SURVIVAL RATE ANALYSIS - UNSECURED PERSONAL LOAN PORTFOLIO

Ian Moore

2025-08-11

SURVIVAL RATE MODELING

Analysis of Unsecured Personal Loan Portfolio Segemented by Credit Score, Rate and Original Loan Amount

This notebook demonstrates advanced survival analysis techniques using Kaplan-Meier and Cox regression models applied to unsecuredpersonal loan portfolio data. We progress from exploratory data analysis through univariate survival modeling to risk-adjusted survival estimates and competing risk analysis. Lastly, we'll compile findings for credit risk management and portfolio optimization applications. Key Analyses:

1) Portfolio EDA and survival model readiness assessment
2) Kaplan-Meier survival estimation with censoring adjustments for young loans
3) Cox Proportional Hazards modeling with time-varying covariates
4) Credit risk management applications (PD modeling, portfolio segmentation, loss forecasting)

1.0 Loan Portfolio Sumamry and Exploratory Data Analysis

1.1 Generate Simulated Loan Portfolio

```
score_bucket
Prime          1560
Super-Prime    1407
Near-Prime     1204
Subprime        284
Name: count, dtype: int64


Loan data generated successfully.


Total Loans: 4,455
Date Range: 2021-02-01 to 2025-01-31
Years Covered: 5
```

Overview

This analysis utilizes a simulated personal loan portfolio designed to replicate realistic market conditions and lending patterns for survival analysis modeling. The dataset provides a robust foundation for examining loan performance, default behavior, and portfolio risk dynamics.

Loan Volume & Timing

- Total Loans Generated: 5,000
- Origination Period: January 1, 2019 – June 30, 2025
- Volume Distribution: Strategically weighted to reflect market conditions
- Higher origination volumes during low-rate environment (2019-2022)
- Reduced origination volumes post-rate increases (2022-2025)

Loan Terms & Structure

- Principal Amount Range: $1,000 – $50,000
- Maturity Terms: 3 to 8 years (variable by loan)
- Product Type: Unsecured personal loans with fixed rates and terms
- Amortization: Standard fixed-payment structure

Credit Risk Profile

- Credit Scores: Randomly distributed across realistic range (350-850)
- 6-Month Refresh Scores: Secondary credit score data point for risk monitoring
- Risk Segmentation: Natural distribution across credit quality spectrum

Rate Environment Stratification

The interest rate structure was strategically designed to mirror actual market conditions.

Pre-Rate Increase Period (2019 – April 2022):

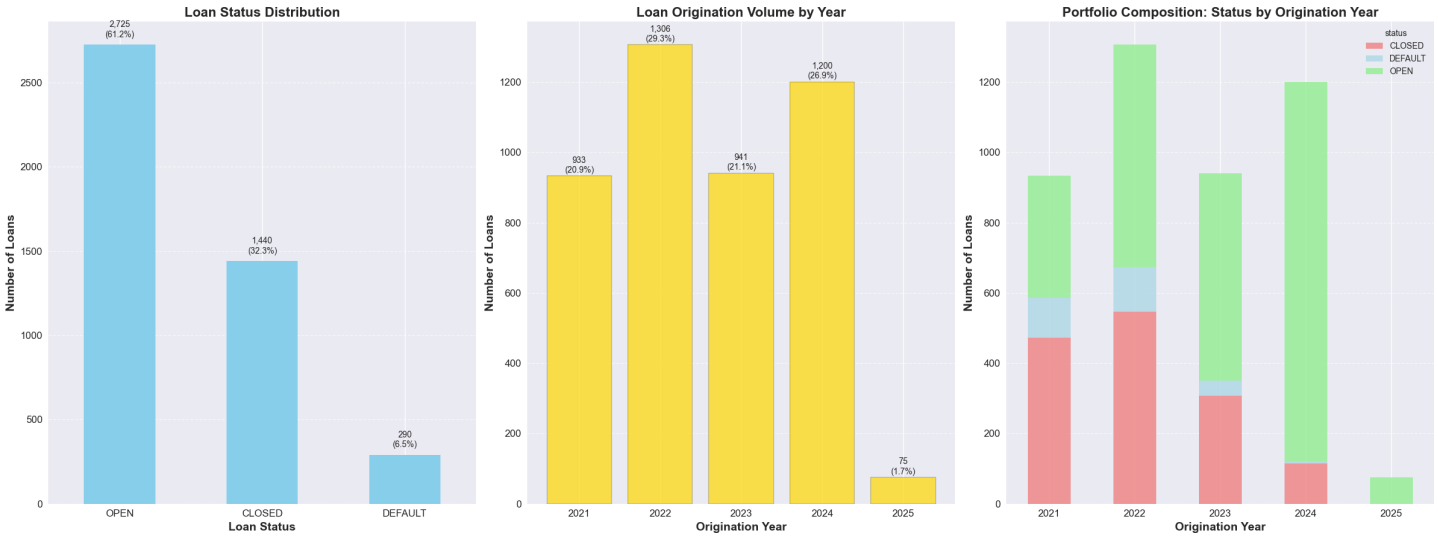
- Lower base interest rates reflecting historical market conditions
- Higher loan origination volumes
- Inverse correlation maintained between credit scores and rates

Rate Increase Period (April 2022 – 2025):

- Progressive rate increases of 5-6 percentage points
- Gradual implementation over 18-24 months
- Corresponding reduction in loan origination volumes
- Maintained credit-risk pricing differentials

1.2 Portfolio Summary and Review

```
Portfolio summary statistics results are the following:
  Total Number of Loans: 4,455
  Average Portfolio Rate: 12.33%
  Average Original Amount: $13,905
  Average Credit Score: 710
```



Maybe add % open, closed, default, total by year as a table

1.3 Key Metrics Summary by Credit Score, Rate, and Opening Amount

```
DEFAULT RATES BY ORIGINATION VINTAGE:
```

open_year	Total_Loans	Defaults	Default_Rate_Pct	Avg_Duration_Months
2021	933	114	12.2	43.3
2022	1306	127	9.7	31.3
2023	941	44	4.7	18.8
2024	1200	5	0.4	7.3
2025	75	0	0.0	0.5

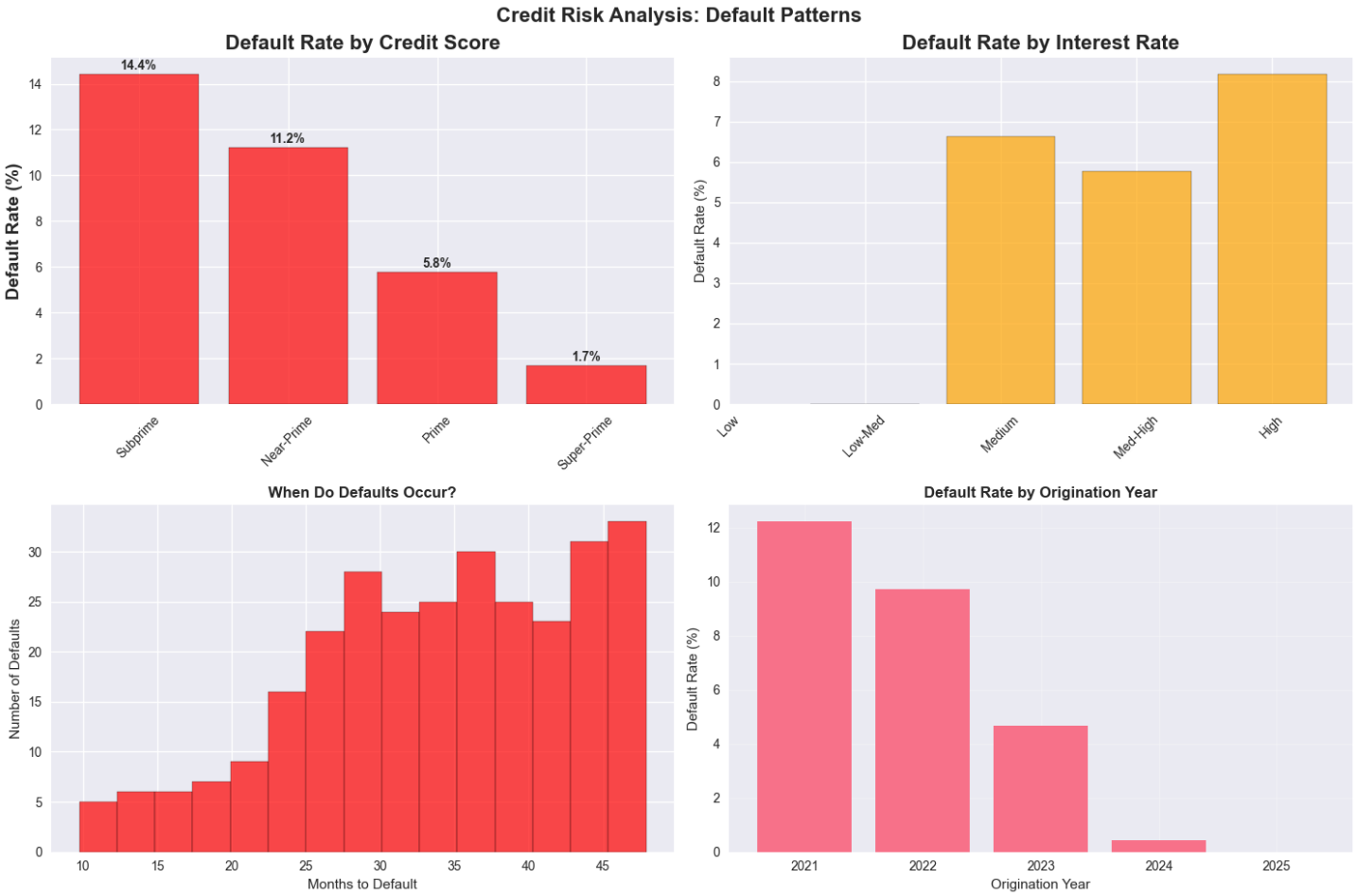
```
DEFAULT RISK ANALYSIS BY KEY SEGMENTS:


1. Default Rates by Credit Score Segment:
              Total_Loans  Defaults  Default_Rate_%
score_bucket
Subprime              284        41            14.4
Near-Prime           1204       135            11.2
Prime                1560        90             5.8
Super-Prime          1407        24             1.7

2. Default Rates by Interest Rate Segment:
             Total_Loans  Defaults  Default_Rate_%
rate_bucket
```

Low	0	0	NaN
Low-Med	1	0	0.0
Medium	2351	156	6.6
Med-High	1610	93	5.8
High	489	40	8.2

3. Risk Differentiation:
   - Credit Score Risk Spread: 12.7 percentage points
   - Highest Risk Segment: 14.4% default rate
   - Lowest Risk Segment: 1.7% default rate

Add some commentary here

VISUALIZE DEFAULT RISK SEGMENTS:



Write a detailed summary about what I'm seeing in the data. May need to add another section on the Rate analysis

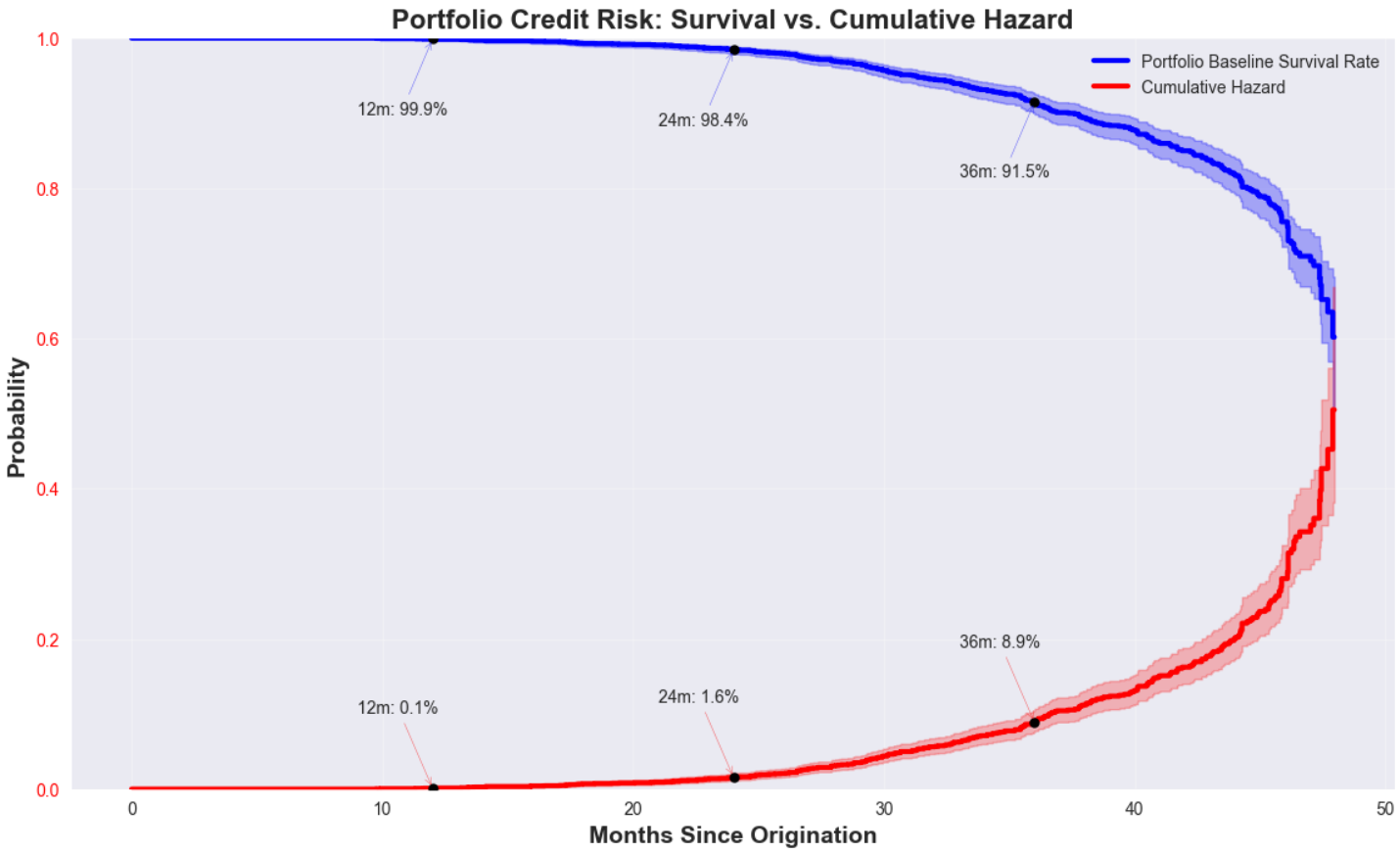
2.0 Kaplan-Meier Survival Rate Analysis

Add a brief writeup of the appraoch. Focus on defaulted. Binary targets: 1 = defaulted, 0 = closed or open. using months as the unit of measurement.

2.1 Baseline Kaplan-Meier Survival Rate and Cumulative Hazard Analysis

```
BASELINE SURVIVAL ANALYSIS - DATA PREPARATION:

  Observation Date: 2025-01-31
  Total Loans: 4,455
  Default Events: 290 (6.5%)
  Censored Observations: 4,165 (93.5%)
  Average Duration: 24.2 months
  Duration Range: 0.0 to 48.0 months
```

2.2 Fit and Visualize the Baseline Model



Provide a summary

2.3 Evaluate Summary Statistics of Baseline Model

```
PORTFOLIO RISK METRICS:

Overall Default Rate: 6.5%
Average Observation Period: 24.2 months
Estimated Annualized Default Rate: 3.2%

BASELINE SURVIVAL STATISTICS:
```

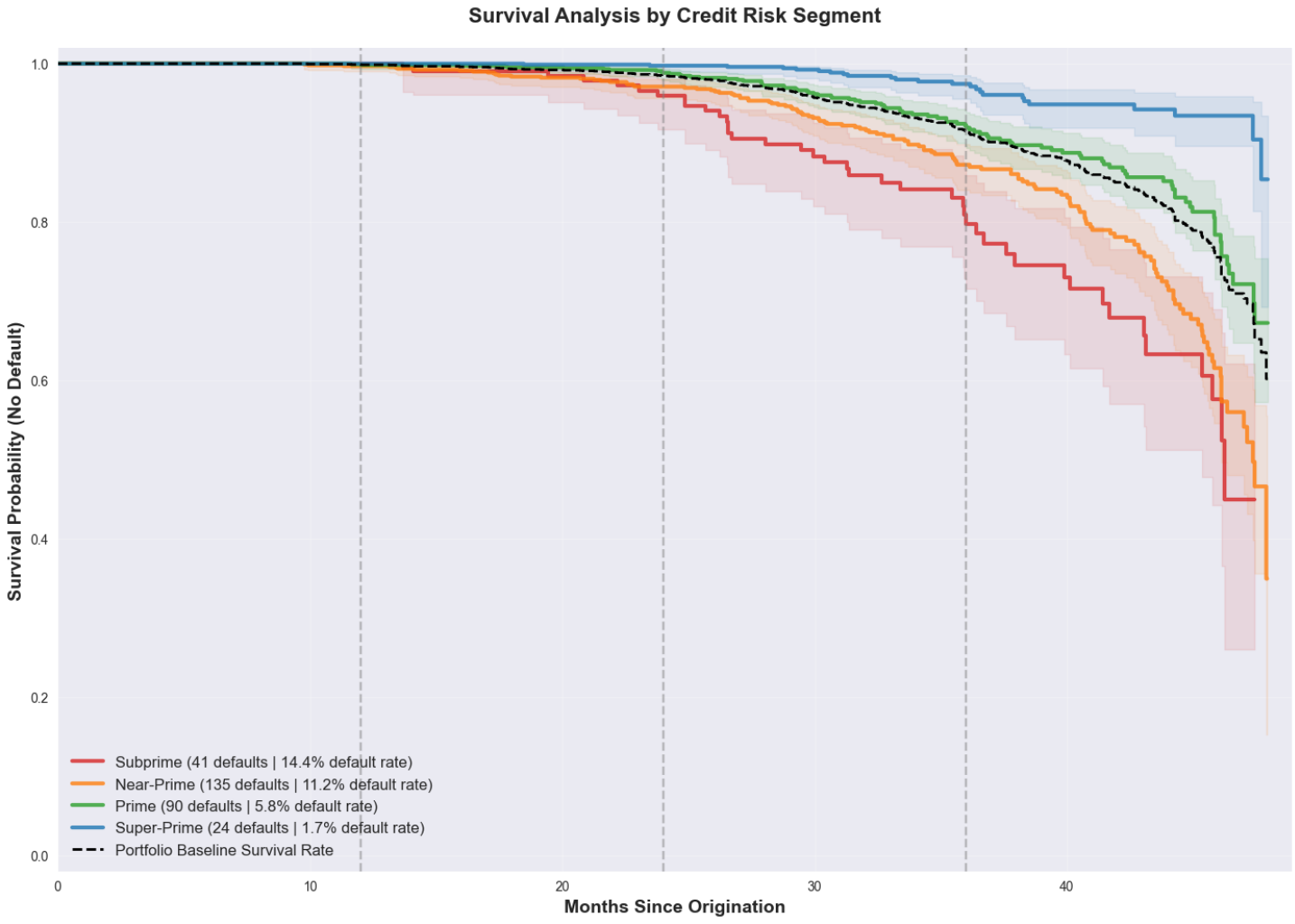
```
 6 months: 100.0% survival | 0.0% default rate | Cumulative Hazard: 0.0%
12 months: 99.9% survival | 0.1% default rate | Cumulative Hazard: 0.1%
18 months: 99.3% survival | 0.7% default rate | Cumulative Hazard: 0.7%
24 months: 98.4% survival | 1.6% default rate | Cumulative Hazard: 1.6%
30 months: 95.7% survival | 4.3% default rate | Cumulative Hazard: 4.4%
36 months: 91.5% survival | 8.5% default rate | Cumulative Hazard: 8.9%

Median time to default (for loans that default): 34.6 months

Average time to default (for loans that default): 33.9 months
```

2.4 Fit and Visualize Models with Risk Segmentation

SEGMENTED CREDIT ANALYSIS:

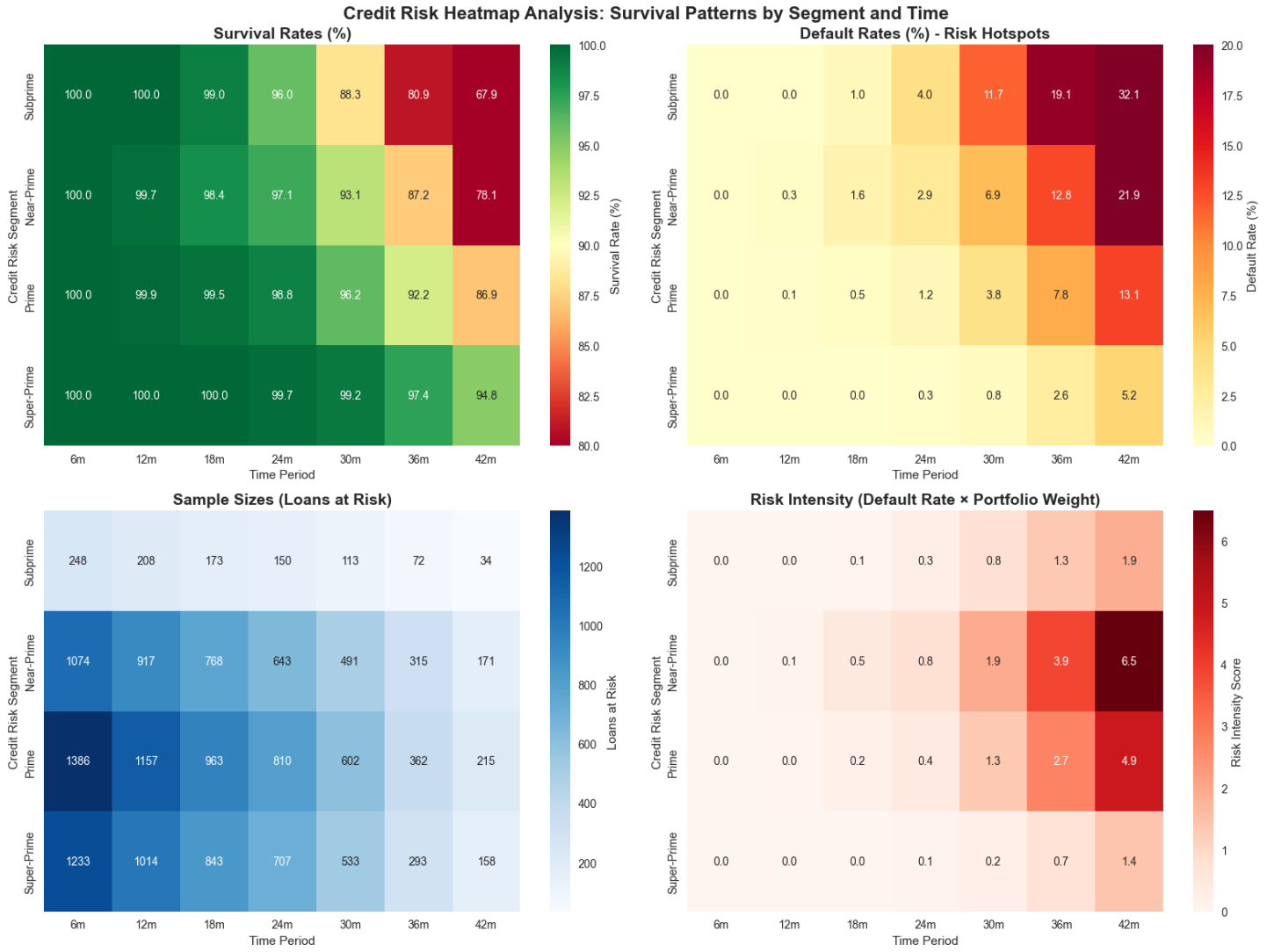
	Risk Segment	Total Loans	Default Loans	Default Rate (%)	Average Duration
0	Subprime	284	41	14.4	24.18
1	Near-Prime	1204	135	11.2	24.97
2	Prime	1560	90	5.8	24.30
3	Super-Prime	1407	24	1.7	23.47

SEGMENTED SURVIVAL RATES:



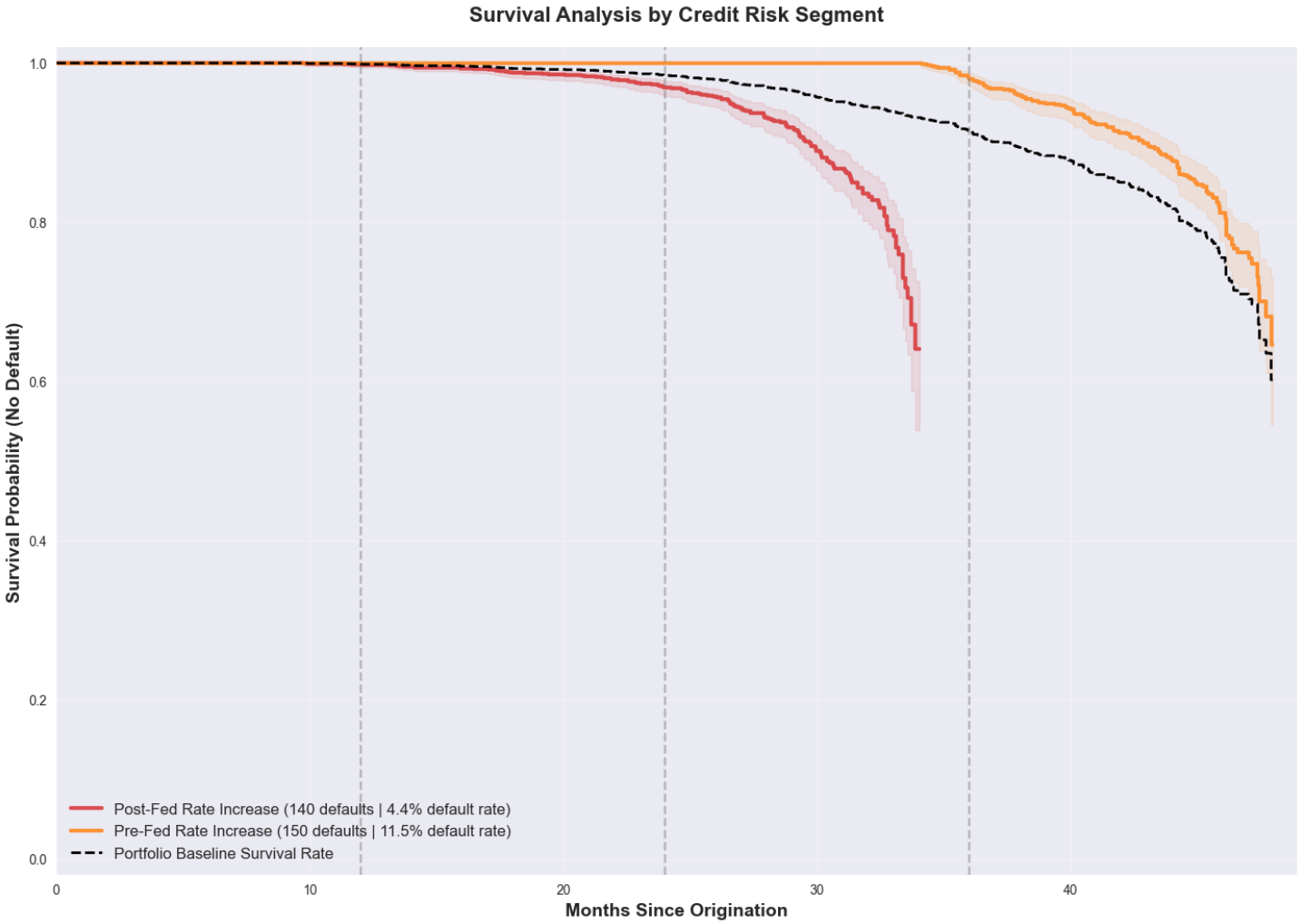
Commentary

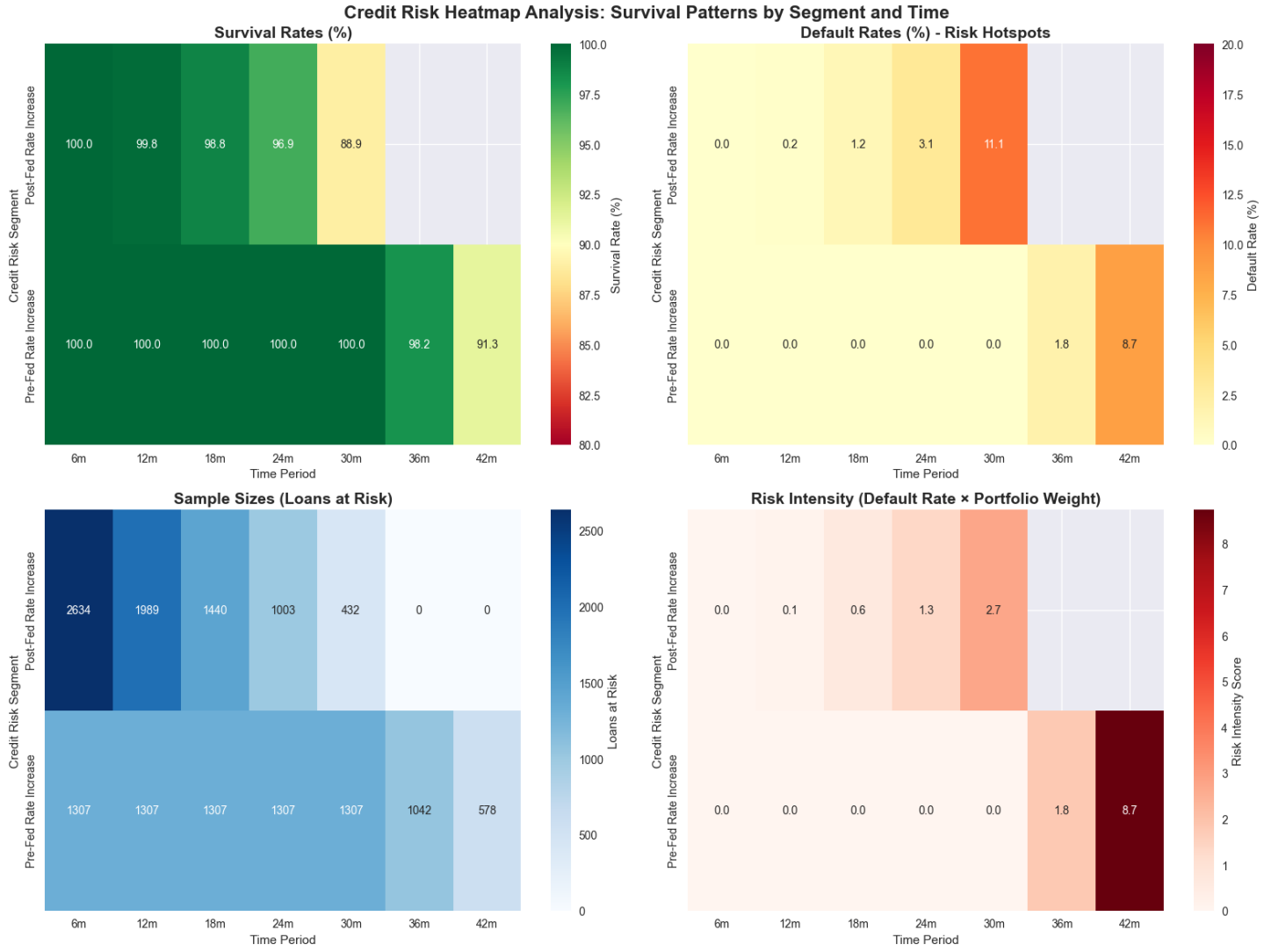
CREDIT RISK HEATMAP ANALYSIS:



Commentary

2.5 Fit and Visualize Model with Interest Rate Period Segmentation


Survival Analysis by Credit Risk Segment
Survival Probability (No Default)
1.0
0.8
0.6
0.4
0.2
0.0
0
10
20
30
40
Months Since Origination
Post-Fed Rate Increase (140 defaults | 4.4% default rate)
Pre-Fed Rate Increase (150 defaults | 11.5% default rate)
Portfolio Baseline Survival Rate

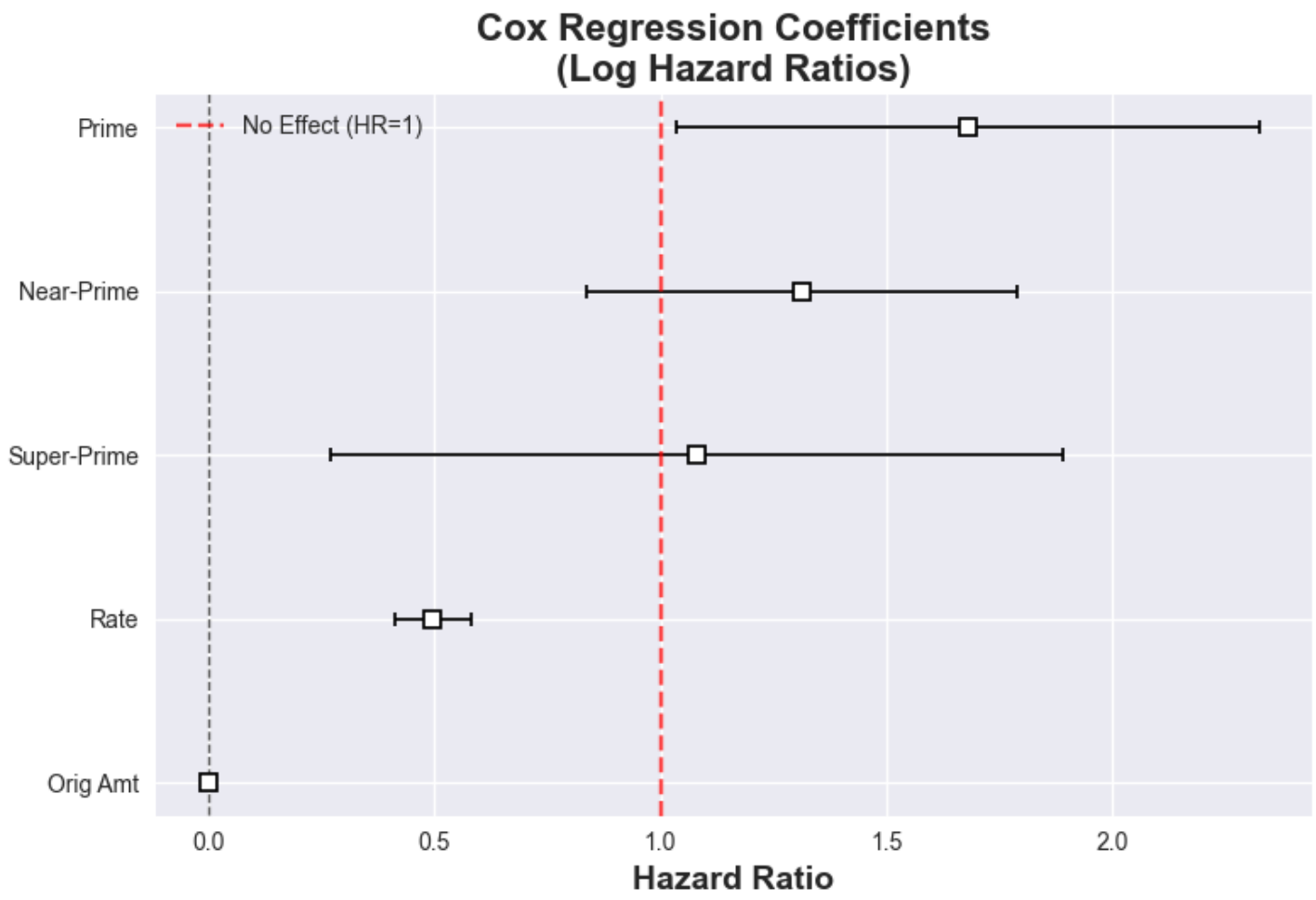


3.0 Cox Proportional Hazards Analysis

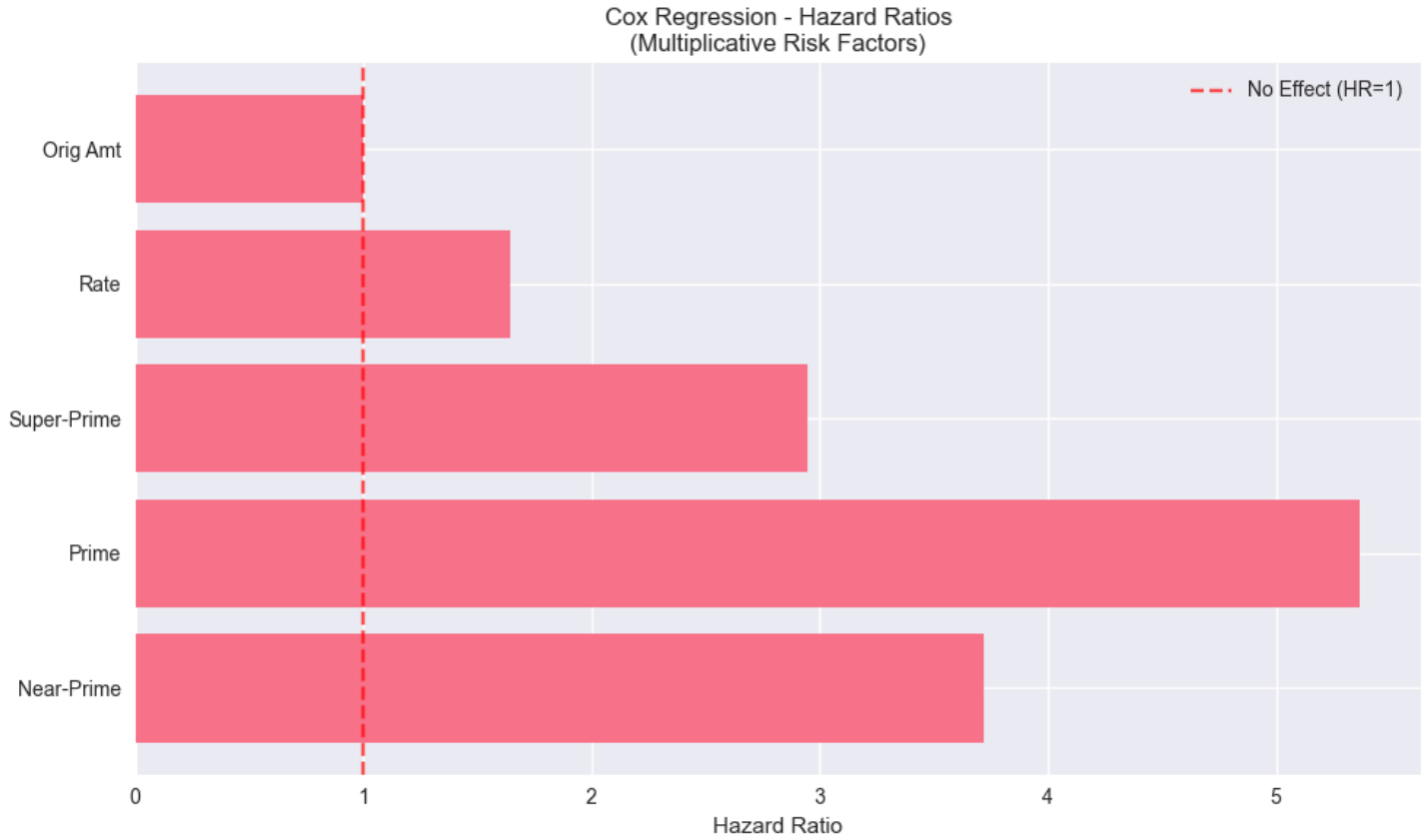
Commentry and approach

3.1 Fit and Assess the COX Model

COX REGRESSION COEFFICIENTS:



Interpretation: Positive coefficients indicate increased risk, negative coefficients indicate decreas



Discuss the paradox - subprime has highest default probability, but 1.00x hazard ratio due to a timing effect.

4.0 Risk Management Applications