Bayesian Time Series Analysis of Mortgage Default Data

Tahir Ekin

ABSTRACT

Residential mortgage default has been a popular subject since the 1960s for scholars, policy analysts and government officials. Understanding the conditions that lead to default and forecasting number of defaults gain more importance because of the recent sub-prime mortgage default crisis. Scholars have identified main factors affecting the default risk by the help of classical sampling theory. This paper proposes an alternative approach by applying Bayesian dynamic linear models to explain the behavior of number of defaults.

Prepared for Summer Paper Requirement
January 15, 2009
Washington D.C., United States

1. INTRODUCTION

Home ownership has been gaining importance as a long term goal for the citizens of United States. According to U.S. Census Bureau statistics, the rate of homeownership has increased to 68.9 % in 2005 from 62.1 % in 1960. One of the main reasons of this increase is the growth in the residential mortgage market. Most residential properties are financed through mortgage loans. There is some risk involved in the mortgage market since some borrowers may fail to repay their mortgage debts and default on their mortgages.

Mortgage default is costly to all parties involved; borrowers, lenders and institutions. Firstly, lenders are affected since they guarantee and insure home mortgages. Increase of mortgage delinquency and foreclosure rates also have a significant impact on the financial markets and the economy. Another cost is linked with the restrictions imposed by regulatory authorities when a history of loan losses impairs the soundness of those institutions (Giliberto and Houston, 1989). Borrowers who default are also affected negatively because of decreases in their credit ratings. Decreased credit scores limit opportunities for future home purchases and even cause emotional distress and a reduction in future employment opportunities (Giliberto and Houston, 1989). In default studies, there is a collective desire to minimize default costs.

In reviewing the literature, scholars have investigated this problem from three main perspectives. Firstly, the mortgage risk is analyzed for different parties spanning borrowers, lenders or institutions. Other than that, the measures used to determine mortgage risk can differ: mortgage interest rate premiums, default rates, delinquency rates, or expected mortgage losses. Finally, the study can either aim to make a theoretical or an empirical contribution to literature (Quercia and

Stegman, 1992).

This paper proposes the use of Bayesian linear dynamic models to explain the behavior of number of defaults. In this paper, the measure to determine the mortgage risk is chosen as aggregate number of defaults. Mortgage risk is analyzed from the perspectives of lenders and institutions. First order polynomial models and dynamic linear models are used in order to investigate the characteristics of number of defaults over time and impact of various macroeconomic factors on number of defaults.

2. LITERATURE REVIEW

There has been extensive research in residential mortgage default literature starting from early 1960s.A detailed review of literature on mortgage default risk between 1960 and 1992 is provided by Quercia and Stegman (1992). In addition to this, an overview of more recent developments can be found in Leece (2004).

Studies dealing with default risk on residential mortgage loans incorporate two theories for default decisions. Jackson and Kasserman (1980) discuss "equity theory of default" (the cash flow approach) and "ability-to-pay theory of default" in detail. "Equity theory" asserts that borrowers base their default decisions on a rational comparison of the prospective cash flows. They default if the market value of the mortgage exceeds house value. On the other hand, according to "Ability-to-pay theory of default"; mortgagors are believed to continue their payments when they have sufficient funds. Equity factors were found to be more important for U.S. market by Jackson and Kasserman, which was later supported by findings of Foster and

Order (1984, 1985) and Waller (1988).

Main research interests in the topic of residential mortgage are; identification of key factors that influence the default decision in aggregate and individual levels, prediction of individual and aggregate default rates and modeling credit risk associated with properties. Nearly all of the previous studies in the literature are based on sampling theory methods for statistical estimation and inference methodologies. There are not many Bayesian works in the residential mortgage default literature. After the introduction of some Bayesian concepts by Herzog (1987), Young and Kazarian (1997) considered a binary regression model and developed a Markov Chain Monte Carlo procedure for Bayesian estimation of the model parameters with disaggregate data. Popova et. al. (2008) proposed a new approach for modeling prepayment rates of individual pools of mortgages by the help of Bayesian forecasting tools. Lastly, Xu(2008) and Xu and Soyer (2008) analyze aggregate mortgage default rates over time via stochastic processes such as logistic beta and markov modulated beta processes. Besides these, there are not any works of Bayesian applications in the mortgage residential default literature. The Bayesian approach in general requires explicit formulation of a model, and conditioning on known quantities, in order to draw inferences about unknown ones (Geweke, 2000). The main contribution of the Bayesian forecasting approach is that it enables representing and measuring all uncertainties by probability.

The main research questions which are analyzed throughout this paper are;

1- Is there evidence of a dynamic relationship among the number of mortgage defaults over time?

2- Is there evidence of a dynamic relationship among the number of mortgage defaults and house price index, interest rate, financial obligation ratio and unemployment rate?
3- What are the influences of the factors such as house price index, cost of index funds, financial obligation ratio and unemployment rate to the number of mortgage defaults?
4- Do proposed Bayesian models provide a good fit retrospectively?
5- Do proposed Bayesian models provide good prediction of the number of mortgage defaults?

In order to answer these research questions, first order polynomial and dynamic linear models are used. We compare the static models and dynamic models to investigate the existence of a dynamic relationship. The parameters of specific explanatory variables are discussed to understand their influence on number of defaults. Error measures which are introduced in *Methods* section are used to assess the fits of retrospective and prospective forecasts.

3. METHODS

Many important underlying concepts and analytic features of dynamic linear models are apparent in the simplest case, the first-order polynomial model. (West and Harrison, 1989) In order to understand the characteristics of number of mortgage defaults over time, a first-order polynomial model is applied first.

In this paper, we concentrate on the dynamic models in which the observational variance, $V_t = V$, is constant and unknown. The constant variance can also be worked by its reciprocal, Φ, which is named as precision. A closed form Bayesian analysis of dynamic linear models with unknown, constant variance V is available if there is a particular structure imposed on W_t sequence and the initial prior. This setting gives the advantage of a derived conjugate sequential updating procedure.

So, the first-order polynomial model is defined by:

Definition 1.1: The first-order polynomial model

Observation equation: $Y_t = \mu_t + v_t \quad v_t \sim N[0, V]$

System equation: $\mu_t = \mu_{t-1} + w_t \quad w_t \sim N[0, V W^*_t]$

Initial information: $(\mu_o \mid D_o, V) \sim N[m_0, V C^*_0]$

$(\Phi \mid D_o) \sim G[n_0/2, d_0/2]$

, where the error sequences v_t and w_t are independent over time and mutually independent conditional on V. In addition they are independent of $(\mu_o \mid D_0)$. D_t represents all available information at time t; $D_t = \{Y_t, D_{t-1}\}$

Initial information is the probabilistic representation of the forecaster`s beliefs and known information at the time before observing a data point. The distributions for different parameters for each time are shown in Definition 1.2 for the specified model in Definition 1.1.

Definition 1.2: Distributional results for the first-order polynomial model

a) *Conditional on V:*

$(\mu_{t-1} \mid D_{t-1}, V) \sim N[m_{t-1}, V C^*_{t-1}]$

$(\mu_t \mid D_{t-1}, V) \sim N[m_{t-1}, V R^*_t]$

$(Y_t \mid D_{t-1}, V) \sim N[f_t, , V Q^*_t]$

$(\mu_t \mid D_t, V) \sim N[m_t, V C^*_t]$

with $R^*_t = C^*_{t-1} + W^*_t$, $f_t = m_{t-1}$, *and* $Q^*_t = 1 + R^*_t$ *,the defining components being updated via*

$m_t = m_{t-1} + A_t e_t$, $C^*_t = R^*_t - A_t^2 Q^*_t$

b) *For precision* $\Phi = V^{-1}$ *:*

$(\Phi \mid D_{t-1}) \sim G[n_{t-1}/2, d_{t-1}/2]$

$(\Phi \mid D_t) \sim G[n_t/2, d_t/2]$

c) *Unconditional on V:*

$(\mu_{t-1} \mid D_{t-1}) \sim Tn_{t-1}[m_{t-1}, C_{t-1}]$

$(\mu t \mid D_{t-1}) \sim Tn_{t-1}[m_{t-1}, R_t]$

$(Yt \mid D_{t-1}) \sim Tn_{t-1}[f_t, , Q_t]$

$(\mu t \mid D_t) \sim Tn_t [m_t, Ct]$

where $C_{t-1} = S_{t-1} C^*_{t-1}$; $R_{t-1} = S_{t-1} R^*_{t-1}$; $Q_t = S_{t-1} Q^*_t$; $S_{t-1} = d_{t-1} / n_{t-1}$

d) *Operational definition of updating equations*

$m_t = a_t + A_t e_t$; $C_t = (S_t / S_{t-1})[R_t - A_t^2 Q_t]$; $S_t = d_t / n_t$; $n_t = n_t + 1$;

$d_t = d_{t-1} + S_{t-1} e_t^2 / Q_t$; $e_t = Y_t - f_t$

e) *Discount factor*

Discount factor, sigma, is also used as an aid to structure and assign values to the sequence W_t. The rate of change in initial uncertainty is (1- δ).The specific structure imposed on the sequence W_t is $W_t = C_{t-1} (1-\delta) / \delta$.

When we want to examine the effect of regressor variables on the response time series, dynamic linear regression models are used. The dynamic linear regression models are systems of equations specifying:

(i) how observations of a process are stochastically dependent on the current process parameters,

(ii) how the process parameters evolve in time (West and Harrison,1989)

Similar to the first-order polynomial model, the assumptions of constant unknown observational variance V and initial gamma priors for V lead to the derivation of conjugate sequential updating procedure which is given in detail in Definitions 1.3 and 1.4.

Definition 1.3: Dynamic Linear Regression Model

Observation equation: $Y_t = \boldsymbol{F_t}' \, \boldsymbol{\theta_t} + v_t \quad v_t \sim N\,[0,V]$

System equation: $\boldsymbol{\theta_t} = \boldsymbol{G}\,\boldsymbol{\theta_{t-1}} + w_t \quad w_t \sim N\,[0,\, V\,\boldsymbol{W}^*_t]$

Initial information: $(\boldsymbol{\theta_o} \mid D_o, \Phi) \sim N\,[m_0, V\,\boldsymbol{C}^*_0]$

$(\Phi \mid D_o) \sim G\,[n_0/2, d_0/2]$

, where the error sequences v_t and w_t are independent over time and mutually independent conditional on V. In addition they are independent of $(\mu_o \mid D_o)$.D_t represents all available information at time t; $D_t = \{Y_t, D_{t-1}\}$, $\Phi = V^{-1}$

Definition 1.4: Distributional results for the dynamic regression model

a) *Conditional on V:*

$(\boldsymbol{\theta_{t-1}} \mid D_{t-1}, V) \sim N\,[\boldsymbol{m_{t-1}}, V\,\boldsymbol{C}^*_{t-1}]$

$(\boldsymbol{\theta_t} \mid D_{t-1}, V) \sim N\,[\boldsymbol{a_t}, V\,\boldsymbol{R}^*_t]$

$(Y_t \mid D_{t-1}, V) \sim N[f_t, V Q^*_t]$

$(\boldsymbol{\theta_t} \mid D_t, V) \sim N[m_t, V \boldsymbol{C^*_t}]$

with $\boldsymbol{a_t} = \boldsymbol{G_t} \boldsymbol{m_{t-1}}$; $\boldsymbol{R^*_t} = \boldsymbol{G_t} \boldsymbol{C^*_{t-1}} \boldsymbol{G'_t} + \boldsymbol{W^*_t}$; $f_t = \boldsymbol{F_t}' \boldsymbol{a_t}$ and $\boldsymbol{Q^*_t} = 1 + \boldsymbol{F_t}' \boldsymbol{R^*_t} \boldsymbol{F_t}$, the defining components being updated via

$\boldsymbol{m_t} = \boldsymbol{a_t} + \boldsymbol{A_t} e_t$, $\boldsymbol{C^*_t} = \boldsymbol{R^*_t} - \boldsymbol{A_t} \boldsymbol{A_t}' Q^*_t$

where $e_t = Y_t - f_t$ and $\boldsymbol{A_t} = \boldsymbol{R^*_t} \boldsymbol{F_t} / Q^*_t$

b) For precision $\Phi = V^{-1}$:

$(\Phi \mid D_{t-1}) \sim G[n_{t-1}/2, \, d_{t-1}/2]$

$(\Phi \mid D_t) \sim G[n_t/2, \, d_t/2]$

c) Unconditional on V:

$(\boldsymbol{\theta_{t-1}} \mid D_{t-1}) \sim Tn_{t-1} [\boldsymbol{m_{t-1}}, \boldsymbol{C_{t-1}}]$

$(\boldsymbol{\theta_t} \mid D_{t-1}) \sim Tn_{t-1} [\boldsymbol{a_t}, \boldsymbol{R_t}]$

$(Y_t \mid D_{t-1}) \sim Tn_{t-1} [f_t, Q_t]$

$(\boldsymbol{\theta_t} \mid D_t) \sim Tn_t [\boldsymbol{m_t}, \boldsymbol{C_t}]$

where $\boldsymbol{C_{t-1}} = S_{t-1} \boldsymbol{C^*_{t-1}}$; $\boldsymbol{R_t} = S_{t-1} \boldsymbol{R^*_t}$; $Q_t = S_{t-1} Q^*_t$; $S_{t-1} = d_{t-1} / n_{t-1}$

d) Operational definition of updating equations

$\boldsymbol{m_t} = \boldsymbol{a_t} + \boldsymbol{A_t} e_t$; $\boldsymbol{C_t} = (S_t / S_{t-1})[\boldsymbol{R_t} - \boldsymbol{A_t} \boldsymbol{A_t}' Q_t]$; $S_t = d_t / n_t$; $n_t = n_t + 1$; $d_t = d_{t-1} + S_{t-1} e_t^2 / Q_t$; $e_t = Y_t - f_t$

where $Q_t = S_{t-1} + F_t' R_t F_t$ and $\boldsymbol{A_t} = \boldsymbol{R_t} \boldsymbol{F_t} / Q_t$

e) Discount factor

$\mathbf{W_t} = \mathbf{C_{t-1}} (1-\delta) / \delta$

Distributional forms and updating equations given in Definition 1.2 and 1.4 can be derived by straightforward applications of Bayes theorem and standard probability calculus. Proofs can be found in West and Harrison (1989). By the help of these dynamic linear regression models, one can examine the influence of the explanatory variables to the dependent variable in a dynamic setting.

The fit of the models in retrospective and prospective manner are also investigated. Mean Absolute Deviation (MAD), Mean Square Error (MSE), Mean Percentage Error (MPE) and Log Likelihood Ratios (LLR) are used as error and fit measures. To evaluate the accuracy the prospective fit, some last values need to be excluded from the model. Parameters of these new models will be different than the former ones since the forecasts do not contain the effects of the predicted value itself.

4. DATA SET and MODELS

There are six models proposed for the mortgages that are started in one specific year. First three are the first order polynomial models that are applied for three different discount factors; 1, 0.9 and 0.8. In the models 4, 5 and 6; dynamic linear regression is applied by using the same discount factors respectively. In this paper, logarithm of number of defaults is used to operationalize the number of defaults instead of using the number itself. This choice results in a big computational advantage with the cost of a limited precision loss, since we deal with normal distribution.

Models 1, 2, 3: for δ values of 1, 0.9 and 0.8 respectively

Y_t = log of number of defaults (N_t)

Observation equation: $Y_t = \mu_t + v_t \quad v_t \sim N[0,V]$

System equation: $\mu_t = \mu_{t-1} + w_t \quad w_t \sim N[0, V\,\boldsymbol{W}^*_t]$

Initial information: $(\mu_o \mid D_o, V) \sim N[m_0, V\,\boldsymbol{C}^*_0]$

$(\Phi \mid D_o) \sim G[n_0/2, d_0/2]$

Models 4, 5, 6: for δ values of 1, 0.9 and 0.8 respectively

Y_t = log of number of defaults (N_t)

Observation equation: $Y_t = \theta_{0,t} + \theta_{CMHPI\,t}\,X_1 + \theta_{COFI,t}\,X_2 + \theta_{FOR,t}\,X_3 + \theta_{u,t}\,X_4 + v_t$

$v_t \sim N[0,V]$

System equation: $\boldsymbol{\theta_t} = \boldsymbol{\theta_{t-1}} + \boldsymbol{w_t} \quad w_t \sim N[0, V\,\boldsymbol{W}^*_t]$

Initial information: $(\boldsymbol{\theta_o} \mid D_o, \Phi) \sim N[\boldsymbol{m_0}, V\,\boldsymbol{C}^*_0]$

$(\Phi \mid D_o) \sim G[n_0/2, d_0/2]$

$\boldsymbol{\theta_t} = (\theta_0, \theta_{CMHPI}, \theta_{COFI}, \theta_{FOR}, \theta_u)$

X_1: Conventional Mortgage Home Price Index (CMHPI)

X_2: Cost of Funds Index

X_3: Financial Obligation Ratio(FOR)

X_4: Unemployment rate

Models 1 and 4 are also named as static models since their discount factor, δ, is equal to 1, which makes it a constant model in terms of initial uncertainty. Static model is similar to classical regression model in that sense. The use of different sigma values enables us to compare the difference between static and dynamic models. In addition, the extent of dynamic behavior in a model can be measured.

In this paper; aggregate count data of the percent of loans past due 90 days are used. They are obtained from individual mortgage loan data by accumulating monthly default occurrences of FRM 30-yr mortgage loans processed in Atlanta region. Loans initiated in 1994 and 2001 are investigated till the end of 2005.These specific years and Atlanta region are chosen because the higher number of mortgage loans available can improve the accuracy of the aggregate data. Individual mortgage loan data is gathered from National Delinquency Survey of Mortgage Banker` s Association. National Delinquency Survey covers approximately 85 percent of more than 50 million outstanding loans in U.S. housing market.

The explanatory variables are chosen according to the main theories for default decisions. Conventional Mortgage Home Price Index (CMHPI) and Federal Cost of Funds Index (COFI), which are provided by Freddie Mac, are chosen according to the "equity theory of default". On the other hand, the determinants of mortgage defaults according to the "ability-to-pay theory" are mortgage FOR (Financial Obligation Ratio) and the U.S. Census South regional unemployment rate. They are provided by U.S. Federal Reserve and U.S. Department of Labor Bureau of Labor Statistics respectively. This choice with respect to both theories enables us to evaluate the determinants of aggregate mortgage defaults from perspectives of lenders and institutions. Borrower related factors are neglected as aggregate data is used.

5. ANALYSIS

In this section of the paper, we use actual data to show the implementation of the proposed models. Firstly, the relationship among the number of mortgage defaults over time is investigated. The error measures and log likelihood ratios are calculated for 1994 and 2001 data which are shown in Table 5.1 and Table 5.2.We can see that static models (first models) have larger error measures and smaller log likelihood ratios. Thus, static models are inferior when we compare all three models. This provides an evidence of dynamic relationship among the number of defaults over time.

Model No	Discounts	MAD	MSE	MPE (%)	LLR
1	$\delta = 1$	45.74	3470.65	163.36	75037.69
2	$\delta = 0.9$	22.41	1396.98	43.38	75192.65
3	$\delta = 0.8$	17.07	827.50	33.01	75163.13

Table 5.1 Summary statistics for Atlanta 1994 data

Model No	Discounts	MAD	MSE	MPE (%)	LLR
1	$\delta = 1$	64.31	15218.69	30.37	9871.61
2	$\delta = 0.9$	46.32	8222.12	22.60	9897.49
3	$\delta = 0.8$	32.81	4444.99	15.82	9908.94

Table 5.2 Summary statistics for Atlanta 2001 data

Secondly, the evidence of dynamic relationship among the number of mortgage defaults and all the explanatory variables is questioned by the help of the dynamic linear regression models

(Table 5.3 and Table 5.4). Similarly, the static models have higher error measures and lower log likelihood ratios compared to dynamic models. The dynamic models have at least two times better fit than static models by all error measures. We can conclude that there is an evidence of dynamic relationship among the number of mortgage defaults and all the explanatory variables.

Model No	Discounts	MAD	MSE	MPE (%)	LLR
4	$\delta = 1$	30.07	2073.03	61.03	75067.39
5	$\delta = 0.9$	16.61	753.56	30.71	75159.56
6	$\delta = 0.8$	15.54	730.73	30.44	75166.05

Table 5.3 Summary statistics for Atlanta 1994 data

Model No	Discounts	MAD	MSE	MPE (%)	LLR
4	$\delta = 1$	35.56	6409.89	13.58	9922.84
5	$\delta = 0.9$	26.22	3921.46	11.08	9926.89
6	$\delta = 0.8$	23.79	3073.04	10.74	9941.74

Table 5.4 Summary statistics for Atlanta 2001 data

Next, the influences of the explanatory variables to the number of mortgage defaults are analyzed. The direction and magnitude of the effect can be understood by specific model parameters.

A negative coefficient of Conventional Mortgage Home Price Index (CMPHI) is expected since people are inclined to default when the price of their houses decrease. Our expectation is met by the dynamic models rather than the static models. (Table 5.5) For 1994, the best fit is provided by Model 5 since the parameter mean is relatively significant and the standard deviation is lower. Besides having a negative mean, the standard deviation of Model 6 for 1994 is very high compared to its mean value. Similar characteristics of CMPHI parameter can be analyzed for 2001 data. Among the models which have negative coefficient of CMPHI, Model 5 is better than Model 6 because of its low variance.

θ_{CMHPI}	Mean	Sd	2.50%	97.50%
Model 4 1994	0.024	0.309	-0.471	0.440
Model 5 1994	-1.803	2.415	-6.534	0.770
Model 6 1994	-1.354	6.026	-16.169	9.977
Model 4 2001	0.278	0.180	-0.151	0.391
Model 5 2001	-0.463	1.135	-3.059	0.380
Model 6 2001	-1.599	2.776	-8.739	0.614

Table 5.5 Posterior Inferences of the model parameter for CMPHI

The operationalization of the interest rate is achieved by Cost of Funds Index (COFI) in this paper. A negative coefficient of COFI is expected as the number of defaults is inversely proportional to the interest rate. When interest rate increases, borrowers benefit from that and become reluctant to default. However, all of the models presented have positive mean values. (Table 5.6) Only the 95% credibility intervals of Model 2 and Model 3 include negative values. This unexpected result can be explained by the aggregate structure of the data. Although an increase in interest rates favors the existing loan borrowers, it may hurt new loan borrowers (Xu, 2008).

θ_{COFI}	Mean	Sd	2.50%	97.50%
Model 4 1994	1.604	0.462	0.274	2.176
Model 5 1994	1.359	1.286	-0.468	4.746
Model 6 1994	1.365	2.986	-6.178	5.104
Model 4 2001	1.086	0.286	0.357	1.369
Model 5 2001	1.829	0.900	0.362	4.106
Model 6 2001	4.155	3.594	0.367	14.110

Table 5.6 Posterior Inferences of the model parameter for COFI

Then, the parameter of Financial Obligation Ratio (FOR) is analyzed for all different models. (Table 5.7) A positive coefficient of Financial Obligation Ratio (FOR) is expected since there is a direct proportionality between the number of defaults and FOR. These expectations are realized in all models. For 1994, Model 5 seems to provide a relatively significant positive parameter mean with the lowest standard deviation. Similarly, Model 5 for 2001 data even does not include any negative values in its 95 % credibility interval and has a lower variance than Model 6.

Θ_{FOR}	Mean	Sd	2.50%	97.50%
Model 4 1994	0.3542	0.142	0.188	0.526
Model 5 1994	1.212	4.176	-6.129	10.599
Model 6 1994	0.432	10.286	-18.199	22.079
Model 4 2001	0.467	0.103	0.314	0.670
Model 5 2001	0.815	0.615	0.316	2.326
Model 6 2001	1.497	3.225	-3.392	10.393

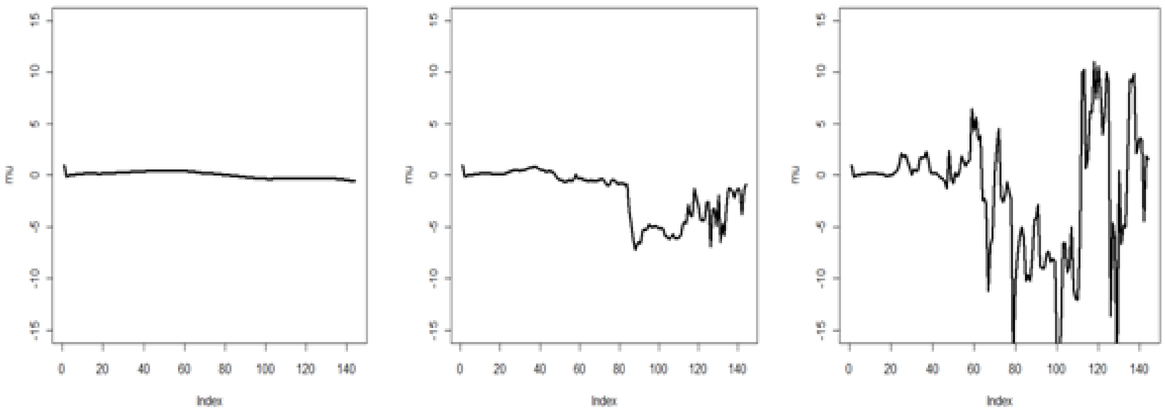
Table 5.7 Posterior Inferences of the model parameter for FOR

The unemployment rate is directly proportional to number of defaults since borrowers cannot pay their mortgage if they do not have any jobs. For 1994 data, dynamic models outperform the static model as their mean is positive and more significant. Similarly, the magnitude of the mean is larger in dynamic models applied to 2001 data. The variances are lower in Model 5 in each year so they seem to explain the relationship between the number of defaults and unemployment rate more clearly.

Θ_{unemp}	mean	Sd	2.50%	97.50%
Model 4 1994	-0.051	0.395	-0.675	0.583
Model 5 1994	0.630	1.607	-1.454	4.611
Model 6 1994	2.090	4.735	-5.427	9.604
Model 4 2001	0.267	0.276	0.013	0.982
Model 5 2001	1.206	1.595	-0.150	4.205
Model 6 2001	1.525	2.412	-0.496	7.477

Table 5.8 Posterior Inferences of the model parameter for unemployment rate

The influence of the explanatory variables change over time in dynamic models as new data becomes available. In static models, the change in the value of parameter is very small relative to the dynamic models since the initial uncertainty stays constant. Dynamic models have increasing uncertainty. The changes in the mean value of the parameter of CMPHI over time for 1994 data is provided as an example.(Figure 5.1) The bigger the discount factor is, the more volatility is realized in the mean of the parameter.



Model4 Model5 Model6

Figure 5.1 Mean values of the parameter of CMPHI over time for 1994 data for all models 4, 5and 6 respectively

We can also compare the proposed models in terms of providing a good fit retrospectively. It can be concluded that the fit for 1994 data is not that good based on the information gathered from Table 5.1 and Table 5.3. The best MPE value is only 30.44 % which is relatively high to mention a good fit. However, all error measures for 1994 data are smaller in dynamic models. Dynamic models provide better fit when more data becomes available. When predictions for number of defaults over time for all models for 1994 are analyzed (See Figure 5.2), the failure of the retrospective fit are seemed to be concentrated in the first 60 observations, 1994-1998 period. After 1998, the actual values of mortgage default are concentrated between 95 % credibility intervals. The insufficient number of observations in the first phases of the model may be the reason for these results.

.

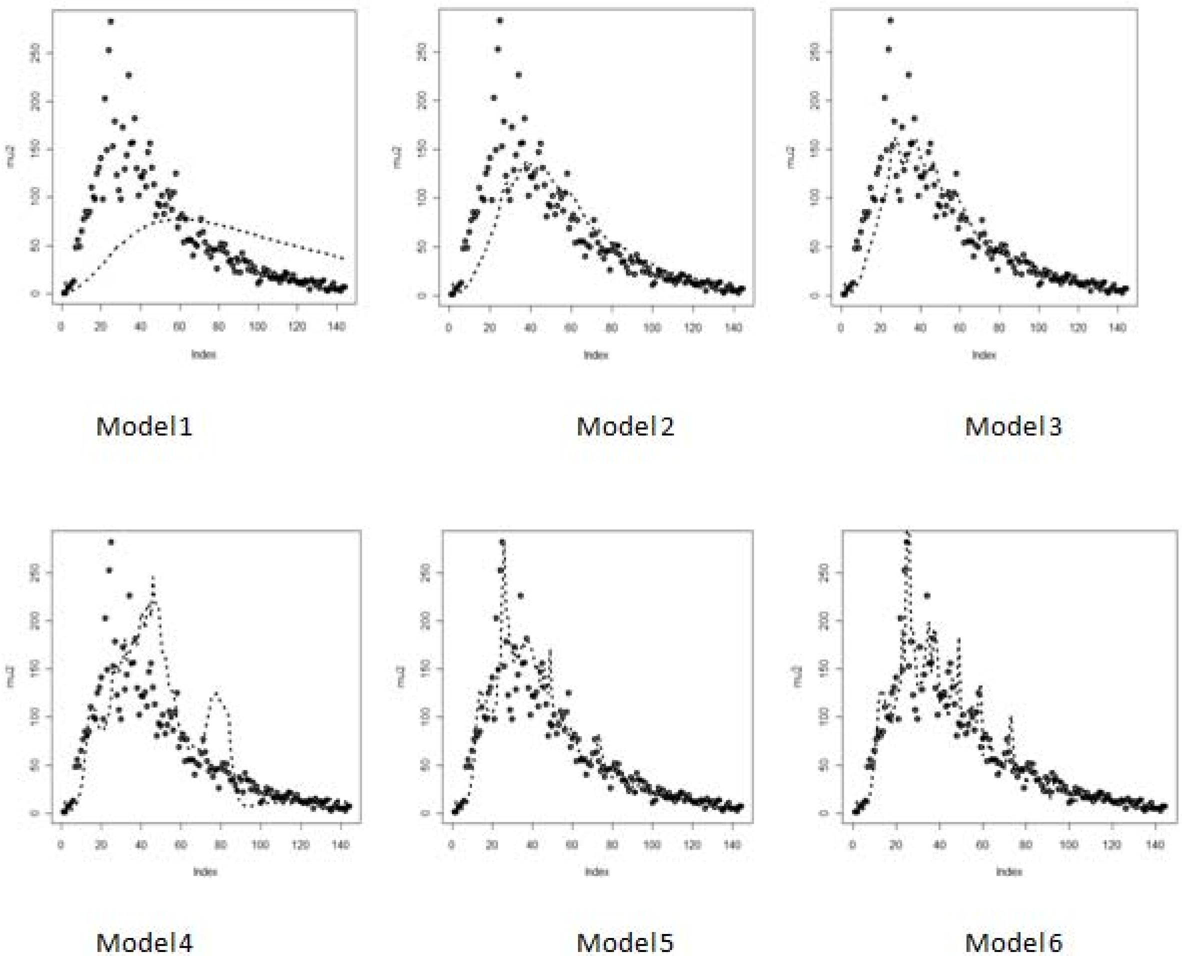


Figure 5.2 Retrospective one-step ahead predictions (dotted line) versus actual number of defaults (bold circles) over time for 1994 data for all models

Proposed models provide a better fit in terms of one step ahead forecasts for 2001 data than 1994 data(See Figure 5.3) The largest value of MPE, 30.37 %, belongs to Model 1.The fit improves with smaller discount factors and dynamic linear models. Mean percentage error values for dynamic linear models lie in the range (10.74 % , 13.58 %) which points out a reasonable fit.

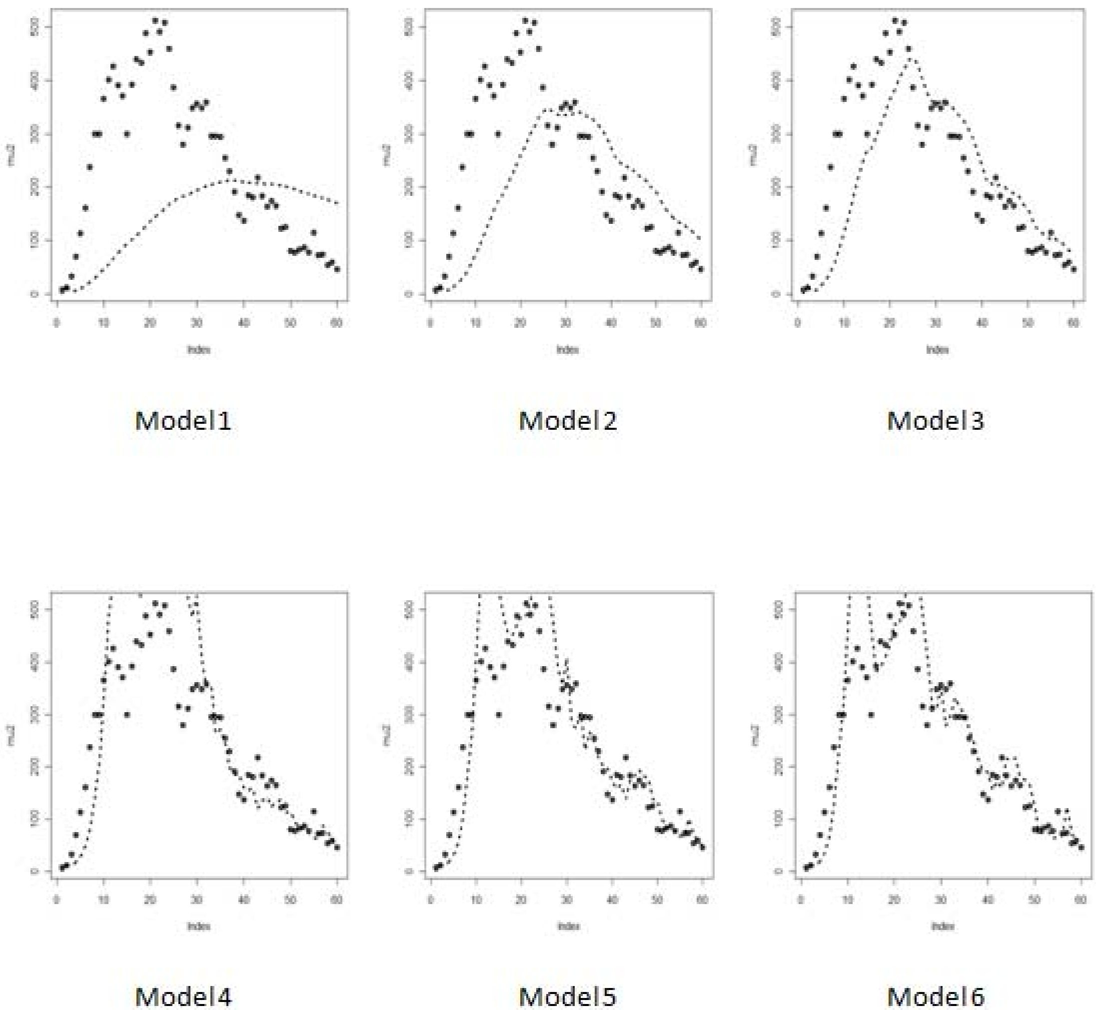


Figure 5.3 Retrospective one-step ahead predictions (dotted line) versus actual number of defaults (bold circles) over time for 2001 data for all models

Lastly, the prospective fit of the models are examined. As we have 144 data points for 1994 data, we have excluded last 10 data points from the models. We have run the models for the first 134 observations and then make a 10-period forecast. The error measures are shown in Table 5.9. Dynamic models have better prospective fit than static models. However, in general we can conclude the models are not that successful in 10-period forecasting. When Figure 5.4 is examined in detail, it can be realized that last data points do not reflect the general behavior of the observations. This can be the reason of bad prospective fit.

Model No	Discounts	MAD	MSE	MPE (%)
1	δ = 1	35.50	1265.69	714.91
2	δ = 0.9	5.12	31.51	116.93
3	δ = 0.8	3.77	17.17	86.97
4	δ = 1	11.62	144.91	242.59
5	δ = 0.9	2.21	7.21	51.64
6	δ = 0.8	2.37	8.18	47.29

Table 5.9 Prospective analysis summary statistics for Atlanta 1994 data

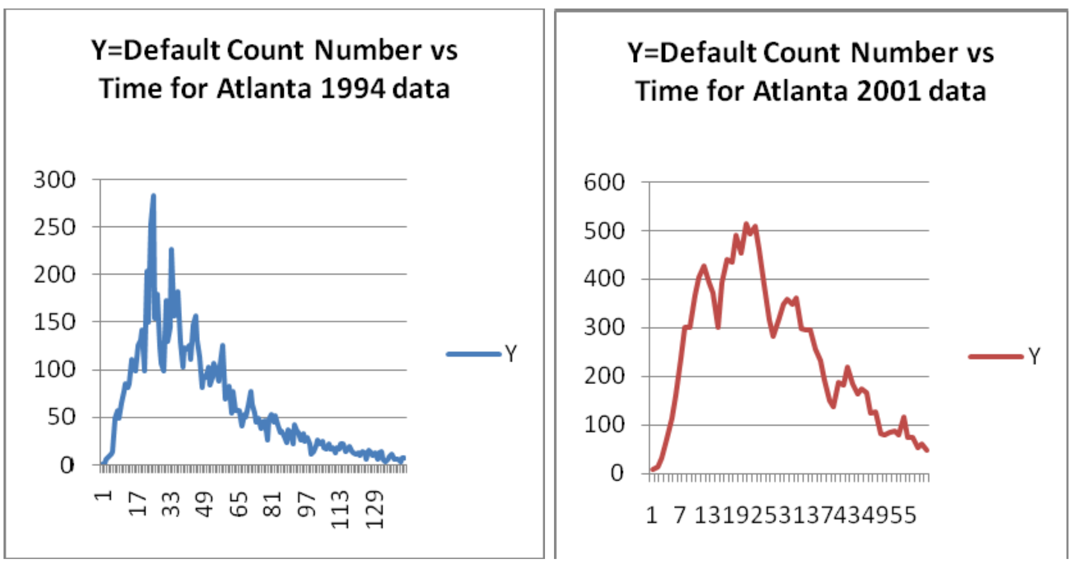


Figure 5.4 Comparison of the time series of default count numbers in 1994 and 2001

Similar to the analysis above, last 8 default counts are excluded from the data set of 2001 Atlanta default data .The predictions are better compared to the 1994 data which may be a result of low volatility among the last 8 default counts(Figure 5.4). Static models may be superior if the data points examined have low volatility. For 2001 data, static dynamic linear model, Model 4, has the lowest MPE value and provides the best prospective fit.

Model No	Discounts	MAD	MSE	MPE (%)
1	$\delta = 1$	116.68	14029.04	179.92
2	$\delta = 0.9$	78.09	6511.41	122.93
3	$\delta = 0.8$	41.53	2121.25	68.82
4	$\delta = 1$	12.65	407.11	15.22
5	$\delta = 0.9$	13.56	415.99	16.98
6	$\delta = 0.8$	17.94	586.35	23.36

Table 5.10 Prospective analysis summary statistics for Atlanta 2001 data

In general, Bayesian methods provide the advantage of defining all uncertainties in terms of probability. Thus, any probabilistic statement for each time can be gathered and analyzed easily. For instance, the gamma distribution of the observational variance, V can be retrieved.

6. CONCLUSION

In this paper, aggregate residential mortgage default data is analyzed using first order polynomial and dynamic linear regression models. Going back to our main research questions, the first question aims to investigate evidence of a dynamic relationship among the number of mortgage defaults over time. Dynamic first-order polynomial models have smaller error measure values and larger log likelihood ratios, which is evidence of better retrospective fit and a dynamic relationship. Then, for the second question we examined the relationship among the explanatory variables and the number of defaults. Dynamic linear models are superior to static models by all error measures. The analyses of these models provide evidence of dynamic relationship in explaining mortgage default behavior. With the help of the third question, we aim to analyze the influences of external macroeconomic factors to the number of defaults. Housing price index is found to be inversely proportional to the number of defaults. However, the interest rate is found to affect the number of defaults directly. This is explained by the aggregate structure of data. Ability to pay factors, financial obligation ratio and unemployment rate, are found to be directly proportional to the default rate.

After that, we analyzed the retrospective and prospective fits. It can be concluded proposed models provide a better retrospective fit in terms of one step ahead forecasts for 2001 data than 1994 data. Prospective fits are not as successful as retrospective fits. The nature of the points in the forecasted period affects the results. Static nature of the forecasted data points in 2001 data leads to superior static models. The predictions for 2001 data are better compared to the 1994 data.

The models used in this paper can be improved by adding more specific characteristics such as seasonal and trend effects. In addition, priors chosen by the help of the expert information may improve the fit of the models slightly. As an extension, proposed models can be applied to different regions and the explanatory variables for number of defaults in different regions can be compared accordingly. On the other hand, there is a potential extrapolation problem. National Delinquency Survey covers approximately 85 percent of the loans in U.S. housing market. This can lead to an extrapolation problem, but it is unavoidable as this survey is the most comprehensive survey at a national U.S. level.

Understanding the determinants of number of mortgage defaults and its behavior over time is very crucial since default affects all parties involved in the financial market. The models proposed by this paper will lead to better policies for institutions via better forecasts and updating procedures. This is one of the first papers using Bayesian approach with aggregate data in residential mortgage default literature. The results of this study will hopefully give birth to other studies.

REFERENCES

Campbell, T., and J. Dietrich. 1983. The Determinants of Default on Insured Conventional Residential Mortgage Loans. Journal of Finance 38 (5):1569-81.

Case K. E., and R.J. Shiller.1995. Mortgage Default Risk and Real Estate Prices: The Use of Index-Based Futures and Options in Real Estate. Journal of Housing Research ,Volume 7 Issue 2.

Ciochetti B.A.,and K.D. Vandell.1999. The Performance of Commercial Mortgages.Real Estate Economics 27 (1): 27–61.

Foster, C., and R. van Order.1984. Option- Based Model of Mortgage Default. Housing Finance Review 3 (4):351-72.

Foster, C., and R. van Order.1985.FHA Terminations: A Prelude to Rational Mortgage Pricing. AREUEA Journal 17 (1) 55-69.

Freddie Mac. 1977-2005.Conventional Mortgage Home Price Index(CMHPI).Quarterly.

Geweke, J.2000. Bayesian econometrics and forecasting. Journal of Econometrics 100: 11–15.

Giliberto, S. M., and A.L. Houston. 1989. Relocation Opportunities and Mortgage Default. AREUEA Journal 13 (3): 273-91.

Green, J., and J.Shoven. 1986. The Effects of Interest Rates on Mortgage Prepayments. Journal of Money, Credit and Banking 18(1) :41-59.

Harrison J. ,and C.F. Stevens.1976. Bayesian forecasting. Journal of the Royal Statistical Society, Series B 38 (3): 205-47.

Harrison J., and M. West.1989. Bayesian Forecasting and Dynamic Models. Springer, New York.

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

Jackson,J., and D. Kasserman. 1980. Default Risk on Home Mortgage Loans: A Test of Competing Hypotheses. Journal of Risk and Insurance 3: 678-90.

Mortgage Bankers Association of America.1977-2005.National Delinquency Survey. Quarterly.

Popova, I., Popova, E. and George, E. "Bayesian Forecasting of Prepayment Rates for Individual Pools of Mortgages" (January 22, 2008). Available at SSRN: http://ssrn.com/abstract=900011

Quercia, R.G. , and M.A.Stegman.1992. Residential Mortgage Default: A Review of the Literature. Journal of Housing Research.

Quigley, J. 1987. Interest Rate Variations, Mortgage Prepayments and Household Mobility. Review of Economics and Statistics 119 (4) :636-643.

Quigley, J., and R. Order.1990. Efficiency in the Mortgage Market: The Borrower's Perspective. Real Estate Economics.

U.S. Bureau of Labor Statistics. 1977-2005. Unemployment rate in the United States.Quarterly.

U.S. Bureau of Economic Analysis. 1977-2005. GDP change in the United States.Quarterly.

Vandell, K. 1978. Default Risk under Alternative Mortgage Instruments. Journal of Finance 33 (5) 1279-96.

von Furstenberg, G. 1969. Default Risk on FHA-Insured Home Mortgages as a Function of the Terms of Financing: A Quantitative Analysis. Journal of Finance 24: 459-77.

von Furstenberg, G., R. J. Green. 1974. Home Mortgage Delinquencies: A Cohort Analysis. Journal of Finance 29: 1545-48.

Waller, N. 1988.Residential Mortgage Default: A Clarifying Analysis.Housing Finance Review 7: 321-33.

Williams A.O., Beranek W., and J.Kenkel.1974. Default Risk in Urban Mortgages: A Pittsburgh Prototype Analysis. Real Estate Economics.

Xu, F., 2008. Bayesian Mortgage Default Models. *Ph.D. Dissertation*. The George Washington University.

Xu,F. and R. Soyer.2008. Discrete Time Bayesian Mortgage Default Models. The Institute for Integrating Statistics in Decision Sciences Technical Report TR-2008-1

Young, M. , and D. Kazarian.1997. Bayesian binary regression with measurement error and auto correction with application to mortgage default models. Working paper series, University of Michigan Business School Research Support 9712-04.