ASSESSMENT OF MORTGAGE DEFAULT RISK VIA BAYESIAN DURATION MODELS

Refik Soyer

Department of Decision Sciences
The George Washington University

Feng Xu

Department of Management
Georgia Southwestern State University

OUTLINE

- Background
  - Mortgage market
  - Default definition

- Issues of interest and previous work

- Connection with reliability modeling

- Modeling duration data
  - mixture models
  - modeling heterogeneity
  - nonmonotonocity of default rates

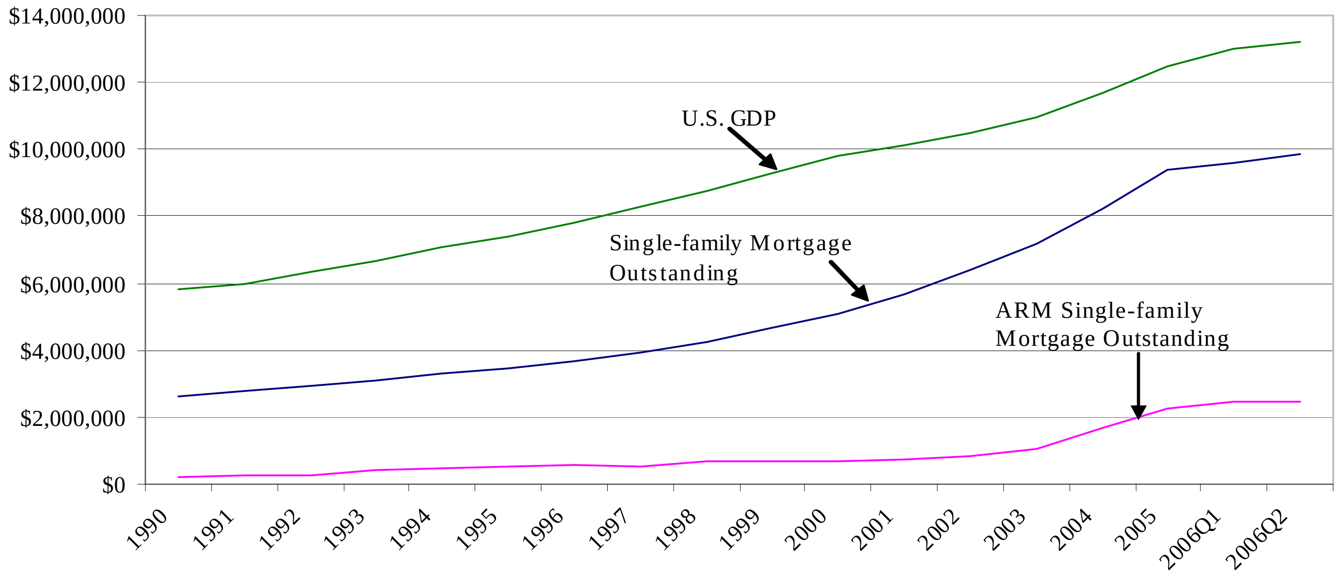
- Illustrations

- Future work: Loan maintenance models

U. S. Residental Mortgage Market

Outstanding debt of single-family mortgage loans has grown from around $2.6 trillion in 1990 to above $9.8 trillion in the second quarter of 2006.
(an increase from 45% to 74.5% as its share in the U.S. GDP)

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



Source: OFHEO & BEA

Mortgage Default

• Legal Definition

"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."

• Definition used in the literature

Being delinquent in mortgage payment for 90 days (3 periods/months).

• Default is costly to all the parties involved

- lenders, borrowers, investors, guarantors

Issues of Interest

- Modeling default rates (nonmonotonic ?)

- Explaining differences in default rates among mortgages

- Identifying key individual borrower, property and loan characteristics affecting likelihood to default (Ability to pay & equity factors)

- Identifying characteristics of "early payment defaults" ($<$ less than 12 months)

- Prediction of individual and aggregate default rates

- Developing loss mitigation options

  $\Rightarrow$ loan modification, pre foreclosure sales, etc.

  $\Rightarrow$ "loan maintenance."

Mortgage Default Models

• Literature starting from 1960s

Quercia and Stegman (1992): A detailed review of literature during 1960-1992
More recent developments can be found in Leece (2004).

• Models

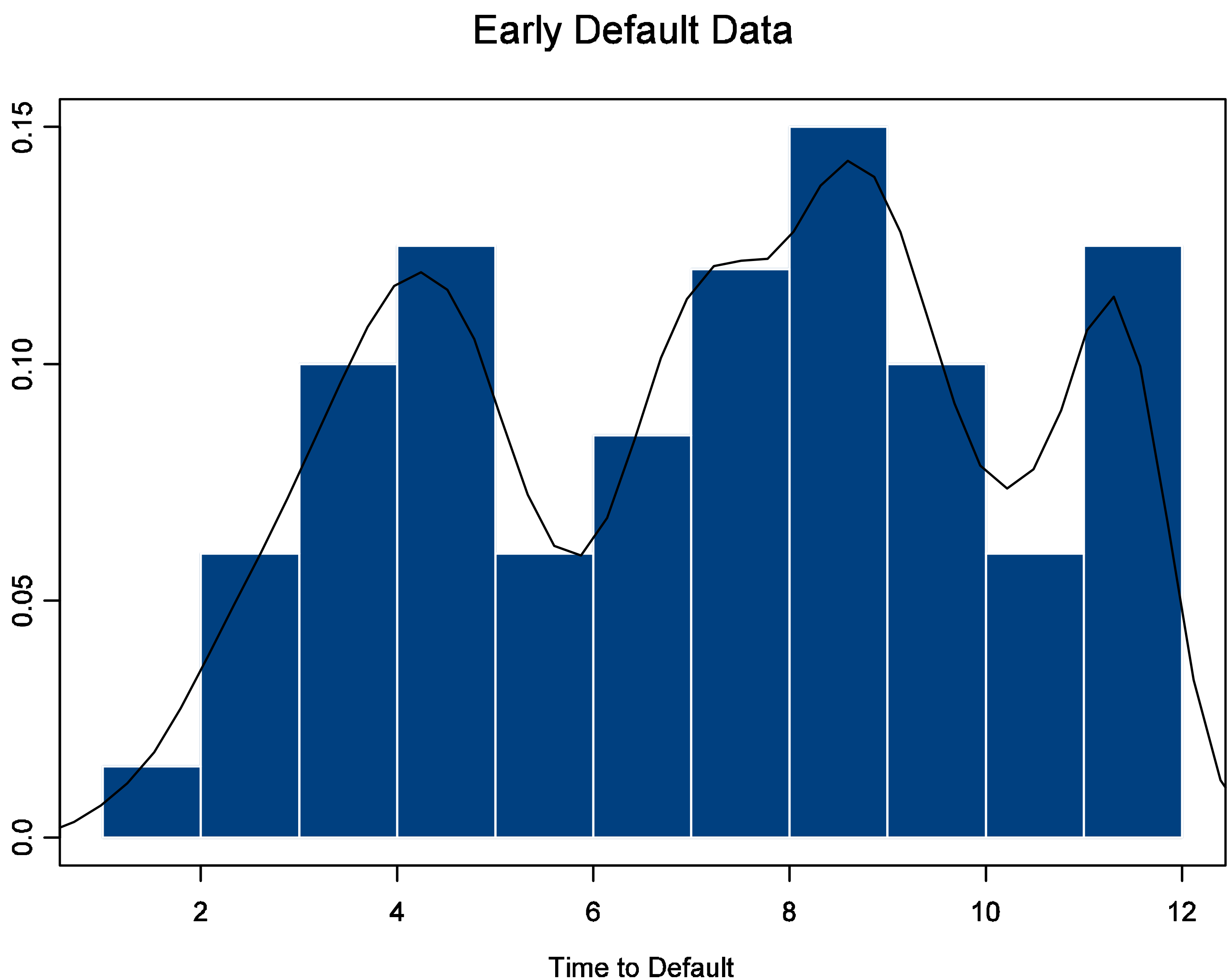
- Option theoretic models (*ruthless default assumption)*
- Direct modeling approach (hazard rate and duration models, competing risk models)

• All based on sampling theory methods

  - Few exceptions: Herzog (1988) and Young and Kazarian (1997) who considered Bayesian binary time-series regression models.
  - Modeling prepayment rates: Popova, Popova and George (2008).

Similarities with Reliability (and Survival) Analysis

- Lynn (2005), Singpurwalla (2007)
- Concept of default is analogous to failure in reliability.
    - modeling default rate (failure rate)
    - modeling time to default (time to failure)
    - default intensity (failure intensity)
- Mortgage duration (Life test duration: censoring)
- Refinancing a loan (replacement: good as new item)
- Prepaying a mortgage (withdrawal ?)



Early Default Data
0.0
0.05
0.10
0.15
2
4
6
8
10
12
Time to Default

Mixture Models

- $K-$ component mixture model for default time T_i

$$f_{T_i}(t) = \sum_{k=1}^{K} \pi_k f_k(t) \quad \text{where} \sum_{k=1}^{K} \pi_k = 1.$$

Do not know from which distribution each t_i is coming from. We define latent variables S_{ik} for each observation such that $\sum_{k=1}^{K} S_{ik} = 1$

$$S_{ik} = \begin{cases} 1 & \text{if } t_i \text{ is from the } kth \text{ component} \\ 0 & \text{otherwise.} \end{cases}$$

If $f_k(t)$ has parameters ϕ_k then

$$S_{ik} = \begin{cases} 1 & \text{if } t_i \sim f_k(t|\phi_k) \\ 0 & \text{otherwise.} \end{cases}$$

Bayesian Analysis

- Independent latent vectors $\boldsymbol{S}_i = (S_{i1} \ldots S_{iK}) \sim Mult(1; \pi_1, \ldots, \pi_K)$.

- Mixing probabilities $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_K)$ can be assumed to have a Dirichlet prior

$$p(\boldsymbol{\pi}) \propto \prod_{k=1}^{K} \pi_k^{\psi_k - 1}$$

- Prior for parameter vector $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_K)$.

Typically independent priors for elements of $\boldsymbol{\phi}$.

- Use of MCMC methods.

Mixtures of PHMs

- Mixtures of Weibull PHMs

For the i^{th} loan the k^{th} component density is given by

$$f_k(t_i|\phi_k, \boldsymbol{X_i}) = \alpha\gamma_k t_i^{\gamma_k - 1} exp(\boldsymbol{\beta}'\boldsymbol{X_i}) exp(-\alpha t_i^{\gamma_k} exp(\boldsymbol{\beta}'\boldsymbol{X_i})).$$

Hurn, Justel and Robert (2003, *JCGS*).

- Bayesian analysis is via MCMC

Typical prior for the covariate parameter vector $\boldsymbol{\beta}$ is multivariate normal independent of ϕ_k's where $\phi_k = (\alpha,\, \gamma_k)$.

Analysis of Early Default Data by Mixture Models

• "Good" versus "bad" loans

• Randomly selected 200 early defaulted mortgage loans originated in 2001.

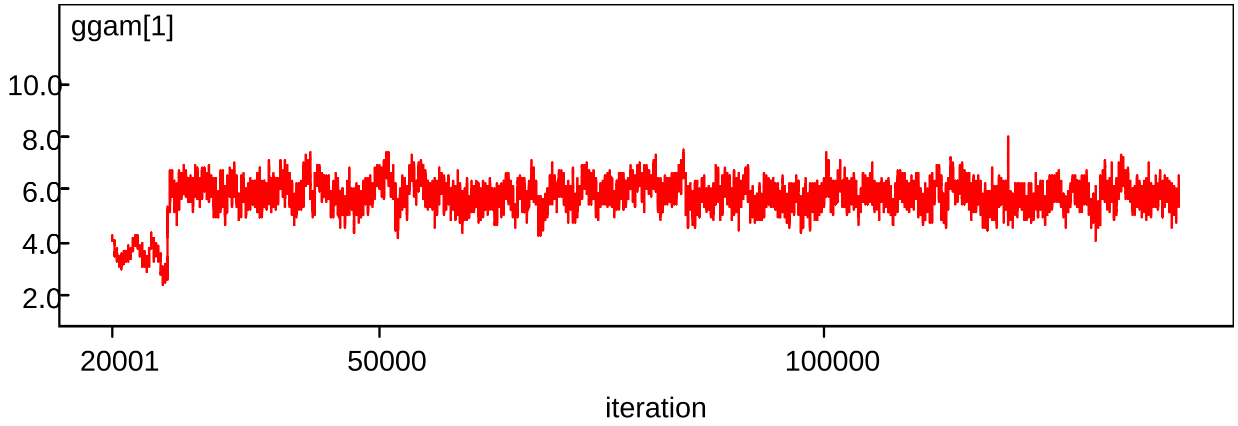
The data is from FHA and it is on default times of 30-year fixed rate single-family mortgage loans.

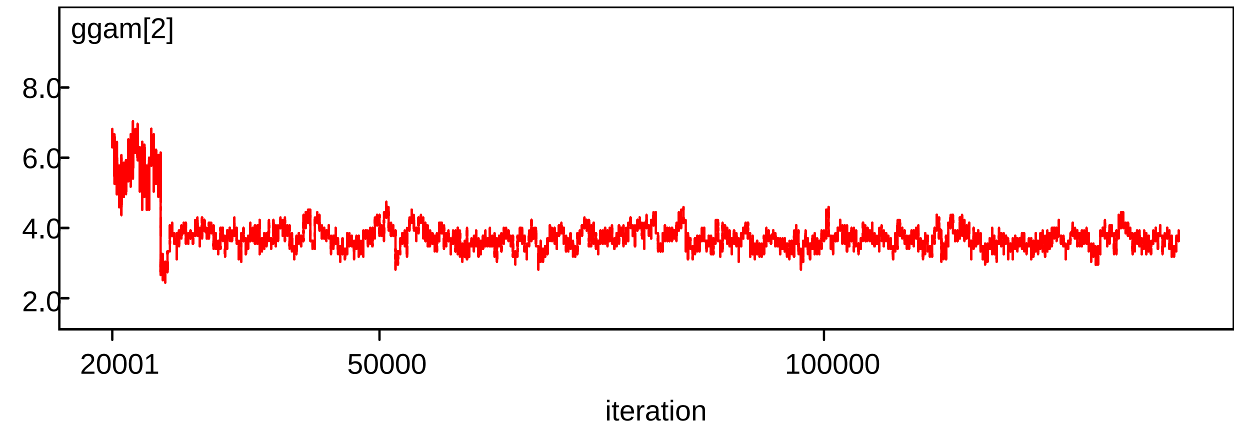
• On each loan we have information:

- loan amount
- interest rate
- borrower's effective household income
- LTV ratio (category),
- borrower's marital status and age

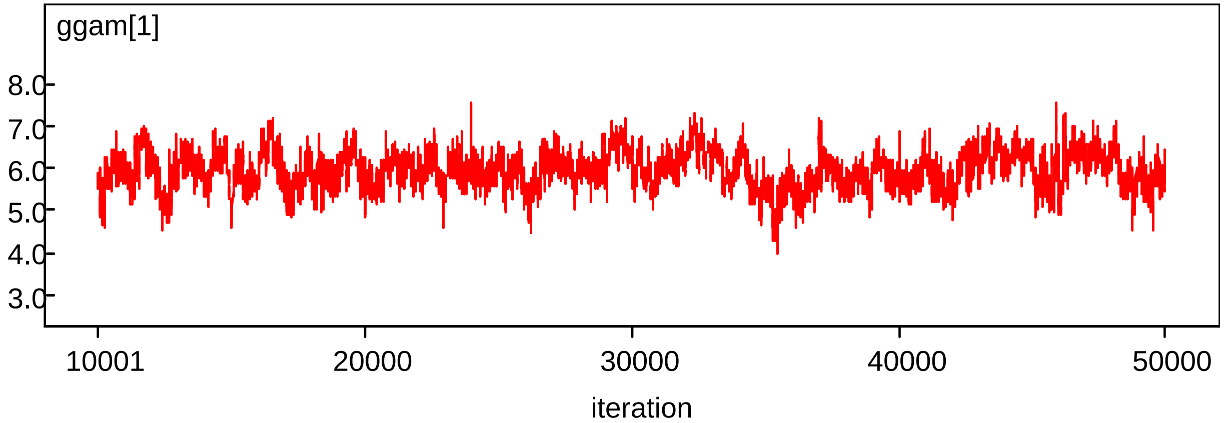
all recorded at loan origination time.

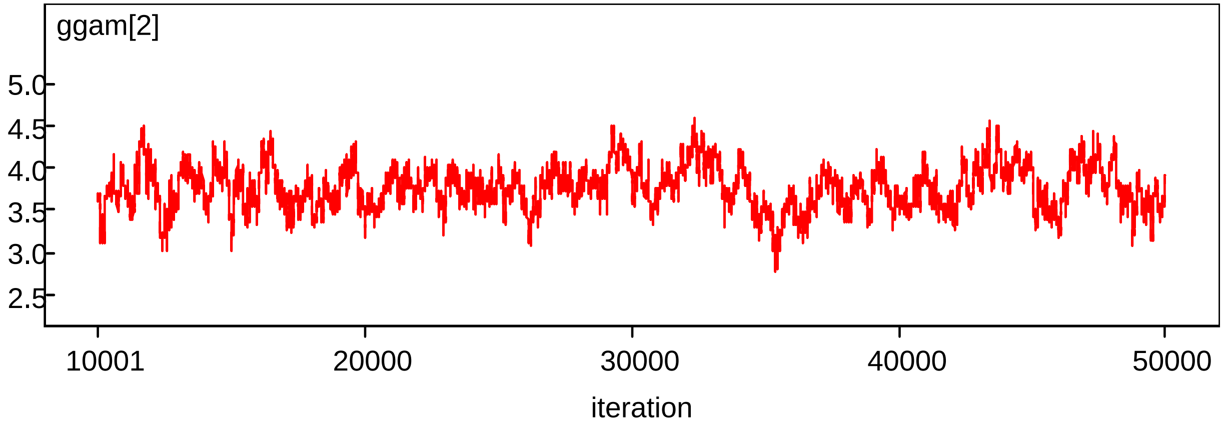
Trace Plots of Shape Parameters Mixture Model (No covariates)



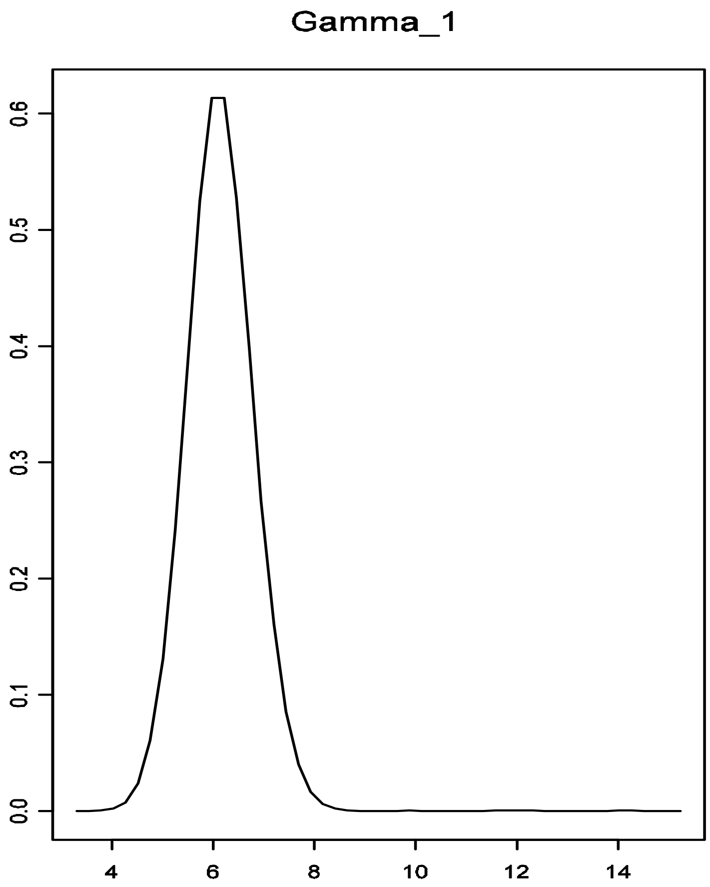


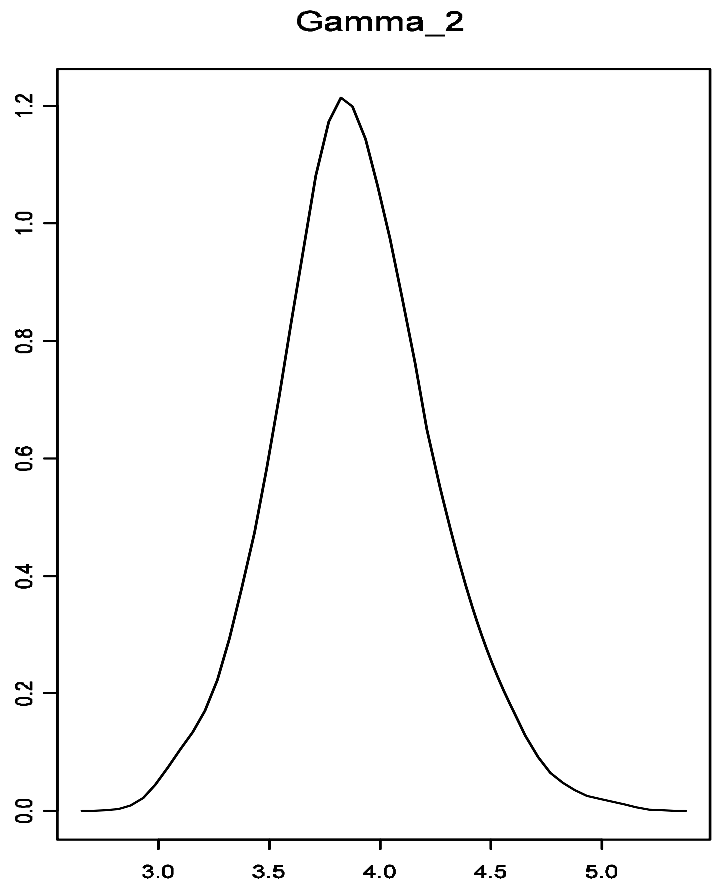
Trace Plots of Shape Parameters
Mixture Model with Covariates





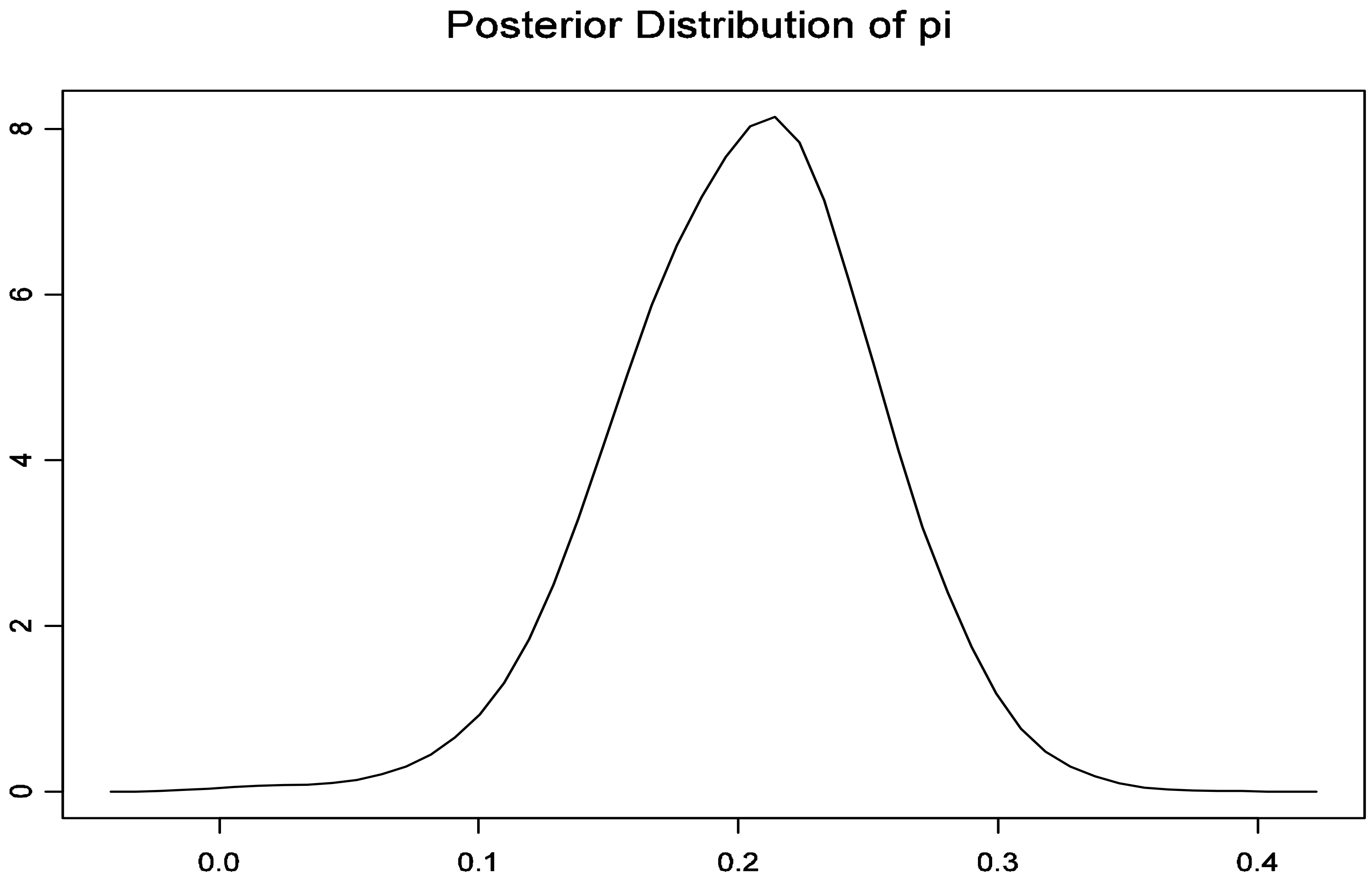
Mixture Model with Covariates



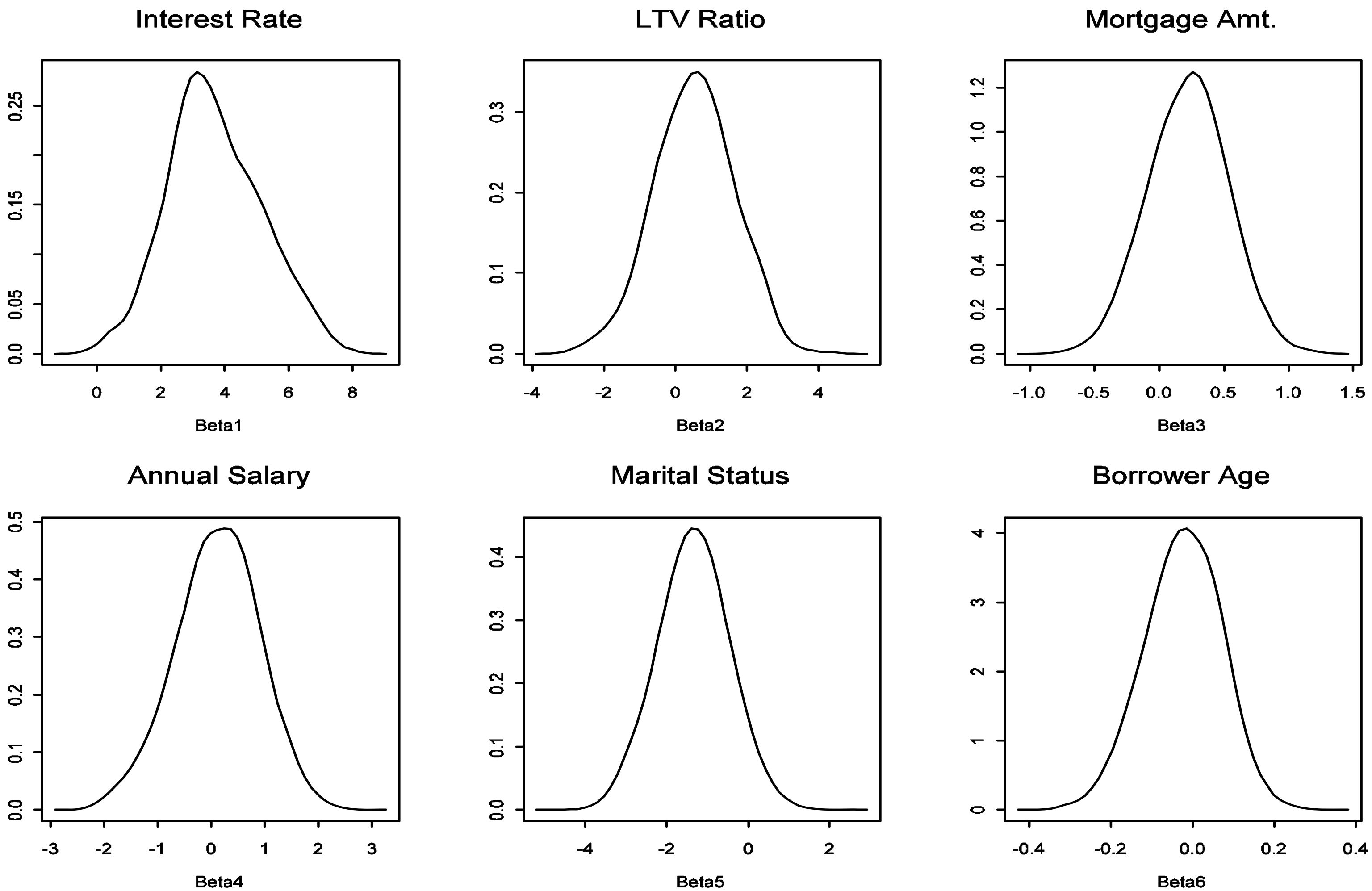


We have $Prob(\gamma_1 > \gamma_2 | D) \approx 1$.

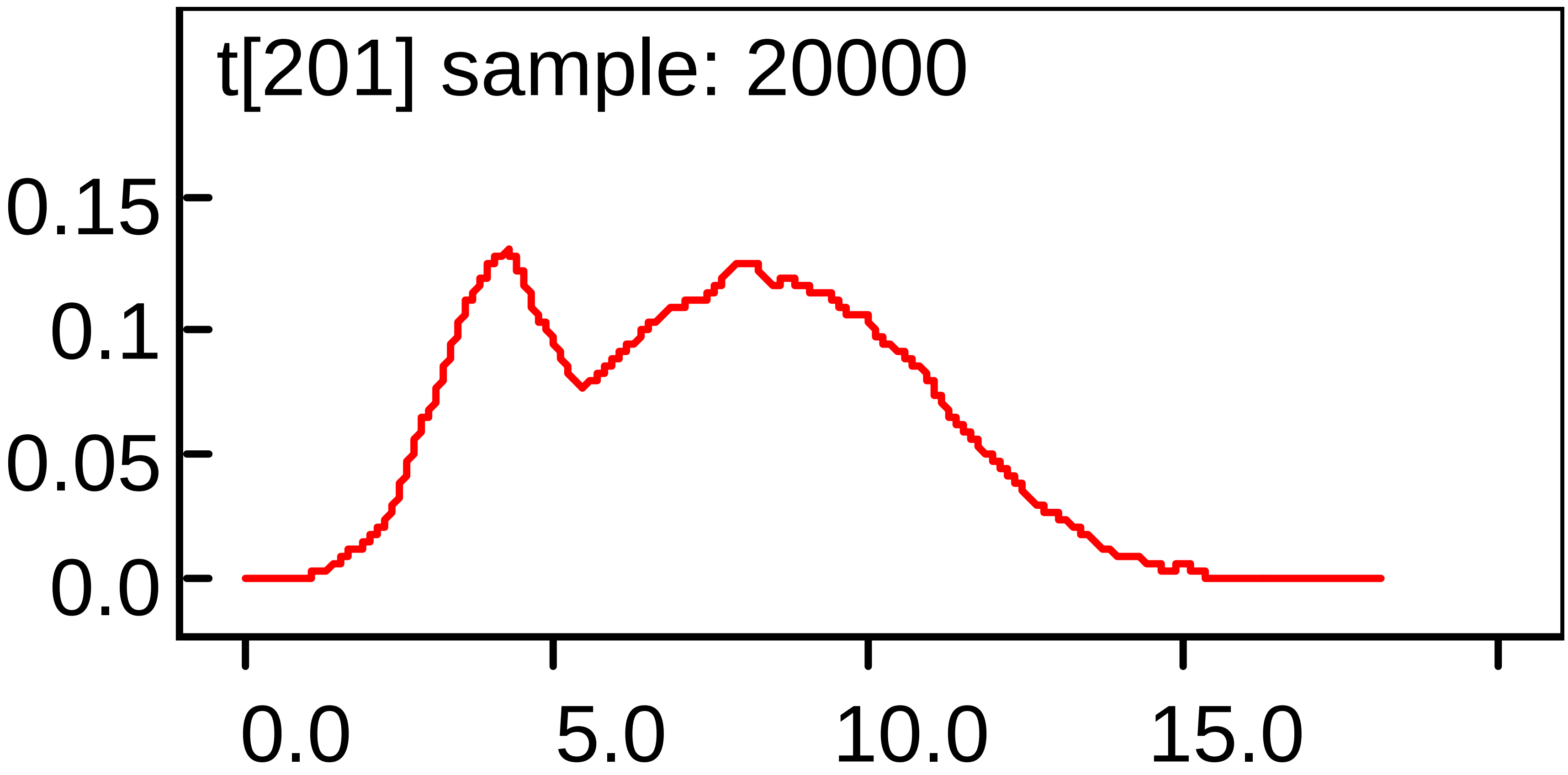
Mixture Model: Posterior Distribution of Mixing Probability $\pi(1)$



Mixture Model: Covariate Effects on Default Rate



Posterior Predictive Distribution
Covariates: (6.25 4 100000 20000 Single 41)



Mixtures with Different Covariate Coefficients

- Mixtures of Weibull PHMs

For the i^{th} loan the k^{th} component density is given by

$$f_k(t_i|\phi_k, \boldsymbol{X}_i) = \alpha\gamma_k t_i^{\gamma_k-1} exp(\boldsymbol{\beta}'_k \boldsymbol{X}_i) exp(-\alpha t_i^{\gamma_k} exp(\boldsymbol{\beta}'_k \boldsymbol{X_i})).$$

Similar results were obtained with again interest rate having a definite effect.

Parameter	Mean	StDev	95% CCI
$\beta_{\text{interest-rate, 1}}$	8.720	3.425	(2.569, 15.820)
$\beta_{\text{interest-rate, 2}}$	3.096	1.626	(0.005, 6.387)
$\beta_{\text{annual-income,1}}$	2.247	2.040	(-2.116, 5.882)
$\beta_{\text{annual-income,2}}$	-0.603	0.906	(-2.401, 1.151)
$\beta_{\text{borrower-age,1}}$	-4.047	2.134	(-8.028, 0.381)
$\beta_{\text{borrower-age,2}}$	0.422	1.058	(-1.702, 2.431)

Modeling Other Default Data

• Heterogeneity in loans ?

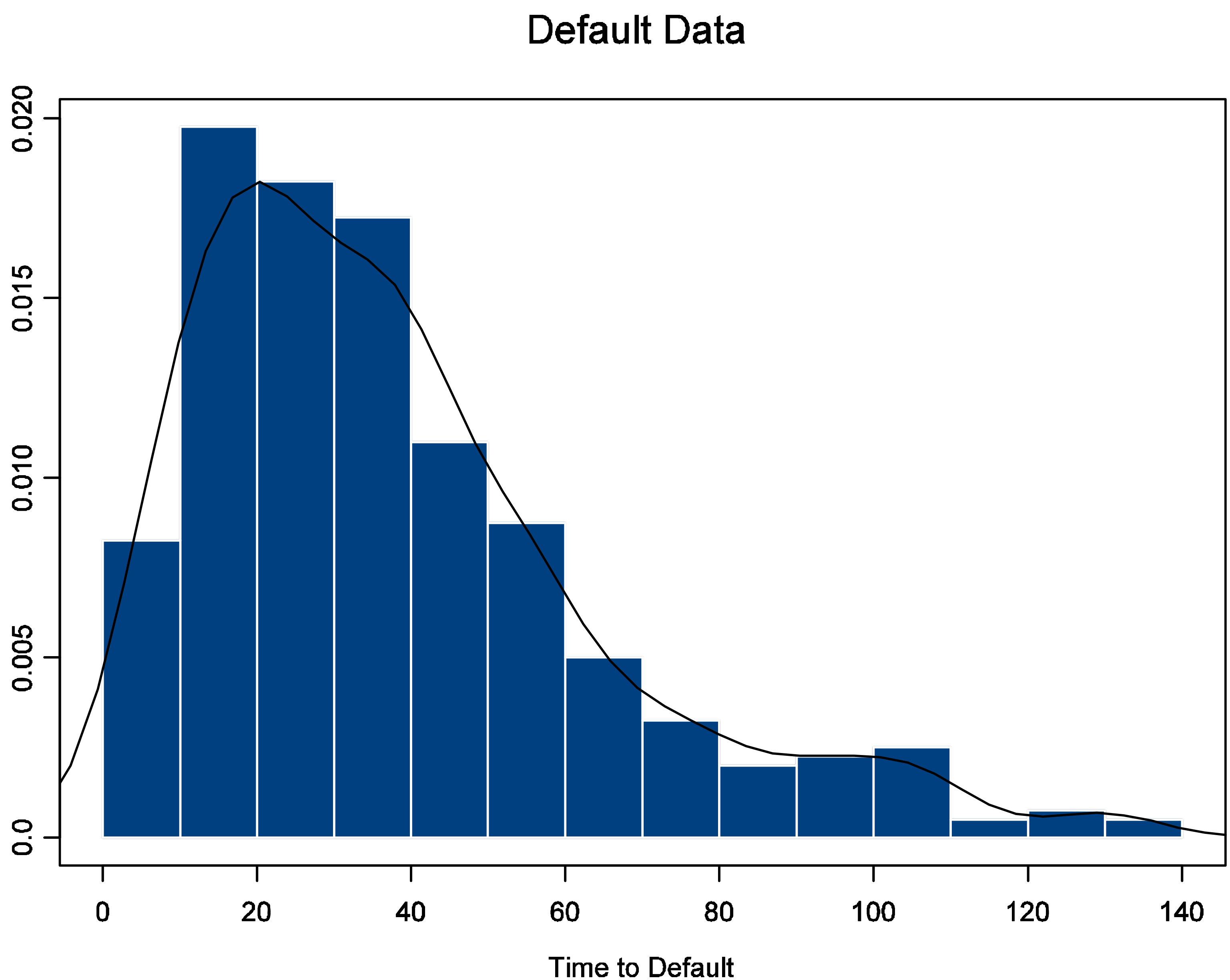
• We consider a Weibull PHM where we incorporate loan heterogeneity by indexing the scale parameter

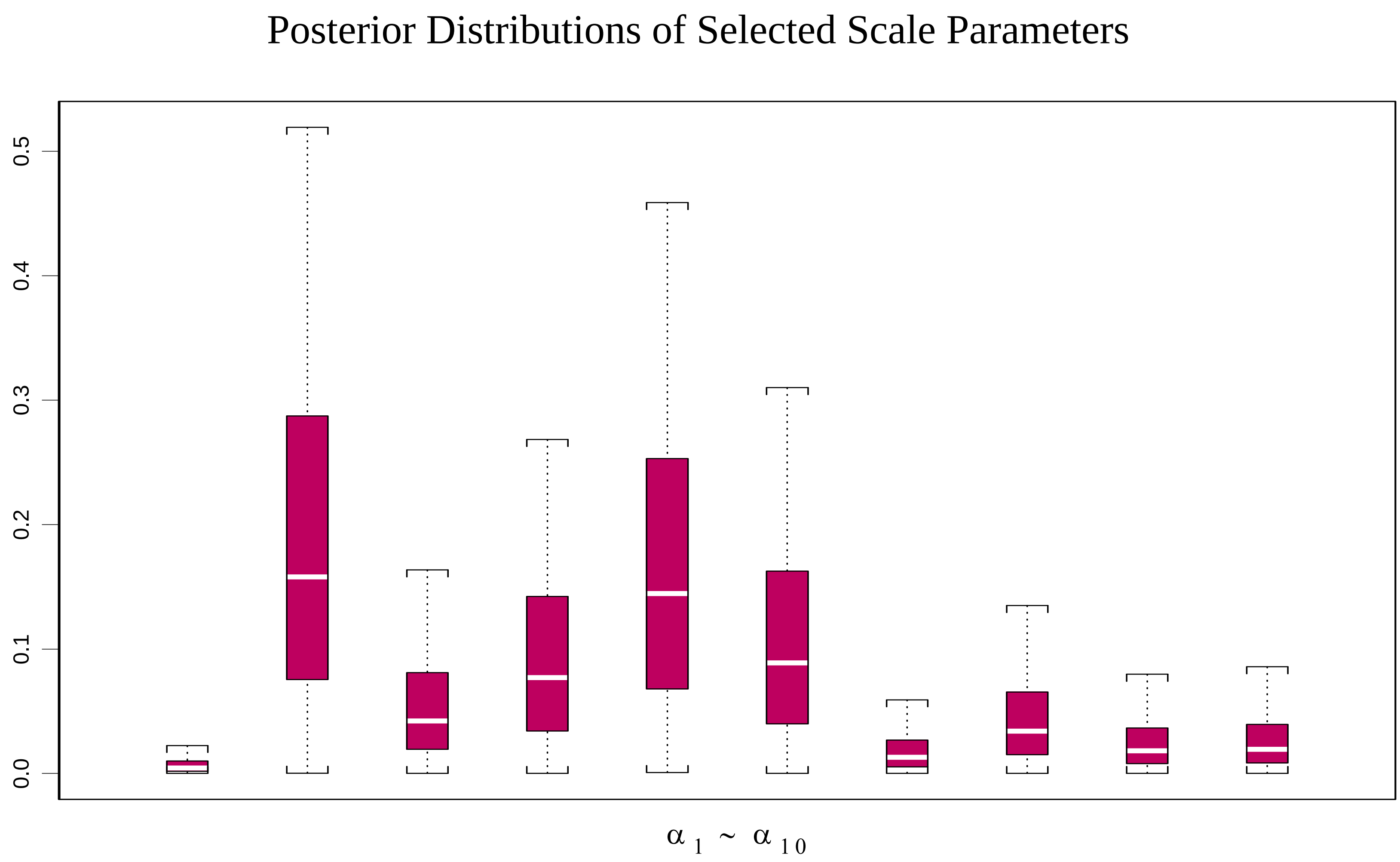
$$f(t_i|\alpha_i,\gamma,\boldsymbol{\beta},\boldsymbol{X_i}) = \alpha_i \gamma t_i^{\gamma-1} exp(\boldsymbol{\beta}'\boldsymbol{X_i}) exp(-\alpha_i t_i^{\gamma} exp(\boldsymbol{\beta}'\boldsymbol{X_i})).$$

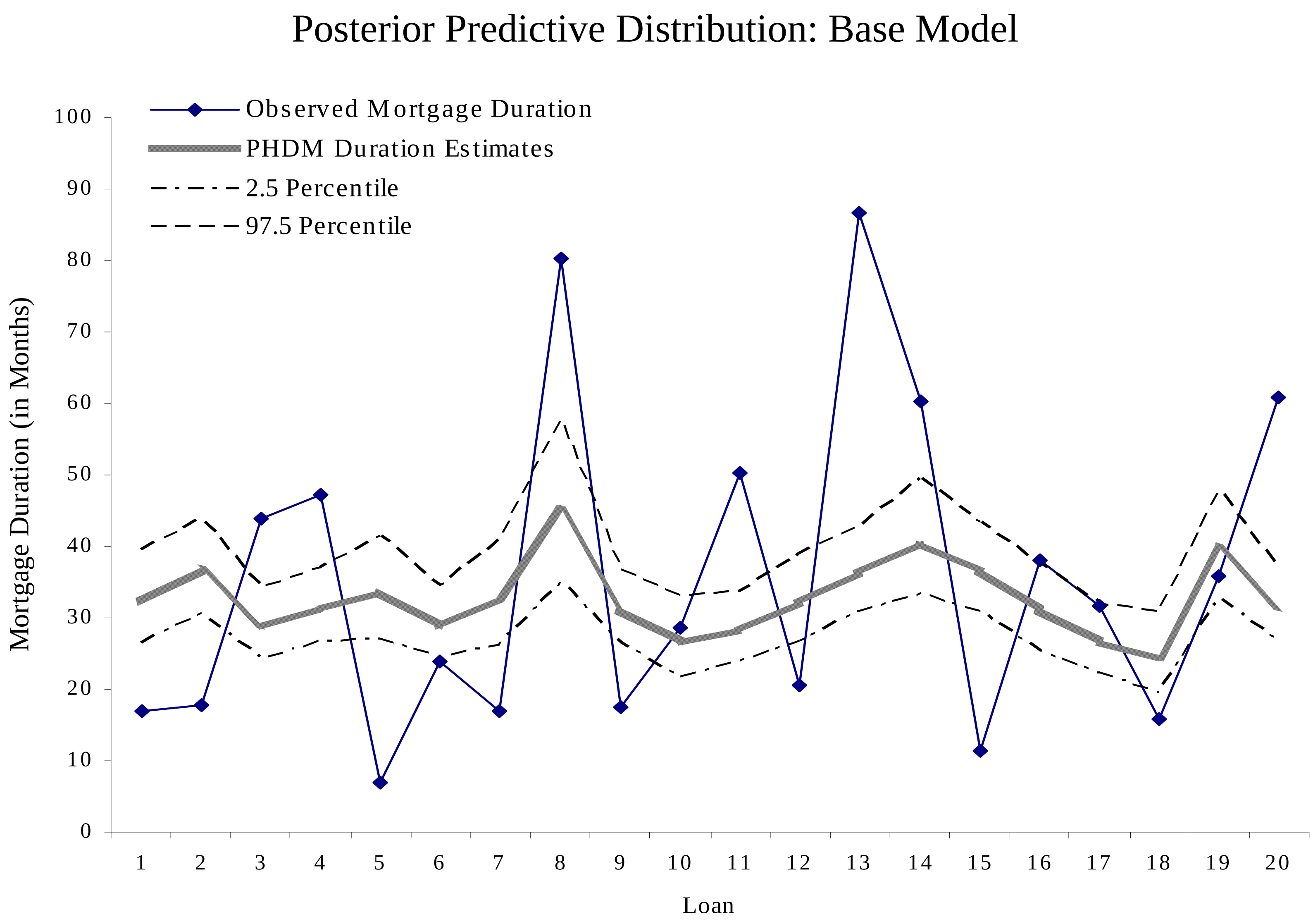
where α_i's are treated as exchangeable random quantities.

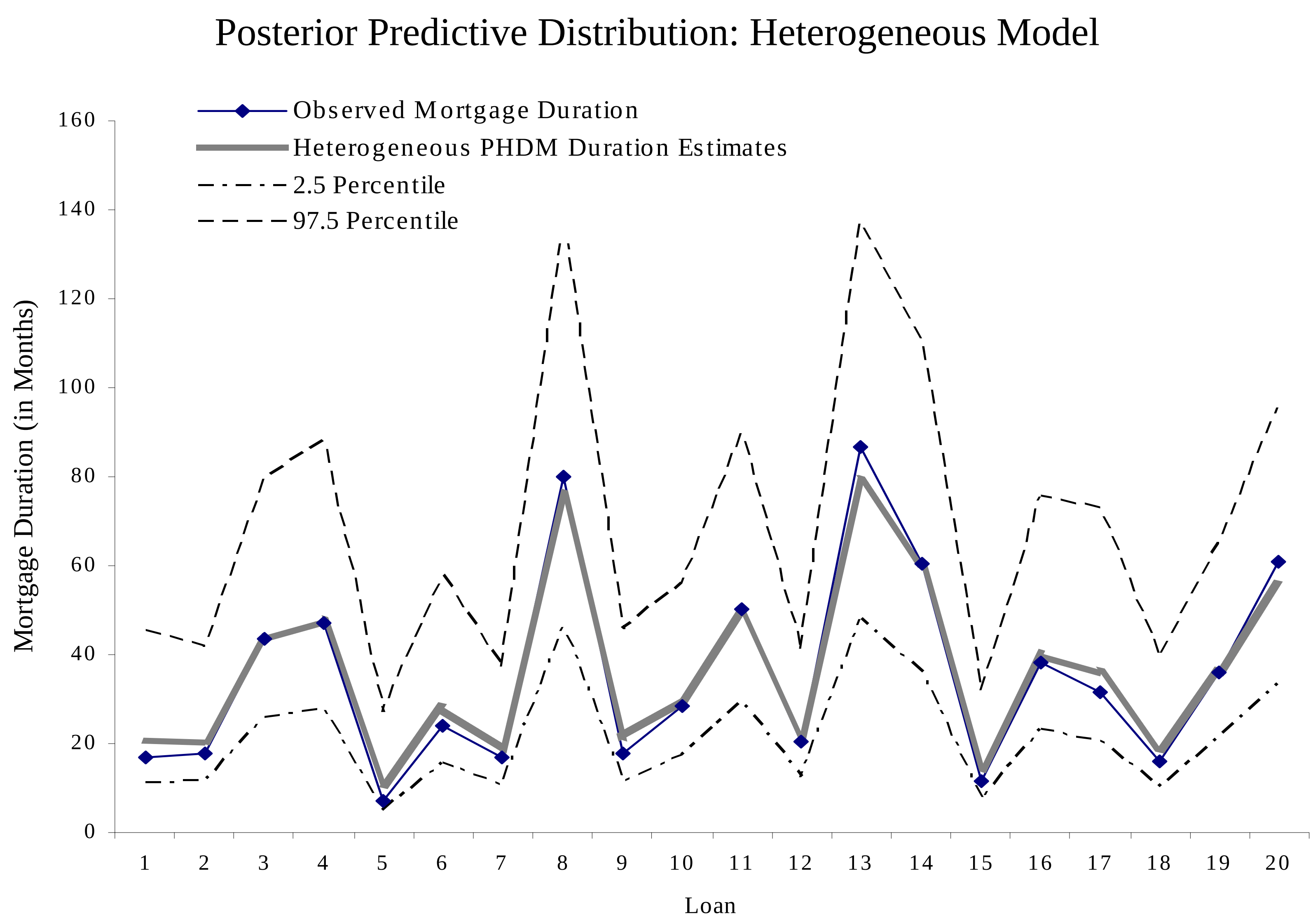
• We also compare the model with the common scale model $\alpha_i = \alpha$ for all i.

• Data is similar to the previous case but now is not limited to early default loans. We have a random sample of 400 defaulted loans from 1994 cohort.

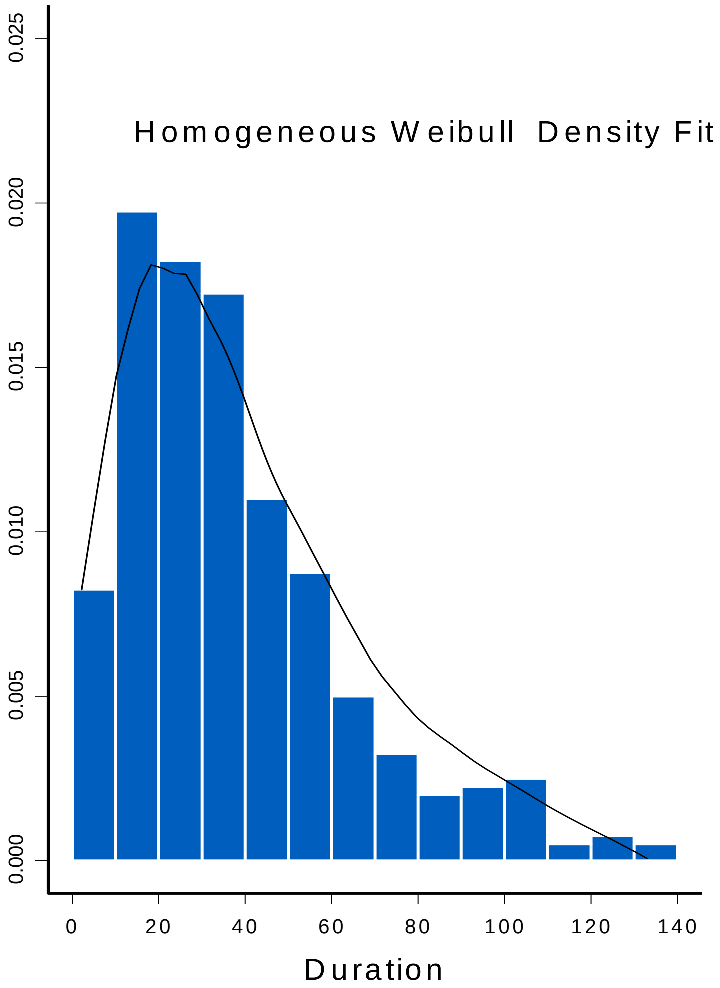

Default Data
0.0
0.005
0.010
0.015
0.020
0
20
40
60
80
100
120
140
Time to Default

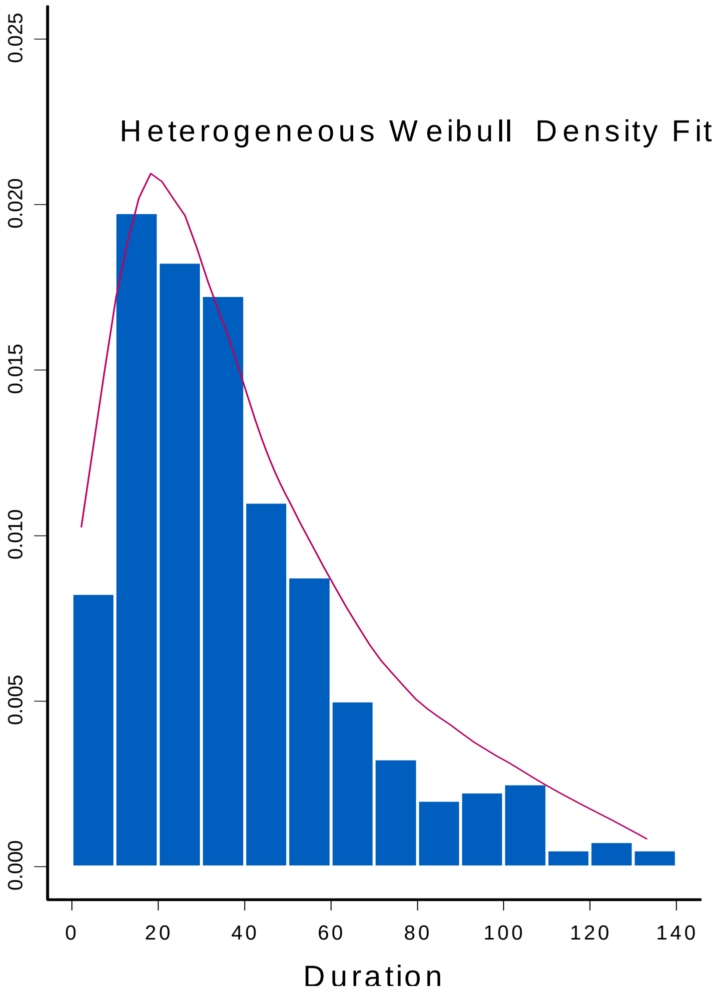

Posterior Distributions of Selected Scale Parameters
0.0
0.1
0.2
0.3
0.4
0.5
$\alpha_1 \sim \alpha_{10}$


Posterior Predictive Distribution: Base Model
Observed Mortgage Duration
PHDM Duration Estimates
2.5 Percentile
97.5 Percentile
Mortgage Duration (in Months)
100
90
80
70
60
50
40
30
20
10
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
Loan


Posterior Predictive Distribution: Heterogeneous Model
Observed Mortgage Duration
Heterogeneous PHDM Duration Estimates
2.5 Percentile
97.5 Percentile
Mortgage Duration (in Months)
160
140
120
100
80
60
40
20
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
Loan

Comparison of Model Fits





Modeling Nonmonotonicity of Default Rate

• Using generalized gamma model for 1994 cohort

$$f_i(t_i \mid \alpha, \gamma, \lambda, \boldsymbol{\theta}) = \frac{\gamma\lambda^{\alpha}\, t_i^{\alpha\gamma-1}}{\Gamma(\alpha)} exp(\alpha\boldsymbol{\theta}'\boldsymbol{X_i})\; exp[(-\,\lambda t_i^{\gamma} exp(\boldsymbol{\theta}'\boldsymbol{X_i})],$$

where

$$h_0(t;\, \alpha, \gamma, \lambda) = \frac{\gamma\lambda^{\alpha}\, t^{\alpha\gamma-1}\; exp\Big\{-\,\lambda t^{\gamma}\Big\}}{\Gamma(\alpha) - \Gamma_{\lambda\, t^{\gamma}}(\alpha)},\;\; \alpha, \gamma, \lambda > 0$$

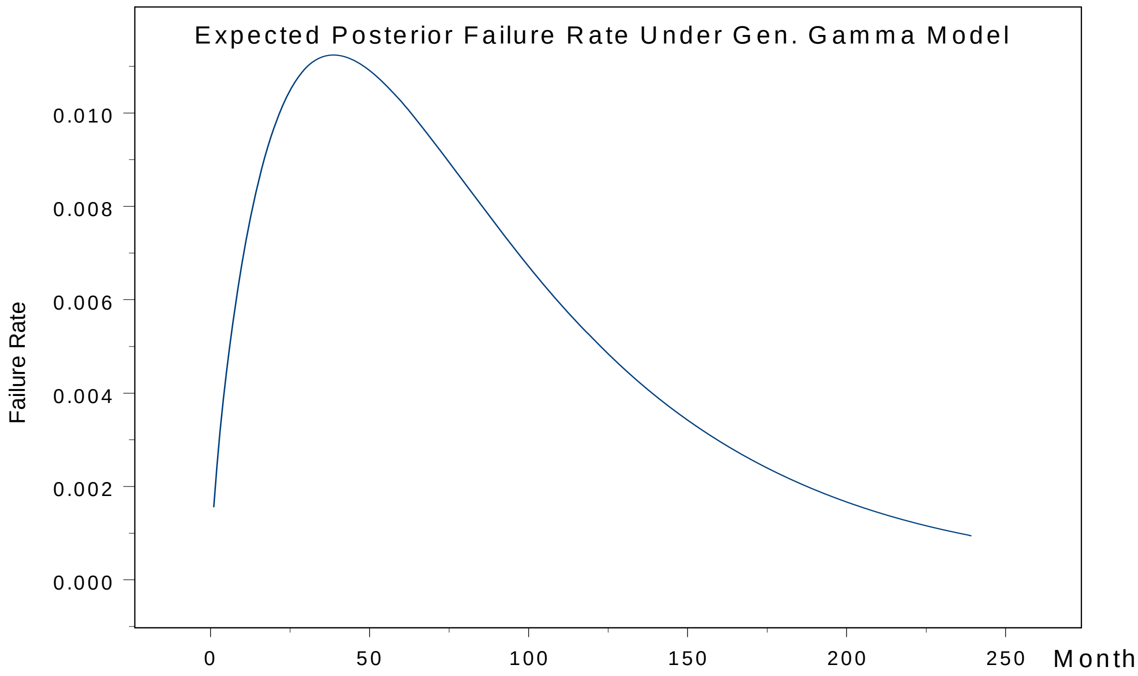
is the baseline default rate.

• Better (posterior) predictive performance than the heterogeneous Weibull model.

- can also consider a heterogeneous version based on the scale λ.

Posterior Failure Rate: Generalized Gamma Model

$$h_0(t|D) = \frac{\int f(t|\alpha,\lambda,\gamma)p(\alpha,\lambda,\gamma|D)d\alpha\, d\gamma\, d\lambda}{\int S(t|\alpha,\lambda,\gamma)p(\alpha,\lambda,\gamma|D)d\alpha\, d\gamma d\lambda}.$$



Posterior Results

Parameter	Mean	StDev	95% CCI
α	1.318	0.444	(0.718, 2.392)
γ	1.422	0.260	(0.953,1.981)
$\beta_{\text{interest-rate}}$	1.449	0.531	(0.526, 2.624)
$\beta_{\text{loan-amount}}$	0.846	0.264	(0.391,1.423)
$\beta_{\text{annual-income}}$	-0.454	0.435	(-1.373, 0.3284)
$\beta_{\text{marital-status}}$	0.146	0.487	(-0.845, 1.077)
$\beta_{\text{borrower-age}}$	-0.820	0.467	(-1.782,0.026)
β_{LTV}	0.031	0.419	(-0.791,0.850)

Posterior Summaries from the Generalized Gamma Model

Posterior Predictive Comparison of Models

- We use 100 randomly selected samples of size 10 from the defaulted loans and compute Predictive Bayes Factors (PBF)

$$PBF = \frac{p(D_F|D_0, M_1)}{p(D_F|D_0, M_2)}$$

where D_F denotes future data and D_0 observed data.

$$p(D_F|D_0, M_i) = \int p(D_F|D_0, \Theta, M_i)p(\Theta|D_0)\, d\Theta$$

Frequency	PBF(WHT,GG)	PBF(WHT, GGHT)	PBF(GG,GGHT)
PBF<1	100	97	8
1<PBF<3	0	2	57
3<PBF<20	0	1	35
20<PBF<150	0	0	0
PBF>120	0	0	0

Bayesian Default Models and Loan Maintenance

In current practice, loss mitigation programs offer options to the borrower after the occurrence of the default.

Given the costs involved, there is much motivation for lenders to take a more active approach in evaluating the loans and offer loan modifications or other options to the borrower before a default actually occurs.

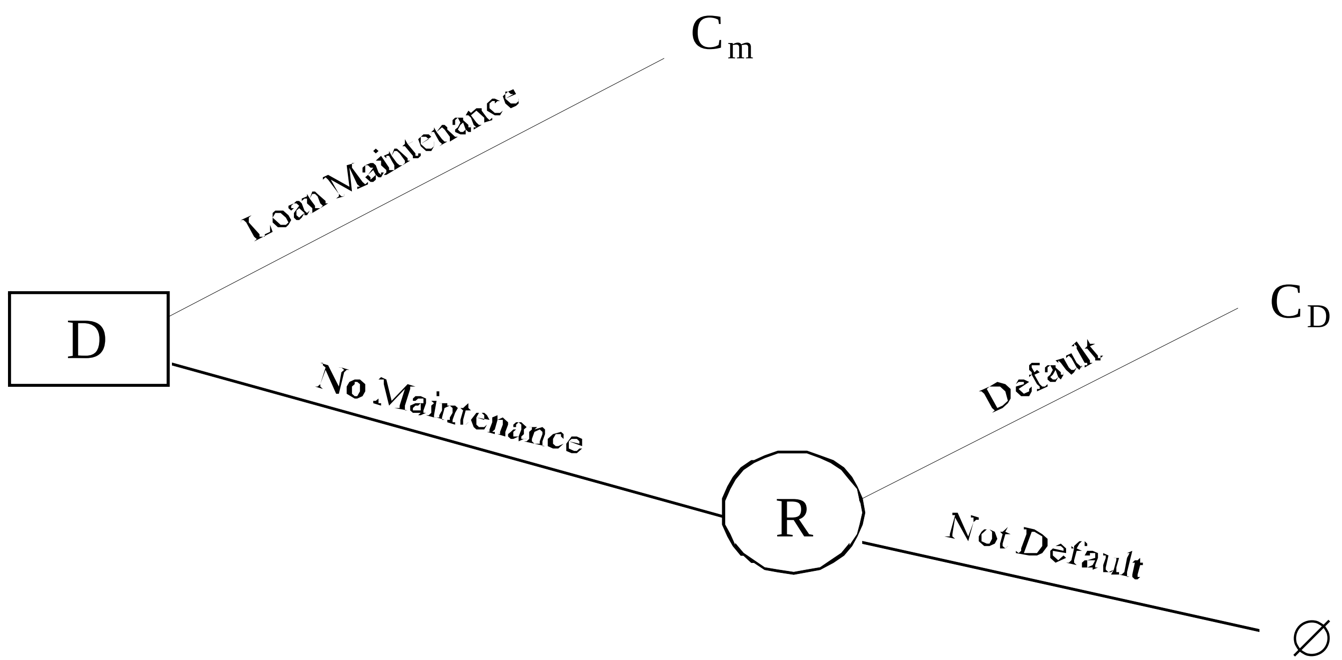
This may be considered as *preventive maintenance* for the loans before the default.

Consider a single loan i which has not defaulted for t units of time and which has a maturity of τ, that is, $t < \tau$.

Assume that at time t the lender considers whether offering a specific loan modification to the borrower

$\Rightarrow$ a decision problem.

Lender's Decision Problem



C_m: cost associated with loan maintenance

C_D: cost associated with default of the loan

A Simple Decision Rule

In the single stage problem the loan will be either defaulted before τ or it will be paid off at τ.

At node R, we need to obtain $p(T_i < \tau | T_i > t)$.

We can obtain the posterior predictive probability $p(T_i < \tau | T_i > t, D)$.

The maintenance option is the optimal strategy if

$$C_m < C_D \, p(T_i < \tau | T_i > t, D)$$

$$\Rightarrow \; (C_m / C_D) < \; p(T_i < \tau | T_i > t, D).$$

Block Loan Maintenance

- Optimal timing of loan maintenance
  - Maintenance on a pool (block) of mortgage loans.

$C_b =$ cost of applying block maintenance on all loans in the mortgage pool
$C_r =$ unit cost of conducting loan modification on a defaulted loan
where $C_b > C_r$

Note: The loan modification cost C_r can be thought as the (minimal) repair cost

We can obtain the cost per unit time as

$$\frac{C_b + C_r\, N(t_B)}{t_B}$$

where t_B is less than maturity time τ.