Time-to-Default Analysis of Mortgage Portfolios

M. Galloway [†], A. Johnson [‡] and A. Shemyakin [†*]

Abstract

Fluctuation in mortgage defaults provides vital information to financial institutions and a key indicator of the state of the economy. Most of these defaults can be attributed to subprime mortgages, which are often issued to borrowers with lower credit ratings. Using mortgage default data for the decade 2001-2010 provided by a major American bank, we develop a statistical model forecasting the probability of defaulting throughout the life of a mortgage. Analyzing time-to-default data over ten-year period allows for the construction of baseline default model less dependent on external covariates. Following the work of Fader and Hardie in customer retention setting, we introduce Weibull-gamma mixture models (WGM) of time-to-default for prime and subprime mortgages. Bayesian parameter estimates are obtained using non-informative priors. Implementation of random walk Metropolis algorithm with normal block updates not only allows for an adequate fit of the data, but also provides diagnostic tools suggesting the use of a simpler Weibull segmentation model (WS), adequately addressing population heterogeneity. Latent classes of risky mortgages characterized by hazard rates increasing with time, are identified in both prime and subprime portfolios.

Keywords: *subprime mortgage, time-to-default, mixture model, latent class, bayesian estimation, random walk Metropolis algorithm.*

1 Introduction

Mortgage defaults played a critical role in the recent financial crisis. Indeed, the majority of these were initiated in the subprime sector of the mortgage market. Accordingly, the practice and foundations of subprime lending have received great attention in the literature. For example, Das and Stein (2009) discuss the impact of underwriting standards whereas Demyanyk and Van Hemert (2011) explore the economical and institutional causes of the subprime mortgage crisis. Crucial to these conversations is an understanding of the inherent *risk* in subprime lending. Bayesian approach to modeling probability of default (PD) is recently becoming popular. A Bayesian reliability model for mortgage default risk in mortgage portfolios was first suggested in Soyer and Xu (2010), and a Bayesian state space model is discussed in Aktekin et al. (2013).

[†]University of St. Thomas, St. Paul, MN, USA
[‡]Macalester College, St. Paul, MN, USA
[*]corresponding author a9shemyakin@stthomas.edu

In the present paper following Soyer and Xu (2010) we suggest addressing default risks by modeling distribution of time-to-default for a mortgage portfolio. Merging default data obtained from ten separate cohorts of mortgages issued by a major American bank each year during the period 2001-2010, we integrate out possible effects of macro-economical covariates and build a model describing *baseline* default risk characterizing the bank's underwriting policies. This is different from cohort-by-cohort approach of Glennon and Nigro (2011) and Aktekin et al. (2013), which emphasizes the short-term role of covariates. We suggest to include external covariates in the further analysis, possibly using the baseline default distribution as a prior.

There are several considerations when conducting a formal *survival analysis* of PD. First, we will consider two different mortgage portfolios separately: prime and subprime. For each of these portfolios, there may exist heterogeneity in default rates among mortgage holders. Further, default rates may vary over time and this variation may be specific for different sectors of the mortgage market. With these properties in mind, we present two strategies for modeling PD. The first, Weibull-Gamma segmentation (WGS,) suggests an infinite mixture model assuming heterogeneity in default risks among individual mortgage holders. We also recognize possible segmentation of both prime and subprime portfolios into two latent classes reflecting the hazard rates either increasing or decreasing with time. The second, a Weibull segmentation model (WS), is a finite mixture model that recognizes segmentation into two latent classes, but assumes a common default risk among individuals of each class. Similar strategies have been used to model heterogeneous survival data in various contexts (see, for example, Marin et al. (2005) and Erisoglu et al. (2011)). Soyer and Xu (2010) considered generalized Gamma as well as the two-component mixture of Weibulls (WS). In fact, both WGS and WS are extensions of the models proposed by Fader and Hardie (2007), Hardie et al. (1998), and Fader et al. (2003) to address population heterogeneity in predicting customer retention in subscription settings. Although mortgage defaults and subscription cancellations differ in nature and consequences, there are some similarities between these two.

Specification of the WGS and WS requires the selection of model parameters reflecting population heterogeneity, denoted θ. In WS model parameters represent shapes and scales of Weibull distributions corresponding to two latent classes and the relative sizes of two classes, while in WGS the scale parameter reflecting population heterogeneity is assumed to have a Gamma distribution with its own shape and scale parameters. The convention in financial and marketing applications is to calculate an estimate of θ from a random sample of data using the *frequentist* maximum likelihood estimation (MLE) approach. However, this ignores potentially powerful prior knowledge and beliefs about θ, knowledge based on past experience and subject-matter expertise. In contrast, we present a *Bayesian* alternative that evaluates θ through a weighted combination of observed data and prior knowledge. Though far from the standard, Bayesian applications in survival analysis settings are increasingly popular (see, for example, Chen et al. (1985), Kiefer (2006, 2008)). A recent paper Popova et al. (2008) applied Bayesian techniques to problems of loan prepayment, somewhat related to default analysis.

We illustrate the Bayesian application of the WGS and WS using empirical data obtained from a major American commercial bank (as reported to FFIEC). These data include the monthly number of defaults in prime and subprime mortgage portfolios, pooled over several cohorts (*vintages*). We show that the simpler WS is superior to the WGS in modeling PD in this setting. Mainly, heterogeneity among default rates within latent classes is insignificant, while segmentation into two latent classes is supported by the results of estimation. Further, in comparison to its MLE competitor, the Bayesian specification of the WS and application of MCMC techniques helps to detect over-parametrization in WGS and enjoys increaed stability in WS.

Our paper is organized as follows. We present our data set and the basics of time-to-default modeling including WGS and WS mixture models of PD in Section 2. We consider specification of WGS model via Bayesian methods and results of MCMC simulation in Section 3.2. Finally, in Section 3.3 we provide a similar Bayesian specification and MCMC simulation for WS model and compare the modeling results in 3.4.

2 Modeling Time-to-Default

2.1 Mortgage Default Data

Our data were provided by a major American bank, as reported to Federal Financial Institution Examination Council (FFIEC). They contained the information on defaults for the aggregated portfolios of residential mortgages on the U.S. market in 2002-2010. Our purpose was to model relatively long-term patterns in PD reflecting the bank's underwriting practices rather than specific economic circumstances. Therefore the data were pooled over several vintages so the time-to-default was measured in months since inception. For vintage-specific analysis see Glennon and Nigro (2011).

Monthly counts of defaults $k_j, j = 1, ..., L$ were observed for $L = 90$ months. The defaults were counted separately for prime and subprime portfolios. Total number of mortgages and loans in each portfolio m was also known.

We analyzed prime and subprime mortgages separately. The emphasis was made on the subprime portfolio, because it was responsible for much higher losses. Figure 1 demonstrates the smoothed histogram of the time-to-default for prime (lower curve) and subprime portfolios (upper curve) scaled to the aggregated portfolio size of 10,000.

Our goal will be to fit a distribution density curve corresponding to one of the models: Weibull-Gamma segmentation (WGS) and Weibull segmentation (WS). One of the important objectives is to be able to fit the tail of the curve, which will allow for prediction of losses related to the long-term behavior of the portfolio.

Let random variable T be the time-to-default of a mortgage portfolio, ie. the number of months from the time of inception to default. Our goal is to characterize the distribution function $F(t|\theta) =$

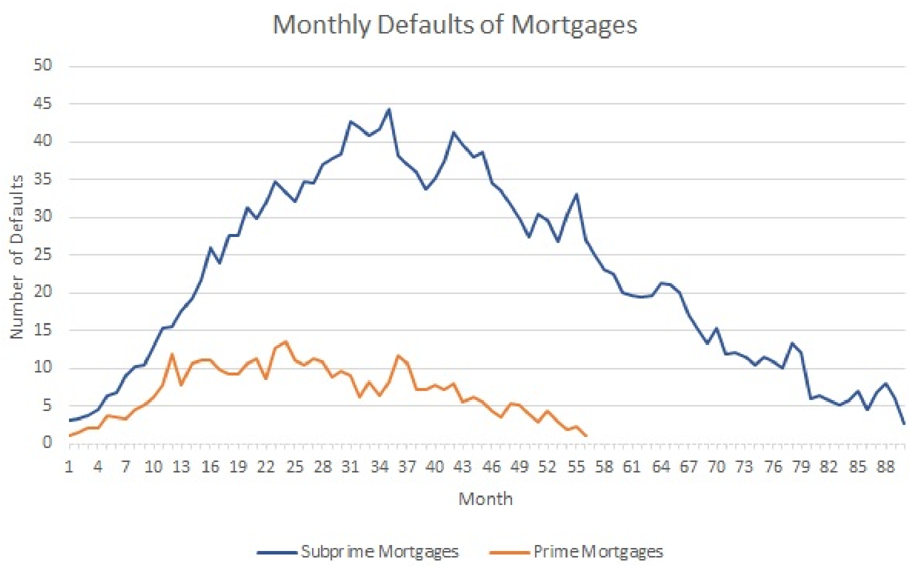

Figure 1: Monthly Defaults for Prime and Subprime Portfolios

$Pr(T \leq t|\theta)$ for $t \geq 0$ and parameters θ, thus capturing the PD within t months of inception. To this end, we provide a short background on similar survival models in Section 2.2 and extend these to the mortgage default setting in Section 2.3.

2.2 Infinite Mixture Models

A simple model for T can be derived from the discrete-time Beta-Geometric (BG) model for customer retention proposed by Fader and Hardie (2007). Their model is based on the following assumptions, stated here in the context of mortgage defaults.

- Defaults are observed at discrete periods, i.e. $T \in \{1, 2, 3, \ldots\}$.
- The probability of default for an individual mortgage holder, θ, remains constant throughout their holding. Thus T is described by the geometric distribution with distribution function

$$F_{BG}(t|\theta) = Pr(T \leq t|\theta) = 1 - (1-\theta)^t \quad \text{for } t \in \{1, 2, ...\} \; .$$

- The probability of default, θ, varies among mortgage holders and can be characterized by the Beta(a, b) distribution with parameters $a, b > 0$ and probability density function

$$f(\theta|a, b) \propto \theta^{a-1}(1-\theta)^{b-1} \quad \text{for } \theta \in [0, 1] \; .$$

Therefore, the model obtained is an infinite mixture of geometric distributions, where parameter θ of the geometric distribution characterizing retention pattern of an individual is random and distributed over the population according to Beta law. This setting addresses the population heterogeneity. The discrete-time Beta-Geometric model assumptions oversimplify the reality of our setting. To begin, T is typically measured on a continuous-time scale, i.e. defaults can occur at any time. A similar Exponential-Gamma model allows for defaults in continuous time. However, with our data being discretized to a month, this does not constitute a big difference.

More importantly, depending on the circumstances of the mortgage portfolio and its holder, the hazard rate may increase or decrease throughout the duration of the holding. To accommodate these features, Fader and Hardie (2007) propose, but do not explore, a continuous-time Weibull-Gamma (WG) model that operates under the following assumptions.

(A1) Defaults can occur at any time, i.e. $T \geq 0$.

(A2) The risk of default for an individual may increase or decrease throughout their holding. Thus T can be characterized by the Weibull distribution with probability density function $f_W(t|\lambda, c)$ and corresponding distribution function

$$F_W(t|\lambda, c) = Pr(T \leq t|\lambda, c) = 1 - e^{-\lambda t^c} \quad \text{for } t \geq 0 \,. \tag{1}$$

Note that the parameter $c > 0$ represents the magnitude and direction of change of the *hazard rate* (risk of default)over time with $c = 1$ indicating a constant hazard rate and $c > 1$ ($c < 1$) indicating an increase (decrease) in the hazard rate. Further, $\lambda > 0$ represents the *scale* of Weibull distribution with larger λ reflecting a larger risk of default.

(A3) The risk of default varies among mortgage holders. To this end, hetereogeneity in scale parameter λ is modeled by a Gamma(α, r) distribution with parameters $\alpha, r > 0$ and probability density function

$$f_G(\lambda|\alpha, r) \propto \lambda^{r-1} e^{-\alpha\lambda} \quad \text{for } \lambda > 0 \,. \tag{2}$$

It follows that the joint Weibull-Gamma distribution of (t, λ) is characterized by density function

$$f_{WG}(t, \lambda|\alpha, r, c) = f_W(t|\lambda, c) f_G(\lambda|\alpha, r) \,.$$

Finally, integrating out λ produces a form of the Burr Type XII distribution:

$$F_{WG}(t|\alpha, r, c) = \int_0^t \int_0^\infty f_{WG}(z, \lambda|\alpha, r, c) d\lambda dz = 1 - \left(\frac{\alpha}{\alpha + t^c}\right)^r \,. \tag{3}$$

In this setting, parameter λ linearly related to the hazard rate of Weibull distribution (1), corresponds to the individual risk factor of a mortgage holder. It is integrated out of the model equation (3) and is replaced by hyperparameters α and r characterizing the entire population. Shape parameter c of Weibull distribution (1) is responsible for hazard rate increasing ($c > 1$) or decreasing ($c < 1$) with time.

2.3 Segmentation

The Weibull-Gamma framework under assumptions (A1)–(A3) provides a foundation for modeling time-to-default with the scale parameter λ responsible for population heterogeneity. However, it does not reflect the different development of risk pattern (hazard rate) with time. The most evident solution would be to assume the split of the mortgage population into two or more separate segments with different shape parameters of the underlying Weibull distributions. From our data it is impossible to determine to which class any particular defaulting mortgage belongs, thus these segments form latent classes, whose sizes have to be determined in process of parametric estimation. We will begin with two classes, keeping in mind that their number can be easily increased if the model so requires.

To this end, let $p \in (0,1)$ be the segmentation weight for the first class of mortgages with shape parameter c_1, and $1-p$ be the weight for the second class of mortgages with shape c_2. Hetereogeneity in λ can be modeled by a Gamma(α, r) distribution with density function (2). In conjunction with (3), it follows that a Weibull-Gamma segmentation mixture model for T has distribution function

$$F_{WGS}(t|\alpha, r, c_1, c_2, p) = 1 - p\left(\frac{\alpha}{\alpha + t^{c_1}}\right)^r - (1-p)\left(\frac{\alpha}{\alpha + t^{c_2}}\right)^r \tag{4}$$

with corresponding density function $f_{WGS}(t|\alpha, r, c_1, c_2, p)$.

If no heterogeneity in λ is detected, then time-to-default T can be characterized by a simpler Weibull segmentation model (WS) with probability density function $f_{WS}(t|\lambda, c_1, c_2, p)$ and corresponding distribution function

$$F_{WS}(t|\lambda, c_1, c_2, p) = 1 - pe^{-\lambda t^{c_1}} - (1-p)e^{-\lambda t^{c_2}} \quad \text{for } t \geq 0 \tag{5}$$

where c_1 and c_2 represent the magnitude and direction of change of hazard rates with time for two segments, respectively.

If default rates are assumed to be constant across mortgage holders, then (5) is a suitable time-to-default model.

3 Parametric Estimation in WGS and WS Models

Specification of the Weibull segmentation model (5) and Weibull-Gamma mixture model (4) requires the selection of parameters (λ, c_1, c_2, p) and (α, r, c_1, c_2, p), respectively. We consider specification of likelihood function first (Section 3.1) followed by Bayesian specification of the priors for WGS (Section 3.2) and WS (Section 3.3).

3.1 Specification of Likelihood Functions

In financial and marketing applications, parameter estimation via maximum likelihood methods is the convention. Let $t_1, t_2, ..., t_n$ be the observed times-to-default for a random sample of n mortgage portfolios. Further, define likelihood functions

$$\mathcal{L}_{WS}(\lambda, c_1, c_2, p \mid t_1, t_2, \ldots, t_n) = \prod_{i=1}^{n} f_{WS}(t_i|\lambda, c_1, c_2, p)$$
$$\mathcal{L}_{WGM}(\alpha, r, c_1, c_2, p \mid t_1, t_2, \ldots, t_n) = \prod_{i=1}^{n} f_{WGM}(t_i|\alpha, r, c_1, c_2, p)\ .$$

Taking into account discretization (defaults recorded at discrete times - monthly) and right censoring of the data (finite period of observation), we will rewrite the likelihood functions. Let m be the total number of mortgages in the portfolio, L the observation period, and k_j the number of defaults recorded at time j for $j = 1, \ldots, L$. Then

$$\mathcal{L}_{WS}(\lambda, c_1, c_2, p \mid m, L, k_1, k_2, \ldots, k_L) = \prod_{j=1}^{L} f_{WS}^{k_j}(j|\lambda, c_1, c_2, p)(1 - F_{WS}(L|\lambda, c_1, c_2, p))^{m - \sum_{j=1}^{L} k_j}$$
$$\mathcal{L}_{WGM}(\alpha, r, c_1, c_2, p \mid m, L, k_1, k_2, \ldots, k_L) = \prod_{j=1}^{L} f_{WGM}^{k_j}(j|\alpha, r, c_1, c_2, p)(1 - F_{WGM}(L|\alpha, r, c_1, c_2, p))^{m - \sum_{j=1}^{L} k_j}\ . \tag{6}$$

Maximum likelihood estimates for the parameters of the Weibull segmentation and Weibull-Gamma mixture models are, respectively

$$(\hat{\lambda}_{\text{MLE}}, \hat{c_1}_{\text{MLE}}, \hat{c_2}_{\text{MLE}}, \hat{p}_{\text{MLE}}) = \underset{\lambda, c_1, c_2, p}{\operatorname{argmax}}\ \mathcal{L}_{WS}(\lambda, c_1, c_2, p \parallel m, L, k_1, k_2, \ldots, k_L); \quad \text{and}$$
$$(\hat{\alpha}_{\text{MLE}}, \hat{r}_{\text{MLE}}, \hat{c_1}_{\text{MLE}}, \hat{c_2}_{\text{MLE}}, \hat{p}_{\text{MLE}}) = \underset{\alpha, r, c_1, c_2, p}{\operatorname{argmax}}\ \mathcal{L}_{WGM}(\alpha, r, c_1, c_2, p \parallel m, L, k_1, k_2, \ldots, k_L)\ .$$

3.2 A Bayesian WGS Model

Bayesian formulations of the Weibull-Gamma segmentation model (4) and Weibull segmentation model (5) require the specification of prior distributions for the relevant parameters. These are chosen to reflect prior knowledge of the parameters based on prior experience and subject-matter expertise. Let $f(\lambda)$ $f(\alpha)$, $f(r)$, $f(c_1)$, $f(c_2)$, and $f(p)$ represent the corresponding prior probability density functions. Further, assume the prior distributions are independent. Further, let $(m, L, k_1, k_2, \ldots, k_L)$ be the observed mortgage default data. Then the posterior distributions for the parameters of the Weibull segmentation and Weibull-Gamma mixture models conditioned on the data are characterized by density functions

$$\begin{aligned} f_{\text{WSpost}}(\lambda, c_1, c_2, p|m, L, k_1, k_2, \ldots, k_L) &\propto \mathcal{L}_{WS}(\lambda, c_1, c_2, p \mid m, L, k_1, k_2, \ldots, k_L) \\ &\quad \cdot f(\lambda)f(c_1)f(c_2)f(p) \\ f_{\text{WGMpost}}(\alpha, r, c_1, c_2, p|m, L, k_1, k_2, \ldots, k_L) &\propto \mathcal{L}_{WGM}(\alpha, r, c_1, c_2, p \mid m, L, k_1, k_2, \ldots, k_L) \\ &\quad \cdot f(\alpha)f(r)f(c_1)f(c_2)f(p) \end{aligned} \tag{7}$$

for $\mathcal{L}_{WS}$ and $\mathcal{L}_{WGM}$ defined by (6).

In the Bayesian setting, posterior expectations provide simple point estimates of the relevant model parameters. For example, the posterior expected value of p conditioned on observed time-to-default data can be calculated by

$$E[p|m, L, k_1, k_2, \ldots, k_L] = \int_0^1 p f_{\text{post}}(p|m, L, k_1, k_2, \ldots, k_L) dp$$

where $f_{\text{post}}(p|m, L, k_1, k_2, \ldots, k_L)$ is the marginal posterior density of p calculated from f_{WSpost} or f_{WGMpost} as appropriate.

Next, consider a Bayesian approach to modeling time-to-default via WGS model. To begin, we place independent, non-informative priors on all relevant parameters $\lambda, \alpha, r, c_1, c_2, p$ specified by prior density functions

$$\begin{aligned} f(\alpha) &\propto \frac{1}{\alpha}, \quad \alpha > 0 \\ f(r) &\propto \sqrt{\sum_{i=0}^{\infty} (r+i)^{-2}}, \quad r > 0 \\ f(c_i) &\propto \frac{1}{c_i}, \quad c_i > 0 \text{ for } i = 1, 2 \\ f(p) &\propto p^{-1/2}(1-p)^{-1/2}, \quad p \in (0, 1) \ . \end{aligned} \tag{8}$$

Note that $f(r)$ and $f(p)$ correspond to $\sqrt{PG(1, r)}$ and Beta$(1/2, 1/2)$ priors on r and p, respectively. Though the priors (8) are themselves simple in structure, the corresponding posterior density functions of the Weibull segmentation and Weibull-Gamma mixture models (7) are analytically intractable. Thus finding closed form solutions for the posterior expected values of the model parameters is prohibitively difficult, if not impossible. In this case, inference requires Markov chain Monte Carlo (MCMC) techniques.

Consider the MCMC analysis of the Weibull segmentation model. The analysis of the Weibull-Gamma mixture model is analogous. To this end, define Markov chain $\Phi = \left\{ \left(\lambda^{(i)}, c_1^{(i)}, c_2^{(i)}, p^{(i)} \right) \right\}_{i=1}^{N}$ for $f_{WSpost}(\lambda, c_1, c_2, p|m, L, k_1, k_2, \ldots, k_L)$. Though there are competing MCMC strategies to consider, we construct Φ using a naive Metropolis algorithm with multivariate Normal block updates.

Specifically, Φ evolves from step i to step $i+1$ as follows:

- Draw independent proposal values of each parameter:

$$\begin{aligned} \lambda' &\sim N(\lambda^{(i)}, \sigma_\alpha^2) \\ c_i' &\sim N(c_i^{(i)}, \sigma_{c_i}^2) \quad \text{for } i = 1, 2 \\ p' &\sim N(p^{(i)}, \sigma_p^2) \end{aligned}$$

  where $(\sigma_\lambda^2, \sigma_{c_i}^2, \sigma_p^2)$ is a fixed set of tuning parameters.

- Calculate acceptance probability

$$\rho = \min\left\{1, \quad \frac{f_{\text{WSpost}}(\lambda', c_1', c_2', p' \mid m, L, k_1, k_2, \ldots, k_L)}{f_{\text{WSpost}}\left(\lambda^{(i)}, c_1^{(i)}, c_2^{(i)}, p^{(i)} \,\middle|\, m, L, k_1, k_2, \ldots, k_L\right)}\right\}.$$

- Set

$$\left(\lambda^{(i+1)}, c_1^{(i+1)}, c_2^{(i+1)}, p^{(i+1)}\right) = \begin{cases} (\lambda', c_1', c_2', p') & \text{with probability } \rho \\ \left(\lambda^{(i)}, c_1^{(i)}, c_2^{(i)}, p^{(i)}\right) & \text{with probability } 1-\rho \end{cases}.$$

Using this algorithm, we produce a Markov chain sample $\Phi = \left\{\left(\lambda^{(i)}, c_1^{(i)}, c_2^{(i)}, p^{(i)}\right)\right\}_{i=1}^N$ of length $N = 3,000,000$. The corresponding acceptance rates for the WGM and WS models were approximately 4.1% and 2.5% which could be related to the block update structure. Though the algorithm could be further tuned in order to increase acceptance rates, we will show below that the accuracy of the estimates of the WG and WS is, indeed, impressive under the current version.

The corresponding Monte Carlo sample averages provide estimates of the posterior expectations of parameters (λ, c_1, c_2, p). For example, we estimate

$$E[p|m, L, k_1, k_2, \ldots, k_L] = \int_0^1 p f_{\text{WSpost}}(p|m, L, k_1, k_2, \ldots, k_L)dp$$

by

$$\hat{p}_{MC} = \frac{1}{N}\sum_{i=1}^N p^{(i)}.$$

The WGS parameter estimates are summarized in Table 1 . Figure 2 demonstrates that the overall fit of WGS model is satisfactory, reasonably capturing the tail behavior.

Figure 3 illustrates the trace plots (consecutive accepted values of MCMC algorithm) for each parameter. Notice that the trace plots for parameters α and r are very unstable and converge slowly. These two parameters represent the shape and scale of the Gamma distribution describing the variation of the individual mortgages' risk of default λ over the

Table 1: Parameter estimates for the Bayesian WGM, subprime portfolios.

Parameter	Estimate	Error	Tuning	Value
α	1.022×10^{15}	1.241×10^{13}	σ_λ	5×10^{12}
c_1	2.179	1.227×10^{-3}	σ_{c_1}	0.05
c_2	9.386×10^{-2}	3.231×10^{-3}	σ_{c_2}	0.2
r	2.422×10^{11}	2.824×10^{9}	σ_r	10^9
p	0.199	1.293×10^{-4}	σ_p	0.007

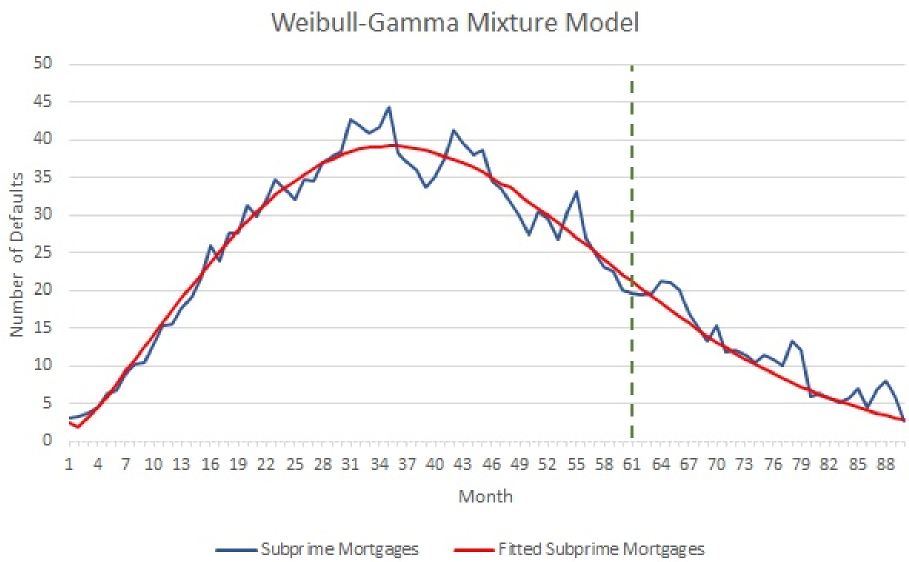

Figure 2: WGM Model for Subprime Portfolio. Bayesian Estimation.

population. This behavior also provides insight into the instability we observed in running MLE estimation for WGS.

Nevertheless, if we closely observe the development of the chains in Figure 3, we see that changes in values of α and r are clearly dependent (which is allowed by the scheme of Normal block updates). Moreover, the values of both parameters tend to climb, while their ratio r/α remains close to constant. This corresponds to $r/\alpha = E(\lambda) \simeq const$, $r/\alpha^2 = Var(\lambda) \simeq 0$, which may be taken as an indication of over-parametrization and suggests the use of a simpler WS model with the risk factor λ staying constant for entire population. That provides a good argument for the use of a simpler WS model.

3.3 A Bayesian WS Model

The Monte Carlo estimates of the WS parameters are summarized in Table 2, with the corresponding model estimate illustrated in Figure 5.

The WS model provides a good overall fit for the subprime portfolio, and the corresponding

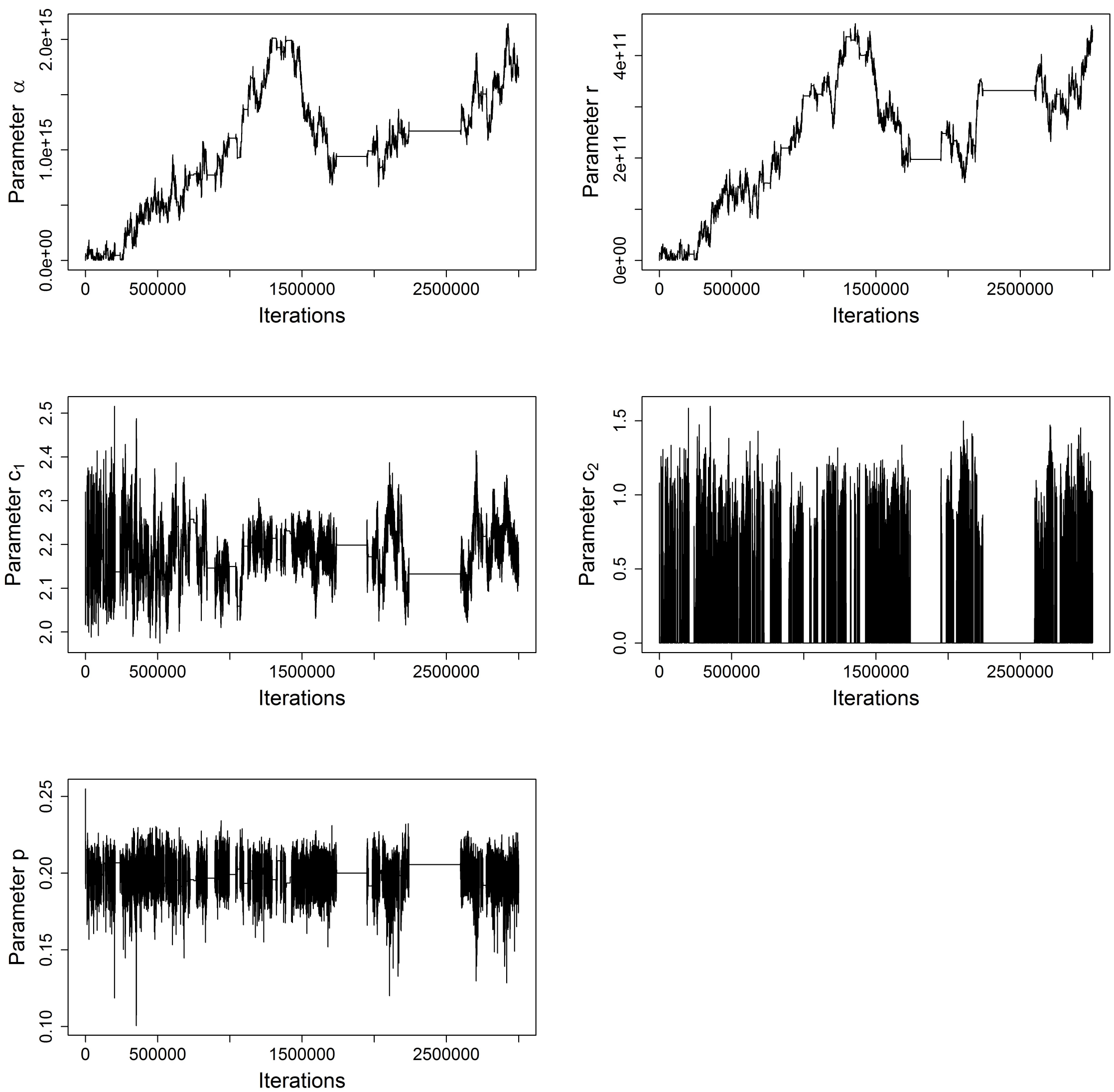

Figure 3: Trace plots for the Metropolis chain for the WGM model.

Table 2: Parameter estimates for the Bayesian WS model, subprime portfolios.

Parameter	Estimate	Error	Tuning	Value
λ	2.381×10^{-4}	9.154×10^{-7}	σ_λ	3×10^{-5}
c_1	2.187	1.286×10^{-3}	σ_{c_1}	0.02
c_2	0.115	5.286×10^{-3}	σ_{c_2}	0.07
p	0.197	2.076×10^{-4}	σ_p	0.004

Markov chain enjoys stable behavior (Figure 6).

Finally, we apply the WS model to the analysis of the prime portfolio. Parameter estimates and the overall fit are illustrated in Table 3 and Figure 7, respectively.

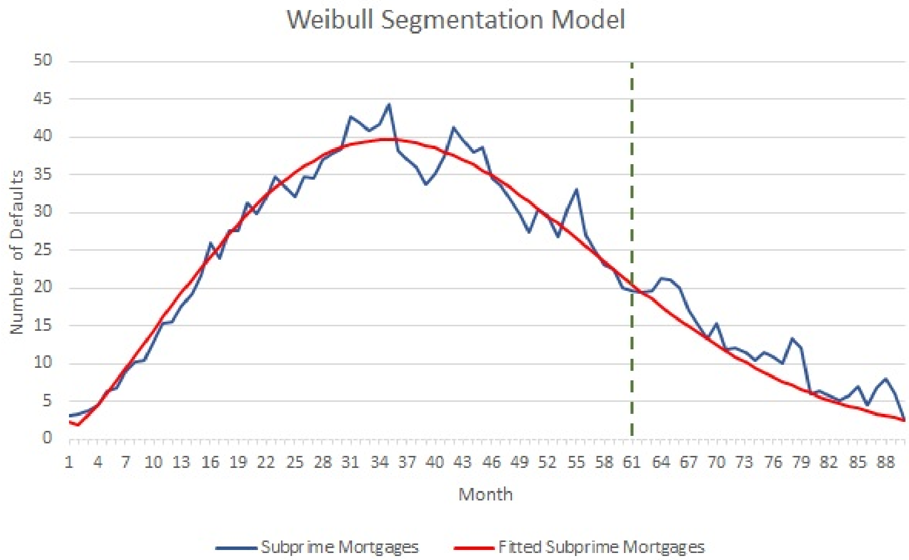

Figure 4: WS model for Subprime Portfolio. Results of Bayesian Estimation.

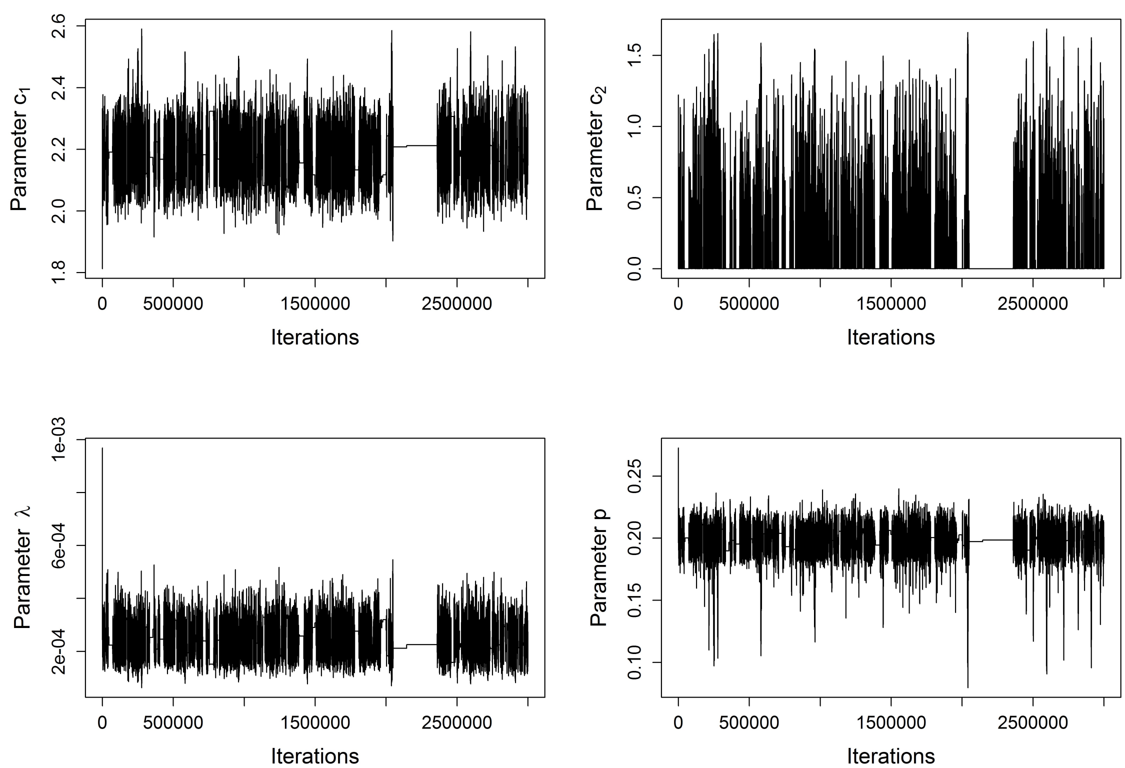

Figure 5: Trace plots of the Metropolis chain for the WS model.

3.4 Conclusions

The use of Bayesian methods involving MCMC allows us to conclude that for our data, the Weibull segmentation model (WS) is sufficient to model defaults for both prime and

Table 3: Parameter estimates for the Bayesian WSM model, prime portfolios.

Parameter	Estimate	Error	Tuning	Value
λ	5.054×10^{-4}	3.554×10^{-6}	σ_λ	3×10^{-5}
c_1	2.206	2.513×10^{-3}	σ_{c_1}	0.02
c_2	0.152	5.350×10^{-3}	σ_{c_2}	0.07
p	4.032×10^{-2}	7.977×10^{-5}	σ_p	0.004

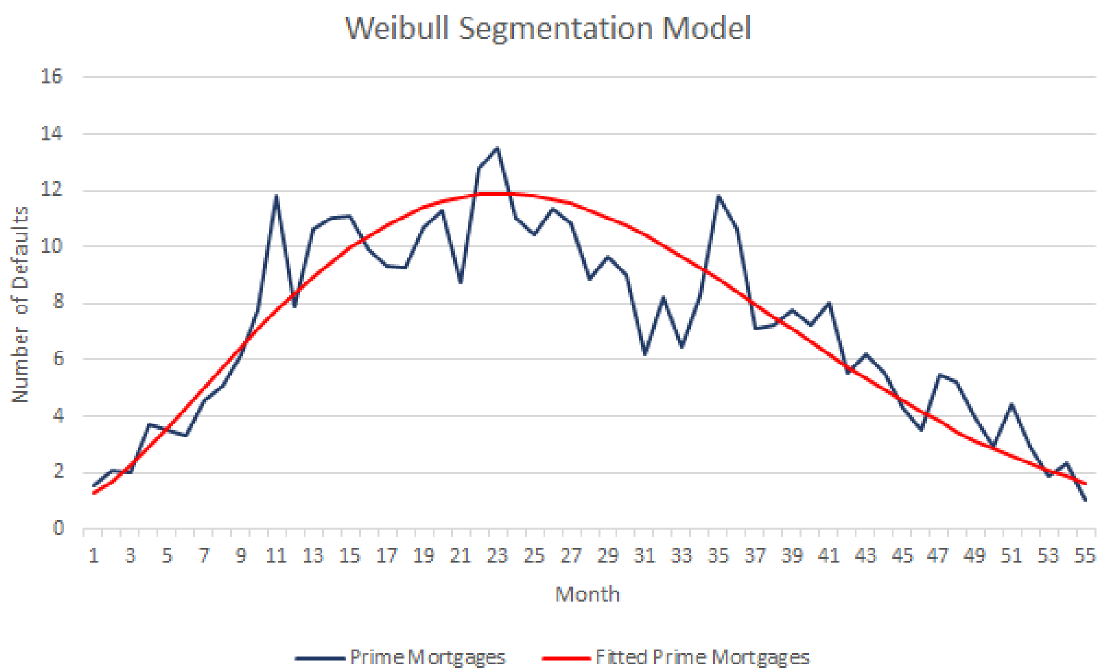

Figure 6: Bayesian WS model for prime portfolios.

subprime portfolios. Hence there is no need to address additional population heterogeneity as suggested by Weibull-Gamma mixture (WGS). Over-parametrization in WGS, which was one of most unexpected results of our study, was suggested by MCMC diagnostics.

Comparing Tables 2 and 3, we can suggest that parameters c_1 and c_2 responsible for temporal changes in hazard rates are quite similar in both cases. The first, larger segment of the portfolios is characterized by a decreasing hazard rate which corresponds to growing mortgagees' equity and increasing financial losses in case of default. As the model suggests, this segment is practically default-free, though it is hard to qualify this statement due to the latent nature of the segmentation. However, the second, smaller segment is characterized by an increasing hazard rate. This might be explained, for instance, by an additional stress on the households signed up in a mortgage contract they cannot afford. For these households the very mortgage payment is a big factor leading to the deterioration of their financial situation.

What constitutes the main difference between prime and subprime portfolios is the relative size of this second segment. As we see from Table 2, for subprime mortgages it is about twenty percent. Two segments of similar proportional sizes (80-20) were detected by Soyer and Xu (2010) in EPD (early default) data. However Table 3 shows that the risky segment constitutes less than five percent of prime mortgages. Another difference between the second segments of two portfolios is in the parameter λ related to the mean lifetime. At this point, we do not have a clear interpretation for this difference.

4 Acknowledgements

The authors wish to thank Ellen Klingner for her research assistance and also NSF CSUMS grant DMS0802959 which made this work possible.

References

Aktekin, T., Soyer, R., and Xu, F. (2013). Assessment of mortgage default risk via bayesian state space models. *Annals of Applied Statistics*, 7,3:1450–1473.

Chen, W. C., Hill, B. M., Greenhouse, J. B., and Fayes, J. V. (1985). Bayesian analysis of survival curves for cancer patients following treatment. *Bayesian Statistics*, 2:299–328.

Das, A. and Stein, R. M. (2009). Underwriting versus economy: a new approach to decomposing mortgage losses. *Journal of Credit Risk*, 15(2):19–41.

Demyanyk, Y. and Van Hemert, O. (2011). Understanding the subprime mortgage crisis. *Review of Financial Studies*, 24(6):1848–1880.

Erisoglu, U., Erisoglu, M., and Erol, H. (2011). A mixture model of two different distributions approach to the analysis of heterogeneous survival data. *International Journal of Computational and Mathematical Sciences*, 5(2):75–79.

Fader, P. S. and Hardie, B. G. S. (2007). How to project customer retention. *Journal of Interactive Marketing*, 21:76–90.

Fader, P. S., Hardie, B. G. S., and Zeithammer, R. (2003). Forecasting new product trial in a controlled test market environment. *Journal of Forecasting*, 22(5):391–410.

Glennon, D. and Nigro, P. (2011). Evaluating the performance of static versus dynamic models of credit default: evidence from long-term Small Business Administration-guaranteed loans. *Journal of Credit Risk*, 7(2):3–35.

Hardie, B. G. S., Fader, P. S., and Wisniewski, M. (1998). An empirical comparison of new product trial forecasting models. *Journal of Forecasting*, 17:209–229.

Kiefer, N. M. (2006). The probability approach to default probabilities. Technical report, Cornell University.

Kiefer, N. M. (2008). Default estimation, correlated defaults, and expert information. Technical report. CAE Working paper #08-02, ISSN 1936-5098.

Marin, J. M., Rodriguez-Bernal, M. T., and Wiper, M. P. (2005). Using Weibull mixture distributions to model heterogeneous survival data. *Communications in Statistics : Simulation and Computation*, 34:673–684.

Popova, I., Popova, E., and George, E. I. (2008). Assessment of mortgage default risk via bayesian reliability models. *Bayesian Analysis*, 3:393–426.

Soyer, R. and Xu, F. (2010). Assessment of mortgage default risk via bayesian reliability models. *Applied Stochastic Models Bus. Ind.*, 26:308–330.