



# 计算金融与仿真

## 课程论文

论文题目：计算金融与仿真

学生姓名：李晶晶 张璐 朱冯婧

指导老师：邓志斌

## 1. 投资组合选择介绍

## 2. 模型建立

### 决策变量

$$x_i \in \{0, 1\}, \quad i = 1, 2, \dots, N$$

其中,

- $x_i = 1$  表示选择资助第  $i$  个贷款对象;
- $x_i = 0$  表示不选择。

### 参数说明

- $A_i$ : 第  $i$  个贷款的金额;
- $r_i$ : 第  $i$  个贷款的利率;
- $P_i$ : 第  $i$  个贷款的违约概率;
- $B$ : 总投资预算;
- $R_{\max}$ : 允许的最大违约风险;
- $G_k$ : 第  $k$  个信用等级的贷款集合;
- $\alpha_k$ : 第  $k$  个信用等级的最大投资比例;
- $m$ : 最多选择的贷款数量 (Top- $m$ )。

### 模型形式

$$\begin{aligned}
 & \max_{x_i \in \{0, 1\}} \sum_{i=1}^N x_i A_i [r_i (1 - P_i) - P_i] \\
 & \text{s.t.} \quad \sum_{i=1}^N x_i A_i \leq B \quad (\text{预算限制}) \\
 & \quad \sum_{i \in G_k} x_i A_i \leq \alpha_k B, \quad \forall k \quad (\text{信用等级比例约束}) \\
 & \quad \sum_{i=1}^N x_i A_i P_i \leq R_{\max} \quad (\text{风险控制}) \\
 & \quad \sum_{i=1}^N x_i \leq m \quad (\text{Top-}m \text{ 选择})
 \end{aligned} \tag{1}$$

$$\begin{aligned}
 & \max_{x_i \in \{0,1\}} \sum_{i=1}^N x_i A_i [r_i(1 - P_i) - P_i] \\
 & \text{s.t.} \quad \sum_{i=1}^N x_i A_i \leq B \quad (\text{预算限制}) \\
 & \quad \sum_{i \in G_k} x_i A_i \leq \alpha_k B, \quad \forall k \quad (\text{信用等级比例约束}) \\
 & \quad \sum_{i=1}^N x_i A_i P_i \leq R_{\max} \quad (\text{期望风险控制}) \\
 & \quad \sum_{i=1}^N x_i \leq m \quad (\text{Top-}m \text{ 选择}) \\
 & \quad \sum_{i=1}^N x_i A_i \cdot \tilde{P}_i^{(s)} - \eta \leq \mathcal{M} z_s, \quad \forall s = 1, \dots, S \quad (\text{VaR 限制}) \\
 & \quad \sum_{s=1}^S z_s \leq (1 - \beta) S \\
 & \quad z_s \in \{0, 1\}, \quad \eta \in \mathbb{R}
 \end{aligned} \tag{2}$$

$$\begin{aligned}
 & \max_{x_i \in \{0,1\}} \sum_{i=1}^N x_i A_i [r_i(1 - P_i) - P_i] \\
 & \text{s.t.} \quad \sum_{i=1}^N x_i A_i \leq B \quad (\text{预算限制}) \\
 & \quad \sum_{i \in G_k} x_i A_i \leq \alpha_k B, \quad \forall k \quad (\text{信用等级比例约束}) \\
 & \quad \sum_{i=1}^N x_i A_i P_i \leq R_{\max} \quad (\text{期望风险控制}) \\
 & \quad \sum_{i=1}^N x_i \leq m \quad (\text{Top-}m \text{ 选择}) \\
 & \quad \xi_s \geq \sum_{i=1}^N x_i A_i \cdot \tilde{P}_i^{(s)} - \eta, \quad \forall s = 1, \dots, S \quad (\text{场景损失}) \\
 & \quad \eta + \frac{1}{S(1 - \beta)} \sum_{s=1}^S \xi_s \leq \text{CVaR}_{\max} \quad (\text{CVaR 限制}) \\
 & \quad \xi_s \geq 0, \quad \eta \in \mathbb{R}
 \end{aligned} \tag{3}$$

### 3. 算法设计

---

**Algorithm 1:** 启发式算法：贷款组合优化（含 CVaR 控制）
 

---

**Input:** 贷款数据  $\{A_i, r_i, P_i\}_{i=1}^N$ ; 预算  $B$ ; Top- $m$  限制; CVaR 上限  $\text{CVaR}_{\max}$ ; 置信水平  $\beta$ ; 模拟场景矩阵  $\tilde{P}_i^{(s)} \in \{0, 1\}^{S \times N}$

**Output:** 最优选择向量  $x^* \in \{0, 1\}^N$

```

1  $x_{\text{best}} \leftarrow 0$ ; // 初始为空解
2 根据评分  $\text{score}_i = A_i[r_i(1 - P_i) - P_i]$  降序排列贷款
3  $\text{total\_budget} \leftarrow 0, \text{total\_selected} \leftarrow 0, x_{\text{current}} \leftarrow 0$ 
4 foreach  $i$  in 排序后的贷款列表 do
5     if  $\text{total\_budget} + A_i > B$  或  $\text{total\_selected} + 1 > m$  then
6         continue;
7      $x_{\text{current}}[i] \leftarrow 1$ ;
8      $\text{total\_budget} \leftarrow \text{total\_budget} + A_i$ ;
9      $\text{total\_selected} \leftarrow \text{total\_selected} + 1$ ;
10 if  $\text{Feasible}(x_{\text{current}})$  then
11      $x_{\text{best}} \leftarrow x_{\text{current}}$ ;
12 Function  $\text{Feasible}(x)$ :
13     for  $s \leftarrow 1$  to  $S$  do
14          $L_s \leftarrow \sum_{i=1}^N x_i A_i \cdot \tilde{P}_i^{(s)}$ ;
15      $\eta \leftarrow \beta$  分位点的  $\{L_s\}$ ;
16     for  $s \leftarrow 1$  to  $S$  do
17          $\xi_s \leftarrow \max(L_s - \eta, 0)$ ;
18      $\text{CVaR}_\beta(x) \leftarrow \eta + \frac{1}{S(1-\beta)} \sum_{s=1}^S \xi_s$ ;
19     return 是否满足  $\text{CVaR}_\beta(x) \leq \text{CVaR}_{\max}$  且满足其他约束;
20 return  $x_{\text{best}}$ ;
    
```

---

### 4. 案例研究

#### 4.1. 数据集描述

本研究使用的数据集来自 Lending Club 平台，原始数据由 Kaggle 网站公开提供<sup>1</sup>。Lending Club 是美国最大的网络借贷平台之一，提供了详尽的个人借款申请及其还款情况的数据，广泛应用于学术界和工业界进行信贷风险评估、违约预测及投资组合优化等研究。

该数据集包含了从 2007 年至 2018 年的借款记录，共计数百万条样本。每条记录对应一笔贷款申请，涵盖了包括贷款金额、利率、借款人信用等级、债务收入比、贷款期限、还款状态、就业年限、收入、地址状态、房屋所有权、FICO 评分区间等在内的多维度信息。

在本研究中，我们主要筛选并保留以下变量用于建模分析：

- **loan\_amnt:** 借款人申请的贷款金额，作为  $A_i$ ;

- **int\_rate**: 借款合同中约定的年利率，用于计算收益率  $r_i$ ;
- **grade**: 借款人的信用等级（A 至 G），用于分组限制;
- **loan\_status**: 实际贷款的还款状态（如 Fully Paid、Charged Off），用于推断违约情况;
- **annual\_inc**: 借款人年收入;
- **dti**: 债务收入比，用于辅助风险刻画;
- **term**: 贷款期限（如 36 months 或 60 months）;
- **emp\_length**: 借款人工作年限;
- **addr\_state**: 借款人所在州;
- **fico\_range\_high, fico\_range\_low**: 借款人 FICO 信用评分区间;

为了满足模型中对违约概率  $P_i$  的需求，我们将“loan\_status”字段中状态为“Charged Off”的贷款视为违约样本，其余如“Fully Paid”、“Current”等状态作为非违约样本，并基于历史频率法估算每一类贷款的违约概率。

此外，为模拟贷款违约的风险场景，我们以借款人的信用等级、FICO 评分和历史违约频率为依据，构建了  $S$  个 Monte Carlo 风险场景，用于后续 CVaR 优化模型的风险评估。

通过上述数据处理步骤，最终形成了一个结构规范、信息完备、适用于组合优化问题的数据集，为后续实证分析与建模提供了坚实的数据基础。

## 参考文献

- [1] 彭桥, 肖尧, 杨宇茜, and 杨沛瑾. 中国新质生产力发展水平测度、动态演化与驱动因素研究. 软科学, 39(04):25–34, 2025. ISSN 1001-8409. doi: 10.13956/j.ss.1001-8409.2025.04.05. URL <https://link.cnki.net/urlid/51.1268.g3.20241125.1056.010>.
- [2] 王珏 and 王荣基. 新质生产力: 指标构建与时空演进. 西安财经大学学报, 37(1):31–47, 2024. doi: 10.19331/j.cnki.jxufe.20231124.001.
- [3] 李光勤 and 李梦娇. 中国省域新质生产力水平评价、空间格局及其演化特征. 经济地理, 44(8): 116–125, 2024. doi: 10.15957/j.cnki.jjdl.2024.08.014.
- [4] 曹东勃 and 蔡煜. 新质生产力指标体系构建研究. 上海财经大学学报, (4):50–62, 2024.
- [5] 张海, 王震, and 李秉远. 新质生产力发展水平、空间差异及动态演进. 统计与决策, (24):11–26, 2024. doi: 10.13546/j.cnki.tjyj.2024.24.002.
- [6] 胡佳霖 and 徐俊. 中国新质生产力: 区域差距、动态演进与跃迁趋势. 统计与决策, (21):5–16, 2024. doi: 10.13546/j.cnki.tjyj.2024.21.001.
- [7] 简新华 and 聂长飞. 中国新质生产力水平测度及省际现状的比较分析. 经济问题探索, (10): 3–20, 2024.
- [8] 冉戎, 花磊, 陈烨靖, and 夏艺嘉. 新质生产力发展潜力测度、时空差异及战略着力点研究. 重庆大学学报(社会科学版), 2024. doi: 10.11835/j.issn.1008-5831.jg.2024.10.002. URL <https://link.cnki.net/urlid/50.1023.C.20241030.1300.006>. 网络首发.
- [9] 颜克高, 王馨悦, and 吴心怡. 中国新质生产力发展的水平测度与区域差异研究. 湖南大学学报(社会科学版), 39(1):10–21, 2025. doi: 10.16339/j.cnki.hdxbskb.2025.01.002.
- [10] 徐波, 王兆萍, 余乐山, and 刘柯. 新质生产力对资源配置效率的影响效应研究. 产业经济评论, (4):35–49, 2024. doi: 10.19313/j.cnki.cn10-1223/f.20240417.001.
- [11] 杨智晨, 涂先青, and 王方方. 我国新质生产力发展的理论基础、时空特征及分异机理. 经济问题探索, (1):50–66, 2025.
- [12] 马大晋, 吴旭辉, and 张博文. 中国新质生产力发展水平区域评价与空间关联网络特征. 财经论丛, pages 1–14, 2025. doi: 10.13762/j.cnki.cjlc.20250103.001. URL <https://doi.org/10.13762/j.cnki.cjlc.20250103.001>.
- [13] 程赛楠 and 冯珍. 数实融合对新质生产力的影响研究. 北京理工大学学报(社会科学版), 26(6): 15–27, 2024. doi: 10.15918/j.jbitss1009-3370.2024.1491.
- [14] 张龙, 申瑛琦, and 张伟琦. 新质生产力的原创价值、统计测度与培育方向. 暨南学报(哲学社会科学版), (11):126–144, 2024. doi: 10.11778/j.jnxb.20240975.
- [15] Younes Ataei, Amin Mahmoudi, Mohammad Reza Feylizadeh, and Deng-Feng Li. Ordinal Priority Approach (OPA) in Multiple Attribute Decision-Making. *APPLIED SOFT COMPUTING*, 86, January 2020. ISSN 1568-4946. doi: 10.1016/j.asoc.2019.105893.