RT&A, No 4 (89)
Volume 20, December 2025

EXPLORING RECIDIVISM THROUGH BAYESIAN SURVIVAL MODELS: A PARAMETRIC APPROACH WITH RSTAN

Fahad Ashraf[1], Mohammad Parvej[2], Athar Ali Khan[3]

•

[1,2,3]Department of statistics and Operations Research
Aligarh Muslim University, Aligarh, India
[1]Corresponding Author e-mail: gj5872@myamu.ac.in
[2]parvezstats97@gmail.com, [3]alkhan.st@amu.ac.in

Abstract

This study adopts a fully Bayesian parametric survival modeling approach to examine recidivism, utilizing the recid dataset from the wooldridge package in R. Leveraging the computational capabilities of RStan, four alternative survival distributions Exponential, Weibull, Lognormal, and Frechet are estimated, incorporating covariates reflecting demographics, criminal history, and post-release conditions. To promote model parsimony and interpretability, covariate selection is guided by variable importance measures derived from a machine learning-enhanced Cox model. Model performance is systematically evaluated using the Leave-One-Out Information Criterion (LOOIC) and the Widely Applicable Information Criterion (WAIC) to ensure robust predictive accuracy. The analysis includes detailed posterior summaries, conditional effect plots to assess covariate influences on time to reoffense, and posterior predictive distribution (PPD) plots for model validation. The findings provide meaningful insights into the timing and risk of reoffending and underscore the value of Bayesian survival models in criminological research.

Keywords: Recidivism, Cox-ML model, Bayesian modeling, Rstan, Survival analysis

1. Introduction

The term recidivism is defined in some specific way, such as rearrest, reconviction, or reimprisonment, of a released prisoner [1]. Recidivism is prevalent in the realm of criminality, is a persistent concern for both criminologists and economists. Research suggests that a small proportion of habitual offenders are responsible for a disproportionately high share of serious crimes [2]. According to recidivism data of 33 countries, the 2-year reconviction rates were found between 18% and 55% [3]. From an economic perspective, criminal behavior and recidivism can be studied within the framework of rational choice theory [4], where individuals weigh the expected costs and benefits of legal and illegal activities. Factors such as employment prospects, sentence length, and prior criminal history shape this calculus, and empirical studies have shown that poor post-release labor market conditions and limited human capital often drive individuals back into criminal activity [5],[6]. Accordingly, modeling time until reoffense provides critical insights into both the effectiveness of penal policies and the socio-economic mechanisms underlying recidivism. Thus it becomes of paramount importance to understand the dynamics of criminal recidivism and factors that influence criminal recidivism in forming policy decisions, designing rehabilitation programs, and allocating criminal justice resources effectively.

Building upon this motivation, statistical analysis and modeling of recidivism has become a crucial area of research, particularly in understanding the time until reoffense after release. [7] is one of the oldest work on recidivism prediction that employed scoring system to predict parole

DOI: https://doi.org/10.24412/1932-2321-2025-489-545-561

outcomes. Currently research on criminal recidivism has spanned a broad range of methods, from descriptive statistics and logistic regression to modern machine learning and Bayesian modeling [8],[6],[9],[10]. Early work focused on describing recidivism patterns, while more recent studies incorporate individual- and contextual-level covariates to predict or understand reoffending behavior. Since recidivism is inherently a time to event phenomenon and censoring is key element as not all offenders reoffends, [11] criticized prevailing approaches for ignoring the timing of reoffenses and emphasized on the need for models that account for varying follow-up periods and censoring core motivations for adopting survival analysis. The earlier works like [12],[13] used exponential model to analyse the recidivism but failed to formally incorporate the censoring information in the model. Later parametric, non parametric, regression, cox survival and split population models were used in recidivism prediction [5],[14] pioneered the use of split-population survival models, allowing for a fraction of offenders who never recidivate. [15] used survival analysis to model time until recidivism among youthful offenders, integrating both static and dynamic covariates. Over time, survival based approaches have emerged as the dominant framework for analyzing recidivism, as they naturally accommodate censoring while capturing the dynamic nature of reoffending. This progression highlights the methodological shift from simple descriptive accounts toward more rigorous and flexible statistical modeling of criminal behavior.

Traditional survival analysis methods, such as Cox models, accelerated failure time model have been widely used [16],[17] however, these approaches often rely on restrictive assumptions and may not fully capture individual-level heterogeneity or uncertainty in parameter estimation. Bayesian approach is a flexible and robust framework for modeling time-to-event data, allowing researchers to incorporate prior knowledge, quantify uncertainty more comprehensively, and explore complex model structures. [18] showcased the use of Bayesian method in criminological data, demonstrating advantages in handling parameter uncertainty and model comparison. Recently there has been a surge in the adoption of Bayesian approaches in all field, In survival analysis, Bayesian methods enable more robust inference on censored time-to-event data, [19],[20] demonstrated the effectiveness of Bayesian method in modeling survival data with generalized distributions through **Rstan**. Despite the methodological advantages of Bayesian approaches, their application to criminal recidivism remains limited. This study addresses that gap by applying fully Bayesian parametric survival models to recidivism data from North Carolina, available via the **wooldridge** package in R. Leveraging machine learning-based variable selection from Cox models, we analyze covariate effects on recidivism timing under four competing survival distributions: Exponential, Weibull, Lognormal, and Frechet. By doing so, we not only identify the most influential socio-demographic and behavioral factors affecting reoffense risk but also evaluate which statistical models best capture the time dynamics of recidivism.

The Cox model combined with techniques from machine learning, especially with the help of Lasso and Elastic Net regularization, makes choosing variables easier because it handles both the issues of many variables and correlated predictors. By assigning penalties to the regression coefficients, these models move minor variable effects besides zero, leading to automatic and all-at-once identification of top features that also prevent the chances of overfitting. With this approach, examples of significant risk factors become easy to spot and the models provide stronger and more accurate outcomes for predicting survival, especially in settings where conventional Cox models might fail [21].

Through Bayesian estimation via **Rstan**, we obtain full posterior distributions of model parameters, enabling nuanced interpretation of covariate effects and credible intervals for uncertainty quantification. Model fit is assessed using the Leave-One-Out Information Criterion (LOOIC) and the Widely Applicable Information Criterion (WAIC), both of which favor the Lognormal model for this dataset. Ultimately, this study highlights the advantages of Bayesian survival models in criminology and the importance of individualized risk assessment for policy formulation and intervention design, thereby bridging the gap between statistical modeling and real-world decision-making.

2. Data Description and Variable Selection

The recidivism data used in this article is available with **wooldrige** package of **R** [22]. The data set contains records of recidivism time along with a number of variables on 1445 inmates released from North Carolina prison. It includes demographic, socioeconomic, and criminal history characteristics such as age, education, prior convictions, substance use, and participation in rehabilitation programs. These variables make the dataset suitable for analyzing the factors influencing the likelihood and timing of reoffending. For more detail about the data one may refer to [23],[2].

Features of the data

- black: = 1 if black
- alcohol: = 1 if alcohol problem
- drugs: = 1 if drug history
- super: = 1 if release supervised
- married: = 1 if married when incarc.
- felon: = 1 if felony sentence
- workprg: = 1 if in N.C. pris. work prg.
- property: = 1 if property crime
- person: = 1 if crime against person
- priors: prior convictions
- educ: years of schooling
- rules: rules violations in prison
- age: in months
- tserved: time served, rounded to months
- follow: length follow period, months
- durat: min(time until return, follow)
- cens: = 1 if duration right censored
- ldurat: log(durat)

Table 1: *Means of Selected Variables from the recid Dataset*

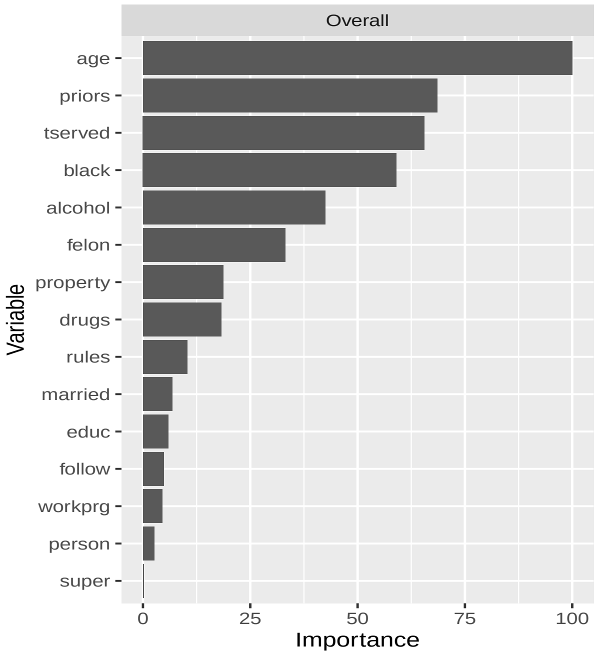
black	alcohol	drugs	super	married	felon	workprg	property	person
0.4851	0.2097	0.2415	0.6941	0.2554	0.3142	0.4651	0.2547	0.05329
priors	educ	rules	age	tserved	follow	durat	cens	ldurat
1.432	9.702	1.185	345.4	19.18	74.89	55.37	0.618	3.745

In Table 1, the mean value of variables with binary response depicts the proportion e.g. the mean of black is 0.4851, it means 48.51% of released individuals are black, for continuous variable mean carry usual interpretation.

There are 15 covariates in the data which may have varying degree of effects on recidivism time of released prisoners, some having large and significant effects while others having small and insignificant effect. To select those with large and significant effects we used variable importance index for selecting the variables.The method is applied through **varimp()** function of **MachineShop** package of **R** software [24]. The variable importance index suggests the importance of a variable on a scale of 0 to 100, with 100 signifying most important variable and 0 the least. For our study we used a rule of thumb to consider any variable with variable importance index greater than 15 as important enough to be selected in the model. This procedure ensures a more parsimonious model by focusing only on influential predictors.

Hence we selected the variable age (of the released prisoner), priors (number of prior convictions), tserved (time served in the prison), black, alcohol, felon, property and drugs as regressor affecting the duration recidivism. Three varibles namely age, tserved and priors are scaled by

1000, 100 and 10 respectively according to the suggestion of [2] to improve the interpretability of coefficients and comparison across covariates.



variables	vi.Overall
age	100.0000000
priors	68.5054989
tserved	65.5856527
black	59.0549773
alcohol	42.4879027
felon	33.1192298
property	18.7498844
drugs	18.2277192
rules	10.3582041
married	6.8459441
educ	5.8767442
follow	4.7962893
workprg	4.4813649
person	2.6101374
super	0.1597608

Figure 1: *Variable Importance Plot*

3. Methods

In this section we discuss the different models and methods used for the analysis, we employed fully parametric survival models and their Bayesian formulations to analyze time-to-event data, allowing flexible modeling of covariate effects under censoring. Model performance and selection were assessed using Bayesian model comparison criteria, including Leave-One-Out Information Criterion (LOOIC) and the Widely Applicable Information Criterion (WAIC).

3.1. Parametric Survival Models

A parametric survival model analyzes time-to-event data by assuming that survival times T follow a specific distribution with a known functional form. This allows direct modeling of key functions such as the survival function $S(t) = P(T > t)$, density $f(t)$, and hazard function $h(t) = \frac{f(t)}{S(t)}$. When covariates $\mathbf{X} \in \mathbb{R}^p$ are included in the model through rate or scale parameter (e.g. in exponential model) using log link function,

$$log(\lambda) = \mathbf{X}'\boldsymbol{\beta},$$
$$\lambda = exp(\mathbf{X}'\boldsymbol{\beta}),$$

and throug lacation parameter (e.g. in lognormal model) using an identity link

$$\mu = \mathbf{X}'\boldsymbol{\beta},$$

where $\boldsymbol{X} = (1, x_1, x_2, \cdots, x_p)$ is the model matrix (set of covariates with interceot column), and $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \cdots, \beta_p)$ represents regression coefficients and θ is a parameter of the distribution, usually it is a scale parameter. Parametric models are particularly useful when extrapolation is required and distributional assumptions are justifiable [25],[26].

3.1.1 Exponential Model

The exponential distribution is one of the simplest and most widely used parametric models in survival analysis. It assumes a constant hazard rate over time, making it suitable for modeling time-to-event data where the event risk does not change as time progresses.Let Let T denote a non-negative continuous random variable representing the time until an event The exponential distribution is defined by the rate parameter $\lambda > 0$ as,

- Probability Density Function (PDF):

$$f(t) = \lambda e^{-\lambda t}, \quad t \geq 0 \tag{1}$$

- Cumulative Distribution Function (CDF):

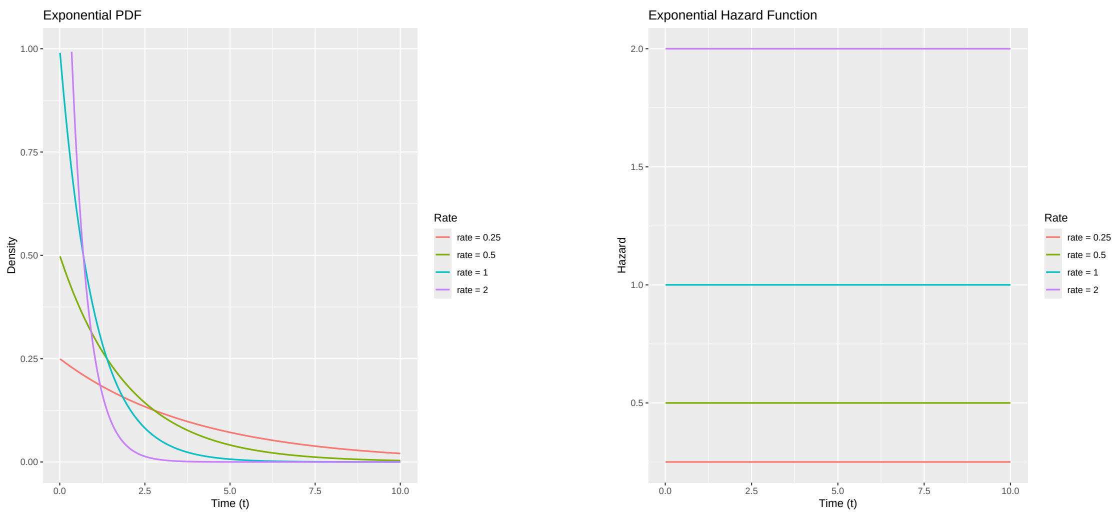
$$F(t) = 1 - e^{-\lambda t} \tag{2}$$

- Survival Function (SF):

$$S(t) = 1 - F(t) = e^{-\lambda t} \tag{3}$$

- Hazard Function (HF):

$$h(t) = \frac{f(t)}{S(t)} = \lambda \tag{4}$$



″ **(a)** *Exponential PDF* **(b)** *Exponential Hazard*

Figure 2: *PDF and hazard function of the Exponential distribution*

3.1.2 Weibull Model

The Weibull distribution is a widely used and flexible parametric model in survival analysis. Unlike the exponential distribution, it allows for a hazard rate that can either increase or decrease over time, making it suitable for a broader range of time-to-event data. Let T be a non-negative continuous random variable representing the time until an event occurs. The Weibull distribution is characterized by two positive parameters: a shape parameter $\alpha > 0$ and a scale parameter $\lambda > 0$.

- Probability Density Function (PDF):

$$f(t) = \frac{\alpha}{\lambda} \left(\frac{t}{\lambda} \right)^{\alpha - 1} \exp\left(- \left(\frac{t}{\lambda} \right)^{\alpha} \right), \quad t \geq 0 \tag{5}$$

- Cumulative Distribution Function (CDF):

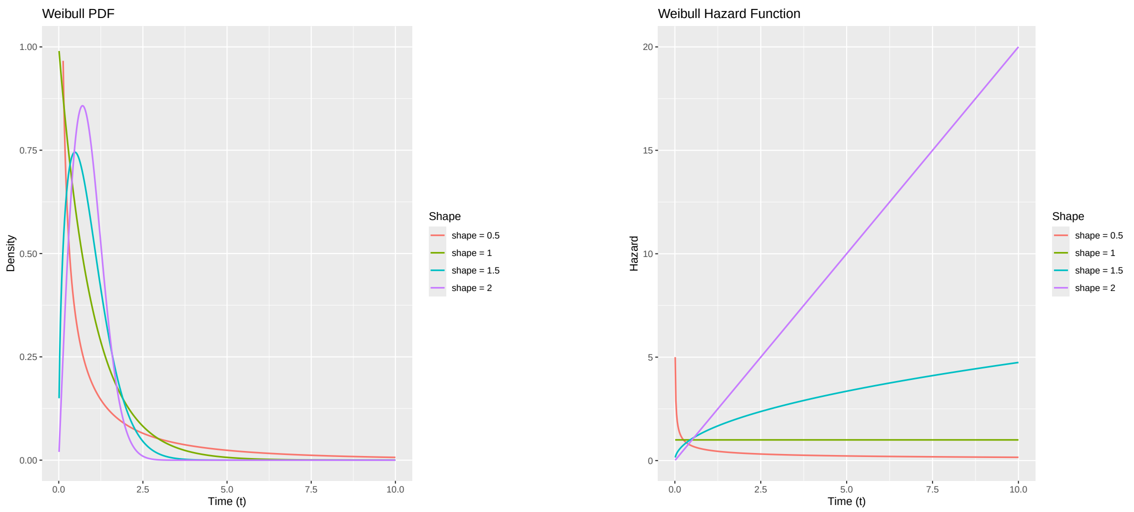
$$F(t) = 1 - \exp\left(-\left(\frac{t}{\lambda}\right)^{\alpha}\right) \tag{6}$$

- Survival Function (SF):

$$S(t) = \exp\left(-\left(\frac{t}{\lambda}\right)^{\alpha}\right) \tag{7}$$

- Hazard Function (HF):

$$h(t) = \frac{\alpha}{\lambda}\left(\frac{t}{\lambda}\right)^{\alpha-1} \tag{8}$$



(a) *Weibull PDF* **(b)** *Weibulll hazard*

Figure 3: *PDF and hazard function of the Weibull distribution*

3.1.3 Frechet Model

The Frechet distribution is part of the extreme value family, used for modeling heavy-tailed survival times or decreasing hazard rates, such as in early failures. Let T be a non-negative continuous random variable representing the time until an event occurs. The Frechet distribution is characterized by two positive parameters: the shape parameter $\alpha > 0$ and the scale parameter $s > 0$. Its probability density function (PDF) is given by:

- Probability Density Function (PDF):

$$f(t) = \frac{\alpha}{s}\left(\frac{t}{s}\right)^{-(\alpha+1)} \exp\left(-\left(\frac{t}{s}\right)^{-\alpha}\right), \quad t > 0 \tag{9}$$

- Cumulative Distribution Function (CDF):

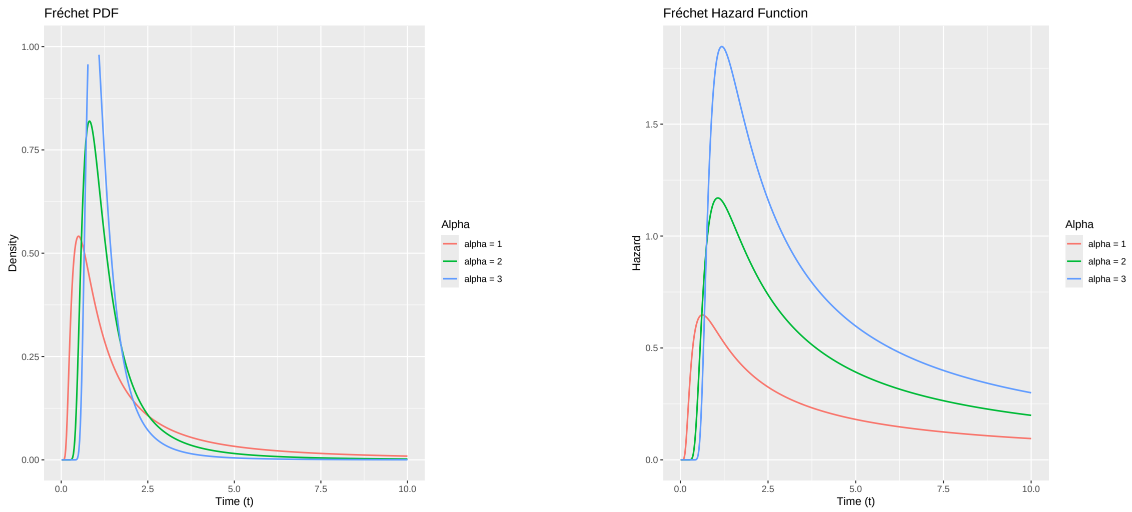
$$F(t) = \exp\left(-\left(\frac{t}{s}\right)^{-\alpha}\right) \tag{10}$$

- Survival Function (SF):

$$S(t) = 1 - \exp\left(-\left(\frac{t}{s}\right)^{-\alpha}\right) \tag{11}$$

- Hazard Function (HF):

$$h(t) = \frac{\alpha}{s}\left(\frac{t}{s}\right)^{-(\alpha+1)} \cdot \left[\frac{1}{1-\exp\left(-\left(\frac{t}{s}\right)^{-\alpha}\right)}\right] \tag{12}$$



(a) *Frechet PDF*

(b) *Frechet Hazard*

Figure 4: *PDF and hazard function of the Frechet distribution*

3.1.4 Lognormal Model

The Log-Normal distribution is a flexible parametric model used in survival analysis, particularly when the hazard function is non-monotonic i.e., it may initially increase and then decrease over time. This makes it suitable for modeling time-to-event data with delayed risk peaks. A random variable T is said to follow a Log-Normal distribution if its logarithm is normally distributed. That is, $\log T \sim \mathcal{N}(\mu, \sigma^2)$, where $\mu \in \mathbb{R}$ and $\sigma > 0$ are the location and scale parameters respectively.

- Probability Density Function (PDF):

$$f(t) = \frac{1}{t\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln t - \mu)^2}{2\sigma^2}\right), \quad t > 0 \tag{13}$$

- Cumulative Distribution Function (CDF):

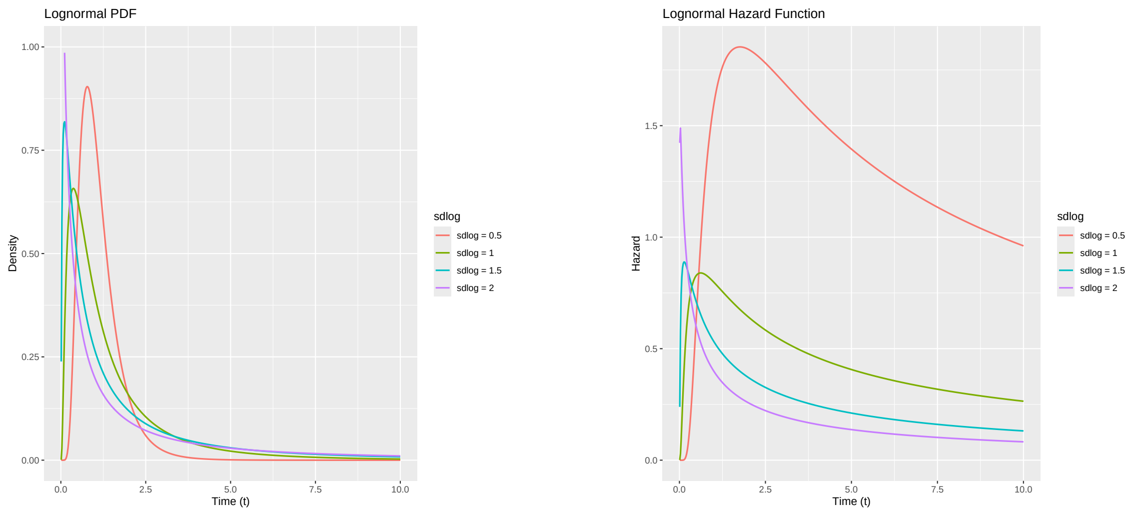
$$F(t) = \Phi\left(\frac{\ln t - \mu}{\sigma}\right) \tag{14}$$

- Survival Function (SF):

$$S(t) = 1 - \Phi\left(\frac{\ln t - \mu}{\sigma}\right) \tag{15}$$

- Hazard Function (HF):

$$h(t) = \frac{1}{t\sigma\sqrt{2\pi}} \cdot \frac{\exp\left(-\frac{(\ln t-\mu)^2}{2\sigma^2}\right)}{1-\Phi\left(\frac{\ln t-\mu}{\sigma}\right)} \tag{16}$$



(a) *Lognormal PDF* **(b)** *Lognoral hazard*

Figure 5: *PDF and hazard function of the lognormal distribution*

3.2. Bayesian Model Formulation

In Bayesian survival analysis, the model is defined through a combination of prior distributions on the parameters and a likelihood function derived from the survival model and observed data. The core idea is to update the prior beliefs in light of the observed data using Bayes' rule:

$$p(\boldsymbol{\theta} \mid \mathcal{D}) \propto p(\mathcal{D} \mid \boldsymbol{\theta}) \cdot p(\boldsymbol{\theta}),$$

where $\boldsymbol{\theta}$ denotes the set of model parameters, $p(\mathcal{D} \mid \boldsymbol{\theta})$ is the likelihood, $p(\boldsymbol{\theta})$ is the prior. This formulation enables us to quantify parameter uncertainty explicitly by working with the entire posterior distribution rather than single point estimates.

3.2.1 Priors

In the Bayesian framework, prior distributions encode our beliefs about the model parameters before observing the data. For regression models [27] recommend the use of weakly informative priors that provide regularization to avoid overfitting, while remaining flexible enough to allow the data to dominate. For a regression model with coefficients $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)$ and a positive scale parameter σ, a typical prior specification is $\beta_j \sim \mathcal{N}(0, 5^2), \quad j = 0, 1, \ldots, p$, and $\sigma \sim \text{half-Cauchy}(0, 25)$. The references [28] and [19] provide detailed information regarding the application of the half-Cauchy prior to shape and scale parameters alongside the Gaussian prior to the regression coefficient.

$$p(\beta) = \frac{1}{\sqrt{(2\pi \times 25)}} \times e^{\frac{-\beta^2}{2\times 25}}, -\infty < \beta < \infty. \tag{17}$$

$$p(\sigma) = \frac{2 \times 25}{\pi(\sigma^2 + 25^2)}, \sigma > 0. \tag{18}$$

These priors reflect prior uncertainty about the magnitude of the coefficients and ensure the scale parameter remains positive and not overly influential. The half-Cauchy prior on σ has heavier tails than the inverse-gamma distribution, offering greater robustness and protection against over shrinkage.

3.2.2 Likelihood

Let t_i be the observed survival time for the i-th individual, and δ_i the event indicator with $\delta_i = 1$ if the event is observed and $\delta_i = 0$ if right-censored. Let $f(t_i \mid \boldsymbol{\theta})$ and $S(t_i \mid \boldsymbol{\theta})$ be the density and survival functions of the survival model parameterized by $\boldsymbol{\theta}$. Then likelihood for right-censored data as given in [26]

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} [f(t_i \mid \boldsymbol{\theta})]^{\delta_i} [S(t_i \mid \boldsymbol{\theta})]^{1-\delta_i}. \tag{19}$$

Now, for the lognormal model where $T \sim lognormal(\mu = X'\beta, \sigma^2)$, substituting the pdf and survival function of lognormal distribution from Equation 13,15 in Equation 19 the likelihood for the log-normal model is obtained as

$$p(t_i \mid \boldsymbol{\beta}, \sigma) = \prod_{i=1}^{n} \left[\left(\frac{1}{t_i \sigma \sqrt{2\pi}} \exp\left(-\frac{(\log t_i - X'\beta)^2}{2\sigma^2} \right) \right)^{\delta_i} \cdot \left(1 - \Phi\left(\frac{\log t_i - X'\beta}{\sigma} \right) \right)^{1-\delta_i} \right] \tag{20}$$

where $\Phi(\cdot)$ cumulative density function (CDF) of the standard normal distribution.

3.2.3 Posterior Distribution

Using Bayes' theorem, the posterior distribution is proportional to the product of the likelihood and priors:

$$p(\boldsymbol{\beta}, \sigma \mid \mathcal{D}) \propto p(\mathcal{D} \mid \boldsymbol{\beta}, \sigma) \cdot p(\boldsymbol{\beta}) \cdot p(\sigma),$$

where $\mathcal{D} = \{(t_i, \delta_i, \boldsymbol{x}_i)\}_{i=1}^{n}$ denotes the observed data. For the log-normal model, so

$$\begin{aligned} p(\boldsymbol{\beta}, \sigma \mid \mathcal{D}) \propto \prod_{i=1}^{n} & \left[\left(\frac{1}{t_i \sigma \sqrt{2\pi}} \exp\left(-\frac{(\log t_i - X'\beta)^2}{2\sigma^2} \right) \right)^{\delta_i} \cdot \left(1 - \Phi\left(\frac{\log t_i - X'\beta}{\sigma} \right) \right)^{1-\delta_i} \right] \\ & \cdot p(\boldsymbol{\beta}) \cdot p(\sigma). \end{aligned} \tag{21}$$

This posterior expression combines information from both censored and uncensored observations, where the likelihood contribution depends on whether the event time was observed or right-censored. The inclusion of prior distributions on β and σ allows incorporation of prior knowledge or regularization to stabilize estimation in the presence of limited or noisy data.

Since the posterior is typically not available in closed form, inference is performed via **stan** modeling language through **(**R) software. Stan is a probabilistic programming language that use No-U-Turn-Sampler (NUTS) an extension of Hamiltonian Monte Carlo (HMC) algorithm to draw samples from the posterior densities where analytical solutions are intractable.This approach relies on advanced Markov Chain Monte Carlo (MCMC) algorithms, which efficiently approximate complex posterior distributions.

3.3. Model Selection Criteria

After model fitting, selection of the best performing model is one of the crucial step in statistical modeling. For this purpose we choose two fully Bayesian criterion, Leave One Out Information Criterion (LOOIC) and Widely Applicable Information Criterion [29],[30]. In recent LOOIC and WAIC have been widely used for Bayesian model comparison due to their principled approach to estimating out-of-sample predictive accuracy [31],[19].

Both WAIC and LOOIC are on the deviance scale, and lower values indicate better predictive performance [32]. In practice, these criteria are often very close numerically, but LOOIC is generally preferred due to its better diagnostic capabilities and reliability, especially in the presence of influential observations. In this work, LOOIC and WAIC are computed using the **loo** package in **R**, which provides efficient implementations of Pareto-smoothed importance sampling and related diagnostics [33].

3.3.1 LOOIC

The Bayesian LOOIC estimate of out-of-sample predictive fit is In leave-one-out cross-validation, the predictive performance is assessed by leaving out each observation one at a time and evaluating how well the model predicts it. Mathematically,

$$\mathrm{LOOIC} = -2 \sum_{i=1}^{n} \log \left(\frac{1}{S} \sum_{s=1}^{S} w_i^s p(y_i \mid \theta^s) \right), \tag{22}$$

where w_i^s are the importance weights from the Pareto Smoothed Importance Sampling (PSIS) approximation. LOOIC provides a robust measure of predictive accuracy and is particularly effective when data points have varying influence on the model.

3.3.2 WAIC

The Watanabe-Akaike information criterion also known as Widely Applicable Information Criterion (WAIC) is a fully Bayesian generalization of the Akaike Information Criterion (AIC)[30]. WAIC evaluates model fit by averaging the pointwise predictive accuracy over the posterior distribution and includes a correction term for effective model complexity. It is calculated as:

$$\mathrm{WAIC} = -2 \left(\sum_{i=1}^{n} \log \left(\frac{1}{S} \sum_{s=1}^{S} p(y_i \mid \theta^s) \right) - \sum_{i=1}^{n} \mathrm{Var}_{s=1}^{S} \left[\log p(y_i \mid \theta^s) \right] \right), \tag{23}$$

where θ^s are posterior samples, and $p(y_i \mid \theta^s)$ is the likelihood for observation y_i. The second term acts as a penalty for model complexity, ensuring that more complex models are not unduly favored.

4. Results and Discussion

This section presents the findings from the Bayesian parametric survival models and interprets the results.

4.1. Model selection and performance

Table 2 reports the LOOIC and WAIC values for each model. Among the four parametric survival models fitted to the data the Lognormal model yielded the lowest values for both the Leave-One-Out Information Criterion (LOOIC = 6328.6) and the Widely Applicable Information Criterion (WAIC = 6328.5). These results indicate that the Lognormal model provides the best out-of-sample predictive accuracy among the candidate models.This suggests that the hazard of reoffending is not monotonic over time, it may initially increase and later decrease which aligns with the intuition that recidivism risk is highest during the early period after release when individuals face the greatest reintegration challenges. The Weibull and exponential models followed, with slightly higher information criteria values, while the Frechet model showed substantially poorer fit, as reflected by its much higher LOOIC and WAIC values. Based on these criteria, the lognormal model was selected for further interpretation and posterior analysis.

Table 2: *LOOIC and WAIC Values of the fitted models*

Model	LOOIC	WAIC
Lognormal	**6328.6**	**6328.5**
Weibull	6402.0	6401.8
Exponential	6435.3	6434.9
Fretchet	7324.6	7325.1

4.2. Posterior estimates

Table 3 presents the posterior summaries of the lognormal survival model. All parameters exhibit good convergence diagnostics, with $\widehat{R}$ values equal to 1.00 and sufficiently large effective sample sizes (N_{eff}), indicating reliable posterior estimation. The intercept estimate is 4.43 with a 95% credible interval [4.01, 4.84], serving as the baseline log survival time. Age (scaled by 1000) has a strong positive effect ($\hat{\beta}$ = 3.97, 95% CI [2.85, 5.16]), suggesting that older individuals tend to have longer survival times (recidivism duration) post-release. In contrast, the number of prior convictions (priors/10) and time served (tserved/100) are both negatively associated with survival time, with posterior means of -1.38 and -2.03, respectively, indicating that more priors and longer incarceration are linked to shorter time to recidivism.

Factor variables in the study also show notable effects. Being black (black1), having a history of alcohol abuse (alcohol1), drug involvement (drugs1), or a property-related offense (property1) are all negatively associated with survival time, with credible intervals excluding zero (significant effect). On the other hand, being a felon (felon1) is positively associated with log survival time ($\hat{\beta}$ = 0.88, 95% CI [0.50, 1.27]), indicating a longer time to reoffending relative to non-felons. The residual standard deviation σ is estimated at 1.83 (95% CI [1.71, 1.95]), capturing the remaining variability in log survival time not explained by the covariates. Overall, the model captures important individual-level heterogeneity in recidivism risk.

Table 3: *Posterior summaries from the lognormal model*

Parameter	mean	*se_mean*	2.5%	97.5%	$\hat{R}$	N_{eff}
Intercept	4.43	0.21	4.01	4.84	1.00	5178
age/1000	3.97	0.59	2.85	5.16	1.00	4212
tserved/100	-2.03	0.30	-2.63	-1.45	1.00	3868
priors/10	-1.38	0.22	-1.80	-0.95	1.00	4085
black1	-0.54	0.12	-0.78	-0.31	1.00	6137
alcohol1	-0.63	0.14	-0.91	-0.35	1.00	5623
felon1	0.88	0.20	0.50	1.27	1.00	3202
property1	-0.62	0.19	-1.00	-0.25	1.00	3684
drugs1	-0.34	0.13	-0.60	-0.08	1.00	5659
sigma(σ)	1.83	0.06	1.71	1.95	1.00	3852

4.3. Posterior density and trace plots

Posterior density plots and trace plots are two crucial visualization tool for assessing the quality and reliability of posterior summaries. The posterior density plots summarize the distribution of the model parameters after updating prior beliefs with the observed data, allowing us to evaluate location, spread, and possible skewness in the estimates. The trace plots display the parameter values sampled across iterations of the Markov Chain Monte Carlo (MCMC) algorithm. A well-mixed trace plots without visible trends or long-term autocorrelation indicate convergence of the chains, while consistent overlap across multiple chains confirms stability of the posterior draws.

Figure 6 displays the marginal posterior distributions of the model parameters. It is evident from the plot that the marginal posterior densities are well-separated from zero indicating significant influence on survival time. This also shows the effectiveness of the variable selection process as all the selected variables have significant effect. Figure 7 confirms the good mixing behavior of MCMC chains, indicating stable inference. Moreover, the trace plots exhibit no apparent trends or drifts, suggesting that the chains have reached stationarity. This strengthens confidence in the reliability of the posterior summaries and the robustness of the estimated effects.

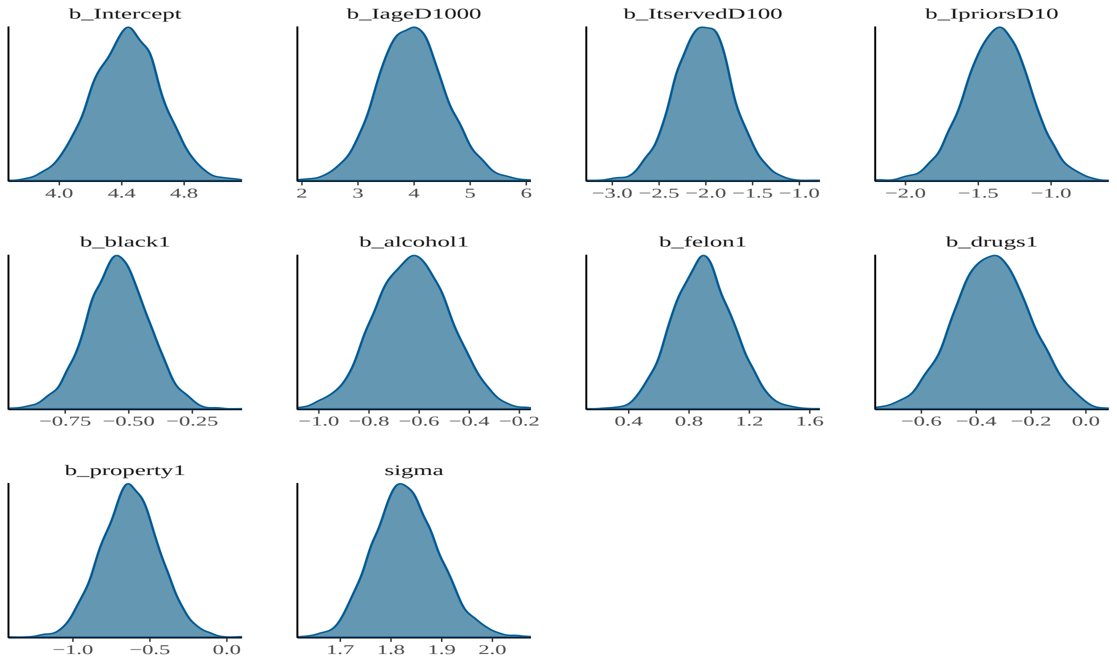


Figure 6: *Posterior distributions of coefficients from the lognormal survival model and scale parameter σ. Each horizontal segment represents the marginal posterior distribution for a specific parameter. This visualization helps identify the magnitude and direction of each predictor's effect on survival time, with zero indicating no effect. Parameters with posteriors concentrated away from zero suggest a meaningful association with time to recidivism.*

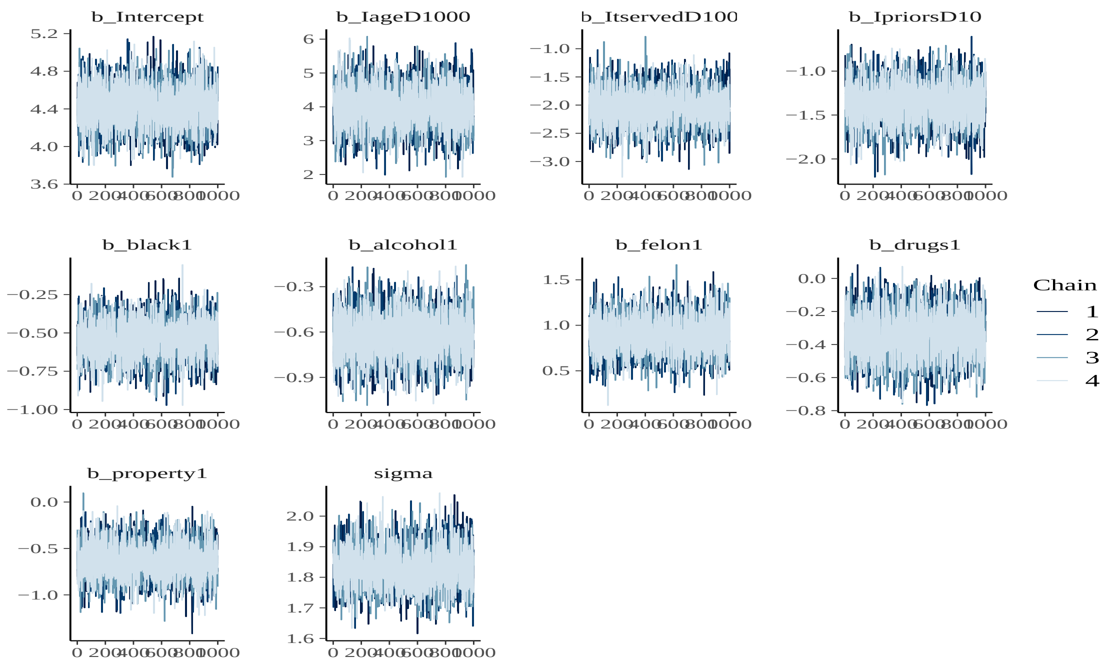


Figure 7: *Trace plots of MCMC chains for the parameters in he lognormal survival model. This shows the MCMC sampling behavior across four chains, well-mixed chains shows consistent behavior across iterations without persistent trends, drifts, or divergence suggests good convergence. The chains here appear to cover the parameter space adequately.*

4.4. Conditional effects plots

A conditional effects plot lets you see the role of a certain factor in the model by keeping the influence of all other factors the same. It demonstrates how the forecasted result responds when the chosen variable's value changes. This allows us to easily pick out how that variable affects the outcome, without other variables confusing the picture. The graph has partial regions to let us judge the confidence in our predictions. For models with many variables, interpreting the results is much simpler with conditional effects plots.

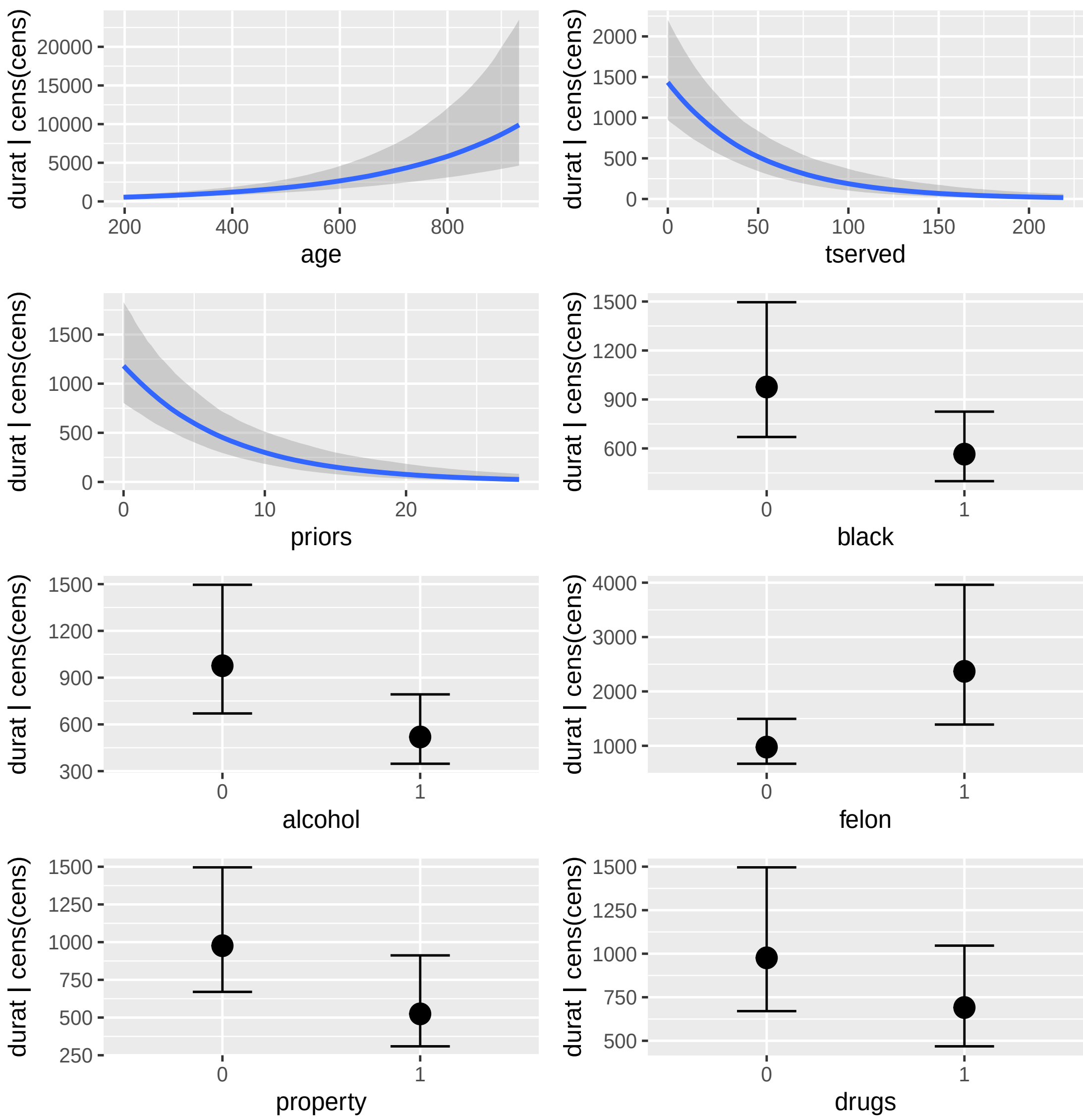


Figure 8: *Conditional effects plots illustrate the conditional effects of each covariate on the expected survival duration, averaged over the posterior distribution and holding other covariates constant at typical values. The vertical axis represents the expected survival time (duration) conditional on no censoring, and the horizontal axis corresponds to the covariate being examined. Each plot includes a fitted line showing the posterior mean along with uncertainty bands (typically 95% credible intervals).*

- **Age**: The plot shows a strong positive association between age and survival duration. As age increases, the expected time to recidivism increases significantly, especially at lower age ranges. This suggests that older individuals tend to remain offense-free for longer periods post-release.

- **Time Served (tserved)**: There is a clear negative relationship, indicating that individuals who served longer prison terms tend to recidivate sooner. Longer incarceration durations are associated with shorter post-release survival times.
- **Number of Priors (priors)**: A higher number of prior convictions is associated with shorter survival times. The decline is steep for lower values of priors, implying diminishing survival time with increasing criminal history.
- **Race (black)**: The binary indicator shows that Black individuals (coded as 1) have shorter expected survival times compared to non-Black individuals. This aligns with the negative regression coefficient and may reflect broader systemic disparities.
- **Alcohol Abuse (alcohol)**: Individuals with a history of alcohol problems tend to have shorter survival times, indicating a higher likelihood of quicker recidivism.
- **Felony Status (felon)**: Contrary to some expectations, individuals identified as felons exhibit longer survival times on average. This may reflect factors such as increased supervision or unobserved protective effects.
- **Property Offense (property)**: Property offenders have slightly shorter survival durations than non-property offenders, though the effect appears modest in magnitude.
- **Drug Involvement (drugs)**: Similar to alcohol, drug involvement is negatively associated with survival time, suggesting a greater likelihood of early reoffending.

Overall, these conditional effect plots visually reinforce the estimated effects from the posterior distributions and provide an interpretable, intuitive summary of how individual covariates influence post-release recidivism risk within the Bayesian framework.

4.5. Posterior predictive density

A posterior density plot is a useful way to visualize the uncertainty around a parameter's value after observing the data, as guided by Bayesian inference. It illustrates the range of likely values a parameter can take, along with their associated probabilities, helping us better understand the model's behavior and compare different components.

In our analysis, we used posterior predictive density (PPD) plots to assess how well each fitted survival model lognormal, Weibull, exponential, and Frechet captures the observed data. These plots allow us to compare the predicted survival times against the actual data, giving a clear picture of how closely each model aligns with reality. By examining these visualizations, we can identify which model offers the best fit and most reliable predictive performance.

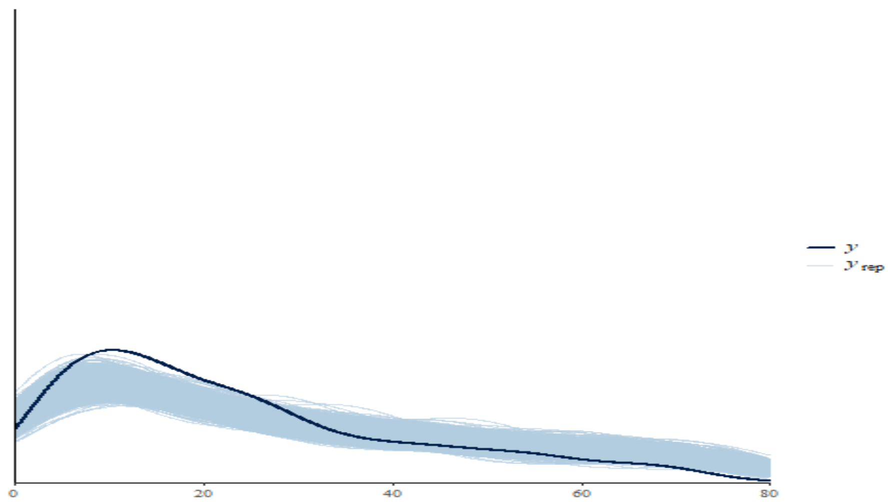


Figure 9: *Lognormal Model*

Figure 10: *Posterior predictive density plots under Lognormal survival model*

Posterior predictive density (PPD) plot are a crucial diagnostic tool in Bayesian modeling to assess the adequacy of the fitted model. Figure 10 presents a density overlay plot comparing the distribution of the observed survival times (y) with replicated data (y_{rep}) generated from the posterior predictive distribution of the Lognormal, Weibull, Exponential, and Fretchet models. The Figure 9 shows that the model-generated data from lognormal model closely track the shape, range, and central tendency of the observed data. The overlap between the distributions suggests that the Lognormal model captures the key features of the observed survival data well. This agreement indicates that the model is capable of generating plausible survival times consistent with the empirical data, providing additional support for its suitability in this context. The lack of significant discrepancy between y and y_{rep} in this plot reinforces the conclusion that the Lognormal model is a good fit for the recidivism data.

5. Conclusion

This study begins from evaluating fifteen covariates of recidivism data for their impact on post-release recidivism time. Using the variable importance index we retained only eight predictors with significant influence, limiting the model to depend on age, priors, time served, race, alcohol and drug involvement, felony status, and property offense.It was evident from a model comparison based on LOOIC and WAIC that the Lognormal survival model offered the best predictive fit, accurately representing the realistic non-monotonic hazard pattern that characterizes recidivism.

From the lognormal Bayesian survival model it was found that older individuals tend to remain offense-free longer, while more prior convictions and longer sentences predict quicker reoffending. Factor variables like race, substance abuse, and offense type also showed significant associations with survival time. Good convergence diagnostics, well-mixed chains, and close alignment of posterior predictive densities with the observed data confirm the model's robustness. Recidivism imposes significant costs on the criminal justice system and society, including incarceration expenses, lost productivity, and broader social impacts. Hence from economic perspective, these findings are important in targeting resources effectively to check reoffending problem. Understanding which factors most strongly predict quicker reoffending can help policymakers allocate funds to programs that address substance abuse, support older ex-offenders, and tailor supervision intensity to individual risk profiles. This evidence based approach can help optimize public spending, reduce the economic burden of crime, and improve reintegration outcomes.

In sum, this analysis demonstrates that carefully selecting covariates and choosing an appropriate survival distribution can yield valuable insights into the dynamics of recidivism. The findings highlight individual level factors that should be considered in designing interventions and policies to support successful reintegration and reduce reoffending.

References

[1] Chung, C. F., Schmidt, P., and Witte, A. D. (1991). Survival analysis: A survey. *Journal of Quantitative Criminology*, 7:59–98.

[2] Schmidt, P. and Witte, A. D. (2012). *Predicting recidivism using survival models*. Springer Science & Business Media.

[3] Yukhnenko, D., Farouki, L., and Fazel, S. (2023). Criminal recidivism rates globally: A 6-year systematic review update. *Journal of Criminal Justice*, 88:102–115.

[4] Becker, G. S. (1968). Crime and punishment: An economic approach. *Journal of Political Economy*, 76(2):169–217.

[5] Schmidt, P. and Witte, A. D. (1984). *An Economic Analysis of Crime and Justice: Theory, Methods and Applications*. Academic Press, Inc.

[6] Kubrin, C. E. and Stewart, E. A. (2006). Predicting who reoffends: The neglected role of neighborhood context in recidivism studies. *Criminology*, 44(1):165–197.

[7] Burgess, E. W. (1928). Factors determining success or failure on parole. *The Workings of the Indeterminate Sentence Law and the Parole System in Illinois*, 221–234.

[8] Durose, M. R., Cooper, A. D. and Snyder, H. N. (2014). Recidivism of prisoners released in 30 States in 2005: Patterns from 2005 to 2010. *US Department of Justice, Office of Justice Programs, Bureau of Justice* 28.

[9] Zeng, J., Ustun, B., and Rudin, C. (2017). Interpretable classification models for recidivism prediction. *Journal of the Royal Statistical Society: Series A*, 180(3):689–722.

[10] Lila, M., Martín-Fernández, M., Gracia, E., López-Ossorio, J. J. and González, J. L. (2019). Identifying key predictors of recidivism among offenders attending a batterer intervention program: A survival analysis. *Psychosocial Intervention*, 28:157–167.

[11] Maltz, M. D. (1984). *Recidivism*. Academic Press, Orlando, FL.

[12] Partanen, J. (1969). On waiting time distributions. *Acta Sociologica*, 12(3):132–143.

[13] Carr-Hill, G. A. and Carr-Hill, R. A. (1972). Reconviction as a process. *British Journal of Criminology*, 12:35.

[14] Schmidt, P. and Witte, A. D. (1989). Predicting criminal recidivism using split population survival time models. *Journal of Econometrics*, 40(1):141–159.

[15] Visher, C. A., Lattimore, P. K. and Linster, R. L. (1991). Predicting the recidivism of serious youthful offenders using survival models. *Criminology*, 29(3):329–366.

[16] Gruenewald, P. J. and West, B. R. (1989). Survival models of recidivism among juvenile delinquents. *Journal of Quantitative Criminology*, 5:215–229.

[17] Arfan, N., Irfanullah, A., Hamidi, M. R. and Mukhaiyar, U. (2025). Comparison of survival analysis using accelerated failure time model and Cox model for recidivist case. *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, 19(1):629–642.

[18] De La Cruz, R., Padilla, O., Valle, M. A. and Ruz, G. A. (2021). Modeling recidivism through Bayesian regression models and deep neural networks. *Mathematics*, 9(6):639.

[19] Parvej, M., Devashish and Khan, A. A. (2024). Enhancing survival analysis with Bayesian modeling: A comparative study of Type II half logistic-Weibull Cox-PH, Weibull Cox-PH and Weibull models. *International Journal of Reliability, Quality and Safety Engineering*, 32(3) 2450038.

[20] Farhin, S., Yousuf, F. and Khan, A. A. (2022). Bayesian survival modeling of Marshal-Olkin generalized-G family with random effects using R and Stan. *Reliability: Theory & Applications*, 17(2):422–440.

[21] Simon, N. and Tibshirani, R. (2011). Discriminant analysis with adaptively pooled covariance. *arXiv preprint arXiv:1111.1687*.

[22] Shea, J. M. (2024). wooldridge: 115 data sets from "Introductory Econometrics: A Modern Approach, 7e" by Jeffrey M. Wooldridge. *R package version 1.4-4*. `https://justinmshea.github.io/wooldridge/`.

[23] Wooldridge, J. M. (2019). *Introductory Econometrics: A Modern Approach*. 7th ed., Cengage Learning.

[24] Smith, B. J. (2024). MachineShop: Machine learning models and tools. *R package version 3.8.0*. `https://brian-j-smith.github.io/MachineShop/`.

[25] Collett, D. (2023). Modelling Survival Data in Medical Research. 4th ed., Chapman and Hall/CRC.

[26] Lawless, J. F. (2011). *Statistical Models and Methods for Lifetime Data*. 2nd ed., John Wiley & Sons.

[27] Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis*, 1(3):515–534. `https://doi.org/10.1214/06-BA117A`.

[28] Parvej, M. and Khan, A. A. (2024). Bayesian extension of the Weibull AFT shared frailty model with generalized family of distributions for enhanced survival analysis using censored data. *Journal of Applied Statistics*, 51(15):3125–3153.

[29] Vehtari, A., Gelman, A., and Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, 27(5):1413–1432.

[30] Watanabe, S. and Opper, M. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, 11(12).

[31] Devashish, Shazia, F., Parvej, M. and Khan, A. A. (2025). Bayesian model inference: Exploring parsimonious models with MCMC and optimization approximations. *Communications in Statistics–Theory and Methods*, 1–26.

[32] Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013). *Bayesian Data Analysis*. 3rd ed., Chapman and Hall/CRC..

[33] Vehtari, A., Gabry, J., Magnusson, M., Yao, Y., Bürkner, P. C., Paananen, T., and Gelman, A. (2024). loo: Efficient leave-one-out cross-validation and WAIC for Bayesian models. *R package version 2.8.0*. `https://mc-stan.org/loo/`.