

基于优化模型的企业信贷策略

摘要

本文研究的主要是中小微企业最优信贷发放策略，建立了信贷发放优化模型，以银行盈利最大为目标，贷款年利率、客户流失率等与信贷风险之间的关系为约束，得到各类情况下银行信贷发放的最优策略。

针对问题一第一小问：我们从“企业信誉”、“企业实力”和“是否违约”这三方面对企业的信贷风险进行量化分析。首先我们从企业的发票信息中提取出“上下游企业数量”等与企业实力和信誉相关的众多因素，对于企业实力，我们以提取出的相关因素作为评价指标，通过**熵信息构权法**建立评价模型来定量评价，而对于企业信誉，我们直接采用附件中“信誉评级”这一指标来进行评价。在此基础上，我们再以企业信誉、企业实力和是否违约这三方面作为评价指标，同样通过熵权法建立评价模型，得到 123 家企业信贷风险的量化得分，作为最终信贷风险评估的结果。

针对问题一第二小问：首先我们根据各企业的“信誉评级”来决定是否为其发放贷款，评级为 D 的不予发放；其次对于符合信誉评级要求的企业均给予 10 万元的基础额度，我们再结合“信贷风险”和“企业实力”这两方面因素建立**贷款额度决定模型**，对其进行综合评价，将剩余总额按各企业综合得分占总得分的比例进行发放，对于超出 100 万贷款限额的企业，将超限的那部分额度，依次发放给按综合得分高至低排序的贷款额度未满足的企业。在确定各企业贷款额度后，我们以银行收益最高为目标，贷款年利率为决策变量，年利率与客户流失率的关系等为约束条件建立**贷款利率优化模型**，通过 Lingo 解得各企业贷款年利率，即可得到各企业的信贷策略。同时，我们考虑到银行短期收益最大会导致客户流失率较大，而不利于银行长期合作，所以我们以银行十年总收益最高为目标，不考虑客户后续的增加等为约束条件，得到各企业每年的信贷策略。

针对问题二：为了得到 302 家企业缺失的“信誉评级”这一数据，我们根据与企业信誉相关的六个因素——每日偷漏税金额、每日收入、上下游企业数量、企业成立时长、有效正数发票数、销项作废发票数，将其作为影响因素，将问题一中的 123 个样本分为训练集和测试集，分别用**判别分析法**和**神经网络算法**对样本进行判别，得到误判率分别为 50.0%，27.8%，说明神经网络的判别效果更优，故采用神经网络算法对 302 家企业的信誉评级进行判别。在得到企业的信誉评级后代入问题一中的贷款利率优化模型，得到具体信贷策略。

针对问题三：由于突发因素对不同行业、企业的影响程度不同，我们先根据各企业的名称对企业所属行业进行划分。对于突发因素，我们分别从“公共卫生突发事件”和“经济安全突发事件”两方面进行分析，分别取 2020 年新冠病毒疫情和 2008 年金融危机作为具体事件，仅从经济角度研究突发事件对各企业的影响。我们收集得到突发事件发生前 5 年各行业的生产总值，通过时间序列模型对突发事件时期内的生产总值进行预测，对预测值与实际值进行分析计算，得到各行业的**相对变化率**，将其作为一项指标，加入贷款额度决定模型，对综合评分重新进行加权计算，并得到各企业的贷款额度，然后利用问题一中的优化模型对年利率进行求解，最终得到新的信贷策略。

关键词：熵信息构权法 优化模型 判别分析法 神经网络算法

一、问题的提出与重述

1.1 问题的提出

贷款是中小微企业解决融资难问题的主要方法之一。由于企业参差不齐，银行的可支配资产也较有限，为了使银行的效益最大，银行通常根据企业的相关指标对企业提供贷款，以确保固定的资金回笼和现金流通率，贷款额度、期限、利率等信贷策略也通常根据企业的类型、规模、实力、信誉等指标进行综合评估，以决定不同企业的信贷方式。

1.2 问题的重述

某银行邀请数模团队解决小微企业的信贷问题，并要求贷款额度在 10 万到 100 万元，年利率为 4%-15%，期限一年。根据相关附件数据解决如下问题：

问题 1：量化分析附件一中企业的各项指标，评估各企业信贷风险，并为银行设计信贷策略。

问题 2：量化分析附件二中企业的各项指标，为银行设计信贷策略使年度信贷总额不超过一亿元。

问题 3：一些不确定因素通常会对不同行业有不同程度的影响。全面考虑各种因素和各企业信贷风险，为银行设计信贷策略使年度信贷总额不超过一亿元。

二、问题的分析

2.1 问题一的分析

对于第一小问，为了对 123 家企业的信贷风险进行量化分析，我们从“企业信誉”、“企业实力”和“是否违约”这三方面出发进行综合分析，通过对题目所给数据的挖掘和分析，我们从企业的发票信息中提取出了与企业实力和企业信誉相关的众多因素，具体如下文表中所示。我们以提取出与企业实力相关性较强的因素作为评价指标，建立相应的评价模型，对企业实力进行评价，而对于企业信誉的评价，我们则直接采用附件中所给的“信誉评级”这一指标作为评价依据。在此基础上，我们再以企业信誉、企业实力和是否违约这三方面作为评价指标，同样通过建立评价模型，得到 123 家企业信贷风险的量化得分，作为最终信贷风险评估的结果。

对于第二小问，我们分别从“贷款额度”和“贷款年利率”两方面进行具体信贷策略的制定。首先，我们根据各企业的“信誉评级”来决定是否为其发放贷款；其次，对于信誉符合要求的企业，我们结合企业的“信贷风险”和“企业实力”这两方面因素对发放贷款的额度进行决策。在确定好对各企业发放的贷款额度后，先通过回归分别得到不同信誉评级企业贷款年利率和客户流失率之间的函数关系，再建立一个以银行利息收益最大为目标，信贷风险和贷款年利率成正比等为约束条件的优化模型，通过求解得到各企业具体的贷款年利率。同时，我们考虑到银行一年内收益最大，会导致客户流失率较大而不利于银行长期合作收益的问题，所以我们以十年为单位，客户量随年递减，重新对收益目标进行计算，得到各企业每年的贷款年利率。

表 1 企业实力、信誉相关因素

企业实力	企业信誉
正数有效发票单数 经营时长、上下游企业数量 每日收入、每日成本 每日利润、每日增值税额等	偷漏税发票数量占比、每日偷漏税金额、负数发票数量占比、进项作废发票数量占比、销项作废发票占比、销项偷漏税数量占比、每日收入、经营时长、有效正数发票数等

2.2 问题二的分析

本题要求在贷款总额为 1 亿元时，对 302 家企业的信贷策略进行制定。相比第一问中的 123 家企业的各项数据，我们发现 302 家企业的“信誉评级”以及“是否违约”这两项指标并未给出，所以我们依次通过判别分析和神经网络算法对 302 家企业的信誉评级进行判别，将两种方法的误判率进行比较，选择误判率较低的那种方法，并根据判别结果以及其他相关因素，利用问题一中的信贷策略模型为 302 家企业制定具体短期和长期收益最大的信贷策略。

2.3 问题三的分析

本题要求我们考虑在突发因素影响下各企业的信贷调整策略，由于突发因素对不同行业、企业的影响程度不同，所以我们先根据各企业的名称对企业所属行业进行划分。对于突发因素，我们分别从“公共卫生突发事件”和“经济安全突发事件”两个方面进行分析，其中“公共卫生突发事件”我们选取 2020 年新冠病毒疫情作为具体事件，“经济安全突发事件”我们选取 2008 年金融危机作为具体事件，我们仅从经济角度研究突发事件对各企业的影响，我们收集得到突发事件发生前 5 年各行业的生产总值，通过时间序列模型对突发事件影响最大的时期内的生产总值进行预测，对预测值与实际值进行分析计算，得到各行业的相对变化率；该系数能直接反映各行业受突发事件影响的程度，同样也能够间接反映属于不同行业的各企业在突发因素影响下的稳定性，将企业稳定性作为一项指标，加入贷款额度决定模型，这样影响企业贷款额度的综合评价得分会受突发因素影响，信贷策略中各企业的贷款额度也会发生变化，在得到各企业贷款额度后利用问题一中的优化模型对利率进行求解，最终得到新的信贷策略。然而考虑到突发因素的影响时长无法预测，我们根据时间序列模型对 2008 年经济危机后各个行业生产总值的预测趋势，与实际值进行对比，定性的分析突发因素对各个行业的长期潜在的影响，并对信贷策略做出一定调整。

三、模型的假设

- (1) 假设附件 1,2 中的企业到目前为止都没有破产。
- (2) 假设各行业生产总值足够描述疫情对经济的总影响。
- (3) 假设未来 10 年银行客户总人数不会增加。
- (4) 假设突发因素对相同行业里的不同企业的影响因素相同。

四、 符号的说明

符号	含义	单位
P_i	综合评价得分	\
w_j	权重系数	\
$Z_{\text{总}b}$	贷款额度得分	\
E	企业还款金额	元
L_b	放贷额度	元
r_b	年利率	\
P_b	企业风险评估得分	\
K	除去基础额度的总额	\
Y_b	客户流失率	\
ℓ_1, ℓ_2	指标权重	\

五、 模型的准备

5.1 数据预处理

对附件数据中许多异常数据、无效数据、无参考价值数据进行预处理，将异常和无参考价值的数据删除，将非数值类数据转换成定量化数值，处理方式如下：

(1) 异常数据

通过对数据样本的分析，我们发现 E27 企业发票号为 82240 的负数发票的金额、税额、价税合计符号不统一，应予剔除。

(2) 无参考价值的数据

与指标无关的无参考价值的数据可以忽略，发票号码无法体现出一个企业的实力、信誉等相关信息，故忽略。

(3) 数据的转换

时间数据的转换：在对企业经营时长的分析中，我们以企业最早开具进项发票的时间为企业成立时间，计算成立距今的时长作为企业经营时长，作为分析企业风险评估指标之一。

税率数据的转换：在对企业所得税率的分析中，我们以销项发票信息中税额与价税合计的比值作为企业所得税率，查找相关资料发现我国企业的最低所得税率为 3%，所以我们将所得税率小于 3% 的票据视为逃漏税单。

六、模型的建立与求解

6.1 问题一的模型建立与求解

6.1.1 信贷风险评价模型的建立与求解

6.1.1.1 信贷风险评价指标体系的建立

对于信贷风险的评价，我们主要从企业实力、企业信誉、是否违约这三方面进行分析，如下图所示：

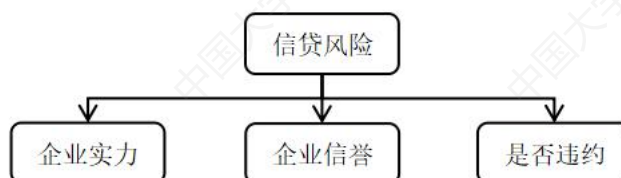


图 1 信贷风险相关因素

(1) 影响企业实力的因素

通过对附件 1 中进、销项发票数据的分析和处理，我们提取出影响企业实力的 6 个因素如下：

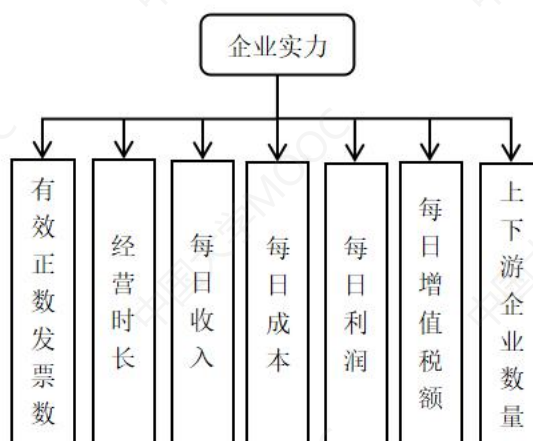


图 2 影响企业实力的因素

有效正数发票数：有效发票的开具次数即一个企业的交易次数，交易次数越多可以在一定程度上体现该公司经营状况良好。

经营时长：我们以企业最早开具进项发票的时间作为企业的成立时间，计算成立距今的时长作为企业经营时长。企业成立时间越长，说明该公司经历了市场的考验，具有较强的实力和稳定性。

每日收入：销项发票价税合计金额的总和与经营时长的比值为每日收入，每日收入越高，说明该企业经济实力较强。

每日成本：进项发票价税合计金额的总和与经营时长的比值为每日成本，每日成本越高，表现该企业可支配资产越多，表明实力越强。

每日利润：销项发票价税合计与进项发票价税合计的差值。每日利润越高，该企业的还贷能力越强。

每日增值税额：销项发票税额与进项发票税额的差值与经营时长的比值为每日增值税额。企业每日需缴纳的税额越高，说明该企业的规模越大。

上下游企业数量：通过统计与企业进行交易对象的数量可以得到上下游企业的数量，上下游企业数量越多，表明该企业供求关系越稳定。

(5) 影响企业信誉的因素

对于企业信誉的评价我们采用附件 1 中所给的“信誉评级”作为具体的企业信誉评价。

通过对附件 1 中进、销项发票数据的分析和处理，我们提取出影响企业信誉因素包括偷漏税发票数量占比、每日偷漏税金额、负数发票数量占比、进项作废发票数量占比、销项作废发票占比、销项偷漏税数量占比、每日收入、经营时长、有效正数发票数等。各因素释义如下：

偷漏税发票数量占比：偷漏税发票数量占比即为税率小于最低所得税税率的发票单数占总单数的比例，该比例越高说明企业的信誉越低。

每日偷漏税金额：偷漏税发票数量占比即为税率小于最低所得税税率的发票金额的总和与运营时长的比值，每日偷漏税金额越大，表明该企业违法行为较严重，信誉较低。

负数发票数量占比：负数发票意味着购方因故取消交易，负数发票数越多，表明企业交易不成功率高、供求关系不稳定。

进项作废发票数量占比：企业因故取消与货家的交易，作废票数越多，表明企业毁约概率越大，供求关系不稳定。

销项作废发票数量占比：购方因故取消交易，发票作废，作废发票越多，表明企业不成功交易率越高。

销项偷漏税数量占比：企业交税率没有到达最低企业所得税标准，可能存在偷漏税行为，表明企业不诚信度越高。

每日收入：企业向消费者销售商品获得的价税合计为收入，收入越高表明该企业经济实力越强。

我们将上述各因素与信誉评级之间进行皮尔逊相关性分析，分析得到相关性较低的几个因素如下表所示：

表 2 皮尔逊相关系数

因素	每日偷漏税金额	负数发票数量占比	进项作废发票数量占比	偷漏税发票数量占比
相关系数	-0.178	-0.003	-0.156	0.203

剔除上述 4 个相关性较低的因素，仅保留以下影响企业信誉的主要因素指标：

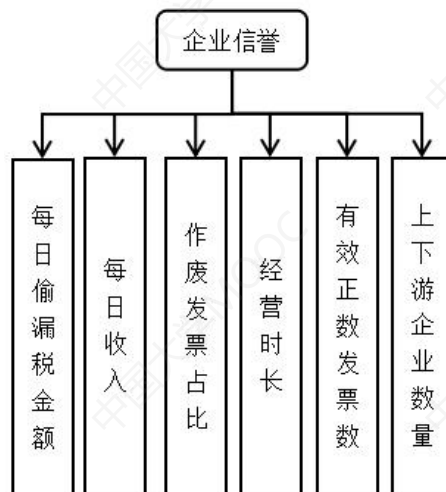


图 3 影响企业信誉的因素

(3) 企业是否违约的概率统计

根据附件一 123 家企业的是否违规情况和信誉评级，整理得到违规企业评级存在 1 个 B、2 个 C、24 个 D，信誉评级为 D 的企业均违规。

我们按不同信誉评级的守约率来赋予相应得分，得到下表不同信誉等级对应的守约得分：

表 3 不同信誉评级的守约得分

信誉评级	A	B	C	D
守约率	100%	97.4%	6.25%	0%
得分	1	0.974	0.0625	0

6.1.1.2 熵权法确定评价指标权重

信贷风险由企业实力、企业信誉、是否违约这三个评价指标决定，企业实力由经营时长等六个评价指标决定，每类指标对评价对象的影响程度不同，故我们引用熵权法，根据指标变异性的确定客观权重。因此我们建立熵值、变异系数、权重系数的模型如下：

$$\text{第}j\text{项的熵值: } I_j = -k \sum_{i=1}^n p_{ij} \ln(p_{ij}) \quad (1)$$

$$\text{其中} k > 0 \text{ 为常数, } k = \frac{1}{\ln(n)}, p_{ij} = \frac{x_{ij}}{\sum_{i=1}^n x_{ij}}, i = 1, 2, \dots, 123; j = 1, 2, 3 \quad (2)$$

$$\text{第}j\text{项指标的变异系数: } r_j = 1 - I_j, j = 1, 2, 3; \quad (3)$$

$$\text{第}j\text{项指标的权重系数: } w_j = \frac{r_j}{\sum_{k=1}^3 r_k}, j = 1, 2, 3; \quad (4)$$

根据上述方法求得指标权重如下：

(1) 影响企业信贷风险评价的 3 个指标的权重（详见程序八）

表 4 影响企业信贷风险的指标权重

	企业实力	信誉等级	是否违约
风险熵值	0.768	0.973	0.994
风险权重	0.876	0.100	0.023

上表可以得到企业实力的权重最大,且远大于信誉等级和是否违约两个因素的权重,故对企业风险评估的决策价值最大。

(2) 影响企业实力的评价的 6 个指标的权重

表 5 影响企业实力的指标权重

	有效正数 发票数	每日收入	每日成本	每日收益	每日增值 税额	上下游 企业数量
实力熵值	0.785	0.994	0.601	0.383	0.998	0.512
实力权重	0.110	0.003	0.203	0.315	0.001	0.249

由上表可知每日收益、上下游企业数量、每日成本这三个指标的权重相对较高。

6.1.1.3 信贷风险综合评价

由评估企业风险因素的权重,可得每个企业的信贷风险评价得分,风险评估得分越高,风险越低。

$$P_i = x'_{ij} * w_j, \quad i = 1, 2, \dots, 123; \quad j = 1, 2, 3; \quad (5)$$

其中, P_i 为综合评价得分, x'_{ij} 为标准化后的数据, w_j 是权重系数。

由上述方法我们得到企业信贷风险得分如下表所示,因数据过多,仅显示评估得分较好和较差的企业:

表 6 企业信贷风险评估得分

企业代号	企业实力评分	企业信誉得分	是否违约	风险评估得分
E1	1.00	1.00	1	1.02
E4	0.63	0.33	1	0.63
E2	0.45	1.00	1	0.54
E8	0.35	1.00	1	0.45
...
E123	0.00	0.25	0	0.05
E101	0.00	0.25	0	0.05
E120	0.00	0.25	0	0.05
E109	0.00	0.25	0	0.05

信贷风险较低的企业是 E1、E4, 信贷风险较高的企业是 E109、E120。根据信贷风险评估得分研究银行对企业的信贷策略,解决是否放贷、放贷金额、年利率、期限等问题。

6.1.2 信贷发放优化模型

根据题意,我们建立的信贷策略主要从是否放贷、贷款额度、年利率这三方面考虑。

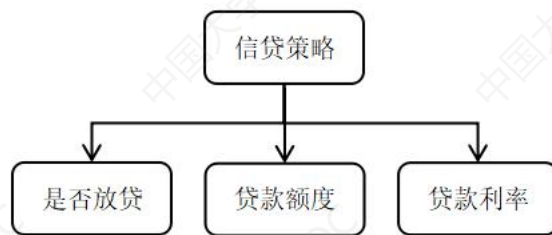


图 4 信贷策略考虑因素

6.1.2.1 客户流失率回归模型

由附件三数据，我们发现贷款年利率与不同信誉评级的客户流失率具有一定的关系，所以我们通过建立回归模型来求得两者之间的函数关系。

(1) 客户流失率回归模型的建立

首先绘制出客户流失率与贷款年利率关系的散点图，可直观看出客户流失率与贷款年利率之间的关系为非线性。

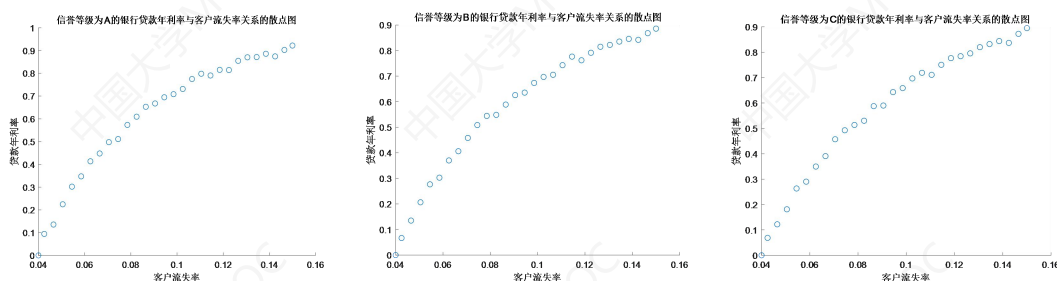


图 5：信誉评级为 A/B/C 的客户流失率与年利率的散点图

故建立一元非线性回归模型：

$$Y_i = \beta_1 r_i + \beta_2 \sqrt{r_i} + \varepsilon_i, i = A, B, C \quad (6)$$

其中， Y_i 为客户流失率， r_i 为年利率， β_i 为待定系数， ε_i 为常数项。

(2) 客户流失率回归模型的求解（详见程序九）

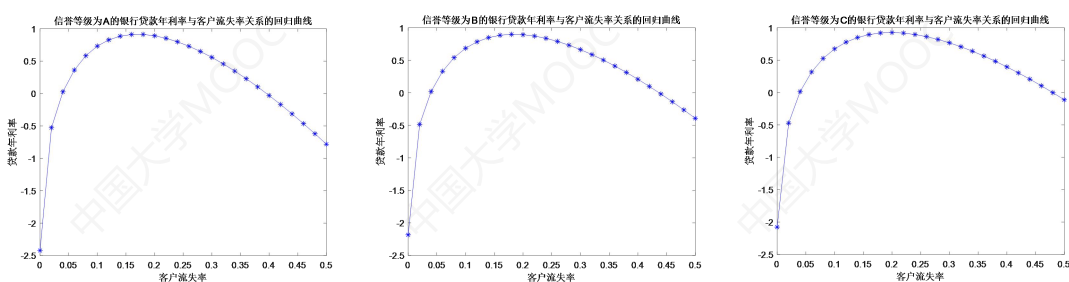


图 6：信誉评级为 A/B/C 的客户流失率与年利率的回归图像

求解得到信誉评级为 A、B、C 的客户流失率与贷款年利率之间的回归模型如下表所示：

表 7 客户流失率回归模型

企业信誉等级	回归模型	r^2
A 评级企业	$Y_A = -2.423 + 16.167\sqrt{r} - 19.581r$	0.9975
B 评级企业	$Y_B = -2.183 + 14.364\sqrt{r} - 16.732r$	0.9982
C 评级企业	$Y_C = -2.077 + 13.486\sqrt{r} - 15.143r$	0.9980

由回归模型的拟合优度可知三种回归模型的回归效果均较好，采用价值较高。

6.1.2.2 信贷发放的具体策略

(1) 是否放贷

我们根据各企业的信誉评级来决定是否发放贷款，对于信誉评级为 D，即信誉极差的企业银行不予放贷。

(2) 贷款额度（贷款额度决定模型）

对于贷款额度的分配我们主要从企业的信贷风险得分和企业实力这两方面进行分析。

Step1: 求出各企业的信贷风险得分与企业实力得分之和，作为决定贷款额度分配的综合得分，综合得分越高可以分配到更高额度的贷款；

$$Z_{\text{总}b} = \ell_1 Z_{\text{信}b} + \ell_2 Z_{\text{实}b}, \quad b = 1, 2, \dots, n \quad (7)$$

Step2: 对于符合信誉评级要求的企业均先给予 10 万元的基础额度；

Step3: 将剩下的总额按各企业综合得分占总得分的比例进行发放；

$$L_b = \frac{Z_{\text{总}b}}{\sum_{b=1}^n Z_{\text{总}b}} \times K, \quad b = 1, 2, \dots, n \quad (8)$$

Step4: 在按得分比例发放后，对于超出 100 万贷款额度限制的企业，我们将超出限制的那部分额度回收；

Step5: 将未达到贷款额度上限的企业按综合得分由高至低排序，将回收的那部分额度按顺序发放，每家企业的总额度不能超过上限，直至回收的额度发放完为止。

其中 $Z_{\text{总}b}$ 表示贷款额度得分， $Z_{\text{信}b}$ 表示企业信贷风险得分， $Z_{\text{实}b}$ 表示企业实

力得分， L_b 表示放贷额度， K 表示除去基础额度的总额， ℓ_1 ， ℓ_2 表示每个指标的权重。

(3) 贷款利率（贷款利率优化模型）

企业信贷风险与年利率、客户流失率和银行收益之间的关系如下图所示：



图 7 各因素联动关系

在确定各企业的贷款分配额度后，我们通过建立贷款利率优化模型来确定各

企业的贷款利率。

目标函数：

为了使银行经济效益最高，就要使还款金额最多，还款金额由还款客户数量、年利率、贷款额度构成。

$$\max E = L_b \times r_b \times (1 - Y_b), \quad b = 1, 2, \dots, 123 \quad (9)$$

其中E表示企业还款金额（即银行收益）， L_b 表示放贷额度， r_b 表示年利率， Y_b 表示客户流失率。

约束条件：

I. 由题目要求，贷款年利率在 4%-15%之间，即我们设定的年利率要满足在此范围内

$$0.04 \leq r_b \leq 0.15, \quad b = 1, 2, \dots, 123 \quad (10)$$

其中 r_b 表示年利率。

II. 企业信贷风险评估得分越高，表明该企业的违约率越低，即银行的损失概率越低，所以银行会给出一定的利率优惠来留下这类企业，故企业风险评估得分与银行年利率成反比

$$\frac{1}{P_b} \propto r_b, \quad b = 1, 2, \dots, 123 \quad (11)$$

其中 P_b 表示企业风险评估得分。

III. 信誉评级为 A、B、C 等级的客户流失率与年利率相关，我们将其关系通过函数表示

$$\text{信誉评级为 A 等级的企业 } Y_A = -2.423 + 16.167\sqrt{r_b} - 19.581r_b \quad (12)$$

$$\text{信誉评级为 B 等级的企业 } Y_B = -2.183 + 14.364\sqrt{r_b} - 16.732r_b \quad (13)$$

$$\text{信誉评级为 C 等级的企业 } Y_C = -2.077 + 13.486\sqrt{r_b} - 15.143r_b \quad (14)$$

综上，获得贷款利率的模型为：

$$\max E = L_b \times r_b \times (1 - Y_b), \quad b = 1, 2, \dots, 123$$

$$\text{s.t.} \begin{cases} b = 1, 2, \dots, 123 \\ 0.04 \leq r_b \leq 0.15 \\ \frac{1}{P_b} \propto r_b \\ Y_A = -2.423 + 16.167\sqrt{r_b} - 19.581r_b \\ Y_B = -2.183 + 14.364\sqrt{r_b} - 16.732r_b \\ Y_C = -2.077 + 13.486\sqrt{r_b} - 15.143r_b \end{cases} \quad (15)$$

6.1.2.3 信贷发放策略的求解（详见程序十）

- ①首先根据企业的信誉评级等级，将评级为 D 的企业剔除
- ②根据企业信贷风险和企业实力的综合得分进行额度分配（本题假设银行年度的信贷总额为 5 千万）
- ③通过 Lingo 对贷款利率优化进行求解，得到各企业的贷款年利率，由于企业数量较多，我们选取部分企业进行展示，如下表所示

表 8 信贷策略（部分）

企业号	贷款金额	利率	企业号	贷款金额	利率
1	1000000	0.04736	77	196691.6867	0.05476
4	1000000	0.04881	94	196177.8457	0.05476
2	1000000	0.04881	105	194829.7315	0.05476
8	1000000	0.04881	86	192585.6143	0.05476
7	1000000	0.04881	69	186426.6771	0.05476
3	1000000	0.04881	110	185050.6935	0.05490
13	1000000	0.04881	96	183603.3924	0.05490
75	1000000	0.04881	68	182084.0036	0.05490
9	1000000	0.04881	104	168323.409	0.05490
6	1000000	0.04881	87	152639.3174	0.05490

按此信贷策略，对应银行最大效益为 2002653 元。

6.1.3 信贷发放策略的拓展

在 6.1.2 的发放策略中我们考虑的主要是银行的短期收益，但是如果以银行的短期收益最大为目标，会导致客户流失率较大，而不利于银行长期合作，所以我们以银行十年总收益最高为目标，不考虑客户后续的增加等为约束条件，得到各企业每年的信贷策略。

表 9 长期信贷策略

企业号	贷款金额	第一年利率	第二年利率	……	第十年利率
1	1000000	0.04	0.04	0.04	0.04736
4	1000000	0.04	0.04	0.04	0.04882
2	1000000	0.04	0.04	0.04	0.04882
8	1000000	0.04	0.04	0.04	0.04882
7	1000000	0.04	0.04	0.04	0.04882
77	196691.686	0.04	0.04	0.04	0.0548
94	196177.845	0.04	0.04	0.04	0.0548

105	194829.731	0.04	0.04	0.04	0.0548
86	192585.614	0.04	0.04	0.04	0.0548
69	186426.677	0.04	0.04	0.04	0.0548

由上表结果可得,对于在银行客户人数不增加且企业每年贷款金额固定的条件下,前面9年所有企业的利率均为0.04,而到了第十年利率才开始变化,这主要是因为对于银行长期的发展而言,客流量是非常重要的,所以它会牺牲前面几年的利率来保持一定的客流量。

6.2 问题二模型的建立与求解

分析附件二中的数据可知缺少了信誉等级和是否违约的相关信息,因此我们考虑通过问题一影响信誉等级的6项指标和123家企业的信誉等级进行判别,从而建立了Fisher判别分析模型和神经网络判别模型进行求解,对附件二中302家企业的信誉等级进行预测。

6.2.1 Fisher 判别模型的建立与求解

Fisher 准则是一种有效的有监督特征选择方法。其目的是从高维特征空间中筛选出鉴别能力强的低维特征,基本思想是在低维特征空间里,相同类别的样本尽可能密集,不同类别的样本距离尽可能远。

Fisher 判别函数:

$$H(x) = \gamma_{11}x_1 + \dots + \gamma_{p1}x_p \quad (16)$$

其中的一个投影方向:

$$(\gamma_{11}, \gamma_{21}, \dots, \gamma_{p1})^T \quad (17)$$

样本投影后的点集:

$$\{\gamma_{11}x_{1i} + \dots + \gamma_{p1}x_{pi}\}_{i=1}^n \quad (18)$$

计算点集内组内方差和、组间方差和:

$$I = \frac{V_0}{V_1} \quad (19)$$

选择给定方向,使 I 值最大,其中 V_0 是组内方差和, V_1 是组间方差和

算法步骤:

Step1: 把来自两类 w_1/w_2 的训练样本集 x 分成 w_1/w_2 两个子集 x_1, x_2

Step2: 计算样本均值向量 $M_i = \frac{1}{n_i} \sum_{x_k \in X_i} x_k, i=1,2$

Step3: 计算各类的类内离散度矩阵 \vec{S}_i , $\vec{S}_i = \sum_{x_k \in X_i} (x_k - M_i)(x_k - M_i)^T$

Step4: 计算类内总离散度矩阵 $\vec{S}_w = \vec{S}_1 + \vec{S}_2$

Step5: 计算 \vec{S}_w 的逆矩阵 \vec{S}_w^{-1}

Step6: 由 $w^* = \vec{S}_w^{-1}(M_1 - M_2)$ 求解向量 w^*

部分企业的求解结果:

表 10 Fisher 判别预测结果

企业代号	信誉等级	预测	企业代号	信誉等级	预测
5	B	B	60	B	D
9	A	A	62	B	C
14	C	A	76	B	B
16	A	C	78	C	C
31	A	A	80	C	C
32	B	C	90	C	B
37	B	B	97	B	D
38	B	B	104	C	D
39	C	D	116	D	D

结论：Fisher 判别的误判率为 50%，误判率较高，原因在于样本指标数据和目标数据之间的关联度太小，选取的指标没有足够的价值。为了优化这一问题，我们选用神经网络判别分析法对企业信誉进行预测。

6.2.1 神经网络判别模型^[1]的建立与求解（详见程序十六）

BP 神经网络是一种智能信息处理方法，主要由输入层、隐层和输出层三部分组成，其主要模型如下：

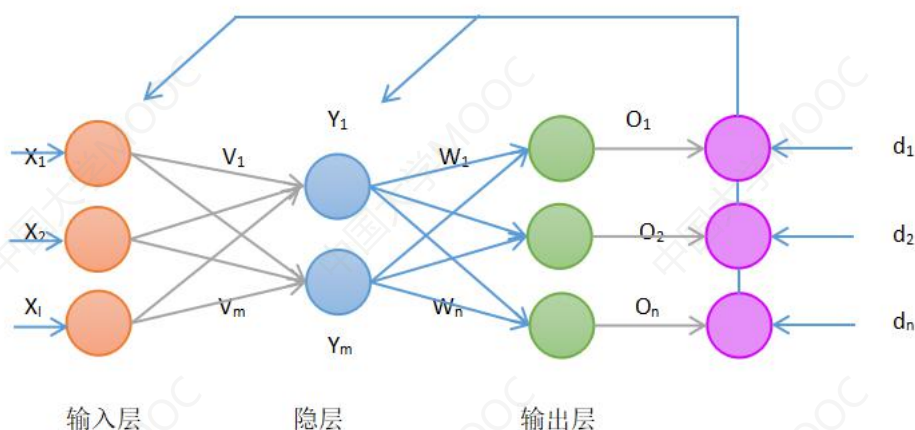


图 8 神经网络模型

其中 $(x_1, x_2, \dots, x_i)^T$ 为输入向量， $(v_1, v_2, \dots, v_m)^T$ 为隐层第 j 个神经元对应的权向量， $(Y_1, Y_2, \dots, Y_m)^T$ 为隐层输出向量， $(W_1, W_2, \dots, W_n)^T$ 为输出层第 k 个神经元对应的权向量， $(O_1, O_2, \dots, O_n)^T$ 为输出层输出的向量， $(d_1, d_2, \dots, d_n)^T$ 为期望输出的向量。

求解神经网络判别模型的步骤如下：

Step1: 载入输入（影响信誉等级的六项指标）和输出数据（123 家企业的信誉等级）并归一化

Step2: 从样本中选取 70% 作为训练样本，15% 作为测试样本

Step3: 选择隐层和输出层的传递函数，分别为 ‘tasing’ 和 ‘purelin’ ；

确定各层神经元的个数，隐层通过输入层节点数根据 $\log_2 n$ 求得

Step4: 构建网络体系，设置网络参数，指定训练参数

Step5: 利用 sim 函数进行计算仿真得到 bp 后的结果，将结果反归一化

Step6:将输入数据的结果反归一化，并作误差计算

Step7:输入需要预测的数据（附件二中 302 家企业），根据上述网络确定的函数进行预测并将 302 家企业的信誉等级进行归类。

具体流程图如下：

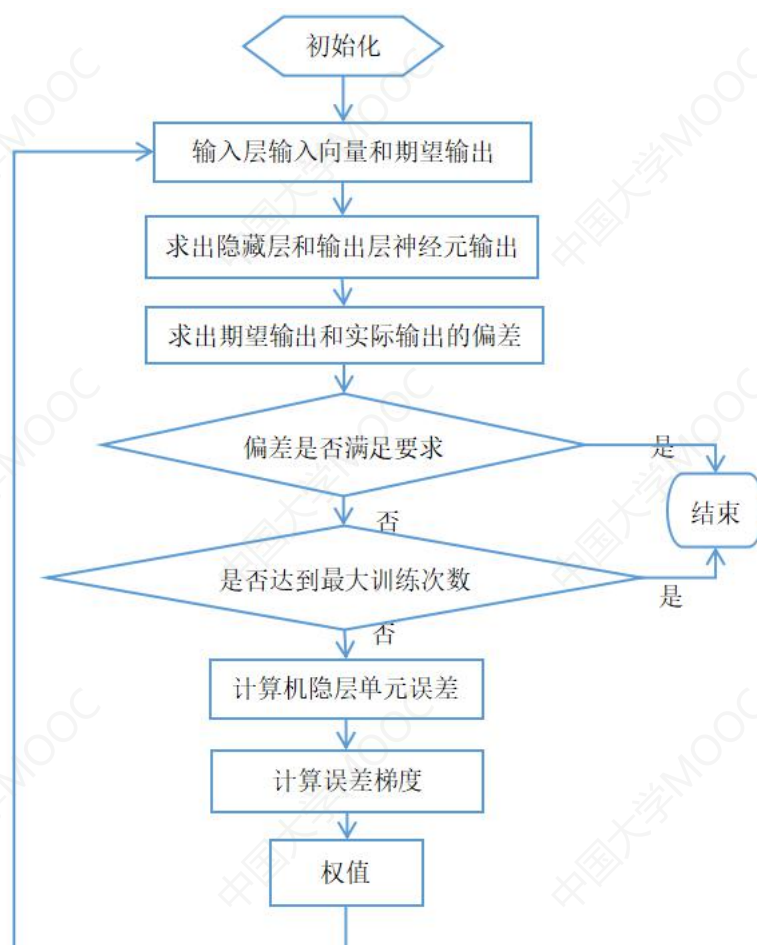


图 9 神经网络流程图

模型求解得到：

部分企业的求解结果：

表 11 神经网络判别预测结果

企业代号	信誉等级	预测	企业代号	信誉等级	预测
5	B	B	60	B	C
9	A	A	62	B	B
14	C	A	76	B	B
16	A	B	78	C	C
31	A	A	80	C	C
32	B	C	90	C	B
37	B	B	97	B	B
38	B	B	104	C	C
39	C	C	116	D	D

结论：由上表测试集预测得到的结果表明神经网络判别模型的误判率为 27.78%，与 Fisher 判别分析比较而言，误判率更低，结果更精确。原因在于神

神经网络对复杂系统有较好的解决方式，所以我们对附件二中的 302 家企业采取神经网络进行判别。

302 家企业 ABCD 信誉等级的占比：

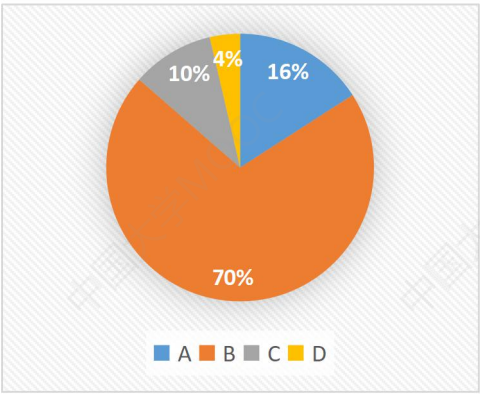


图 10 信誉等级分布

302 家企业(部分)信誉等级如下表所示：

表 12 企业信誉等级

企业代号	信誉等级	企业代号	信誉等级
187	D	194	B
188	A	195	B
189	B	196	B
190	B	197	B
191	B	198	B
192	C	199	A
193	B	200	B

6.2.3 模型的求解

(1) 熵权法求解（风险评估得分）

对 302 家无信贷风险的企业同样进行数据预处理和熵信息构权法分析，求出企业实力、企业信誉、是否违约的权重系数：

表 13 无贷款企业风险各指标熵值

	企业实力	企业信誉	是否违约
风险熵值	0.954	0.989	0.052
风险权重	0.768	0.180	0.142

根据企业实力得分和企业信誉得分按一定熵权比值算出风险评估得分，风险评估得分越高，表现该企业违约的概率越小。

表 14 企业风险评估得分

企业代号	企业实力得分	企业信誉得分	风险评估得分
E214	1.00	0.50	1
E349	0.74	1.00	0.869
E193	0.71	0.50	0.725
E273	0.61	1.00	0.751
...
E335	0.03	0.33	0.028

E377	0.02	0.33	0.018
E399	0.02	0.33	0.0162
E397	0.00	0.33	0

(2) 信贷策略的求解

①对 D 评级的 10 个企业不发放贷款

②确定其他企业的贷款金额和利率

对 302 家企业各项指标进行分析, 通过 matlab 求解得到下表, 以贷款金额排序, 左边一列企业贷款金额排前 10, 右边一列企业贷款金额排名后 10。

表 15 信贷策略 (部分)

企业号	贷款金额	利率	企业号	贷款金额	利率
214	1000000.00	0.0402	413	158819.90	0.0778
349	1000000.00	0.0402	392	155366.84	0.0778
193	1000000.00	0.0402	368	154299.25	0.0778
273	1000000.00	0.0546	411	144842.21	0.0778
194	1000000.00	0.0546	395	142380.31	0.0778
336	1000000.00	0.0546	407	138937.11	0.0778
125	904490.11	0.0546	335	136276.87	0.0778
414	723696.88	0.0546	377	123520.26	0.0778
192	706056.78	0.0546	399	120496.24	0.0778
238	701739.36	0.0546	397	100000.00	0.0778

③银行最大效益 3902259 元。

6.3 问题三模型的建立与求解

突发事件包括自然灾害、事故灾难、公共卫生事件和社会安全事件, 通常会影响企业的生产经营和经济效益。影响程度会根据企业类别、规模等发生改变。现我们对附件 2 中 302 家企业进行分类, 分成以下六个行业:

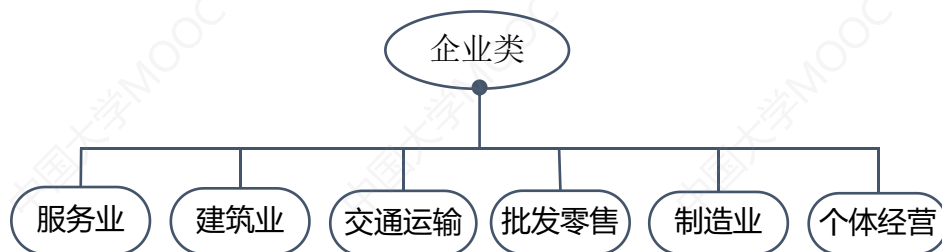


图 11 企业类型

新冠疫情和经济危机对各行业的冲击程度不同, 现仅选择经济方面的指标对各行业冲击程度进行评定。

为展现突发事件对经济的影响, 对于新冠疫情, 查找有关前 5 个行业 2015 年-2020 年第一季度的生产总值, 并运用 SARIMA 模型对 2020 年第一季度的生产总值进行预测。对于经济危机, 我们选取 2004 年-2009 年的企业生产总值。

我们引入“相对变化率”表现企业生产总值的变化率, 变化率越大, 表现该企业受突发事件的冲击力越大, 企业经营不稳定性越大。

对于个体经营产业, 种类多但是规模小, 所以单独分为一类, 它的影响程度

为前 5 个行业的平均影响程度。因此，银行放贷的决定因素改变为企业实力、企业信誉、突发事件影响程度。解题流程图为：

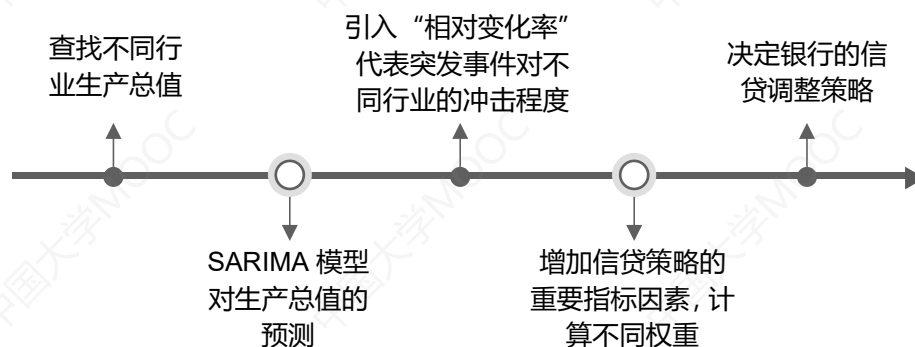


图 12 流程图

6.3.1 时间序列模型的建立与求解（详见代码十九——二十五）

由于行业种类太多，我们仅选取制造业和交通运输业进行展示。

· 无突发事件影响下的各行业生产总值预测模型

结合数据图像分析，我们发现每个行业的产值都具有明显的季节性，且数据均为时间序列，SARIMA(p, q)模型通用表达式为

$$Y_t = \sum_{i=1}^p \varphi_i Y_{t-i} + \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \mu_t \quad (20)$$

其中 Y_t 为 t 期时间序列变量， φ_i 为自回归参数， θ_j 为滑动平均参数， p 为自回归阶数， q 为滑动平均次数， μ_t 为随机变量， ε_{t-j} 为 $t-j$ 期时间序列变量。

· 建立 SARIMA 模型分析时序步骤：

Step1: 画出各行业生产总值的时序图，判断该时间序列是否为平稳状态

Step2: 将非平稳序列通过差分处理变为平稳序列

Step3: 画出差分后的生产总值的时序图、相关图，算出单位根，检验该序列是否平稳，如果还是非平稳，再进行一次差分。

Step4: 根据相关图、AIC 准则选定 SARIMA 模型，算出结果，并画出时序图。

Step5: 画出预测时序图。

· 原始数据的时序图

SARIMA 模型的处理对象必须是平稳的，我们分别对新冠疫情（2015-2020 年）影响下各行业产值作图观察，对经融危机（2004-2008 年）影响下各行业产值作图观察，如下图所示（选取部分）：

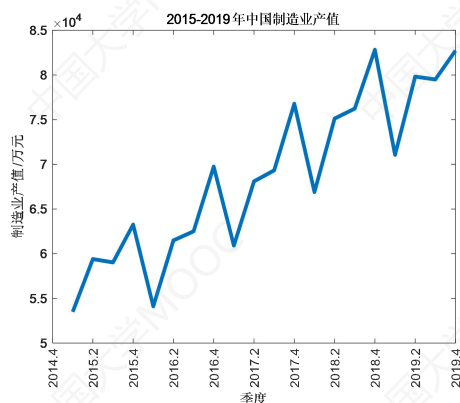


图 13: 2015-2020 年制造业产值

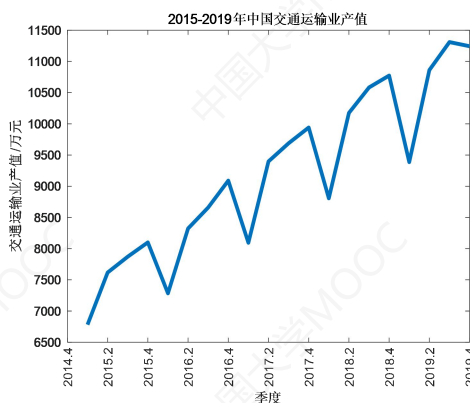


图 14: 2015-2019 年交通运输业产值

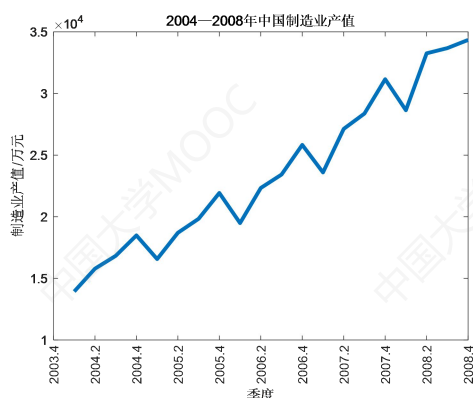


图 15: 2004-2008 年制造业产值

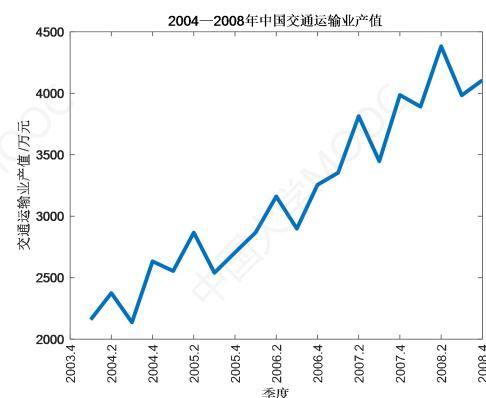


图 16: 2004-2008 年交通运输业产值

从上图可以看出制造业和交通运输业产值随时间变化而变化，总体呈现波动上升趋势且具有明显季节性，说明时间序列并非稳定序列。

· 通过变换获得平稳序列

由于各行业产值随时间是不平稳变化的，因此需要将原始时间序列按照一定方法重新组合，形成新的时间序列。

我们采用差分分析法对数据进行重新组合，其中季节差分表达式为：

$$Y_t = X_{n+t} - X_t, t=1,2,\dots \quad n \text{ 为季节性周期} \quad (21)$$

一阶差分表达式为：

$$Q_t = Y_{t+1} - Y_t, t=1,2,\dots \quad (22)$$

在 MATLAB 中对各行业产值进行差分分析，包括季节差分和一阶差分，结果显示数据需要在季节差分后再进行二阶差分。

· 确定变换后平稳序列的 SARIMA 模型

要对时间序列进行预测，需分别确定各行业 GDP 的变动函数，对此通过 SARIMA 模型求解得到。其中的工作任务流程为：

1、确定 p 和 q ，确立 SARIMA (p, q) 模型；

2、确定参数，构建函数关系式。

i. p, q 确定

由于样本数据不够大，适合采用 AIC 准则。AIC 表达式为

$$AIC = \log \sigma^2 + \frac{m}{n} \log n \quad (23)$$

其中 σ^2 是残差的方差, m 是自变量个数, 即 $m = p + q$, n 是样本个数。

初步确定 p, q 范围为 $p = 1, 2, 3; q = 1, 2, 3$, 通过热图可以清晰的观察并估计预测模型的阶数。

ii. 参数确定

在建立 SARIMA (p, q) 模型的基础上, 求解出相关的模型参数, 结果如表所示:

表 16 新冠肺炎下交通运输业的模型参数

参数类型	系数	标准误差	概率 P 值
AR{1}	1.6004	0.21189	4.2547e-14
AR{2}	-0.69664	0.1605	1.4217e-05
MA{1}	-1	0.40464	0.013461

表 17 经融危机下交通运输业的模型参数

参数类型	系数	标准误差	概率 P 值
AR{1}	0.79552	0.079452	1.3425e-23
MA{2}	-1	0.14746	1.1897e-11

根据检验值 P 值, 可以分析得到 SARIMA 模型的显著性较好。

· 进行预测

通过 SARIMA (p, q) 模型, 对疫情影响 (2020 年第一季度) 情况下各行业产值进行预测, 对经融危机 (2009-2011 年) 情况下各行业产值进行预测, 预测结果如下表所示 (仅部分):

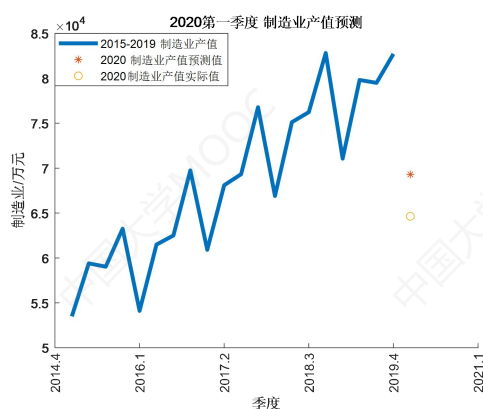


图 17: 2020 第一季度制造业产值预测

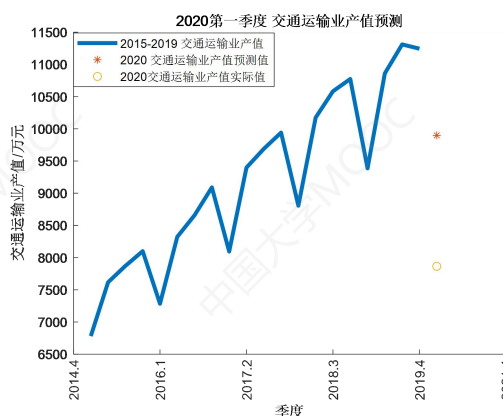


图 18: 2020 第一季度交通运输业产值预测

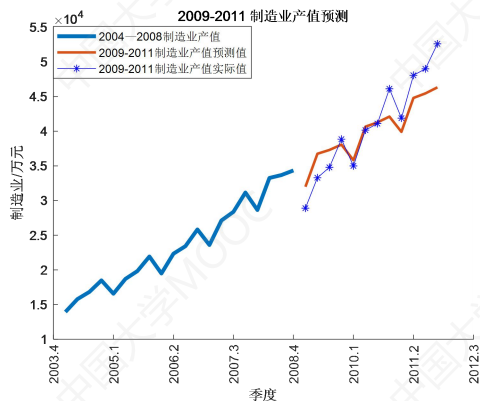


图 19: 2009-2011 制造业产值预测

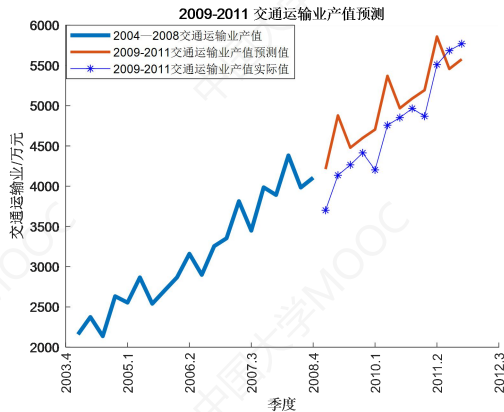


图 20: 2009-2011 交通运输业产值预测

在 6 个大类行业中，通过预测图来看，各个行业的现象比较类似，所以我们仅选取制造业、交通运输业作为主要研究对象，分别得出新冠疫情和经融危机影响下这两个行业的预测产值与实际产值。

表 18 突发事件对行业产值的预测（万元）

	新冠疫情		金融危机	
	预测产值	实际产值	预测产值	实际产值
制造业	69300	64642	31994.6176	28905.4
交通运输业	9898.2	7865.1	4212.0758	3701.1

6.3.2 相对变化率

用 ARIMA 模型算出各行业不受突发事件影响行业的预测产值，与受影响后的实际行业产值进行对比。我们引入“相对变化率”表现生产总值的相对变化程度，相对变化率越高表明企业受突发事件的影响程度越大，企业发展越不稳定，违约概率越大。

相对变化率：

$$\xi = \frac{|\delta_{\text{实}} - \delta_{\text{预}}|}{\delta_{\text{预}}} \quad (24)$$

其中， $\delta_{\text{实}}$ 表示实际的生产总值， $\delta_{\text{预}}$ 表示预测的生产总值。

相对变化率代表的是每个行业的不稳定程度，所以这个行业包含的所有企业相对变化率相同，即发展稳定性得分相同。

求解得到各行业的相对变化率为：

表 19 各行业相对变化率

新冠疫情影响（2015-2020）		经济危机（2004-2008）	
行业	相对变化率	行业	相对变化率
建筑业	0.209	建筑业	0.4266
交通运输业	0.2054	交通运输业	0.1213
批发零售业	0.2183	批发零售业	0.1010
制造业	0.0672	制造业	0.0966
个体经营	0.1431	个体经营	0.1501
服务业	0.0154	服务业	0.0005

6.3.3 信贷调整策略（详见程序二十六）

信贷调整策略不仅考虑企业实力、信贷风险，还加入了相对变化率作为决定贷款额度的标准，由此各企业的贷款额度便会发生变化。

· 贷款额度

贷款额度原本根据企业信贷风险得分和实力得分进行评判，在信贷调整策略中我们加入了相对变化率这一因素，将这三者因素结合重新求得一综合得分。

$$Z_{\text{总}b} = \ell_1 Z_{\text{信}b} + \ell_2 Z_{\text{实}b} + \ell_3 Z_{\text{变}b}, \quad b = 124, 125, \dots, 425 \quad (25)$$

对某家企业的放贷额度

$$L_b = \frac{Z_{\text{总}b}}{\sum_{b=124}^{425} Z_{\text{总}b}} \times K, \quad b = 124, 125, \dots, 425 \quad (26)$$

其中 $Z_{\text{总}b}$ 表示贷款额度得分， $Z_{\text{信}b}$ 表示企业风险得分， $Z_{\text{实}b}$ 表示企业实力得分， L_b 表示放贷额度， K 表示除去基础额度的总额， $Z_{\text{变}b}$ 表示相对变化率得分， ℓ_1, ℓ_2, ℓ_3 表示每个指标的权重。

在贷款额度发生变化后，将变化后的贷款额度代入问题一中的贷款利率优化模型便可得到新的信贷策略。下表为调整后的信贷策略：

a. 新冠疫情下信贷调整策略

以贷款金额排序，左边一列为金额排名前 10 的企业，右边一列是金额排名后 10 的企业。

表 20 新冠疫情信贷调整策略

企业号	贷款金额	利率	企业号	贷款金额	利率
349	815043.1753	0.0474	276	111697.118	0.0532
194	775014.4206	0.0517	420	132660.1201	0.0532
167	730803.6127	0.0517	357	115453.672	0.0532
134	719079.3617	0.0517	378	111105.644	0.0532
171	718335.5176	0.0517	388	196623.3788	0.0532
360	711967.0101	0.0517	338	195993.5284	0.0532
268	708618.6625	0.0517	391	194702.0497	0.0532
248	702516.2976	0.0517	335	107933.9631	0.0549
225	700310.2995	0.0517	399	106188.2579	0.0549
330	698989.5038	0.0517	397	100000	0.0549

银行最大效益：4026031 元。

b. 经融危机下信贷调整策略

以贷款金额排序，左边一列为金额排名前 10 的企业，右边一列是金额排名后 10 的企业。

表 21 经融危机信贷调整策略

企业号	贷款金额	利率	企业号	贷款金额	利率
349	815043.1753	0.0474	338	109458.1204	0.0531
194	782676.5032	0.0516	386	109251.8002	0.0531
167	746928.2834	0.0516	276	109087.7741	0.0535
134	737448.2236	0.0516	391	108979.8631	0.0535
171	736846.7621	0.0516	413	108560.3594	0.0535
360	731697.279	0.0516	411	108051.0714	0.0549
268	728989.8536	0.0516	407	107006.8005	0.0549
248	724055.5696	0.0516	335	106415.2878	0.0549
225	722271.8315	0.0516	399	101595.8943	0.0549
330	721203.8551	0.0516	397	100000	0.0549

银行最大效益：3864359 元

七、模型的检验

本文各企业的产值均构成时间序列。对于一个时间序列的分析，首先要判断是否是平稳时间序列，即看均值和方差是否随时间的变化而变化，对原数据进行季节差分单位根检验。 $h=0$ 表示接受原假设，即 X 不平稳； $h=1$ 表示拒绝原假设，即可以认为 X 平稳； p 为统计量的相伴概率；若不平稳则需进行差分计算，继续检验。我们以零售业为例：

表 22 模型的检验

2015-2019 零售	h	p
原序列	0	0.9634
一阶季阶差分	0	0.7177
一阶差分	1	0.0027
白噪声	0	0.9952

由以上差分计算后 h 值得到零售业产值差分检验结果都是平稳的，故零售业产值构成时间序列模型。白噪声序列 $h=0$ 表示接受原假设，零售业的时间残差序列均为白噪声序列，不存在自相关性。

八、模型的评价

8.1 模型的优点

- 1、Fisher 模型的优点在于没有做分布假设，模型适用度高。
- 2、引入“相对变化率”，较好地体现突发事件对行业生产总值的影响程度，模型简单，可操作性强，且具有较强的适用性、灵活性，易于实现。

8.2 模型的缺点

- 1、信贷风险评估模型的指标很多，但是指标的相关性都不高，指标代表性不强。
- 2、由于新冠疫情的全方面性，我们仅选取各行业生产总值作为影响经济的指标，将模型处于局限状态，影响力评估存在一定的不合理性。

九、 模型的应用与推广

本文主要通过对企业交易票据信息的分析,建立了企业风险评估和信贷策略模型,研究了使银行获益最大和客户流失率越少的双目标信贷策略规划问题。此类方法不仅适用于银行向中小微企业放贷的研究,也适用于对大型企业贷款和个人贷款方面的研究。

参考文献

[1] 米帅, 军习勤, 概率神经网络在判别分析中的比较优势[J]. 统计与决策, 2010, (2):17-19

附录 matlab2018a lingo11

代码:

问题一:

程序一、废票进项占比:

```
clc;clear;
data=xlsread('shuju1.xlsx','进项替换后');
x=data(:,1);
y=data(:,9);
sumcount=zeros(123,1);
zerocount=zeros(123,1);
for j=1:123
    for i=1:length(data)
        if x(i)==j
            sumcount(j)=sumcount(j)+1;
            if y(i)==0
                zerocount(j)=zerocount(j)+1;
            end
        end
    end
end
result=zerocount./sumcount;
```

程序二、废票销项占比:

```
clc;clear;
data=xlsread('shuju1.xlsx','销项替换后');
x=data(:,1);
y=data(:,9);
sumcount=zeros(123,1);
zerocount=zeros(123,1);
for j=1:123
    for i=1:length(data)
        if x(i)==j
            sumcount(j)=sumcount(j)+1;
            if y(i)==0
                zerocount(j)=zerocount(j)+1;
            end
        end
    end
end
result=zerocount./sumcount;
```

程序三、企业经营时长:

```
clc;clear;
data=xlsread('各企业各单距今天数.xlsx');
```

```

x=data(:,1);
y=data(:,2);
yy=[];
mm=zeros(123,1);
for j=1:123
    for i=1:length(data)
        if x(i)==j
            yy=[yy;y(i)];
            mm(j)=max(yy);
        end
    end
    yy=[];
end
end

```

程序四、税收额：

```

clc;clear;
data1=xlsread('shuju1.xlsx','进项替换后');
data2=xlsread('shuju1.xlsx','销项替换后');
x=data1(:,1);a=data2(:,1);
y=data1(:,9);b=data2(:,9);
x1=data1(:,6);
x2=data2(:,6);
sumshuishou1=zeros(123,1);
sumshuishou2=zeros(123,1);
for j=1:123
    for i=1:length(data1)
        if x(i)==j&&y(i)~=0
            sumshuishou1(j)=sumshuishou1(j)+x1(i);
        end
    end
end
for j=1:123
    for i=1:length(data2)
        if a(i)==j&&b(i)~=0
            sumshuishou2(j)=sumshuishou2(j)+x2(i);
        end
    end
end
sumshuishoue=sumshuishou2-sumshuishou1;
k=find(sumshuishoue<0);
sumshuishoue(k)=0;

```

程序五、交易单数：

```

clc;clear;

```

```

data1=xlsread('各企业进销项有效正(1).xlsx','进项有效正');
data2=xlsread('各企业进销项有效正(1).xlsx','销项有效正');
x1=data1(:,1);
x2=data2(:,1);
count1=zeros(123,1);
count2=zeros(123,1);
for j=1:123
    for i=1:length(data1)
        if x1(i)==j
            count1(j)=count1(j)+1;
        end
    end
end
for j=1:123
    for i=1:length(data2)
        if x2(i)==j
            count2(j)=count2(j)+1;
        end
    end
end
end

```

程序六、总成本：

```

clc;clear;
data=xlsread('shuju1.xlsx','进项替换后');
x=data(:,1);
y=data(:,9);
chengben=data(:,7);
sumchengben=zeros(123,1);
for j=1:123
    for i=1:length(data)
        if x(i)==j&& y(i)~=0
            sumchengben(j)=sumchengben(j)+chengben(i);
        end
    end
end
end

```

程序七、总收入/收益：

```

%clc;clear;
%在总成本求得以后
data=xlsread('shuju1.xlsx','销项替换后');
x=data(:,1);
y=data(:,9);
shouru=data(:,7);

```

```

sumshouru=zeros(123,1);ll=[];mm=zeros(123,1);ss=zeros(123,1);
for j=1:123
    for i=1:length(data)
        if x(i)==j&&y(i)~=0
            sumshouru(j)=sumshouru(j)+shouru(i);
            ll=[ll;shouru(i)];
        end
    end
    mm(j)=mean(ll);
    ss(j)=sum((ll-mean(ll)).^2);
    ll=[];
end
shouyi1=sumshouru-sumchengben;%绝对收益
shouyi2=(sumshouru-sumchengben)./sumchengben;%相对收益

```

程序八、熵信息构权（实力和风险评分）：

```

clc;clear;
data=xlsread('附件一整合后数据.xlsx');
zhibiao=data(:,[3:5,7,9,12,20]);
zhibiao=(zhibiao-min(zhibiao))./(max(zhibiao)-min(zhibiao));
%% 熵信息构权 实力
x=zhibiao;
[n,p]=size(x);
pp=x./sum(x);
k=1/log(n);%n 表示数据个数，p 是指标个数
p1=zeros(n,p);
for i=1:n
    for j=1:p
        if pp(i,j)~=0
            p1(i,j)=pp(i,j)*log(pp(i,j));
        end
    end
end
I=-k*sum(p1);
r=1-I;
w=r./sum(r);
data1=w.*zhibiao;
data2=sum(data1,2);
[Y,M]=sort(data2);
zzz=(data2-min(data2))./(max(data2)-min(data2));
%% 风险评估
zhibiao2=[zzz,1./data(:,21),data(:,22)];
x=zhibiao2;
[n,p]=size(x);

```

```

pp=x./sum(x);
k=1/log(n);%n 表示数据个数， p 是指标个数
p1=zeros(n,p);
for i=1:n
    for j=1:p
        if pp(i,j)~=0
            p1(i,j)=pp(i,j)*log(pp(i,j)) ;
        end
    end
end
I=-k*sum(p1);
r=1-I;
w=r./sum(r);
data1=w.*zhibiao2;
data2=sum(data1,2);
[Y,M]=sort(data2);
zz=(data2-min(data2))./(max(data2)-min(data2));

```

程序九、回归：

```

clc;clear;
data=xlsread('附件 3： 银行贷款年利率与客户流失率关系的统计数据.xlsx');
y=data(:,1);
x=data(:,2:4);
titleName=["信誉等级为 A 的银行贷款年利率与客户流失率关系的散点图","信誉等级为 B 的银行贷款年利率与客户流失率关系的散点图","信誉等级为 C 的银行贷款年利率与客户流失率关系的散点图"];
for i=1:3
    figure,scatter(y,x(:,i));
    title(titleName(i));
    xlabel('客户流失率');
    ylabel('贷款年利率');
end

Y=[ones(29,1),sqrt(y),y];
[b,bint,r,rint,stats]=regress(x(:,3),Y);
i=0:0.02:0.5;
z=b(1)+b(2)*sqrt(i)+b(3)*i;
figure,plot(i,z,'b*-');
title('信誉等级为 C 的银行贷款年利率与客户流失率关系的回归曲线');
xlabel('客户流失率');
ylabel('贷款年利率');

```

程序十、信贷策略：

Lingo 代码见支撑材料中 ques_1_1 和 ques_1_2

问题二：

程序十一、废票进项占比：

```
clc;clear;
data=xlsread('aaa.xlsx','进项发票替换后');
x=data(:,1);
y=data(:,6);
sumcount=zeros(302,1);
zerocount=zeros(302,1);
for j=124:425
    for i=1:length(data)
        if x(i)==j
            sumcount(j-123)=sumcount(j-123)+1;
            if y(i)==0
                zerocount(j-123)=zerocount(j-123)+1;
            end
        end
    end
end
result=zerocount./sumcount;
```

程序十二、废票销项占比：

```
clc;clear;
data=xlsread('aaa.xlsx','销项发票替换后');
x=data(:,1);
y=data(:,7);
sumcount=zeros(302,1);
zerocount=zeros(302,1);
for j=124:425
    for i=1:length(data)
        if x(i)==j
            sumcount(j-123)=sumcount(j-123)+1;
            if y(i)==0
                zerocount(j-123)=zerocount(j-123)+1;
            end
        end
    end
end
result=zerocount./sumcount;
```

程序十三、总成本：

```
clc;clear;
```

```

data=xlsread('shuju1.xlsx','进项替换后');
x=data(:,1);
y=data(:,9);
chengben=data(:,7);
sumchengben=zeros(123,1);
for j=1:123
    for i=1:length(data)
        if x(i)==j&&y(i)~=0
            sumchengben(j)=sumchengben(j)+chengben(i);
        end
    end
end
end

```

程序十四、总收入/收益:

```

%clc;clear;
%在总成本求得以后
data=xlsread('aaa.xlsx','销项发票替换后');
x=data(:,1);
y=data(:,7);
shouru=data(:,5);
sumshouru=zeros(302,1);ll=[];mm=zeros(302,1);ss=zeros(302,1);
for j=124:425
    for i=1:length(data)
        if x(i)==j&&y(i)~=0
            sumshouru(j-123)=sumshouru(j-123)+shouru(i);
            ll=[ll;shouru(i)];
        end
    end
    mm(j-123)=mean(ll);
    ss(j-123)=sum((ll-mean(ll)).^2);
    ll=[];
end
shouyi1=sumshouru-sumchengben;%绝对收益
shouyi2=(sumshouru-sumchengben)./sumchengben;%相对收益

```

程序十五、税收额

```

clc;clear;
data1=xlsread('aaa.xlsx','进项有效');
data2=xlsread('aaa.xlsx','销项有效');
x=data1(:,1);a=data2(:,1);
y=data1(:,6);b=data2(:,6);
x1=data1(:,4);
x2=data2(:,4);
sumshuishou1=zeros(302,1);

```

```

sumshuishou2=zeros(302,1);
for j=124:425
    for i=1:length(data1)
        if x(i)==j&&y(i)~=0
            sumshuishou1(j-123)=sumshuishou1(j-123)+x1(i);
        end
    end
end
for j=124:425
    for i=1:length(data2)
        if a(i)==j&&b(i)~=0
            sumshuishou2(j-123)=sumshuishou2(j-123)+x1(i);
        end
    end
end
sumshuishoue=sumshuishou2-sumshuishou1;
k=find(sumshuishoue<0);
sumshuishoue(k)=0;

```

程序十六、神经网络：

```

clc
clear all
close all
data=xlsread('附件一整合后数据.xlsx');
p=data(:,[3:5,15,18,20]);
t=data(:,21);
p=p';
t=t';
[p1,ps]=mapminmax(p);
[t1,ts]=mapminmax(t);
[tsample.p,vsample.p,testsamle.p]=dividerand(p,0.7,0.15,0.15);
[tsample.t,vsample.t,testsamle.t]=dividerand(t,0.7,0.15,0.15);

suanfa1='tansig';suanfa2='purelin';
net=newff(p,t,3,{suanfa1 suanfa2},'traingdm');%网络创建,3 为隐层神经元的个数,
根据 log2n, n 为输入层节点数
%网络参数的设置
net.trainParam.epochs=10000;%训练次数设置
net.trainParam.goal=1e-7;%训练目标设置
net.trainParam.lr=0.01;%学习率设置,应设置为较少值,太大虽然会在开始加快收敛速度,但临近最佳点时,会产生动荡,而致使无法收敛
net.trainParam.mc=0.9;%动量因子的设置,默认为 0.9
net.trainParam.show=25;%显示的间隔次数

```



```

net.trainFcn='trainlm';
[net,tr]=train(net,tsample.p,tsample.t);

[normtrainoutput,trainPerf]=sim(net,tsample.p,[],[],tsample.t);%训练的数据，根据
BP 得到的结果
[normvalidateoutput,validatePerf]=sim(net,vsample.p,[],[],vsample.t);%验证的数据，
经 BP 得到的结果
[normtestoutput,testPerf]=sim(net,testsample.p,[],[],testsample.t);%测试数据，经 BP
得到的结果
%将所得的结果进行反归一化，得到其拟合的数据
toutput=mapminmax('reverse',normtrainoutput,ts);
voutput=mapminmax('reverse',normvalidateoutput,ts);
testoutput=mapminmax('reverse',normtestoutput,ts);
%正常输入的数据的反归一化的处理，得到其正式值
tvalue=mapminmax('reverse',tsample.t,ts);%正常的验证数据
vvalue=mapminmax('reverse',vsample.t,ts);%正常的验证的数据
testvalue=mapminmax('reverse',testsample.t,ts);%正常的测试数据
error=(testoutput'-testvalue');
result=[testvalue',testoutput',error];
meanerror=sum(abs(result(:,3)))/length(result);
%% 做预测，输入要预测的数据 pnew
zzz=xlsread('附件二整合后数据.xlsx');
shuru=zzz(:,[2:4,13:15]);
shurgy=mapminmax(shuru);
shuchugy=sim(net,shurgy);
shuchufgy=mapminmax('reverse',shuchugy,ts);

```

程序十七、熵信息构权（实力，风险）：

```

clc;clear;
data=xlsread('附件二整合后数据.xlsx');
zhibiao=data(:,[2:4,6,8,10,12,15]);
zhibiao=(zhibiao-min(zhibiao))./(max(zhibiao)-min(zhibiao));
%% 熵信息构权 实力
x=zhibiao;
[n,p]=size(x);
pp=x./sum(x);
k=1/log(n);%n 表示数据个数，p 是指标个数
p1=zeros(n,p);
for i=1:n
    for j=1:p
        if pp(i,j)~=0
            p1(i,j)=pp(i,j)*log(pp(i,j));
        end
    end
end

```

```

end
I=-k*sum(p1);
r=1-I;
w=r./sum(r);
data1=w.*zhibiao;
data2=sum(data1,2);
[Y,M]=sort(data2);
zzz=(data2-min(data2))./(max(data2)-min(data2));
%% 风险评估
zhibiao2=[zzz,1./data(:,16),1./data(:,17)+1)];
x=zhibiao2;
[n,p]=size(x);
pp=x./sum(x);
k=1/log(n),%n 表示数据个数，p 是指标个数
p1=zeros(n,p);
for i=1:n
    for j=1:p
        if pp(i,j)~=0
            p1(i,j)=pp(i,j)*log(pp(i,j));
        end
    end
end
end
I=-k*sum(p1);
r=1-I;
w=r./sum(r);
data1=w.*zhibiao2;
data2=sum(data1,2);
[Y,M]=sort(data2);
zz=(data2-min(data2))./(max(data2)-min(data2));

```

程序十八、信贷策略：

Lingo 代码见支撑材料 ques_2_2

问题三：

时间序列模型：

新冠影响下：

程序十九、服务业：

```

clc
clear
y1=xlsread('几大行业产值(1).xlsx',4);
y=y1(1:20,2);
figure(1),
plot(y,'linewidth',3)
set(gca,'Xticklabel',{'2014.4','2015.2','2015.4','2016.2','2016.4','2017.2','2017.4','2018.

```

```

2','2018.4','2019.2','2019.4'});
set(gca,'XTickLabelRotation',90);
xlabel('季度');
ylabel('服务业产值/万元');
title('2015-2019 年中国服务业产值');
[h,p]=adftest(y)
diff1_y=y(5:end)-y(1:end-4);
figure(2),
plot(diff1_y,'linewidth',3)
set(gca,'Xticklabel',{'2014.4','2015.2','2015.4','2016.2','2016.4','2017.2','2017.4','2018.
2','2018.4','2019.2','2019.4'});
set(gca,'XTickLabelRotation',90);
xlabel('季度');
ylabel('服务业产值/万元');
title('2015-2019 年中国服务业产值一阶季节差分');
[h,p]=adftest(diff1_y)

diff_y=diff(diff1_y);
figure(3),
plot(diff_y,'linewidth',3)
set(gca,'Xticklabel',{'2014.4','2015.2','2015.4','2016.2','2016.4','2017.2','2017.4','2018.
2','2018.4','2019.2','2019.4'});
set(gca,'XTickLabelRotation',90);
xlabel('季度');
ylabel('服务业产值万元');
title('2015-2019 年中国服务业产值一阶差分');
[h,p]=adftest(diff_y)

maxLags = 3;
AICSet = zeros(maxLags, maxLags);
for i = 1:maxLags
for j = 1:maxLags
mdl = arima('ARLags', [1:i], 'MALags', [1:j]);
[EstMdl, EstParamCov, LogL, info] = estimate(mdl, diff1_y);
AICSet(i, j)= aicbic(LogL, length(info.X), length(diff1_y));
end
end
figure(4), heatmap(AICSet/1000)
title('2015-2019 服务业产值热图');

mdl = arima('Constant',0,'Seasonality',4,'D', 0,'MALags',2);
[EstMdl] = estimate(mdl, y);

res = infer(EstMdl,y);

```

```

figure(6),
%subplot(2,2,1), plot(res./sqrt(EstMdl.Variance))
qqplot(res)
%subplot(2,2,3), autocorr(res)
%subplot(2,2,4), parcorr(res)
[h, p] = lbqtest(res)

[yF, yMSE] = forecast(EstMdl,1,'Y0',y);
UB = yF + 1.96*sqrt(yMSE); % 95%置信区间上限
LB = yF - 1.96*sqrt(yMSE); % 95%置信区间下限
figure(7), hold on
plot(y,'linewidth',3)
hold on
plot(21,yF,'*');
hold on,plot(21,y1(21,2),'o');
legend('2015-2019 服务业产值','2020 服务业产值预测值','2020 服务业产值实际
值')
xlabel('季度')
ylabel('服务业/万元')
set(gca,'Xticklabel',{'2014.4','2016.1','2017.2','2018.3','2019.4','2021.1'});
set(gca,'XTickLabelRotation',90);
title('2020 第一季度 服务业产值预测')
chazhi=abs(y1(21,2)-yF);
difference=chazhi/yF;%0.0154

```

程序二十、建筑业：

```

clc
clear
y1=xlsread('几大行业产值(1).xlsx');
y=y1(1:20,2);
figure(1),
plot(y,'linewidth',3)
set(gca,'Xticklabel',{'2014.4','2015.2','2015.4','2016.2','2016.4','2017.2','2017.4','2018.
2','2018.4','2019.2','2019.4'});
set(gca,'XTickLabelRotation',90);
xlabel('季度');
ylabel('建筑业产值/万元');
title('2015-2019 年中国建筑业产值');
[h,p]=adftest(y)
diff1_y=y(5:end)-y(1:end-4);
figure(2),
plot(diff1_y,'linewidth',3)
set(gca,'Xticklabel',{'2014.4','2015.2','2015.4','2016.2','2016.4','2017.2','2017.4','2018.
2','2018.4','2019.2','2019.4'});

```

```

set(gca,'XTickLabelRotation',90);
xlabel('季度');
ylabel('建筑业产值/万元');
title('2015-2019 年中国建筑业产值一阶季节差分');
[h,p]=adftest(diff1_y)

maxLags = 2;
AICSet = zeros(maxLags, maxLags);
for i = 1:maxLags
    for j = 1:maxLags
        mdl = arima('ARLags', [1:i], 'MALags', [1:j]);
        [EstMdl, EstParamCov, LogL, info] = estimate(mdl, diff1_y);
        AICSet(i, j)= aicbic(LogL, length(info.X), length(diff1_y));
    end
end
figure(4), heatmap(AICSet/1000)
title('2015-2019 建筑业产值热图');

mdl = arima('Constant',0,'Seasonality',4,'D', 0,'ARLags',1:2)
[EstMdl] = estimate(mdl, y);

res = infer(EstMdl,y);
figure(6),
%subplot(2,2,1), plot(res./sqrt(EstMdl.Variance))
qqplot(res)
%subplot(2,2,3), autocorr(res)
%subplot(2,2,4), parcorr(res)
[h, p] = lbqtest(res)

[yF, yMSE] = forecast(EstMdl,1,'Y0',y);
UB = yF + 1.96*sqrt(yMSE); % 95%置信区间上限
LB = yF - 1.96*sqrt(yMSE); % 95%置信区间下限
figure(7), hold on
plot(y,'linewidth',3)
hold on
plot(21,yF,'*');
hold on,plot(21,y1(21,2),'o');
legend('2015-2019 建筑业产值','2020 建筑业产值预测值','2020 建筑业产值实际值')
xlabel('季度')
ylabel('建筑业/万元')
set(gca,'Xticklabel',{'2014.4','2016.1','2017.2','2018.3','2019.4','2021.1'});
set(gca,'XTickLabelRotation',90);
title('2020 第一季度 建筑业产值预测')

```

```

chazhi=abs(y1(21,2)-yF);
difference=chazhi/yF;%0.2090
程序二十一、交通运输业：
clc
clear
y1=xlsread('几大行业产值(1).xlsx',3);
y=y1(1:20,2);
figure(1),
plot(y,'linewidth',3)
set(gca,'Xticklabel',{'2014.4','2015.2','2015.4','2016.2','2016.4','2017.2','2017.4','2018.
2','2018.4','2019.2','2019.4'});
set(gca,'XTickLabelRotation',90);
xlabel('季度');
ylabel('交通运输业产值/万元');
title('2015-2019 年中国交通运输业产值');
[h,p]=adftest(y)
diff1_y =y(5:end)-y(1:end-4);
figure(2),
plot(diff1_y,'linewidth',3)
set(gca,'Xticklabel',{'2014.4','2015.2','2015.4','2016.2','2016.4','2017.2','2017.4','2018.
2','2018.4','2019.2','2019.4'});
set(gca,'XTickLabelRotation',90);
xlabel('季度');
ylabel('交通运输业产值/万元');
title('2015-2019 年中国交通运输业产值一阶季节差分');
[h,p]=adftest(diff1_y)

diff_y=diff(diff1_y);
figure(3),
plot(diff_y,'linewidth',3)
set(gca,'Xticklabel',{'2014.4','2015.2','2015.4','2016.2','2016.4','2017.2','2017.4','2018.
2','2018.4','2019.2','2019.4'});
set(gca,'XTickLabelRotation',90);
xlabel('季度');
ylabel('交通运输业产值/万元');
title('2015-2019 年中国交通运输业产值一阶差分');
[h,p]=adftest(diff_y)

maxLags = 3;
AICSet = zeros(maxLags, maxLags);
for i = 1:maxLags
for j = 1:maxLags
mdl = arima('ARLags', [1:i], 'MALags', [1:j]);
[EstMdl, EstParamCov, LogL, info] = estimate(mdl, diff1_y);

```

```

AICSet(i,j)=aicbic(LogL, length(info.X), length(diff1_y));
end
end
figure(4), heatmap(AICSet/1000)
title('2015-2019 交通运输业产值热图');

mdl = arima( 'Seasonality',4,'D', 0,'ARLags',1:2,'MALags',1);
[EstMdl] = estimate(mdl, y);

res = infer(EstMdl,y);
figure(6),
%subplot(2,2,1), plot(res./sqrt(EstMdl.Variance))
qqplot(res)
%subplot(2,2,3), autocorr(res)
%subplot(2,2,4), parcorr(res)
[h, p] = lbqtest(res)

[yF, yMSE] = forecast(EstMdl,1,'Y0',y);
UB = yF + 1.96*sqrt(yMSE); % 95%置信区间上限
LB = yF - 1.96*sqrt(yMSE); % 95%置信区间下限
figure(7), hold on
plot(y,'linewidth',3)
hold on
plot(21,yF,'*');
hold on,plot(21,y1(21,2),'o');
legend('2015-2019 交通运输业产值','2020 交通运输业产值预测值','2020 交通运输业产值实际值')
xlabel('季度')
ylabel('交通运输业产值/万元')
set(gca,'Xticklabel',{'2014.4','2016.1','2017.2','2018.3','2019.4','2021.1'});
set(gca,'XTickLabelRotation',90);
title('2020 第一季度 交通运输业产值预测')
chazhi=abs(y1(21,2)-yF);
difference=chazhi/yF;%0.2054

```

程序二十二、零售业：

```

clc
clear
y1=xlsread('几大行业产值(1).xlsx',2);
y=y1(1:20,2);
figure(1),
plot(y,'linewidth',3)
set(gca,'Xticklabel',{'2014.4','2015.2','2015.4','2016.2','2016.4','2017.2','2017.4','2018.2','2018.4','2019.2','2019.4'});

```

```

set(gca,'XTickLabelRotation',90);
xlabel('季度');
ylabel('零售业产值/万元');
title('2015-2019 年中国零售业产值');
[h,p]=adftest(y)
diff1_y=y(5:end)-y(1:end-4);
figure(2),
plot(diff1_y,'linewidth',3)
set(gca,'Xticklabel',{'2014.4','2015.2','2015.4','2016.2','2016.4','2017.2','2017.4','2018.2','2018.4','2019.2','2019.4'});
set(gca,'XTickLabelRotation',90);
xlabel('季度');
ylabel('零售业产值/万元');
title('2015-2019 年中国零售业产值一阶季节差分');
[h,p]=adftest(diff1_y)

diff_y=diff(diff1_y);
figure(3),
plot(diff_y,'linewidth',3)
set(gca,'Xticklabel',{'2014.4','2015.2','2015.4','2016.2','2016.4','2017.2','2017.4','2018.2','2018.4','2019.2','2019.4'});
set(gca,'XTickLabelRotation',90);
xlabel('季度');
ylabel('零售业产值/万元');
title('2015-2019 年中国零售业产值一阶差分');
[h,p]=adftest(diff_y)

maxLags = 2;
AICSet = zeros(maxLags, maxLags);
for i = 1:maxLags
    for j = 1:maxLags
        mdl = arima('ARLags', [1:i], 'MALags', [1:j]);
        [EstMdl, EstParamCov, LogL, info] = estimate(mdl, diff1_y);
        AICSet(i, j)= aicbic(LogL, length(info.X), length(diff1_y));
    end
end
figure(4), heatmap(AICSet/1000)
title('2015-2019 零售业产值热图');

mdl = arima('Seasonality',4,'D', 0,'ARLags',2,'MALags',1);
[EstMdl] = estimate(mdl, y);

res = infer(EstMdl,y);
figure(6),

```



```

%subplot(2,2,1), plot(res./sqrt(EstMdl.Variance))
qqplot(res)
%subplot(2,2,3), autocorr(res)
%subplot(2,2,4), parcorr(res)
[h, p] = lbqtest(res)

[yF, yMSE] = forecast(EstMdl,1,'Y0',y);
UB = yF + 1.96*sqrt(yMSE); % 95%置信区间上限
LB = yF - 1.96*sqrt(yMSE); % 95%置信区间下限
figure(7), hold on
plot(y,'linewidth',3)
hold on
plot(21,yF,'*');
hold on,plot(21,y1(21,2),'o');
legend('2015-2019 零售业产值','2020 零售业产值预测值','2020 零售业产值实际
值')
xlabel('季度')
ylabel('零售业/万元')
set(gca,'Xticklabel',{'2014.4','2016.1','2017.2','2018.3','2019.4','2021.1'});
set(gca,'XTickLabelRotation',90);
title('2020 第一季度 零售业产值预测')
chazhi=abs(y1(21,2)-yF);
difference=chazhi/yF;%0.2183

```

程序二十三、制造业：

```

clc
clear
y1=xlsread('几大行业产值(1).xlsx',5);
y=y1(1:20,2);
figure(1),
plot(y,'linewidth',3)
set(gca,'Xticklabel',{'2014.4','2015.2','2015.4','2016.2','2016.4','2017.2','2017.4','2018.
2','2018.4','2019.2','2019.4'});
set(gca,'XTickLabelRotation',90);
xlabel('季度');
ylabel('制造业产值/万元');
title('2015-2019 年中国制造业产值');
[h,p]=adftest(y)
diff1_y=y(5:end)-y(1:end-4);
figure(2),
plot(diff1_y,'linewidth',3)
set(gca,'Xticklabel',{'2014.4','2015.2','2015.4','2016.2','2016.4','2017.2','2017.4','2018.
2','2018.4','2019.2','2019.4'});
set(gca,'XTickLabelRotation',90);

```

```

xlabel('季度');
ylabel('制造业增加值/万元');
title('2015-2019 年中国制造业产值一阶季节差分');
[h,p]=adftest(diff1_y)

```

```

diff2_y=diff(diff1_y);
figure(3),
plot(diff2_y,'linewidth',3)
set(gca,'Xticklabel',{'2014.4','2015.2','2015.4','2016.2','2016.4','2017.2','2017.4','2018.2','2018.4','2019.2','2019.4'});
set(gca,'XTickLabelRotation',90);
xlabel('季度');
ylabel('制造业产值万元');
title('2015-2019 年中国制造业产值一阶差分');
[h,p]=adftest(diff2_y)

```

```

diff_y=diff(diff2_y);
figure(3),
plot(diff_y,'linewidth',3)
set(gca,'Xticklabel',{'2014.4','2015.2','2015.4','2016.2','2016.4','2017.2','2017.4','2018.2','2018.4','2019.2','2019.4'});
set(gca,'XTickLabelRotation',90);
xlabel('季度');
ylabel('制造业产值/万元');
title('2015-2019 年中国制造业产值二阶差分');
[h,p]=adftest(diff_y)

```

```

maxLags = 3;
AICSet = zeros(maxLags, maxLags);
for i = 1:maxLags
    for j = 1:maxLags
        mdl = arima('ARLags', [1:i], 'MALags', [1:j]);
        [EstMdl, EstParamCov, LogL, info] = estimate(mdl, diff1_y);
        AICSet(i, j)= aicbic(LogL, length(info.X), length(diff1_y));
    end
end
figure(4), heatmap(AICSet/1000)
title('2015-2019 制造业产值热图');

```

```

mdl = arima('Constant',0,'Seasonality',4,'D', 0,'ARLags',1:2,'MALags',1);
[EstMdl] = estimate(mdl, y);

```

```

res = infer(EstMdl,y);
figure(6),

```

```

%subplot(2,2,1), plot(res./sqrt(EstMdl.Variance))
qqplot(res)
%subplot(2,2,3), autocorr(res)
%subplot(2,2,4), parcorr(res)
[h, p] = lbqtest(res)

[yF, yMSE] = forecast(EstMdl,1,'Y0',y);
UB = yF + 1.96*sqrt(yMSE); % 95%置信区间上限
LB = yF - 1.96*sqrt(yMSE); % 95%置信区间下限
figure(7), hold on
plot(y,'linewidth',3)
hold on
plot(21,yF,'*');
hold on,plot(21,y1(21,2),'o');
legend('2015-2019 制造业产值','2020 制造业产值预测值','2020 制造业产值实际
值')
xlabel('季度')
ylabel('制造业/万元')
set(gca,'Xticklabel',{'2014.4','2016.1','2017.2','2018.3','2019.4','2021.1'});
set(gca,'XTickLabelRotation',90);
title('2020 第一季度 制造业产值预测')
chazhi=abs(y1(21,2)-yF);
difference=chazhi/yF;%0.0672

```

程序二十四、熵信息构权（综合得分）：

```

clc;clear;
data=xlsread('综合评价.xlsx','que_33');
data(:,7)=1./data(:,7);
zhibiao=data(:,[2,5,7]);
zhibiao=(zhibiao-min(zhibiao))./(max(zhibiao)-min(zhibiao));

```

%% 风险评估

```

x=zhibiao;
[n,p]=size(x);
pp=x./sum(x);
k=1/log(n);%n 表示数据个数，p 是指标个数
p1=zeros(n,p);
for i=1:n
    for j=1:p
        if pp(i,j)~=0
            p1(i,j)=pp(i,j)*log(pp(i,j));
        end
    end
end

```

```
end
I=-k*sum(p1);
r=1-I;
w=r./sum(r);
data1=w.*zhibiao;
data2=sum(data1,2);
[Y,M]=sort(data2);
zz=(data2-min(data2))./(max(data2)-min(data2));
```

程序二十五、2008 年经济危机影响下的各行业产值的预测代码见支撑材料

程序二十六、信贷策略 lingo 代码见支撑材 ques_3_1,ques_3_2,ques_3_3