

# 基于改良后的 SEIR 模型对我国经济预测分析

## 摘要

中小微企业，作为中国经济的“毛细血管”，为我国贡献了一半以上的税收、近六成的国民生产总值与超过八成的城镇劳动就业，有着举足轻重的作用。然而，由于企业规模小、抵押资产少等特有性质，中小微企业融资难已成为困扰众多经营者的问题。商业银行面向中小微企业提供间接融资时，在控制信贷风险的前提下，如何精准判断企业信贷等级、确定放贷额度与利率等亦是困扰银行管理者所在。

针对该现象，我们通过三个模型进行数据分析，分别是多元逻辑回归模型、基于逻辑回归模型优化的 FAHP 模型与基于 LASSO 选择的支持向量回归模型。最终，我们选择准确率最高的模型三——支持向量回归模型进行后续的分析与解答。

对于问题二，我们基于题目中对于放贷额度与利率的限制外，加入了对企业流水与营业收入的限制，使得问题更加贴近实际。

最后，我们用新冠疫情代表突发情况、用行业指数代表行业或者个体面对突发情况的风险来确定企业信贷等级的变化。最终使用前述最优模型进行求解，得到最优的放贷策略。

**关键字：** 归一化 LASSO 逻辑回归 模糊层次分析法 计算机仿真 微分方程模型

# 目录

一、 问题重述 .....	3
1.1 问题背景 .....	3
1.2 问题提出 .....	3
二、 问题分析 .....	3
2.1 问题一 .....	3
2.2 问题二 .....	4
2.3 问题三 .....	4
三、 基本假设 .....	4
四、 符号说明 .....	4
五、 模型的建立、检验与选择 .....	5
5.1 确定指标 .....	5
5.2 选择模型 .....	7
5.2.1 模型一：多元逻辑回归模型 .....	7
5.2.2 模型二：基于逻辑回归模型优化的 FAHP 模型 .....	9
5.2.3 模型三：基于 LASSO 选择的支持向量回归模型 .....	11
六、 问题求解 .....	12
6.1 问题一求解 .....	12
6.2 问题二求解 .....	13
6.3 问题三求解 .....	14
6.3.1 计算机仿真模型 .....	17
6.3.2 通过股票指数反应突发因素对企业发展状况影响 .....	18
七、 模型的评价 .....	18
7.1 模型优点 .....	18
7.2 模型缺点 .....	18
7.3 模型缺点 .....	19
7.4 模型推广 .....	19

# 一、问题重述

## 1.1 问题背景

中小微企业在我国经济中发挥中举足轻重的作用，具有数量多、活力强、调整灵活等特点。但是，在实际中，由于此类企业规模相对较小，同时缺少抵押资产，因此银行通常是依据信贷政策、企业的交易票据信息和上下游企业的影响力，向实力强、供求关系稳定的企业提供贷款，并可以对信誉高、信贷风险小的企业给予利率优惠。银行首先需要搭建一套准确的风险评价模型，根据中小微企业的实力、信誉对其信贷风险做出评估，然后依据信贷风险等因素来确定是否放贷及贷款额度、利率和期限等信贷策略。此外，面对 2020 年年初发生的新冠疫情，我们在此类风险模型中加入了对于突发因素与极端情况的考虑，使得模型的适用范围更广阔。

## 1.2 问题提出

某银行对确定要进行放贷企业的额度控制在 10 万至 100 万元之间，年利率为 4% 至 15%，贷款期限为 1 年。该银行需要根据 123 家有具体信贷记录企业与 302 家无信贷记录企业相关数据，同时考虑贷款利率与客户流失率关系，通过建立数学模型研究并形成对中小微企业的信贷策略，主要解决下列问题：

1. 对附有信贷记录的 123 家企业的风险程度进行量化分析，给出该银行在年度信贷总额固定时对这些企业的信贷策略；
2. 基于上述问题，对 302 家无信贷记录企业的信贷风险进行量化分析，并给出该银行在年度信贷总额为 1 亿元时对这些企业的信贷策略；
3. 基于上述条件与解决策略，考虑小概率的突发因素对不同行业、不同类别的企业的影响，进而综合考虑各企业的信贷风险后给出该银行在年度信贷总额为 1 亿元时的信贷调整策略。

# 二、问题分析

## 2.1 问题一

问题一要求基于 123 家有信贷记录的公司相关数据进行量化分析，给出总额固定时的贷款方案。首先，我们将 123 家公司随机分成两部分——含有 100 家公司的第一部分作为量化分析模型的训练集，使用训练集进行建模与测试。剩余 23 家作为检验集，代入建立的模型观察预测效果。由于其财务数据较多，其过程中考虑使用不同的模型来拟合以求出最优模型。

对于给定的贷款总额，我们将利用附件中给出信息，平衡流失率与贷款年利率之间的得失，求出银行的最优策略。

## 2.2 问题二

基于问题一求解出的最优模型，我们将代入剩余三百家公司的各项指标进行回归判断。此时，由于模型并非导出离散型结果，我们将会使用线性加和的方法对银行的预期收益率进行调整，从而得到调整后的最优放贷策略。

## 2.3 问题三

# 三、基本假设

- 假设该银行所处市场为有效市场且银行属于纯粹的商业银行，不考虑政府各类政策与补贴的影响；
- 假设企业票据信息无造假且体现了企业于该时期所有的资金流情况；
- 假设

# 四、符号说明

符号	定义
$I$	Industry, 行业信誉均值或赋分
$T$	Tendency, 企业盈利发展趋势
$A$	Addition, 企业产品附加值
$U$	Unavailable, 作废发票
$Z$	Money, 资金
$B_{bs}$	Buy-side bargain ability, 买方议价能力
$B_{ss}$	Sell-side bargain ability, 卖方议价能力
$R_{in}$	Real in, 企业净购进率
$R_{out}$	Real out, 企业净销出率

## 五、模型的建立、检验与选择

### 5.1 确定指标

我们通过分析与挖掘企业相关的大量数据，从企业经营实力与企业在产业供应链中影响力等方面，计算并挑选出 8 项能够识别企业信贷风险的指标：行业信誉度（信贷风险大小程度） $I$ ，企业盈利增长趋势  $T$ ，发票作废率  $U$ ，企业产品增值度  $A$ ，企业作为买方时议价能力  $B_{bs}$ ，企业作为卖方时议价能力  $B_{ss}$ ，企业净进率  $R_{in}$  和企业净销率  $R_{out}$ 。

我们在分析指标时发现，不同的指标往往有不同的量纲，所以我们由公式：

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

将所有的指标的初始值或者运算值进行归一化处理，作为最终代入模型进行回归的数值。

#### • 行业信誉度 $I$

根据 2017 年 6 月 30 日由中华人民共和国国家质量监督检验检疫总局与中国国家标准化委员会联合发布的《国民经济行业分类》，我们将所有企业（附件 1 与附件 2 中共计 425 家企业）划分为 11 类企业，部分企业划分见下表：

表 1 11 类行业划分

公司代码	公司名称	所在行业
E179	*** 园林有限公司	农、林、牧、渔
E394	*** 环保设计研究院	科学研究和技术服务业
E21	*** 建设工程有限公司	建筑业
E60	*** 机械租赁有限公司	租赁和商业服务业
E57	*** 机械设备有限公司	制造业
E65	*** 商贸有限公司	批发与零售业
E14	个体经营 E14	个体经营
E3	*** 电子有限公司	信息传输、软件和信息技术服务业
E46	*** 广告传媒有限公司	文化、体育和娱乐业
E61	*** 调味品有限公司	住宿和餐饮业
E83	*** 社会福利院	其他

分类后，为 A、B、C、D 四个等级赋分为 90、70、50 和 30 分。通过计算行业内包含各个等级公司的分数进行求平均值作为行业信贷等级，再进行归一化处理，用  $I$  代表行业的信誉程度。

- 企业净进率  $R_{in}$  和企业净销率  $R_{out}$

在企业给出的票据信息中，有效发票中包含两类：一类是正数的进项、销项发票，一类是负数的、因购方退款而产生的退货退款的发票。我们在判别企业真正的进货量与销售量时需要辨别清真实的数量，所以需要将正数值与负数值加和，除以正数有效发票之和。

$$\begin{aligned} R_{in} &= \frac{\sum \text{正数有效发票购入额} + \sum \text{负数有效发票购入额}}{\sum \text{正数有效购入额}} \\ R_{out} &= \frac{\sum \text{正数有效发票销售额} + \sum \text{负数有效发票销售额}}{\sum \text{正数有效购入额}} \end{aligned} \quad (2)$$

- 盈利增长趋势  $T$

我们将有效发票中正、负数相加后，得到了企业净销售额度和企业净购入额度（税前），求其差值作为企业挣得利润。通过计算，我们可以算出企业披露各个财年利润的增长率。我们运用几何平均法求出过往数年企业的利润平均增长率。结合利润增长率的方差  $\bar{r}$ ，我们对标准差  $\sigma$  进行归一化处理后得到  $\bar{r}'$  与  $\sigma'$ ，将其二者进行作差，再进行归一化处理后得到企业盈利增长趋势指标  $T$ 。

$$T' = \bar{r}' - \sigma' \quad (3)$$

- 作废率  $U$

我们观察到企业票据中除了有效票据外还存在着作废发票，其为在为交易活动开具发票后，因故取消了该项交易，使发票作废。与净销率和净进率类似，我们认为一家企业频繁取消或者被取消交易意味着其产品市场认可度低，公司信贷风险升高。

$$U = \frac{\sum \text{作废发票额度}}{\sum \text{作废发票额度} + \sum \text{有效发票额度}} \quad (4)$$

- 增值额  $A$

我国增值税是以商品在流转过程中产生的增值额作为计税依据而征收的一种流转税。因此，计算一家企业的增值税额度能够综合度量其产品增值能力与生产能力。我们将企业销项总税额减去进项总税额后经过归一化得到增值额  $A$ 。

- 买方议价力  $B_{bs}$ 、卖方议价力  $B_{ss}$

企业票据上登记有贸易公司的购方或者销方代码，我们对每一家企业的贸易往来对象进行统计分析，按照我方销售量与购买量降序排列，选取往来金额最大的前 10% 的上、下游企业，统计其 10% 客户或商家总量占总贸易量的百分比  $B'_{bs}$ ， $B'_{ss}$ ，再进行归一化处理得到  $B_{bs}$ ， $B_{ss}$ 。

$$B'_{bs} = \frac{\sum \text{前 10\% 供货商供应额}}{\sum \text{供应商供应额}} \quad B'_{ss} = \frac{\sum \text{前 10\% 客户购买量}}{\sum \text{客户购买量}} \quad (5)$$

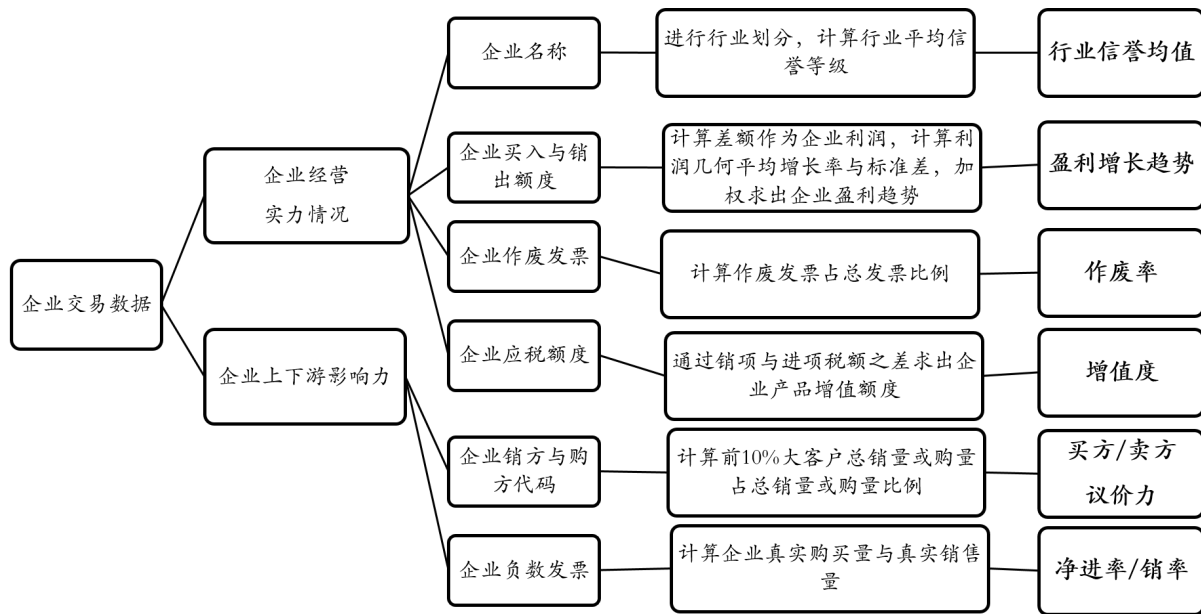


图 1 企业相关信息的挖掘与模型指标定义

## 5.2 选择模型

基于上述分析挖掘的八项指标，我们将运用三种模型：多元逻辑回归模型、基于逻辑回归优化的 FAHP 模型和基于 LASSO 选择的支持向量回归模型对附件 1 中 123 家企业的相关信息进行分析。我们将随机抽取 100 家企业信息进行分析建模，最后使用 23 家企业信息进行检验，筛选匹配率最高的模型进行进一步分析与运用。

### 5.2.1 模型一：多元逻辑回归模型

多元逻辑回归模型，又名多元对数几率回归模型，能够更深一层地拟合属性变量之间的函数关系，以描述变量之间的互相影响。它利用列联表中优势的对数，当概率在 0 ~ 1 取值时逻辑函数可以取任意实数，因而避免了线性概率模型的结构缺陷 [1]。

假设因变量  $y$  是离散型变量且能够取值为 1,2,3... $k$  ( $k>2$ )，则有其结果概率分布为：

$$P(Y = 1) = p_1, P(Y = 2) = p_2, \dots, P(Y = k) = p_k \quad \left( \sum_{j=1}^k p_j = 1 \right) \quad (6)$$

假设我们有自变量  $x_1, x_2, \dots, x_n$ ，且我们希望有  $P(y = j) = p_j$  为含  $x_i$  的方程。现在我们选取一个变量  $x_1$  作为基准变量，有：

$$\log\left(\frac{p_{ij}}{p_{1j}}\right) = \beta_{0j} + \beta_{1j}x_1 + \beta_{2j}x_2 + \dots\beta_{nj}x_n \quad (7)$$

因此，如果结果一共有  $k$  种可能，那么该多元逻辑回归模型会有  $k - 1$  条方程式。

参考附件一中的企业信贷等级划分共有 A、B、C 和 D 四种等级，我们将得到三条方程式，并且每条方程均可以求出与 D 等级的发生概率之比，多元逻辑线性回归恰好

能给出每一种等级发生的可能性，需求贴合度高。

运用 R Studio 软件，我们将归一化处理后的各项指标数据导入，首先建立初始模型，使用 R “Stats” 包中 “step” 函数进行逐步回归分析，求得最佳模型。由模型上述

**表 2 最佳模型变量回归系数**

等级	(Intercept)	A	I	$B_{bs}$	U
C	-5.0374	0.14965	0.0205	0.0598	-0.0295
B	-4.6545	0.1733	0.0342	0.0493	-0.0602
A	-5.5024	0.1869	0.0402	0.0474	-0.0638

**表 3 最佳模型变量标准误**

等级	(Intercept)	A	I	$B_{bs}$	U
C	1.7524	0.1076	0.0161	0.0183	0.0211
B	1.7076	0.1077	0.0171	0.0174	0.0307
A	1.9339	0.1073	0.0182	0.0195	0.03852

数据算出模型各个指标 p 值均小于 0.05，满足 95% 的显著性检验。求得模型最终方程为：

$$\begin{cases} \ln\left(\frac{P(y=C)}{P(y=D)}\right) = -5.0374 + 0.14965A + 0.0205I + 0.0598B_{bs} - 0.0295U \\ \ln\left(\frac{P(y=B)}{P(y=D)}\right) = -4.6545 + 0.1733A + 0.0342I + 0.0493B_{bs} - 0.0602U \\ \ln\left(\frac{P(y=A)}{P(y=D)}\right) = -5.5024 + 0.1869A + 0.0402I + 0.0474B_{bs} - 0.0638U \end{cases} \quad (8)$$

接下来我们将该模型运用于剩余企业数据中进行训练预测，由于附件 1 中企业等级已知，所以我们将模型对企业在其真实等级的概率进行加和，除以企业数目，计算得到模型一的准确率：

**表 4 最佳模型预测效果（部分）**

序号	A	B	C	D	真实等级
E1	0.74	0.10	0.12	0.04	A
E2	0.69	0.20	0.06	0.05	A
E3	0.19	0.35	0.30	0.16	C
E4	0.26	0.28	0.30	0.16	C



$$\theta_1 = \frac{\sum P_{\text{真实等级对应概率}}}{\text{测试企业数目}} = 43.67\% \quad (9)$$

### 5.2.2 模型二：基于逻辑回归模型优化的 FAHP 模型

美国运筹学家 A.L.Saaty 教授于 20 世纪 70 年代提出层次分析法，是一种定性分析和定量分析相结合的系统分析方法 [2]。层次分析法通过明确问题, 建立层次分析结构模型, 构造判断矩阵, 层次单排序和层次总排序五个步骤计算各层次构成要素对于总目标的组合权重, 从而得出不同可行方案的综合评价值, 为选择最优方案提供依据。AHP 的关键环节是建立判断矩阵, 判断矩阵是否科学、合理直接影响到 AHP 的效果, 通过分析, 我们发现其有若干缺点:

1. 当判断矩阵不具有 consistency 时需要调整判断矩阵的元素, 使其具有 consistency, 这有可能需要经过若干次调整、检验、再调整、再检验的过程才能使判断矩阵具有 consistency;
  2. 检验判断矩阵是否具有 consistency 的判断标准:  $CR < 0.1$  缺乏科学依据, 缺少数据的支撑。
- 为了解决上述问题, 我们引进了模糊一致矩阵的概念。下面先介绍模糊一致矩阵的定义及其性质。

模糊一致矩阵有如下概念:

**定理一** 设矩阵:  $R = (r_{ij})_{n \times n}$  若能够满足:  $0 \leq r_{ij} \leq 1, (i = 1, 2, \dots, n; j = 1, 2, \dots, n)$  则称  $R$  是模糊矩阵。

**定理二** 若模糊矩阵  $R = (r_{ij})_{n \times n}$  满足:  $r_{ij} + r_{ji} = 1, (i = 1, 2, \dots, n; j = 1, 2, \dots, n)$  则称模糊矩阵  $R$  是模糊互补矩阵。

**定理三** 若模糊互补矩阵  $R = (r_{ij})_{n \times n}$  满足任意  $i, j, k$ :  $r_{ij} = r_{ik} - r_{jk} + 0.5$  则称模糊矩阵  $R$  是模糊一致矩阵。

**定理四** 模糊一致矩阵  $R = (r_{ij})_{n \times n}$  是模糊一致矩阵的充要条件是任意指定两行的对应元素之差为常数。

综上所述, 以及模糊一致矩阵的性质知, 用模糊一致矩阵  $R = (r_{ij})_{n \times n}$  表示论域  $U = (a_1, a_2, \dots, a_n)$  上的模糊关系 “...比...重要得多” 是合理的。由此, 我们可采用如下的 0.1 至 0.9 标度给予数量标度。

标度	定义	说明
0.5	同等重要	两元素相比, 同等重要
0.6	稍微重要	两元素相比, 一元素比另一元素稍微重要
0.7	明显重要	两元素相比, 一元素比另一元素明显重要
0.8	重要得多	两元素相比, 一元素比另一元素重要得多
0.9	极端重要	两元素相比, 一元素比另一元素极端重要
0.1、0.2、0.3、0.4 反比较 若元素与另一元素比较得到判断 $r$ , 则两元素相比较得到判断为 $r(ij) = 1 - r(ji)$		

由上方表格标度，我们得出图二中的模糊互补矩阵。

但是，由于我们得到的是模糊互补矩阵，需要调整为模糊一致矩阵。调整方法如下 [2]:

1. 确定一个同其余元素的重要性相比较得出的判断有把握的元素，不失一般性，设决策者认为对判断  $r_{11}, r_{12}, r_{13} \dots r_{1n}$  比较有把握;
2. 用 R 的第一行元素减去第二行对应元素，若所得的 n 个差为常数，则不需调整第二行元素。否则，对其调整。
3. 同理，用 R 的第一行元素减去第三行的对应元素，若所得的 n 个差为常数，则不需调整第三行元素。否则，对其调整。
4. 同理，用 R 的第一行元素减去第 k(k=2,3,...,n) 行的对应元素，若所得的 n 个差为常数，则不需调整第 k 行元素。否则，对其调整。

同时，经过模型一多元逻辑回归模型的搭建，我们发现在进行二元逻辑回归时有一些变量与模型结果呈现较强的相关性。于是，在我们经过共计 28 次两两搭配进行逻辑回归后，我们对于显著性较强的模型，选取其两个自变量回归系数，其比值作为模糊一致矩阵上确定参数的有力参考。最终，我们得到模糊一致矩阵，见图 3。

	1	2	3	4	5	6	7	8
1	0.5	0.3	0.3	0.4	0.6	0.8	0.3	0.3
2	0.7	0.5	0.7	0.7	0.8	0.9	0.7	0.6
3	0.7	0.3	0.5	0.6	0.7	0.9	0.6	0.4
4	0.6	0.3	0.4	0.5	0.7	0.9	0.4	0.3
5	0.4	0.2	0.3	0.3	0.5	0.8	0.3	0.2
6	0.2	0.1	0.1	0.1	0.2	0.5	0.3	0.1
7	0.7	0.3	0.4	0.6	0.7	0.7	0.5	0.4
8	0.7	0.4	0.6	0.7	0.8	0.9	0.6	0.5

图 2 模糊互补矩阵

	1	2	3	4	5	6	7	8
1	0.5	0.3	0.3	0.4	0.6	0.8	0.3	0.3
2	0.7	0.5	0.5	0.6	0.8	1	0.5	0.5
3	0.7	0.5	0.5	0.6	0.8	1	0.5	0.5
4	0.6	0.4	0.4	0.5	0.7	0.9	0.4	0.4
5	0.4	0.2	0.2	0.3	0.5	0.7	0.2	0.2
6	0.2	0	0	0.1	0.3	0.5	0	0
7	0.7	0.5	0.5	0.6	0.8	1	0.5	0.5
8	0.7	0.5	0.5	0.6	0.8	1	0.5	0.5

图 3 模糊一致矩阵

最终，将模糊一致矩阵元素代入下列公式：

$$W = \left[ \frac{\sum_{j=1}^n r_{1j}}{\sum_{i=1}^n \sum_{j=1}^n r_{ij}}, \frac{\sum_{j=1}^n r_{2j}}{\sum_{i=1}^n \sum_{j=1}^n r_{ij}}, \dots, \frac{\sum_{j=1}^n r_{nj}}{\sum_{i=1}^n \sum_{j=1}^n r_{ij}} \right]^T \quad (10)$$

我们计算得到：

表 5 权重向量

w	1	2	3	4	5	6	7	8	总和
权重	10.94%	15.94%	15.94%	13.4%	8.44%	3.44%	15.94%	15.94%	100%

最终，我们将检验集数据代入模型，最终算出各个企业离散的预测信贷分数，按照 80 至 100 分划分为 A 级、60 至 80 分划分为 B 级..... 以此类推，得到此模型预测后

的各企业信贷等级，与真实信息进行对比，得出准确率：

$$\theta_2 = 40.89\% \quad (11)$$

### 5.2.3 模型三：基于 LASSO 选择的支持向量回归模型

Lasso 回归方法属于正则化方法的一种，是以缩小特征集（降价）为思想的一种收缩估计方法。它通过构造一个惩罚函数，得到一个较为精炼的模型，可以将特征系数进行压缩并使某些回归系数变为 0，保留了子集收缩的优点，从而实现特征选择的目的。模型选择本质上寻求模型稀疏表达的过程，而这种过程可以通过优化一个“损失”+“惩罚”的函数问题来完成。

Lasso 参数估计定义如示：

$$\hat{\beta} = \arg \min_{\beta} \|y - \sum_{j=1}^p x_j \beta_j\|^2 + \lambda \sum_{j=1}^p |\beta_j| \quad (12)$$

其中， $\lambda$  为非负正则参数，控制着模型的复杂程度， $\lambda$  越大对特征较多的线性模型的惩罚力度就越大，从而获得一个特征较少的模型， $\lambda \sum_{j=1}^p |\beta_j|$  代表惩罚项。调整参数  $\lambda$  的确定可以采用交叉验证法，选取交叉验证误差最小的值。我们得到参数  $\lambda=0.9885$ 。

最后用全部数据重新拟合模型即可，如下图中表格所示。

Profit tendency	Additional value	Buy-side Bargain ability	Sell-side Bargain ability
0.08967	-0.23883	-0.24154	0.02705
industry credit	Real Buy-in ratio	Real Sell-out ratio	Unavailable receipt
0.23337	-0.44769	0.4945	0.3711

图 4 模型系数表

因此，我们可以看出八个指标之间是非共线性的。所以我们通过 LASSO 模型选取八项特征代入支持向量回归模型。

支持向量回归模型，是采用支持向量的思想，来对数据进行回归分析。给定训练数据集  $T = (x_1, y_1), \dots, (x_n, y_n)$ ，其中  $x_i = (x_i^{(1)}, \dots, x_i^{(n)})^T$ ，对于样本  $(x_1, y_1)$ ，根据模型输出  $f(x_i)$  与真实值  $y_i$  之间差别计算损失。差别最多有  $\varepsilon$  偏差，用公式表达为：

$$\min \frac{1}{2} \|\vec{w}\|_2^2 + C \sum_{i=1}^n L_{\varepsilon}(f(\vec{x}_i) - y_i) \quad (13)$$

其中  $C \geq 0$  为罚项系数， $L$  为损失函数。为进一步，我们引入松弛变量  $a_i$ 、 $b_i$ ，则新的最优化问题可表达为：

$$\min \frac{1}{2} \|\vec{w}\|_2^2 + C \sum_{i=1}^n (a_i - b_i) \quad (14)$$

$$s.t. = \begin{cases} f(\vec{x}) - y_i \leq \varepsilon + a_i \\ y_i - f(\vec{x}) \leq \varepsilon + b_i \\ a_i, b_i \geq 0, i = 1, 2, \dots, n \end{cases} \quad (15)$$

由于支持向量机拥有完善的理论和良好特性，不仅适用于线性模型，也能很好地抓住数据和特征之间非线性关系，可以避免局部极小化问题，提高泛化性能，是可以使电影数据中异常点所引起偏差更小。

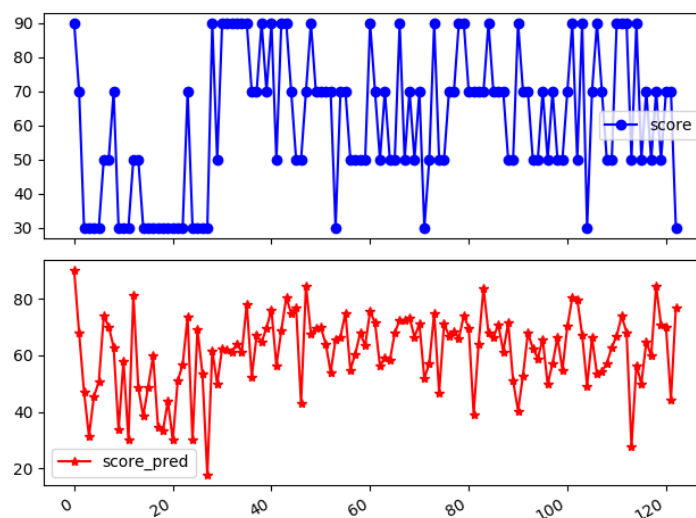


图5 评分真实值与预测值对比图

平均绝对误差	均方误差	R平方值
9.9052	115.27626	0.6678

图6 模型评价指标

综上，我们利用该模型进行检测，测算得准确率为：

$$\theta_3 = 58.19\% \quad (16)$$

由此可见，基于 LASSO 选择的支持向量回归模型是我们三项模型中预测准确度最高的一项。在接下来求解问题二与问题三中，我们会选择使用其对附件 2 中的企业进行信贷等级的预测与划分。

## 六、问题求解

### 6.1 问题一求解

经过上述建模过程，我们选定了模型三作为我们往后的信贷评判模型。根据附件三，我们可以看出任何一种潜在的借贷客户（D 类客户除外）的流失率随着贷款利率

的上升而下降，所以我们通过分析，计算出以流失率与贷款利率为基础的银行收益方程，其中违约率为附件 1 中各信贷等级违约比例：

$$\begin{aligned} Z \times R(i) &= Z \times r_i \times (1 - p_i) \times (1 - f_{ij}) - Z \times p_i \\ R(i) &= r_i \times (1 - p_i) \times (1 - f_{ij}) - p_i \end{aligned} \quad (17)$$

故，通过以上公式，可以得到按照评级和年贷款利率所算出的最大可能收益率并进行排序，最终可以得到针对各个评级，在一特定贷款利率下可以获得最大收益，如下图中表格所示：

	贷款年利率	预期收益率
信誉评级 A	0.0465	0.040188686
信誉评级 B	0.0585	0.013392339
信誉评级 C	0.0585	-0.019742176

图 7 信誉评级最大化收益表

同时考虑到题目中对于贷款额度与贷款利率的限制，我们得出如下结论：

- 在额度一定的请款下，首先满足附件 1 中 A 等级企业的贷款需求，利率定为年利率 4.65%，且单一企业放贷不得超过 100 万且不能低于 10 万；
- 当附件 1 中 A 等级企业所有放款需求都得到满足后，银行才能考虑 B 等级企业的需求。银行给予 B 等级企业的年利率为 5.85%，且单一企业放贷不得超过 100 万且不能低于 10 万；
- 在同时满足所有 A、B 等级企业后，由于 C 等级企业给银行带来的预期收益为负，就算仍有余额，银行也不应该考虑向 C 等级企业进行放贷。

## 6.2 问题二求解

利用我们建立的基于 LASSO 选择的支持向量回归模型，我们将附件 2 中 302 家企业的数据进行相关必要的清洗、处理归一化后，进行信贷等级预测。部分结果如下图中表格所示：

企业代码	得分（预测）	企业代码2	得分（预测）2
E124	49.486	E132	80.091
<b>E125</b>	<b>40.710</b>	<b>E133</b>	<b>79.331</b>
E126	69.544	E134	81.807
<b>E127</b>	<b>69.823</b>	<b>E135</b>	<b>83.580</b>
E128	74.468	E136	83.369
<b>E129</b>	<b>51.271</b>	<b>E137</b>	<b>91.262</b>
E130	73.095	E138	99.476
<b>E131</b>	<b>70.015</b>	<b>E139</b>	<b>71.100</b>

图 8 部分信贷等级预测结果

在我国，一般贷款额度不会超过企业过去 3 年平均营业收入额度的 40%[3]，所以在题目原有的限制下，我们又加入了一条限制以更好地贴近实际：

1. 放贷额度不会超过企业过去 3 年平均净销总额的 40%；
2. 放贷利率处于 4% 至 15% 之间；
3. 放贷额度处于 10 万至 100 万。

由于模型得出信贷等级预测评分为连续型数值，我们对于每一企业的等级评分做线性加成：

$$\theta_i = 1 - \frac{\text{预测评分与预期等级距离}}{\text{邻近等级间距离}} \quad E(R) = \theta_i R(i) + (1 - \theta_i) R(j) \quad (18)$$

最终，我们可以得到每一个企业的预期收益率，将其进行降序排序，根据前文所述的放贷限制条件，我们的策略是：

- 按照预期收益率从高到低进行放款，额度为 100 万与过去 3 年平均净销总额的 40% 之较低者。

企业代码	可分配额度	预期收益率	优先程度
E357	1000000	0.040188686	1
<b>E237</b>	<b>1000000</b>	<b>0.040188686</b>	<b>1</b>
E419	1000000	0.040188686	1
<b>E314</b>	<b>1000000</b>	<b>0.040188686</b>	<b>1</b>
E189	679292.744	0.040188686	1
...	...	...	...
E258	1000000	0.034842592	100
<b>E261</b>	<b>1000000</b>	<b>0.034812632</b>	<b>101</b>
E195	320707.256	0.034796352	102

图 9 贷款发放情况表（部分）

### 6.3 问题三求解

根据观察与经验，突发因素往往具有：开始 → 发展 → 高潮 → 衰退 → 结束此五种过程。为了研究突发因素对于不同行业 and 不同类别企业的不同影响，我们选取了最近的“新冠疫情”作为突发因素的代表展开研究。

突发因素的严重程度，即“新冠疫情”的严重程度，可以用当前余存患者人数来衡量，据此，我们引入了 SEIR 模型来对此次新冠疫情的传染过程进行分析。但是由于新冠的特殊性，传统的 SEIR 模型不能很好地模拟此次的传播方式，所以我们在原有模型的基础上增加了 3 个特征，使得分析更加贴合实际。

传统的 SEIR 模型将研究对象分为 S、E、I、R 四种类型：

- S 代表易感人群（Susceptible），表示潜在的可感染人群。
- E 代表潜伏者（Exposed），表示已经被感染但是没有表现出感染症状的人群。



- $I$  代表感染者 (Infected)，表示已经被感染而且表现出感染症状的人群。
- $R$  代表移除者 (Removed)，表示死亡人群。

结合本次新冠的实际情况，我们基于原有的 SEIR 模型进行改良，增加了 3 个新的特征：

1. 感染者  $I$  会被立刻隔离治疗，所以只有处于潜伏期后期的潜伏者  $E$  才具备传染能力。且因假设所有的潜伏者最终都会转变为感染者，所以每日具有传染性人群数量为潜伏者  $E$  乘以概率  $r$ ， $r$  是潜伏者平均发病时间的倒数。
2. 由于国家与社会对疫情防控的重视，人们逐渐减少外出、重视卫生情况，病毒的传染力  $\beta$  随时间推移越来越低。所以在改良模型中，易感人群的感染率由常数改为方程  $\beta(t)$ 。
3. 康复者会重新得病（核酸检测复阳），并且康复者得病概率与易感者  $S$  相同。

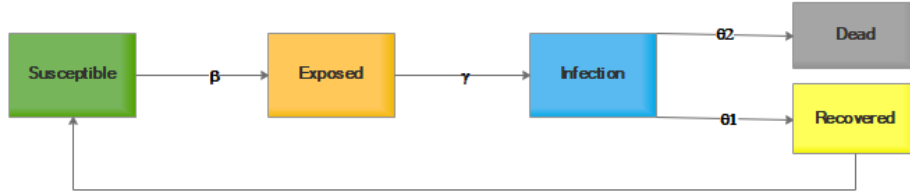


图 10 改良后 SEIR 模型结构图

记  $S(t)$ 、 $E(t)$ 、 $I(t)$ 、 $R(t)$  分别为时刻  $t$  的易感人群数量、潜伏者数量、感染者数量和移除数量（死亡人数）。显然， $S(t) + E(t) + I(t) + R(t) = N$ ，其中  $N$  为总人群数量。

对于死亡人数  $R$ ，则在感染者  $I$  中会有  $\theta_2$  的死亡率：

$$\frac{dR}{dt} = \theta_2 \cdot I \quad (19)$$

对于感染人数  $I$ ，则由潜伏者  $E$  存在  $r$  概率转化为感染者，同时也存在  $\theta_1 + \theta_2$  的概率痊愈或者死亡：

$$\frac{dI}{dt} = Er - (\theta_1 + \theta_2)I \quad (20)$$

相应的，潜伏者  $E$  的个数因为感染易感者  $S$  被感染而上升，但是存在  $r$  比例的潜伏者转化为感染者  $I$ 。假设病毒成功传染给易感者的概率为与时间相关的递减函数  $\beta(t)$ ，一位潜伏者  $E$  在具有传染性后平均每天密切接触的人数设为  $n$ ，在接触到的人中又有  $(S + \theta_1 I)/N$  比例属易感染者：

最后，易感者  $S$  的数量由于转化为潜伏者  $E$  而下降，同时有感染者  $I$  痊愈后被重新归类于易感人群  $S$ ：

整个改良后的 SEIR 模型的表达式为:

$$\begin{cases} \frac{dS}{dt} = -Enr \cdot \frac{S + \theta_1 I}{N} \cdot \beta(t) + \theta_1 I \\ \frac{dE}{dt} = Enr \cdot \frac{S + \theta_1 I}{N} \cdot \beta(t) - Er \\ \frac{dI}{dt} = Er - (\theta_1 + \theta_2) I \\ \frac{dR}{dt} = \theta_2 \cdot I \end{cases} \quad (21)$$

感染率  $\beta$  的计算方法为每日新增患者人数除以密切接触者，治愈率  $\theta_1$  和死亡率  $\theta_2$  的则为当日新增出院人数和新增死亡人数与上一天感染者的比率。我们利用已知疫情数据对感染率进行指数拟合，对治愈率和死亡率进行高斯曲线拟合（以 2020 年 1 月 21 日为开始的第一天，记  $t = 1$  为 2020 年 1 月 21 日），结果如下所示：

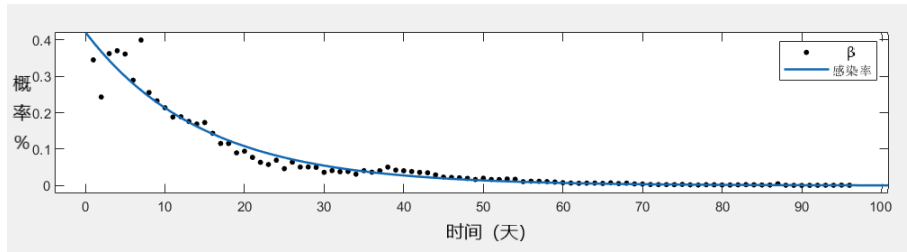


图 11 感染率  $\beta(t)$  拟合图

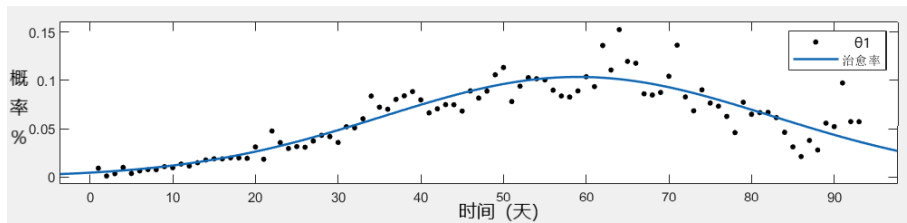


图 12 治愈率  $\theta_1$  拟合图

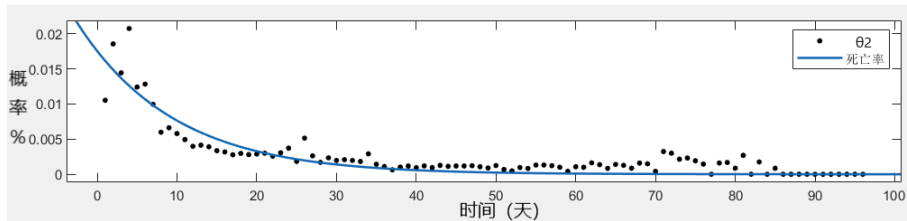


图 13 死亡率  $\theta_2$  拟合图



表 6 感染率、治愈率和死亡率的拟合结果

拟合方程	可决系数 $R^2$	调整可决系数
$\theta_1 = 0.1035e^{-(\frac{t-0.4453}{1.235})^2}$	0.8266	0.8227
$\theta_2 = 3.341 \times 10^{-8}e^{-(\frac{t-22.4}{4.246})^2}$	0.8019	0.7976
$\beta(t) = 0.422e^{-.006809t}$	0.9422	0.9416

### 6.3.1 计算机仿真模型

可收集的数据均以天为单位，因此我们考虑离散变化的情况，相关微分方程转化为如下的差分方程：

$$\begin{cases} S_t = S_{t-1} - E_{t-1}nr \cdot \frac{S_{t-1} + \theta_1 I_{t-1}}{N} \cdot \beta_t + \theta_1 I_{t-1} \\ E_t = E_{t-1} + E_{t-1}nr \cdot \frac{S_{t-1} + \theta_1 I_{t-1}}{N} \cdot \beta_t - E_{t-1}r \\ I_t = I_{t-1} + E_{t-1}r - (\theta_1 + \theta_2)I_{t-1} \\ R_t = R_{t-1} + \theta_2 I_{t-1} \end{cases} \quad (22)$$

由国家卫生健康委员会官方网站 [5] 提供的数据分析，我们得出 SEIR 模型中  $E_0$ 、 $I_0$ 、 $R_0$  的具体数值：

$$\begin{cases} E_0 = 1757 \\ I_0 = 431 \\ R_0 = 9 \end{cases} \quad (23)$$

我们接着使用计算机进行模拟仿真，通过递推得到了潜伏者人数  $E$ 、感染者人数  $I$ 、死亡人数  $R$ ，见下图：

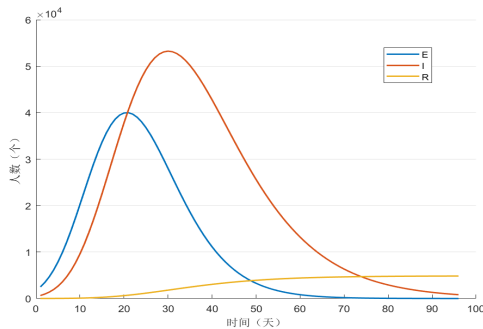


图 14 计算机拟合结果

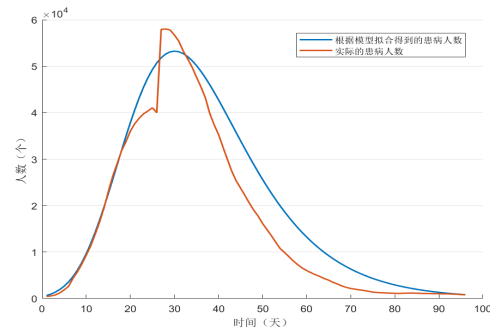


图 15 拟合结果与实际数量对比

对于感染人数  $I$ ，我们计算  $I_{\text{实际}}$  和  $I_{\text{预测}}$  的相关系数，得到： $r = 0.9760$ 。上述研究结果表明，我们的拟合结果非常接近真实情况，具有较高的可信度。

我们发现，新冠疫情突发因素与一般突发因素一样都有开始、发展、高潮、衰退

和结束等五个过程。所以可以不失一般性地用新冠疫情这一突发因素来代表所有的突发因素。

6.3.2 通过股票指数反应突发因素对企业发展状况影响

因为某行业的股价与该行业当前的发展状况存在着正相关关系，且该行业的发展状况是该行业股价变化的重要影响因素.[5] 所以我们采用该行业的平均股价来代表该行业当前的发展状况进而反应出该行业的信誉等级。

我们通过对附件 2 中 302 家企业进行分类，总共有如下几类：

制造业
金融业
房地产业
交通运输、仓储和邮政业
住宿和餐饮业
信息传输、软件和信息技术服务业
电力、热力、燃气及水生产和供应业
公共管理、社会保障和社会组织
采矿业
教育业
科学研究和技术服务业
农、林、牧、渔
批发与零售业
租赁和商务服务业
文化、体育和娱乐业

图 16 行业类别图

将每个行业的股价与当前突发因素的严重程度进行作图分析。之后按照突发因素前期分为五个阶段，前期、高潮前一段时间、高潮后一段时间，后期，长后期。对刚刚得到的各个行业的图表中的这五段分别计算股票价格曲线与突发因素严重程度的相关系数。

七、模型的评价

7.1 模型优点

1. 对所有指标进行归一化处理，把有量纲表达式变为无量纲表达式，加快了梯度下降求最优解的速度，提升模型的收敛速度。2. 对问题进行多个模型分析和比较，得出最优模型，使模型更具说服力。

7.2 模型缺点

1. 由于原始数据有部分缺失对模型的评估有一定的影响。

1. 在第一问寻求短期经济与疫情关系的模型中，我们将定性的指标进行量化，使得问题求解的过程更为严谨，比如经济发展状况以平均 A 股股票价格等指数来衡量，疫情程度以当前的余存患者数来表示。
2. 在量化疫情程度时，我们对原有 SEIR 模型进行改良，添加了“复阳”“有症状感染者被立即隔离，无症状感染者存在感染能力”等与现实贴切的因素，使得模型预测结果更加接近真实情况。
3. 在预测中长期 GDP 时，我们选择了 Holt-Winters 乘法模型，因为该模型同时考虑 GDP 数据成正比趋势性和随季节性变化的两个特征，所以最终预测结果相比简单平滑指数模型更为精确。

### 7.3 模型缺点

1. 由于疫情数据有存在补报的情况，会有一定程度对模型准确性判断有偏差。
2. 在对中长期的宏观经济预测中，由于突发事件的存在（譬如此次新冠疫情），我们只能挑选数量有限的角度和较重要的影响因素进行分析，存在一定的片面性。

### 7.4 模型推广

在发生重大公共卫生事件，对事件的发展程度以及对经济影响起到分析和预测的作用。

## 参考文献

- [1] E. M. Hashimoto, E. M. M. Ortega, G. M. Cordeiro, et al. The multinomial logistic regression model for predicting the discharge status after liver transplantation: estimation and diagnostics analysis. 2020, 47(12):2159-2177.
- [2] 陶余会. 如何构造模糊层次分析法中模糊一致判断矩阵 [J]. 四川师范学院学报 (自然科学版), 2002(03):71-74.
- [3] 卢超. 银行贷款业务财务风险管控刍议 [J]. 会计师, 2020(14):33-34.
- [4] Wind-经济数据库. 股票数据. (2020-04-30)[2020-04-30]. <http://www.wind.com.cn/newsite/edb.html>
- [5] 刘家树. 股市波动与经济增长关系的实证分析%Analysis of the Relationship between Stock Fluctuation and Economic Growth[J]. 安徽工业大学学报 (社会科学版), 2008, 025(002):42-44.