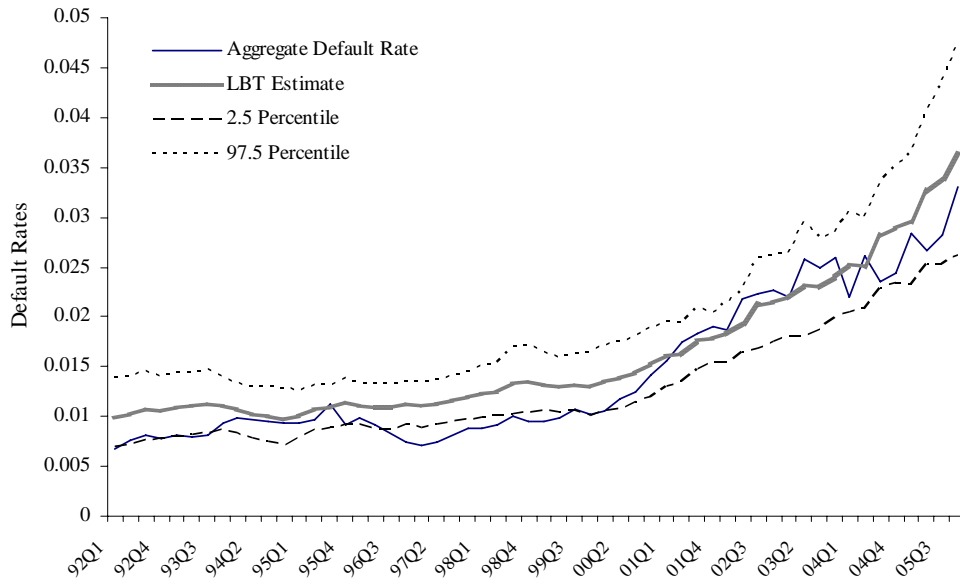
Figure 3. LBT Estimates vs. Real Default Rates

	mean	sd	2.50%	median	97.50%
θ_0	-5.405	0.9961	-7.341	-5.422	-3.419
$\theta_{interest\ rate}$	-4.338	7.501	-18.94	-4.322	10.52
θ_{HPI}	0.006172	0.001685	0.002858	0.006183	0.009516
$\theta_{unemployment\ rate}$	3.642	6.49	-8.951	3.639	16.3
$\theta_{FOR\ Mortgage}$	-2.367	9.258	-20.59	-2.287	15.83
κ	230.7	45.7	149.1	228.5	327.1
τ	31.88	13.41	12.61	29.49	63.85
ε_{16}	-0.00517	0.1835	-0.3775	-0.00371	0.3597
ε_{32}	-0.01293	0.1808	-0.369	-0.01341	0.3426
ε_{48}	0.03048	0.1749	-0.3136	0.03065	0.3793

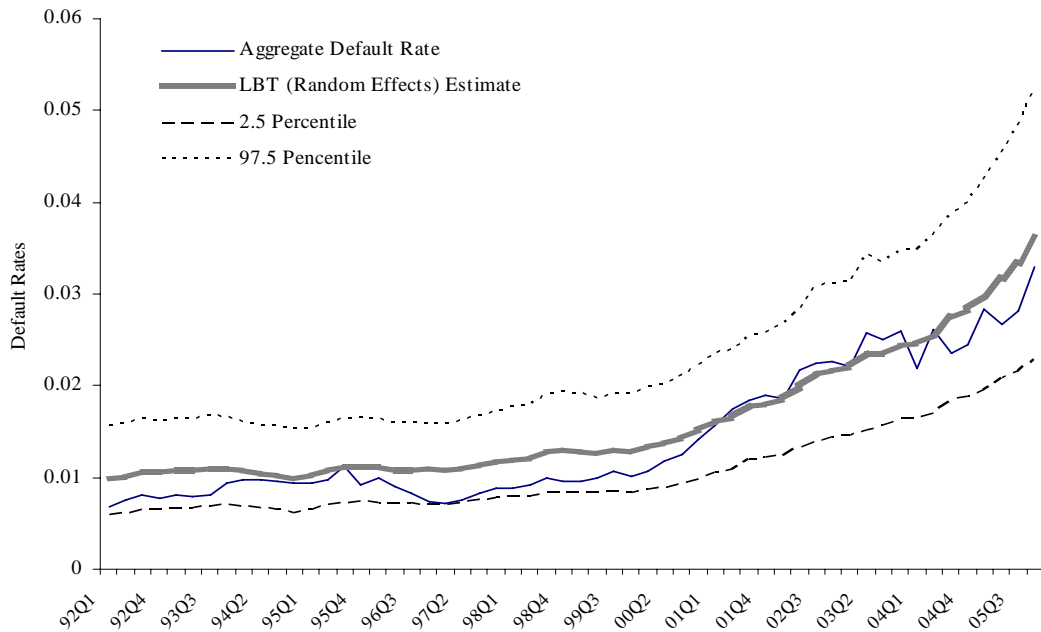Table 2. Posterior Inferences of LBT Model with Random Effects

Figure 4. LBT (with Random Effects) Estimates vs. Real Default Rate

The same data was analyzed using the MMBP model with no covariates included. The posterior results were obtained for a three-state Markov model assuming diffused priors. Figure 5 shows the actual and the estimamted default rates based on the MMBP model. We note that the MMBP model seems to provide a better fit but also yields wider posterior probability intervals.

To be able to assess the predictive performance of these two models we can use a cross validation criterion by excluding some of the data and repeating the analysis. This can be done by excluding the last four quarters and then use the posterior inferences to predict the default rates for those four periods. In Table 3, we present the cross validation results for the three models. Note that the MMBP model provides the best point forecasts while having higher posterior predictive variation, represented by larger standard deviation values in the parentheses. These are also shown in Figure 6.
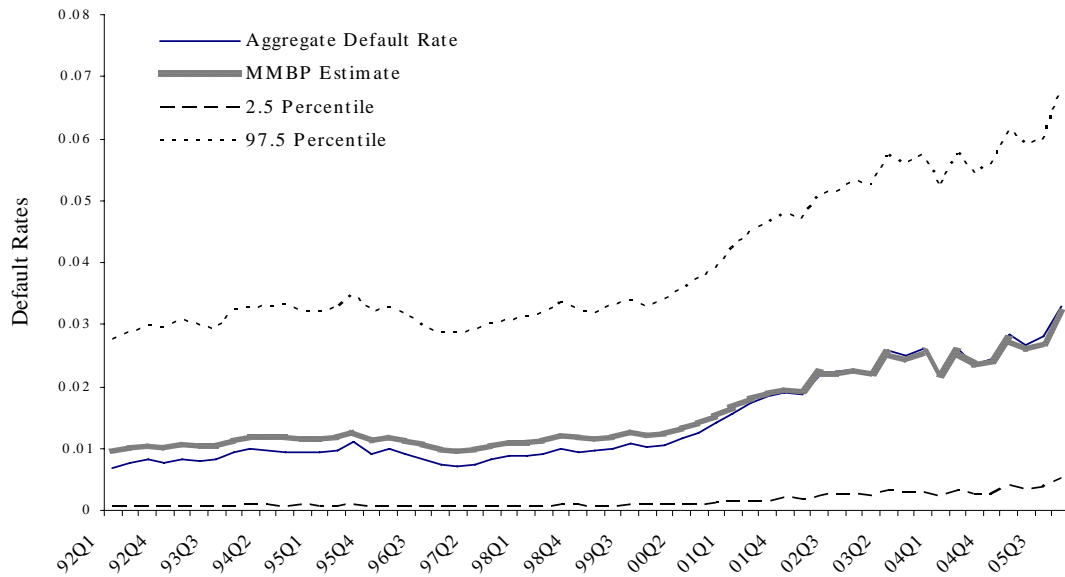
Figure 5. MMBP Estimates vs. Real Default Rates

Quarter	Real Default Rate	LBT Forecast	LBT Forecast (with Random Effects)	MMBP Forecast
05Q1	0.0284	0.033237	0.031766	0.024419
		(0.006111)	(0.007722)	(0.013878)
05Q2	0.0266	0.035877	0.035156	0.027629
		(0.006167)	(0.008113)	(0.014895)
05Q3	0.0282	0.036026	0.034588	0.025947
		(0.006387)	(0.008042)	(0.014389)
05Q4	0.033	0.036584	0.035819	0.031737
		(0.006404)	(0.008194)	(0.016415)

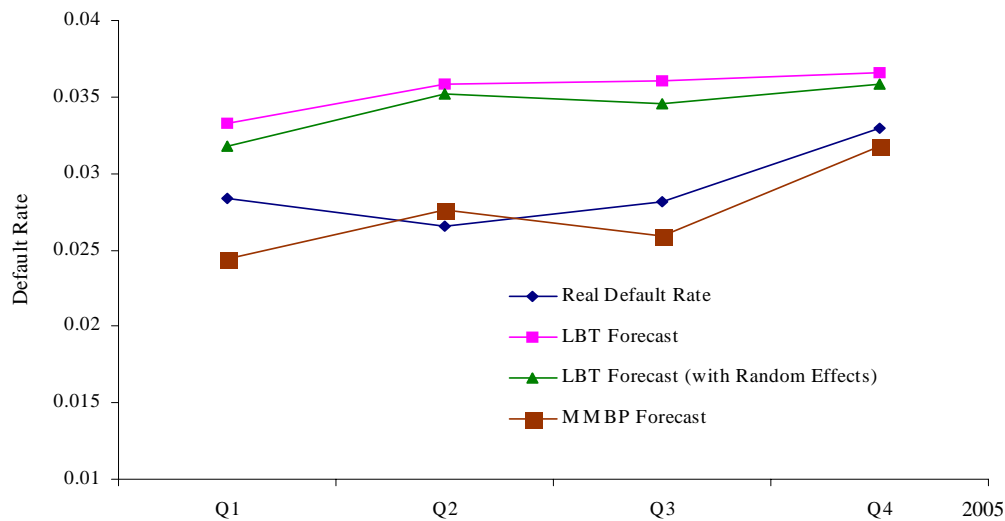Table 3: Comparison of Forecasts for 2005 FHA-FRM Default Rate

Figure 6. Forecast Comparison of the Three Models

References

Aitchison, J., and S. M. Shen, 1980. Logistic-Normal Distributions: Some Properties and Uses. *Biometrika*, 67 (2): 261-72.

Ambrose, B. W., and C. A. Capone, Jr. 1998. Modeling the Conditional Probability of Foreclosure in the Context of Single-Family Mortgage Default Resolutions. *Real Estate Economics* 26 (3):391-429.

Bernardo, J. M., and A. F. M. Smith, 1994. *Bayesian Theory*. New York: John Wiley and Sons.

Calhoun, C. A., and Y. Deng 2002. A Dynamic Analysis of Fixed and Adjustable-rate Mortgage Terminations. *Journal of Real Estate Finance and Economics* 24 (1/2):9-33.

Capone, C. A., Jr. 2002. Research Into Mortgage Default and Affordable Housing: A Primer. March 2002. *http://www.lisc.org/resources/2002/05/affordable_793.shtml?Affordable+Housing*.

Capone, C. A., Jr., and A. Metz 2003. Mortgage Default and Default Resolutions: Their Impact on Communities. Federal Reserve Bank of Chicago, Conference on Sustainable Community Development, Washington, DC, March 27, 2003.

Clap, J. M., G. M. Goldberg, J. P. Harding, and M. LaCour Little 2001. Movers and Shuckers: Interdependent Prepayment Decisions. *Real Estate Economics* 29(3):411-50.

Deng, Y. 1997. Mortgage Termination: An Empirical Hazard Model with a Stochastic Term Structure. *The Journal of Real Estate Finance and Economics* 14(3):309-32.

Gardner, M. J. and D. L. Mills 1989. Evaluating the Likelihood of Default on Delinquent Loans. *Financial Management* 18(4):55-63.

Giliberto, S. M. and A. L. Houston 1989. Relocation Opportunities and Mortgage Default. *AREUEA Journal* 17(1):55-69.

Green, J., and J.B. Shoven 1986. The effects of interest rates on mortgage prepayments, Journal of Money. *Credit and Banking* 18(1): 41-59.

Grunwald, G., A. Raftery, and P. Guttorp, 1993. Time series for continuous proportions. *Journal of the Royal Statistical Society, Ser. B*, 55: 103-16.

Herzog, J. P., and J. S. Earley 1970. Home Mortgage Delinquency and Foreclosure. *National Bureau of Economic Research* (New York).

Kau, J. B., D. Keenan, and T. Kim 1991. Default Probabilities for Mortgages. *Department of Insurance, Legal Studies and Real Estate*, Terry College of Business, University of Georgia. Mimeo.

Kau, J. B., D. C. Keenan, and T. Kim 1993. Transaction Costs, Suboptimal Termination and Default Probabilities. *AREUEA Journal* 21(3):247-63.

Lambrecht, B., W. Perraudin, and S. Satchell 1997. Time to default in the UK mortgage market. *Economic Modelling* 14:485-99.

Leece, D. 2004. Economics of the Mortgage Market: Perspectives on Household Decision Making. *Blackwell Publishing Ltd.*

Lekkas, V., J. M. Quigley, and R. Van Order 1993. Loan Loss Severity and Optimal Mortgage Default. *AREUEA Journal* 21(4):353-71.

Ozekici, S., and R. Soyer 2003. Bayesian Analysis of Markov Modulated Bernoulli Processes. *Mathematical Methods of Operations Research* 57:125-40.

Pavlov, A. D. 2001. Competing Risks of Mortgage Termination: Who Refinances, Who Moves, and Who Defaults?. *Journal of Real Estate Finance and Economics* 23(2):185-211.

Quercia, R. G., and M. A. Stegman, 1992. Residential Mortgage Default: A Review of the Literature. *Journal of Housing Research* 3 (2): 341-79.

Quintana, J. M., and M. West, 1988. Time series analysis of compositional data. in *Bayesian Statistics 3*, J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith, Eds.: 747-56, Oxford University Press.

Robert, C. P., G. Celeux, and J. Diebolt, 1993. Bayesian Estimation of Hidden Markov Chains: A Stochastic Implementation. *Statistics and Probability Letters*, 16: 77-83.

Ross, S., 1976. The Arbitrage Theory of Capital Asset Pricing. *Journal of Economic Theory* 13: 341-60.

Ryden, T., T. Terasvirta, and S. Asbrink, 1998. Stylized facts of daily return series and the hidden Markov model. *Journal of Applied Econometrics* 13: 217-44.

Sandor, R., and H. Sosin 1975. The Determinants of Mortgage Risk Premiums: A Case Study of the Portfolio of A Savings and Loan Association. *The Journal of Business* 48(1):27-38.

Schwartz, E.S., and W.N. Torous 1989a Valuing stripped mortgage-backed securities, *Housing Finance Review* 8(4):241-51.

Spiegelhalter, D., A. Thomas, N. Best, and W. Gilks, 1996. *Bayesian Inference Using Gibbs Sampling Manual (version ii)*. MRC Biostatistics Unit, Cambridge University.

van Order, R. 1990. The Hazards of Default. *Secondary Mortgage Markets*, Fall:29-31.

von Furstenberg, G. M. 1969. Default Risk on FHA-Insured Home Mortgages as a Function of the Terms of Financing: A Quantitative Analysis. *Journal of Finance* 24:459-477.

Williams, A. O., W. Beranek, and J. Kenkel 1974. Default Risk in Urban Mortgages: A Pittsburgh Prototype Analysis. *AREUEA Journal* 2(2):101-2.