

基于多元线性回归的无人承运人平台定价优化模型

戚升力 钱佳希 唐代林

June 22, 2020

摘要

随着我国无车承运行业逐步兴起, 承运线路科学定价问题亟待解决。基于此, 通过结合采用主成分分析和相关性矩阵分析方法, 研究分析了影响无车承运人平台进行货运线路定价的主要因素。然后采用模糊综合评价模型评价已成交货运线路历史交易数据中的定价, 另外针对注意到的不合理定价提出了相应定价优化方案; 基于上述研究, 采用多元线性回归模型并结合定价优化方案, 建立了完整线路定价数学模型并预测了附件二的线路定价。最后根据以上模型主要结论对无车承运人平台提出了相关建议。

首先通过数据质量分析和数据预处理工作, 对附件数据进行了相应的剔除、修正和整合。分析线路总成本、线路指导价和线路价格三者关系, 发现线路总成本与线路指导价拟合度很高, 线路价格为线路指导价与调价比例之积。因而只需分析线路总成本和调价比例的影响因素即可。结合采用相关性矩阵和主成分分析方法得出影响线路总成本和线路指导价的因素为里程和计划发车到车时间差; 而调价比例与需求紧急程度、调价紧急程度、业务类型和调价类型有关。最终得出线路价格主要影响因素为里程、计划发车到车时间差、需求紧急程度、调价紧急程度、业务类型和调价类型。

基于货运路线历史交易数据, 采用 SPSS 对计划时间差、线路总成本、调价比例、调价类型、需求紧急程度等因素进行相关性分析, 得出与线路定价的相关性大小, 然后确定该因素对评价定价的权重。基于此建立模糊综合评价集, 并确定各因素对评价集的隶属度函数, 建立模糊评价矩阵并选取了其中六组数据进行评价分析。另外, 注意到实际发车到达时间差与计划发车到达时间差存在较大偏差, 但较多定价并未反映实际耗时间成本远高于计划时间成本的溢价, 容易影响交易成功率。为平滑比值以平滑价格, 选取实际和计划时间差比值的平方根作为新的调价比例, 对定价模型进行了优化。

根据以上研究量化建立线路定价数学模型, 包括总成本定价、第一次报价、第一次调价比例及得到的第二次报价、第二次调价比例及得到的第三次报价。基于附件一线路总成本与总里程、计划时间差存在很高相关性, 建立总里程与计划时间差关于线路总成本定价的多元回归方程并求得总成本定价。基于线路指导价和线路总成本的拟合关系, 根据总成本定价得到第一次报价。同样, 基于业务类型、调价类型、需求紧急程度、调价紧急程度得到关于调价比例的多元回归方程, 然后计算得到附件二的第一次调价比例及第二次报价。根据问题二提及的定价模型优化策略, 建立关于实际时间差的多元回归方程, 以预测实际时间差和计划时间差比值的算术平方根作为第二次调价比例, 从而得到第三次报价。最终采用模糊综合评价模型对研究出的定价进行评价。

最后, 基于研究出的定价方法, 阐述定价策略以及该定价策略能顾弥补的不足和带来的利益, 为无车承运人平台提出相关建议。

关键词: 主成分分析; 相关性分析; 模糊综合评价; 多元线性回归

Contents

一、问题的重述	3
1.1 问题的背景	3
1.2 问题的提出	3
二、问题的分析	3
2.1 问题一的分析.....	3
2.2 问题二的分析.....	3
2.3 问题三的分析.....	4
2.4 问题四的分析.....	4
三、模型假设	4
四、符号说明	4
五、模型的建立与求解	5
5.1 问题一的模型建立与求解.....	5
5.1.1 数据处理.....	5
5.1.2 模型的准备	5
5.1.3 模型的介绍	6
5.1.4 模型的建立	6
5.1.5 结果分析	11
5.2 问题二的模型建立与求解.....	12
5.2.1 模型的准备	12
5.2.2 模型的介绍	12
5.2.3 模糊综合评价模型的建立.....	13
5.2.4 定价优化模型	18
5.3 问题三的模型建立与求解.....	20
5.3.1 数据处理.....	20
5.3.2 模型的建立	21
5.3.3 模型介绍.....	22
5.3.4 定价模型的建立.....	22
5.4 问题四的建议.....	26
六、模型的评价与推广	27
6.1 模型的优点	27
6.2 模型的缺点	27
6.3 模型的推广	27
七、参考文献	28
八、附录.....	28

一、问题的重述

1.1 问题的背景

无车承运人平台是将货主和承运人进行衔接的一个平台，将需要承运的线路任务以一定价格提前发布到网络平台上供承运端司机浏览并决定是否承运该运输任务。若在给定的时间内，该任务没有司机接单，则该线路就可以进行调价。每条线路任务最多允许发布 3 次价格。随着我国无车承运行业的逐步兴起，承运线路的科学定价问题是众多无车承运人平台亟待解决的问题，同时无车承运人平台在当前阶段较为关注的目标是快速促进成交和较低的承运成本。

1.2 问题的提出

通过数学建模的方法帮助某无车承运人平台解决以下问题：

问题 1：通过定量分析的方法，研究影响无车承运人平台进行货运线路定价的主要因素有哪些，并说明理由。

问题 2：根据附件 1 数据，通过建立数学模型，对已经成交货运线路历史交易数据中的定价进行评价。

问题 3：建立关于线路定价的数学模型，给出附件 2 的线路任务的三次报价以及总成本定价，并填充在附件 3 的表格中；给出你们的调价策略；评价你们对附件 2 的线路任务所给出的定价。其中附件 3 的表格以 Excel 文件形式，连同论文答卷一起上传至参赛系统，请勿改变附件 3 中各任务 ID 的原有顺序。附件 3 将用于测试报价的准确性，对于某个确定的任务，三次报价中有一次成交，则后续价格将不再考虑。

问题 4：根据你们的研究，给无车承运人平台写一封不超过一页的建议信。

二、问题的分析

2.1 问题一的分析

问题一要求通过定量分析方法，研究影响无车承运人平台进行货运线路定价的主要因素。首先查找相关的文献以及对附件一中的数据进行处理和分析，挖掘数据内在信息，将数据可视化以图表的形式展现。发现线路定价、线路指导价和线路总成本三者之间存在一定关系，所以通过分析线路总成本、线路指导价与线路定价的关系以及分析线路总成本、线路指导价的影响因素来研究影响线路定价的主要因素，从而全面地分析线路定价的主要影响因素。此处利用 SPSS 进行主成分分析，获得线路总成本、线路指导价的影响因素，从而分析线路定价的主要影响因素。

2.2 问题二的分析

问题二要求对历史交易数据定价进行评价，由问题一可知定价与线路总成本、线路指导价、总里程、计划时间差、需求紧急程度、调价紧急程度、调价比例存在相关性，本文对线路定价的评价建立模糊综合评价模型，为更合理的评价线路定价，这里选取线路总成本，计划时间差，调价比例，调价类型以及需求紧急程度为指标，通过这五项指标与线路定价的相关性大小，来确定五项指标的权重大小，由于数据量过多，本文随机选取了六组历史交易数据，通过隶属度函数建立模糊评价矩阵，最终得到六组数据的线路价格评分，评分越高说明该定价越为合理，评分偏低则说明定价存在一定的不合理因素。接着通过附件 1 数据发现，实际发车和到达时间差与计划时间存在较大偏差，但是

线路定价却未反映出该溢价，存在一定的不合理，于是通过相关性分析以及多元回归建立预测实际时间差模型，同时以预测实际时间差与计划时间差的比例作为优化比例，对线路定价进行优化。

2.3 问题三的分析

建立关于线路定价的数学模型，首先通过问题一得到线路总成本定价与里程，计划时间差存在相关性，于是通过里程和计划时间差建立关于线路总成本定价的多元回归方程；通过附件 1 线路总成本与线路指导价数据比较发现存在一定线性关系，利用拟合得出线路指导价，即为第一次定价。第一次定价通过第一次调价比例可得到第二次报价，而第一次调价比例由第一问可知与业务类型，调价类型，需求紧急程度以及调价紧急程度相关，同样建立关于调价比例的多元回归方程，进而得到第二次报价。对于第二次调价比例的设计，考虑到问题二提及的优化策略，为了反映实际耗费时间差相对计划耗费时间差的溢价，以预测的实际时间差与计划时间差的比值为参考，设计第二次调价比例策略，最终得到第三次报价。

2.4 问题四的分析

针对问题四，综合前三问所得出的模型和结论，在原有结论的基础上，考虑其他条件对无车承运平台的影响，为无车承运平台的提高提出若干建议。

三、模型假设

1. 假设各任务之间不存在相互影响的关系，每个任务都是相互独立的。
2. 假设用于问题分析的数据的来源真实可靠。
3. 假设运输过程中没有发生任何货物的损失。
4. 假设每一个无车承运人都能将货物顺利运达目的地并交易成功。
5. 假设单一因素的缺失不会对整体评价造成较大影响，因素之间有一定的独立性。

四、符号说明

符号	说明
u_1	计划发车和到达时间差
u_2	线路总成本
u_3	调价比例
u_4	调价类型
u_5	需求紧急程度
A	权重向量
B_i	模糊集合
S_i	评分结果
R_i	单因素评判矩阵
V	评价集
v_i	评分
R	拟合优度

五、模型的建立与求解

5.1 问题一的模型建立与求解

5.1.1 数据处理

为了降低异常数据对整体模型的影响，以及提升分析结果的准确性并缩短优化模型的过程，我们针对附件 1 的数据进行预处理。通过对附件 1 的检查，我们发现附件 1 存在的问题如下：

（1）在所给的数据中存在许多数据的缺失，我们无法通过缺失的数据去验证与线路总成本的相关性，同时没有相关的原始数据可以弥补缺失的数据，我们也无法通过预测的方法去补全数据，因此将的一部分进行剔除。

（2）在附件 1 给出的众多因素中，存在因素与线路定价、线路总成本、线路指导价变化无关的情况，所以剔除此类不影响结果准确性的因素。同时为了简化数据，我们对存在线性关系的两个变量进行二选一的处理，在不影响误差的情况下，剔除其中一个因素。

（3）由于到车时间和发车时间对线路总成本无影响，但之间的时间差与线路总定价、线路总成本、线路指导价存在关联，因而计算出时间差在表中列示。

（4）附件中部分因素所给出的数据以文字型的方式出现，为后期的模型建立带来不便，所以这里将其数字化，如以 0/1 数值表示不同的状态。

对附件一的数据处理如下：

（1）删除线路成本和调价紧急程度中的缺失数据。

（2）对数据进行筛选，剔除不影响结果准确性的因素，同时在不影响结果准确性的情况下对存在线性关系的两个变量：车辆长度与车辆吨位进行二选一处理，保留因素车辆长度。

（3）将到达时间和发车时间做差，记录下时间差。

（4）将业务类型、是否续签、打包类型、调价类型、调价紧急程度、地区类型、需求紧急程度这些字符型数据进行数字化，如 0 代表不续签，1 代表续签。

表 1 附件 1 数据处理

数据表	附件 1
数据总量	1009008 条
因素处理	512512 条
缺失值处理	128183 条
剩余数据	368313 条
实际数据清除率（缺失值/数据总量）	12.7%

5.1.2 模型的准备

问题一要求研究影响无车承运人平台进行货运线路定价的主要因素，由于单从对数据的直接分析难以找出线路定价与各因素之间的关系，不能全面分析影响线路定价的主要影响因素，而我们通过对数据的预处理，发现线路总成本、线路指导价与线路定价存

在联系，同时筛选出总里程、业务类型、线路编码、是否续签、交易成功时长、车辆长度、打包类型、地区类型、调价类型、计划时间差、实际时间差、需求紧急程度、调价紧急程度、调价比例这些因素。所以从线路总成本、线路指导价和线路定价的关系出发，利用 SPSS 建立主成分分析模型分析在筛选出的各类因素中线路总成本、线路指导价的主要影响因素，从而全面分析出线路定价的主要影响因素。

5.1.3 模型的介绍

1. 主成分分析的概念

主成分分析是设法将原来众多具有一定相关性(比如 P 个指标)，重新组合成一组新的互相无关的综合指标来代替原来的指标。通常数学上的处理就是将原来 P 个指标作线性组合，作为新的综合指标。经典的做法就是用 F_1 (选取的第一个线性组合，即第一个综合指标)的方差来表达，即 $Var(F_1)$ 越大，示 F_1 包含的信息越多。因此在所有的线性组合中选取的 F_1 应该是方差最大的，故称 F_1 为第-主成分。如果第一主成分不足以代表原来 P 个指标的信息，考虑选取 F_2 即选第二个线性组合，为了有效地反映原来信息 F_1 已有的信息就不需要再出现在 F_2 中，用数学语言表达就是要求 $Cov(F_1, F_2)=0$ ，则称 F_2 为第二主成分，依此类推可以构造出第三、第四，...第 P 个主成分。

2. 主成分分析数学模型为：

$$\begin{cases} F_1 = a_{11}ZX_1 + a_{21}ZX_2 + \dots + a_{p1}ZX_p \\ F_2 = a_{12}ZX_1 + a_{22}ZX_2 + \dots + a_{p2}ZX_p \\ \dots \\ F_n = a_{1n}ZX_1 + a_{2n}ZX_2 + \dots + a_{pn}ZX_p \end{cases}$$

其中 $a_{1i}, a_{2i}, \dots, a_{pi}(i=1, \dots, m)$ 为 X 的协方差阵 Σ 的特征值多对应的特征向量， ZX_1, ZX_2, \dots, ZX_p 是原始变量经过标准化处理的值因为在实际应用中，往往存在指标的量纲不同，所以在计算之前须先消除量纲的影响，而将原始数据标准化。

5.1.4 模型的建立

1. 线路总成本的相关性矩阵与相关性检验

如表 2，将处理好的数据导入 SPSS 中，通过相关性分析方法得到相关性矩阵，分析线路总成本与总里程、业务类型、线路编码、是否续签、交易成功时长、车辆长度、打包类型、调价类型、地区类型、计划时间差、实际时间差、需求紧急程度、调价紧急程度这些因素之间的相关性。由表 3 中 KMO 和巴特利特检验可知，KMO 检验用于检查变量间的偏相关性，巴特利特检验用于检验相关性矩阵是否是单位阵，而图中 KMO 值为 0.714 大于 0.7 说明各因素之间的偏相关性的效果比较好，有研究的意义，同时巴特利特检验中 $P < 0.01$ 说明各项指标之间并非独立，具有相关性。再从表 1 中可以看出总里程、计划时间差与线路总成本之间的相关性超过 0.9 接近 1，说明总里程、计划时间差与线路总成本的相关性很高。

表 2 线路总成本的相关性矩阵

		总里程	业务类	线路编	是否续	交易成	车辆长	打包类	调价类	地区类	计划时	实际时	需求紧	调价紧	线路总
			型	码	签	功时长	度	型	型	型	间差	间差	急程度	急程度	成本
相关性	总里程	1.000	-.222	.004	-.010	.051	.761	.007	.181	.248	.998	.252	.078	.031	.997
	业务类型	-.222	1.000	-.230	.375	.119	.272	-.033	.551	-.439	-.219	.067	.041	.075	-.208
	线路编码	.004	-.230	1.000	-.516	-.088	.107	.186	-.296	.717	-.016	-.041	.076	.021	-.003
	是否续签	-.010	.375	-.516	1.000	.127	.111	.148	.355	-.383	.000	.062	.092	.237	-.005
	交易成功时长	.051	.119	-.088	.127	1.000	.079	-.063	.133	-.019	.051	.033	-.015	-.017	.766
	车辆长度	.761	.272	.107	.111	.079	1.000	.002	.501	.192	.748	.250	.161	.107	.765
	打包类型	.007	-.033	.186	.148	-.063	.002	1.000	-.049	.129	.003	.005	.240	.460	.003
	调价类型	.181	.551	-.296	.355	.133	.501	-.049	1.000	-.327	.182	.147	.099	.086	.182
	地区类型	.248	-.439	.717	-.383	-.019	.192	.129	-.327	1.000	.230	-.010	.114	.047	.243
	计划时间差	.998	-.219	-.016	.000	.051	.748	.003	.182	.230	1.000	.253	.081	.027	.995
	实际时间差	.252	.067	-.041	.062	.033	.250	.005	.147	-.010	.253	1.000	.023	.018	.253
	需求紧急程度	.078	.041	.076	.092	-.015	.161	.240	.099	.114	.081	.023	1.000	.598	.084
	调价紧急程度	.031	.075	.021	.237	-.017	.107	.460	.086	.047	.027	.018	.598	1.000	.034
	线路总成本	.997	-.208	-.003	-.005	.058	.766	.003	.182	.243	.995	.253	.084	.034	1.000

表 3 KMO 和巴特利特检验

KMO 取样适切性量数		.714
巴特利特球形度检验		近似卡方
		250777.546
		自由度
		91
		显著性
		.000

表 4 总方差解释

成分	初始特征值			提取载荷平方和		
	总计	方差百分比	累积%	总计	方差百分比	累积%
1	5.885	34.618	34.618	5.885	34.618	34.618
2	3.421	20.126	54.744	3.421	20.126	54.744
3	1.984	11.671	86.416	1.984	11.671	86.416
4	1.217	7.157				
5	.975	5.737				
6	.908	5.341				
7	.788	4.633				
8	.542	3.190				
9	.475	2.795				
10	.338	1.986				
11	.197	1.158				
12	.146	.857				
13	.093	.550				
14	.016	.092				
15	.010	.058				
16	.004	.023				
17	.001	.008				

如表 4 和表 5，采用主成分分析方法进一步检验和验证相关性矩阵的分析结果，总方差解释中，成分 1、成分 2 和成分 3 特征值大于 1，总合计能解释 86.416%的方差，分析结果比较不错，所以我们保留成分 1、成分 2 和成分 3 作为主成分，其他成分包含信息较少，故舍去。再分析图 3 的成分矩阵，观察成分 1 中的相关系数，绝对值越大，关系越密切，故得到与线路总成本相关性密切的主要因素：总里程、计划发车和到达时间差、调价比例、线路价格和线路指导价。其中，调价比例、线路价格和线路指导价具有内部勾稽关系，故而舍去，总里程、计划发车和到达时间差为外部影响因素，需保留。所以综合两者分析交叉验证，得到线路总成本最主要的影响因素为总里程与计划发车和到达时间差。

表 5 成分矩阵

	成分		
	1	2	3
总里程	.985	-.101	-.064
业务类型	-.116	.732	.064
线路编码	.009	-.669	.294
是否续签	.043	.682	.170
交易成功时长	.084	.189	-.078
车辆长度	.840	.200	.088
打包类型	.023	-.046	.694
调价类型	.306	.765	.024
地区类型	.241	-.727	.258
计划时间差	.981	-.090	-.071
实际时间差	.313	.120	-.022
需求紧急程度	.138	.076	.748
调价紧急程度	.083	.152	.848
调价比例	.939	.835	.023
线路价格	.982	-.058	-.071
线路指导价	.979	-.104	-.064
线路总成本	.985	-.094	-.063

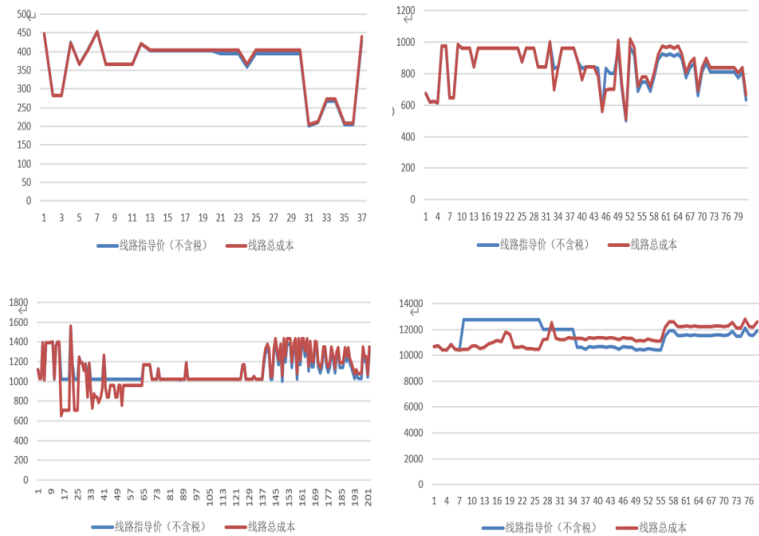


图 1 线路指导价和线路总成本折线图

2. 线路指导价和线路总成本相关性分析

将数据可视化，以图表的形式展现，如上图 1 所示，发现线路总成本和线路指导价的拟合程度很高，几乎保持一致。说明线路总成本与线路指导价的相关性很强，以及线路指导价的主要影响因素与线路总成本的主要影响因素具有相似性，可能都受总里程和计划时间差的影响。下面通过相关性矩阵进一步验证。

3. 线路指导价的相关性矩阵与相关性检验

如表 6，将处理好的数据导入 SPSS 中，通过相关性分析方法，分析线路指导价与总里程、业务类型、线路编码、是否续签、交易成功时长、车辆长度、打包类型、调价类型、计划时间差、实际时间差、需求紧急程度、调价紧急程度这些因素之间的相关性，由表 7 知 KMO 值为 0.724 大于 0.7 说明各因素之间的偏相关性的效果比较好，同时巴特利特检验中 $P < 0.01$ 说明各项指标之间并非独立，具有相关性。表 6 可以分析出总里程、计划时间差与线路指导价的相关性大于 0.9 接近 1，说明总里程和计划时间差是线路指导价的主要影响因素，线路指导价与线路总成本的主要影响因素是相同的。

表 6 线路指导价的相关性矩阵

	总里程	业务类型	线路编码	是否续签	交易成功时长	车辆长度	打包类型	调价类型	地区类型	计划时间差	实际时间差	需求紧急程度	调价紧急程度	线路指导价
总里程	1.000	-.222	.004	-.009	.050	.761	.008	.181	.248	.998	.252	.078	.030	.991
业务类型	-.222	1.000	-.231	.375	.119	.272	-.033	.551	-.439	-.220	.067	.042	.076	-.208
线路编码	.004	-.231	1.000	-.516	-.088	.107	.185	-.296	.717	-.015	-.041	.075	.020	-.000
是否续签	-.009	.375	-.516	1.000	.127	.111	.149	.355	-.383	.000	.062	.093	.238	-.012
交易成功时长	.050	.119	-.088	.127	1.000	.079	-.063	.133	-.019	.050	.033	-.015	-.018	.060
车辆长度	.761	.272	.107	.111	.079	1.000	.003	.501	.192	.748	.250	.161	.107	.761
打包类型	.008	-.033	.185	.149	-.063	.003	1.000	-.049	.128	.004	.005	.242	.461	.006
调价类型	.181	.551	-.296	.355	.133	.501	-.049	1.000	-.327	.181	.147	.098	.085	.170
地区类型	.248	-.439	.717	-.383	-.019	.192	.128	-.327	1.000	.230	-.010	.113	.047	.245
计划时间差	.998	-.220	-.015	.000	.050	.748	.004	.181	.230	1.000	.253	.080	.027	.989
实际时间差	.252	.067	-.041	.062	.033	.250	.005	.147	-.010	.253	1.000	.023	.018	.253
需求紧急程度	.078	.042	.075	.093	-.015	.161	.242	.098	.114	.080	.023	1.000	.599	.083
调价紧急程度	.030	.076	.020	.238	-.018	.107	.461	.085	.047	.027	.018	.599	1.000	.031
线路指导价	.991	-.208	.000	-.012	.060	.761	.006	.170	.245	.989	.253	.083	.031	1.000

表 7 KMO 和巴特利特检验

KMO 取样适切性量数。		.711
巴特利特球形度检验	近似卡方	231972.081
	自由度	91
	显著性	.000

4. 调价比例的相关性矩阵与相关性检验

在数据的处理过程中发现线路价格=线路定价=线路指导价*调价比例，所以为了更

加全面地分析线路定价的主要影响因素，需从调价比例着手，分析其主要影响因素。表 8 为调价比例的相关性矩阵，分析调价比例与总里程、业务类型、线路编码、是否续签、交易成功时长、车辆长度、打包类型、调价类型、计划时间差、实际时间差、需求紧急程度、调价紧急程度这些因素之间的相关性，用同样的分析方法，可以分析出需求紧急程度、调价紧急程度、业务类型、调价类型与调价比例的相关性很高。进一步地，采用主成分分析方法进一步验证。

表 8 调价比例的相关性矩阵

	总里程	业务 类型	线路 编码	是否 续签	交易成 功时长	车辆 长度	打包 类型	调价 类型	地区 类型	计划时 间差	实际时 间差	需求紧 急程度	调价紧 急程度	调价 比例
总里程	1.000	-.222	.004	-.009	.050	.761	.008	.181	.248	.998	.252	.078	.030	.132
业务类型	-.222	1.000	-.231	.375	.119	.272	-.033	.551	-.439	-.220	.067	.042	.076	.951
线路编码	.004	-.231	1.000	-.516	-.088	.107	.185	-.296	.717	-.015	-.041	.075	.020	-.423
是否续签	-.009	.375	-.516	1.000	.127	.111	.149	.355	-.383	.000	.062	.093	.238	.466
交易成功时长	.050	.119	-.088	.127	1.000	.079	-.063	.133	-.019	.050	.033	-.015	-.018	.138
车辆长度	.761	.272	.107	.111	.079	1.000	.003	.501	.192	.748	.250	.161	.107	.364
打包类型	.008	-.033	.185	.149	-.063	.003	1.000	-.049	.128	.004	.005	.242	.461	-.012
调价类型	.181	.551	-.296	.355	.133	.501	-.049	1.000	-.327	.181	.147	.098	.085	.825
地区类型	.248	-.439	.717	-.383	-.019	.192	.128	-.327	1.000	.230	-.010	.113	.047	-.462
计划时间差	.998	-.220	-.015	.000	.050	.748	.004	.181	.230	1.000	.253	.080	.027	.136
实际时间差	.252	.067	-.041	.062	.033	.250	.005	.147	-.010	.253	1.000	.023	.018	.125
需求紧急程度	.078	.042	.075	.093	-.015	.161	.242	.098	.114	.080	.023	1.000	.599	.988
调价紧急程度	.030	.076	.020	.238	-.018	.107	.461	.085	.047	.027	.018	.599	1.000	.923
调价比例	.132	.951	-.423	.466	.138	.364	-.012	.825	-.462	.136	.125	.988	.923	1.000

表 9 KMO 和巴特利特检验

KMO 取样适切性量数。		.659
巴特利特球形度检验	近似卡方	189638.364
	自由度	91
	显著性	.000

表 10 总方差解释

成分	初始特征值			提取载荷平方和		
	总计	方差百分比	累积%	总计	方差百分比	累积%
1	5.885	34.618	34.618	5.885	34.618	34.618
2	3.421	20.126	54.744	3.421	20.126	54.744
3	1.984	11.671	88.769	1.984	11.671	88.769
4	1.217	7.157				
5	.975	5.737				
6	.908	5.341				
7	.788	4.633				
8	.542	3.190				

9	.475	2.795	
10	.338	1.986	
11	.197	1.158	
12	.146	.857	
13	.093	.550	
14	.016	.092	
15	.010	.058	
16	.004	.023	
17	.001	.008	

如表 10 和表 11，采用主成分分析方法进一步检验和验证相关性矩阵的分析结果，总方差解释中，成分 1、成分 2 和成分 3 特征值大于 1，总合计能解释 88.769% 的方差，分析结果比较不错，所以我们保留成分 1、成分 2 和成分 3 作为主成分，其他成分包含信息较少，故舍去。得到与线路总成本相关性密切的主要因素：需求紧急程度、调价紧急程度、业务类型、调价类型。综合两者得到的结果，可以说明需求紧急程度、调价紧急程度、业务类型、调价类型是影响调价比例的主要因素。

表 11 成分矩阵

	成分		
	1	2	3
总里程	.985	-.101	-.064
业务类型	-.116	.732	.064
线路编码	.009	-.669	.294
是否续签	.043	.682	.170
交易成功时长	.084	.189	-.078
车辆长度	.840	.200	.088
打包类型	.023	-.046	.694
调价类型	.306	.765	.024
地区类型	.241	-.727	.258
计划时间差	.981	-.090	-.071
实际时间差	.313	.120	-.022
需求紧急程度	.138	.076	.748
调价紧急程度	.083	.152	.848
调价比例	.939	.835	.023
线路价格	.982	-.058	-.071
线路指导价	.979	-.104	-.064
线路总成本	.985	-.094	-.063

5.1.5 结果分析

图 2 为模型建立后得到的逻辑结果图，最终的结果总结如表 12。我们通过主成分分析法从线路总成本、线路指导价、调价比例着手，分析三个指标的主要影响因素，从而得出线路定价的主要影响因素，于是通过分析可以发现影响无车承运人平台进行货运线路定价的主要因素是总里程、计划时间差、需求紧急程度、调价紧急程度、调价类型和业务类型。

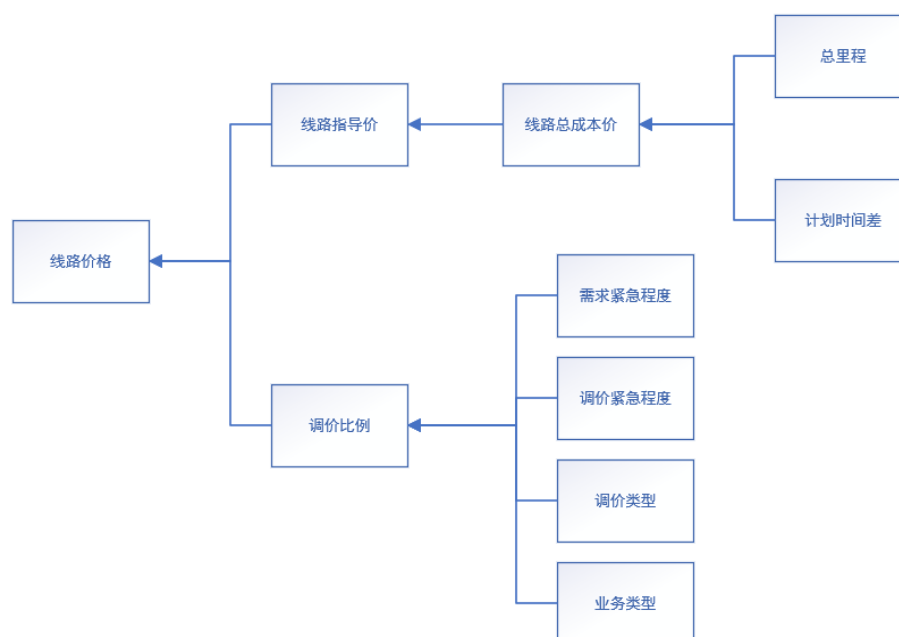


图 2 模型建立逻辑结果图

表 12 影响无车承运人平台进行货运线路定价的主要因素总结

研究指标	总结
线路总成本主要影响因素	线路总成本与其他各因素之间都具有相关性，而其主要的影响因素是总里程和计划时间差。
线路指导价的主要影响因素	线路指导价与线路总成本的拟合度很高，同时线路指导价与其他因素也具有相关性，其中主要影响因素是总里程和计划时间差。
调价比例的主要影响因素	调价比例的主要影响因素是需求紧急程度、调价紧急程度、调价类型和业务类型。

5.2 问题二的模型建立与求解

5.2.1 模型的准备

通过问题一的相关性分析，筛选出与线路价格具有相关性的因素后，为了对历史数据中的定价进行评价，选用模糊综合评价法。因为数据量过大，无法全部进行评价，故选择其中六组具有代表性的数据使用模糊综合评价法进行举例，对其归一化处理的隶属度进行打分，并依此评价历史定价的好差。另外，注意到实际发车和到达时间差与计划发车和到达时间差存在较大偏差，说明实际中耗费巨大时间成本，但是收益较低，存在定价不合理的情况，因而提出优化方案。

5.2.2 模型的介绍

1. 模糊综合评价法的概念

模糊综合评价是借助模糊数学的一些概念，对实际的综合评价问题提供一些评价的方法。具体地说，模糊综合评价就是以模糊数学为基础，应用模糊关系合成的原理，将一些边界不清、不易定量的因素定量化，从多个因素对被评价事物隶属等级状况进行综

合性评价的一种方法。

2. 模糊综合评价法的基本原理

首先确定被评判对象的因素（指标）集和评价（等级）集；再分别确定各个因素的权重及它们的隶属度向量，获得模糊评判矩阵；最后把模糊评判矩阵与因素的权向量进行模糊运算并进行归一化，得到模糊评价综合结果。

5.2.3 模糊综合评价模型的建立

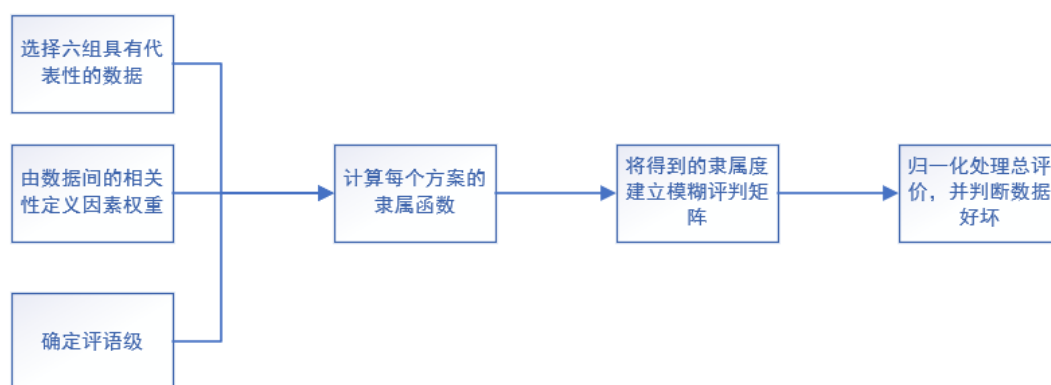


图 3 模型建立流程图

1. 建立综合评价的因素集

因素集是以影响评价对象的各种因素为元素所组成的一个普通集合，通常用 U 表示， $U = (u_1, u_2, \dots, u_m)$ ，其中元素 u_i 代表影响评价对象的第 i 个因素。这些因素，通常都具有不同程度的模糊性。

这里设定价的指标集为 $U = (u_1, u_2, u_3, u_4, u_5)$ ， u_1 表示为计划发车和到达时间差， u_2 表示线路总成本， u_3 表示调价比例， u_4 表示调价类型， u_5 表示需求紧急程度。

2. 确定相应权重向量

由问题一可知定价与计划发车和到达时间差，线路总成本，调价比例，调价类型，需求紧急程度存在相关性，相关性矩阵表如表 13。

表 13 相关性矩阵表

		线路总成本	需求紧急程度	调价比例	计划时间差	调价类型	线路价格
相关性	线路总成本	1.000	.081	.125	.995	.186	.987
	需求紧急程度	.081	1.000	.091	.077	.099	.474
	调价比例	.125	.091	1.000	.133	.824	.508
	计划时间差	.995	.077	.133	1.000	.186	.987
	调价类型	.186	.099	.824	.186	1.000	.523
	线路价格	.987	.474	.508	.987	.532	1.000
显著性	线路总成本		.000	.000	.000	.000	.000
	需求紧急程度	.000		.000	.000	.000	.000
	调价比例	.000	.000		.000	.000	.000

计划时间差	.000	.000	.000	.000	.000
调价类型	.000	.000	.000	.000	.000
线路价格	.000	.000	.000	.000	.000

这里以各因素与定价的相关性大小占所有因素相关性大小之和的百分比作为评价价格的权重，因此得到权重向量为 $A = (0.25, 0.25, 0.13, 0.13, 0.12)$ 。

3. 建立综合评价的评价集

评价集是评价者对评价对象可能做出的各种结果所组成的集合，通常用 V 表示， $V = (v_1, v_2, \dots, v_j)$ ，其中元素 v_j 代表第 j 种评价结果，可以根据实际情况的需要，用不同的等级、评语或数字来表示。

这里设定评价集 $V = (v_1, v_2, v_3)$ ， v_1, v_2, v_3 分别表示评分 0, 1, 2，根据因素集通过对附件 1 的数据筛选，观察发现大部分数据的所在区间如表 14 所示，其中对于计划时间差，线路总成本，对这两种因素设置的评判标准是越低相应得分越高，对于调价比例，调价类型，以及需求紧急程度则相应标准是越高评分越高。

表 14 数据所在区间

分数	计划时间差	线路总成本	调价比例	调价类型	需求紧急程度
2	0.83-16.83	192.29-565.19	1.56-2.03	3	3
1	16.83-33.83	565.19-1569.31	1.04-1.56	2	2
0	33.83-50.67	10421.17-12720.98	0.62-1.04	1	1

4. 单因素模糊评价，获得评价矩阵

若因素集 U 中第 i 个元素对评价集 V 中第 1 个元素的隶属度为 r_i ，则对第 i 个元素单因素评价的结果用模糊集合表示为： $R_i = (r_{i1}, r_{i2}, \dots, r_{im})$ 以 m 个单因素评价集 R_1, R_2, \dots, R_m 为行组成矩阵 $R_{m \times n}$ ，称为模糊综合评价矩阵。

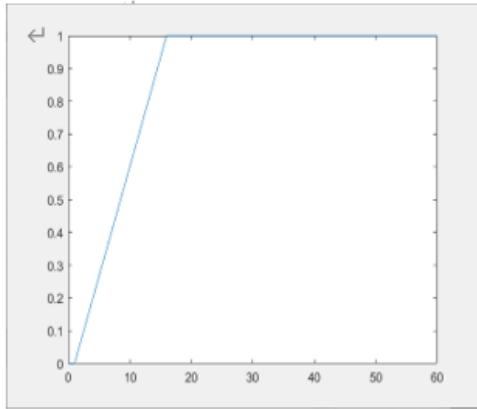
本文确定各个因素对评语集的隶属度函数，由于计划时间差，线路成本以及调价比例均为分段函数，这里选择梯形隶属度函数；调价类型，需求紧急程度为单值函数，这里采用三角形隶属度函数，本文以下只展示了计划发车和到达时间差以及需求紧急程度的隶属函数表达式及相应的隶属函数图（图 4, 图 5），相应的 MATLAB 代码见附录。

计划时间差：

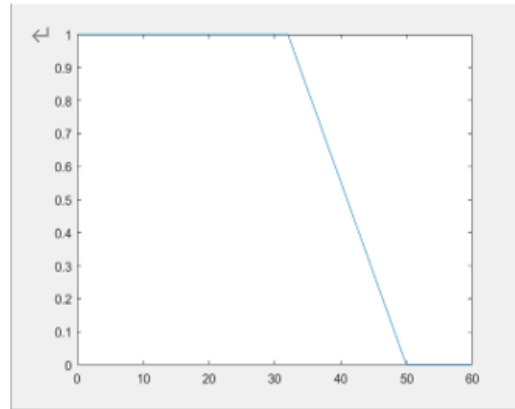
$$C_0(u_1) = \begin{cases} 1, & 32.78 \leq u_1 \\ \frac{u_1 - 16.83}{32.78 - 16.83}, & 16.83 \leq u_1 < 32.78 \\ 0, & u_1 < 16.83 \end{cases}$$

$$C_1(u_1) = \begin{cases} \frac{u_1 - 0.83}{16.83 - 0.83}, & 0.83 \leq u_1 < 16.83 \\ 1, & 16.83 < u_1 \leq 32.78 \\ \frac{50.67 - u_1}{50.67 - 32.78}, & 32.78 \leq u_1 \leq 50.67 \\ 0, & u_1 \leq 0.83 \text{ 或 } u_1 > 50.67 \end{cases}$$

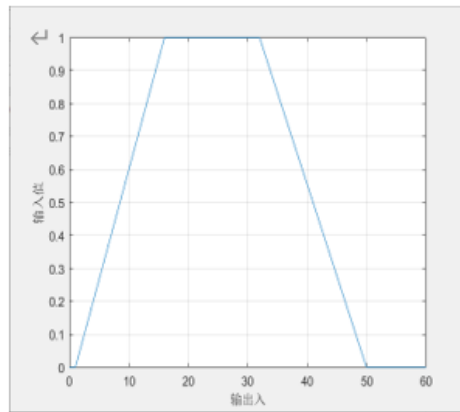
$$C_2(u_1) = \begin{cases} 1, & 0.83 < u_1 \leq 16.83 \\ \frac{32.78 - x}{32.78 - 16.83}, & 16.83 \leq u_1 \leq 32.78 \\ 0, & u_1 > 32.78 \end{cases}$$



a



b



c

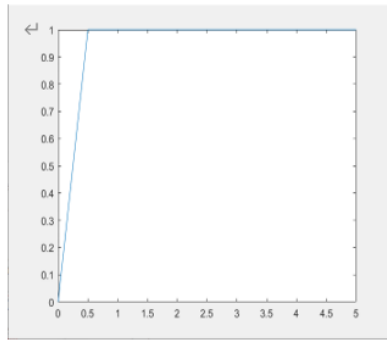
图 4 计划时间差对评语集的隶属度函数图

需求紧急程度:

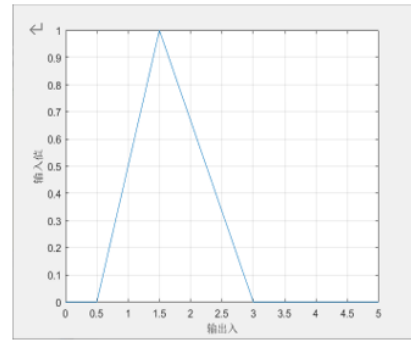
$$C_0(u_5) = \begin{cases} 1, & u_5 \leq 1.5 \\ \frac{2-u_5}{2-1.5}, & 1.5 < u_5 \leq 2 \\ 0, & u_5 > 2 \end{cases}$$

$$C_1(u_5) = \begin{cases} 1, & u_5 \leq 1.5 \text{ 或 } u_5 > 2.5 \\ \frac{u_5-1.5}{2-1.5}, & 1.5 < u_5 \leq 2 \\ \frac{2-u_5}{2.5-2}, & 2 < u_5 \leq 2.5 \end{cases}$$

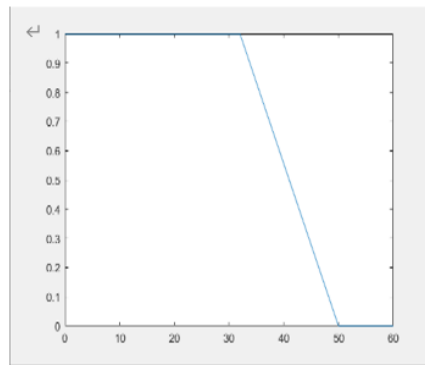
$$C_2(u_5) = \begin{cases} 1, & 3 \leq u_5 \\ \frac{u_5-2.5}{3-2.5}, & 2.5 \leq u_5 < 3 \\ 0, & u_5 < 2.5 \end{cases}$$



a



b



c

图 5 需求紧急程度对评语集的隶属度函数

5. 建立模糊评价矩阵

由于数据量巨大，这里只挑选六组历史数据进行评价，六组数据的相关因素数据如下图所示，由上面所建立的隶属度函数，以任务 4713 为例，得到 u_1 表示为计划发车

和到达时间差, u_2 表示线路总成本, u_3 调价比例, u_4 表示调价类型, u_5 表示需求紧急程度分别对评价集即评分 0, 1, 2 的隶属度如下:

表 15 抽取的六组评价数据

编号	计划时间差	线路总成本	调价比例	调价类型	需求紧急程度
4713	2.43	281.82	0.84	1	1
4721	3.67	718.32	1.50	3	2
4722	3.57	955.53	1.62	3	3
4725	1.00	1278.67	0.87	1	3
12888	1.00	1287.51	1.53	3	3
5093	1.00	404.53	1.51	3	1

表 16 任务 4713 对评价集的隶属度

	v_1	v_2	v_3
u_1	0	0.0056	1
u_2	0	0.24	1
u_3	1	0.524	0
u_4	1	0	0
u_5	1	0	0

可得到任务 4713 的模糊评价矩阵如下:

$$R_1 = \begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ r_4 \\ r_5 \end{bmatrix} = \begin{bmatrix} 0 & 0.0056 & 1 \\ 0 & 0.24 & 1 \\ 1 & 0.524 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

类似的可得到其他五组历史数据定价的模糊评价矩阵

$$R_2 = \begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ r_4 \\ r_5 \end{bmatrix} = \begin{bmatrix} 0 & 0.095 & 1 \\ 0 & 1 & 0.89 \\ 0.12 & 1 & 0.88 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \quad R_3 = \begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ r_4 \\ r_5 \end{bmatrix} = \begin{bmatrix} 0 & 0.104 & 1 \\ 0 & 1 & 0.88 \\ 0 & 0.87 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \quad R_4 = \begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ r_4 \\ r_5 \end{bmatrix} = \begin{bmatrix} 0 & 0.1 & 1 \\ 0 & 1 & 0.95 \\ 1 & 0.904 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$R_5 = \begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ r_4 \\ r_5 \end{bmatrix} = \begin{bmatrix} 0 & 0.1825 & 1 \\ 0 & 1 & 0.45 \\ 0.058 & 1 & 0.92 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \quad R_6 = \begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ r_4 \\ r_5 \end{bmatrix} = \begin{bmatrix} 0 & 0.1 & 1 \\ 0 & 1 & 0.97 \\ 0.096 & 1 & 0.904 \\ 0 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix}$$

6. 模糊综合评判

在上述确定单因素评判矩阵 R 和因素权向量 A 之后, 通过模糊变化将 U 上的模糊

向量 A 变为 V 上的模糊向量 B ，计算总评价 $B=A*R$ ，为了兼顾各个因素，我们采用 $M(+)$ 算子，并对 B 进行归一化处理，MATLAB 程序代码见附件，最终六组历史数据的模糊向量如表 17 所示。

表 17 六组历史数据的模糊向量

向量名称	大小	大小	大小
B_1	1.1400	0.4328	1.8600
B_2	0.0468	1.7684	2.4909
B_3	0	1.3660	2.8884
B_4	0.3900	1.7656	2.1735
B_5	0.0226	1.4897	2.4573
B_6	0.3974	1.4130	2.5747

7. 结果分析

在得到六组数据的模糊向量后，最终的评分结果等于 $S_i=B_i \times V_i$ ($i=1,2,\dots,6$) 六组数据的评分结果如表 18 所示，其中评分越高代表定价合理性越高，评分偏低则说明定价存在一定的不合理因素。

表 18 六组历史数据的评分结果

名称	评分
S_1	1.1528
S_2	6.7501
S_3	7.1428
S_4	6.1126
S_5	3.4043
S_6	6.5623

5.2.4 定价优化模型

通过上述分析对定价评分发现部分定价存在不合理的因素，通过附件 1 的历史数据观察发现，发现实际发车和到达时间差与计划发车和到达时间差存在较大偏差，说明实际中耗费巨大时间成本，但是收益较低，存在定价不合理的情况，而调价比例在实际情况中是由于司机因价格不合理等原因而进行适当调价，影响最终实际时间差的大小，因此本文以实际时间差与计划时间差为比例作为调价比例，继而对调价比例进行优化，使得定价趋于合理。

1. 调价比例的优化

本文以预测实际发车和到达的时间差与计划发车和到达的时间差的比例作为调价比例，通过对实际发车和到达时间差用主成分分析，由分析可得预测时间差与总里程，线路价格，线路总成本，计划发车和到达时间差值存在相关性，相关矩阵如表 19。

通过对预测发车和到达时间差与相关因素建立多元回归分析，可得到预测发车和到达时间差的多元回归方程如下：

$$\text{预测实际时间差} = -1.085 - 0.0235 \times \text{总里程} + 1.942 \times \text{计划发车和到达时间差} + 0.003 \times \text{线路总成本}$$

由已知的计划发车和到达时间差，然后选取其平方根作为优化后的调价比例为：

$$\text{调价比例} = (\text{预测实际发车和到达时间差} / \text{计划发车和到达时间差})^{1/2}$$

这样的好处在于可以平滑比值参数以平衡价格。

表 19 相关性矩阵表

相关性矩阵 ^a														
	总里程	业务类型	需求类型2	线路价格（不含税）	线路编码	线路指导价（不含税）	是否续签	交易成功时长	调价紧急程度	线路总成本	需求紧急程度	调价比例	交易时长	计划发车和到达时间差
相关性	总里程	1.000	-.228	.131	.989	.004	.991	.001	.044	.028	.997	.075	.132	.108
	业务类型	-.228	1.000	.082	-.202	-.228	-.214	-.372	.123	.075	-.214	.041	.508	-.070
	需求类型2	.131	.082	1.000	.111	.009	.130	-.363	-.030	.639	.136	.391	.129	-.183
	线路价格（不含税）	.989	-.202	.111	1.000	-.017	.979	-.004	.054	.029	.987	.074	.190	.109
	线路编码	.004	-.228	.009	-.017	1.000	.000	.513	-.082	.021	-.003	.076	-.425	-.039
	线路指导价（不含税）	.991	-.214	.130	.979	.000	1.000	.003	.053	.030	.992	.080	.106	.115
	是否续签	.001	-.372	-.363	-.004	.513	1.000	-.129	-.236	-.005	-.089	-.464	-.048	-.008
	交易成功时长	.044	.123	-.030	.054	-.082	.053	1.000	-.018	.050	-.016	.131	.005	.044
	调价紧急程度	.028	.075	.639	.029	.021	.030	-.236	1.000	.031	.597	.124	-.221	.025
	线路总成本	.997	-.214	.136	.987	-.003	.992	-.005	.050	.031	1.000	.081	.127	.116
	需求紧急程度	.075	.041	.391	.074	.076	.080	-.089	-.016	.597	.081	1.000	.089	-.105
	调价比例	.132	.508	.129	.190	-.425	.106	-.464	.131	.124	.127	.089	1.000	-.013
	交易时长	.108	-.070	-.183	.109	-.039	.115	-.048	.005	-.221	.116	-.105	-.013	1.000
	计划发车和到达时间差	.998	-.225	.127	.987	-.015	.989	-.008	.044	.025	.995	.077	.135	.110
	实际发车和到达时间差	.846	.064	.038	.825	-.041	.257	-.063	.030	.018	.792	.023	.126	.023
显著性（单尾）	总里程		.000	.000	.000	.299	.000	.464	.000	.000	.000	.000	.000	.000
	业务类型		.000		.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	需求类型2		.000	.000		.000	.132	.000	.000	.000	.000	.000	.000	.000
	线路价格（不含税）		.000	.000	.000		.017	.000	.293	.000	.000	.000	.000	.000
	线路编码		.299	.000	.132	.017		.495	.000	.004	.344	.000	.000	.026
	线路指导价（不含税）		.000	.000	.000	.000	.495		.336	.000	.000	.000	.000	.000
	是否续签		.464	.000	.000	.293	.000	.336		.000	.266	.000	.000	.152
	交易成功时长		.000	.000	.000	.000	.000	.000		.014	.000	.025	.000	.258
	调价紧急程度		.000	.000	.000	.000	.004	.000	.014		.000	.000	.000	.001
	线路总成本		.000	.000	.000	.000	.344	.000	.266	.000		.000	.000	.000
	需求紧急程度		.000	.000	.000	.000	.000	.000	.025	.000	.000		.000	.000
	调价比例		.000	.000	.000	.000	.000	.000	.000	.000	.000	.000		.053
	交易时长		.000	.000	.000	.000	.000	.258	.000	.000	.000	.053		.000
	计划发车和到达时间差		.000	.000	.000	.000	.026	.000	.152	.000	.001	.000	.000	.000
	实际发车和到达时间差		.000	.000	.000	.000	.000	.000	.000	.012	.000	.002	.000	.002

决定因子 = 3.652E-10

2. 对历史定价进行优化

在得到优化后的调价比例后，由问题一分析可知线路定价等于线路指导价与调剂比例的乘积，便可得到与基于实际发车和到达时间差的时间成本的优化后的线路定价。优化后的部分价格如表 20 所示。

表 20 优化后部分价格举例表

任务 id	计划时间差	预测实际时间	优化调价比例	线路价格	预测线路价格	线路指导价（不含税）
4713	0.92	0.78	0.91	256.4568	233.23	282
4721	2.35	2.11	0.95	637.0795	603.97	672
4722	2.50	3.43	1.17	797.6356	934.25	681
4723	2.50	3.43	1.17	797.6356	934.25	681
4724	2.50	3.43	1.17	797.6356	934.25	681
4727	2.43	2.11	0.93	576.8468	537.56	619
4725	2.43	2.11	0.93	576.8468	537.56	619
4726	2.43	2.11	0.93	576.8468	537.56	619
4728	2.50	2.25	0.95	591.5994	561.78	623

4714	0.92	0.48	0.73	204.7264	148.63	282
4729	2.50	2.73	1.04	466.7964	487.47	447
4730	2.50	2.73	1.04	466.7964	487.47	447
4715	0.92	0.78	0.91	256.4568	233.23	282
5455	2.50	2.22	0.94	577.1222	544.23	612
12888	3.75	3.90	1.02	1142.941	1165.31	1121
4716	0.92	0.78	0.91	256.4568	233.23	282
4720	1.05	0.74	0.84	237.0686	199.30	282
4718	1.00	0.94	0.97	274.1295	266.48	282
4717	1.00	0.94	0.97	274.1295	266.48	282
4719	1.00	0.94	0.97	278.0179	270.26	286
5067	2.43	2.08	0.93	563.5864	521.56	609
5318	0.92	0.78	0.91	256.4568	233.23	282
5496	0.92	0.78	0.91	256.4568	233.23	282
5351	0.92	0.78	0.91	256.4568	233.23	282
4740	1.00	0.64	0.80	226.4725	181.88	282
5093	2.43	2.08	0.93	563.5864	521.56	609
5264	0.92	0.78	0.91	256.4568	233.23	282
5322	0.92	0.78	0.91	256.4568	233.23	282
4823	1.00	0.64	0.80	226.4725	181.88	282

5.3 问题三的模型建立与求解

5.3.1 数据处理

为了简化数据以及缩减建模的程序，提高结果的准确性，对附件 2 中的数据进行处
理，通过对附件 2 的检查，我们发现附件 2 存在的问题如下：

（1）附件 2 中有很多数据存在缺失，由于这些缺失无法通过预测求得同时与解题无
关，不影响结果的准确性，所以将这些数据剔除。

（2）在附件 2 给出的众多数据中存在与解题无关以及不影响结果准确性的数据，为了
简化数据，将这些不影响结果准确性以及和解题无关的数据剔除。

（3）附件 2 中的部分数据以字符型的方式出现，为了方便后期的模型建立和求解，我
们将这些数据进行量化。

对附件 2 的数据处理如下：

（1）将缺失的数据进行剔除。

（2）将不影响结果准确性以及和解题无关的数据筛选出来，并予以剔除。

（3）将业务类型、是否续签、打包类型、调价类型、调价紧急程度、地区类型、需求
紧急程度这些字符型数据进行量化。

处理后的结果见表 21。

表 21 附件 2 数据处理图

数据表	附件 1
数据总量	56582 条
剔除数据	11912 条
数据筛选	25313 条
剩余数据	37225 条
数据清除率	65.8%

5.3.2 模型的建立

建立关于线路定价的数学模型，首先通过里程和计划时间差建立关于线路总成本定价的多元回归方程；利用对总成本定价的拟合得出线路指导价，即第一次定价，第一次定价通过第一次调价可得到第二次报价，第一次调价比例由业务类型，调价类型，需求紧急程度以及调价紧急程度建立关于调价比例多元回归方程，进而得到第二次报价，对于第二次调价比例策略，本文选取预测的实际发车和到达时间差与计划发车和到达时间差比值的平方根为第二次调价比例，最终得到第三次报价，线路定价模型建立的流程图如图 6。

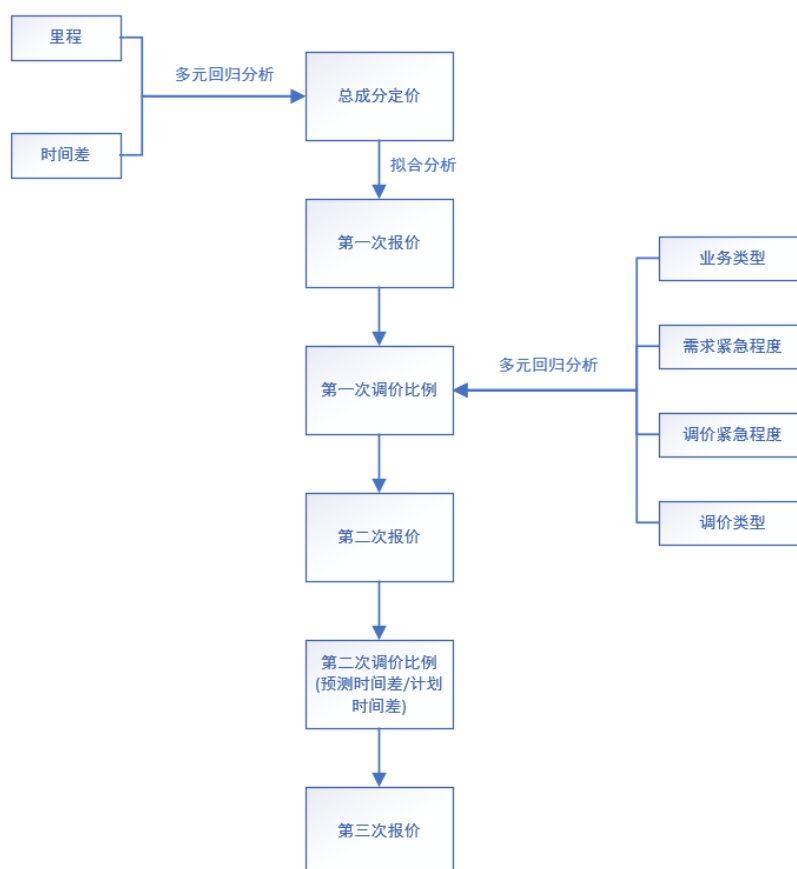


图 6 模型建立流程图

5.3.3 模型介绍

多元回归分析是指在相关变量中将一个变量视为因变量，其他一个或多个变量视为自变量，建立多个变量之间线性或非线性数学模型数量关系式并利用样本数据进行分析的统计分析方法。多元回归模型的数学形式：

设因变量为 Y ，影响因变量的 k 个自变量分别为 X_1, X_2, \dots, X_k ，假设每一个自变量对因变量 Y 的影响都是线性的，也就是说，在其他自变量不变的情况下， Y 的均值随着自变量 X_i 的变化均匀变化，这时我们把：

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon$$

称为总体回归模型，把 $\beta_0, \beta_1, \beta_2, \beta_k$ 称为回归参数。式中 β_0 为常数项，它表示当所有自变量取值均为零时因变量的估计值， β_k 为偏回归系数，表示当其他自变量取值固定时自变量 x_i 每改变一个单位时，决定 y 的变化量式。

5.3.4 定价模型的建立

1. 总成本定价

在第一问中，我们得到线路总成本定价与里程，计划发车和到达时间差具有相关性，利用 SPSS 软件进行多元回归分析，得到如下结果， R 表示拟合优度，用来衡量模型的拟合程度，由结果可知 $R=0.997$ ，说明拟合程度很高；Anova 表示方差分析结果，由 F 值对应的显著性可知，显著性小于 0.05，说明该多元回归方程结果有效，从而得到总成本计算公式如下，并将预测的附件 2 的线路总成本价填充至附件 3 中。

$$\text{线路总成本} = 120.952 + 4.245 \text{ 总里程} + 10.867 \text{ 计划发车和到达时间差}$$

表 22 模型摘要

模型	R	R 方	调整后 R 方	标准估算的错误
1	.997	.994	.994	217.89637

a. 预测变量: (常量), 计划发车和到达时间差, 总里程

表 23 ANOVA

模型		平方和	自由度	均方	F	显著性
1	回归	133401962348.103	2	66700981174.051	1404857.395	.000 ^b
	残差	759898625.774	16005	47478.827		
	总计	134161860973.877	16007			

a. 因变量: 线路总成本

b. 预测变量: (常量), 计划发车和到达时间差, 总里程

表 24 系数 a

模型		未标准化系数		标准化系数	t	显著性
		B	标准错误	Beta		
1	(常量)	120.952	1.891		63.969	.000
	总里程	4.245	.039	.953	108.421	.000

	计划发车和 到达时间差	10.867	2.159	.044	5.034	.000
--	----------------	--------	-------	------	-------	------

2. 第一次定价

第一次定价即指的是的线路指导价，在问题一中通过线路总成本与线路指导价的差值散点图，说明线路总成本与线路指导价并不属于等价关系，但通过线路总成本与线路指导价的曲线分析图观察发现，二者存在一定的线性关系，这里通过 SPSS 软件曲线拟合得到线路指导价与线路总成本的线性方程，结果如下图所示，由 $R=0.992$ ，可说明拟合程度很高，由线性图（图 7）可看出，数值基本在直线附近波动或者重合，说明该方程具有较高参考性，因此可得第一次定价表达式如下：

$$\text{第一次报价价格（线路指导价）} = 1.037 * \text{线路总成本} - 23.701$$

表 25 模型摘要

R	R 方	调整后 R 方	标准估算的错误
.992	.985	.985	372.750

自变量为 线路总成本。

表 26 系数

	未标准化系数		标准化系数	<i>t</i>	显著性
	<i>B</i>	标准 错误	Beta		
线路总成本	1.037	.001	.992	1019.078	.000
(常量)	-23.701	3.283		-7.220	.000

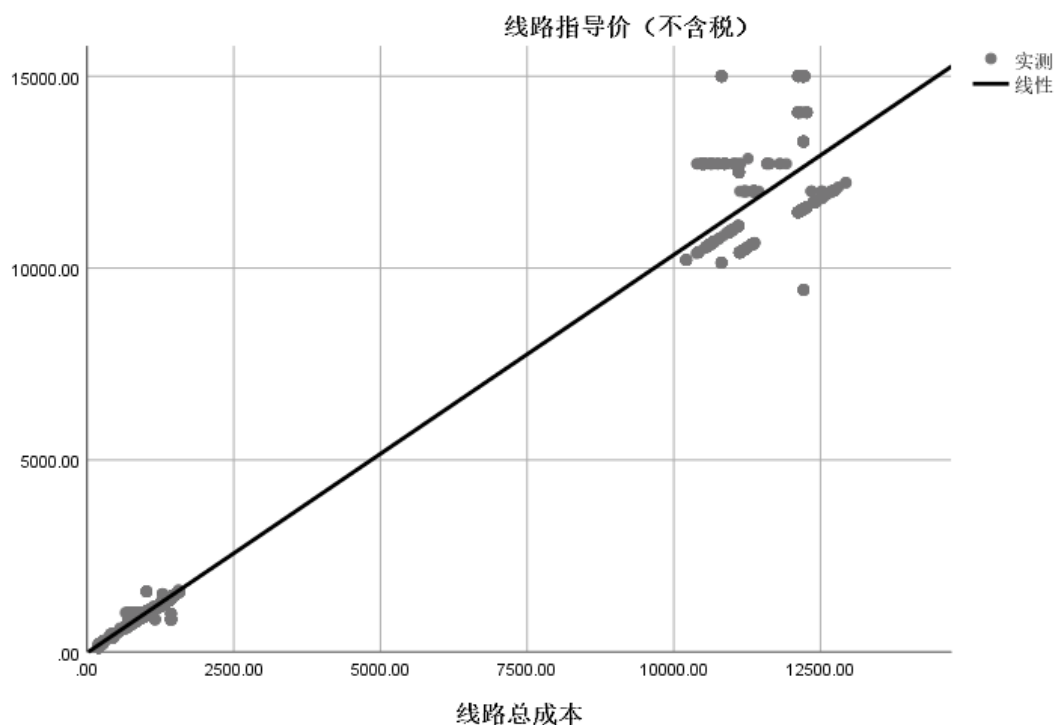


图 7 线路指导价和线路总成本的线性图

3. 第一次调价比例

在问题一中，得到调价比例与以下四个因素存在相关性，分别是业务类型，调价类型，需求紧急程度，调价紧急程度，在第一次调价中，本文由以上得到的四个因素，利用 SPSS 软件进行多元线性分析，可得到计算结果如下， $R=0.83$ ，说明拟合程度较好，显著性小于 0.05，说明该方程可适用于计算调价比例，由此可到第一次调价比例为：

$$\text{调价比例} = 0.237 \times \text{业务类型} + 0.089 \times \text{调价类型} + 0.116 \times \text{需求紧急程度} + 0.065 \times \text{调价紧急程度} + 0.839$$

观察附件 2 发现其中缺失调价类型和调价紧急程度两项指标，但数据预处理中我们将调价类型用 1, 2, 3 分别用来表示未调整，调低，调高，调价紧急程度用 1, 2, 3 分别表示常规，紧急，非常紧急，两项指标对最终调价比例的影响在 0.154 到 0.462 之间，通过附件 1 对于大部分数据来看对最终调价比例结果影响在 0.154 到 0.236 之间，由此可知即便附件 2 中缺失调价类型和调价紧急程度两项指标，对最终调价比例的影响有限，因此该调价比例策略适用于第一次调价比例的计算。

表 27 模型摘要

模型	<i>R</i>	<i>R</i> 方	调整后 <i>R</i> 方	标准估算的错误
1	.830 ^a	.689	.688	.15194

a. 预测变量: (常量), 调价紧急程度, 业务类型, 调价类型, 需求紧急程度

表 28 ANOVAa

模型		平方和	自由度	均方	<i>F</i>	显著性
1	回归	814.639	4	203.660	8821.784	.000 ^b
	残差	368.545	15964	.023		
	总计	1183.184	15968			

a. 因变量: 调价比例

b. 预测变量: (常量), 调价紧急程度, 业务类型, 调价类型, 需求紧急程度

表 29 系数

模型		未标准化系数		标准化系数	<i>t</i>	显著性
		<i>B</i>	标准 错误	Beta		
1	(常量)	.893	.002		382.669	.000
	业务类型	.237	.003	.077	14.707	.000
	调价类型	.089	.002	.781	147.866	.000
	需求紧急程度	.116	.003	-.036	-6.451	.000
	调价紧急程度	.065	.005	.070	12.561	.000

4. 第二次报价

由前面计算得到第一次报价和第一次调价比例的计算结果，本文建立的第二次报价价格是基于得到的第一次报价价格与第一次调价比例的基础上所得到第二次报价，可直接得到第二次报价的最终结果即等于：

$$\text{第二次报价价格} = \text{第一次报价价格} \times \text{第一次调价比例}$$

5. 第二次调价比例

第二次调价比例本文根据附件 2 中数据通过预测，得到预测的实际发车和到达时间差值，在问题二中对定价进行优化分析中我们得到优化调价比例，即通过预测实际发车和到达时间差值与计划发车和到达时间差值的比值，去其算术平方根作为调价比例，同样也适用于第二次调价比例的建立，计算公式如下：

$$\text{第二次调价比例} = (\text{预测实际发车和到达时间差值} / \text{计划发车和到达时间差值})^{1/2}$$

6. 第三次报价

由前面计算得到第二次报价和第二次调价比例的计算结果，本文建立的第三次报价价格是基于得到的第二次报价价格与第二次调价比例的基础上所得到第三次报价，第三次报价确立的公式如下所示，相关完整数据见附件 3。

$$\text{第三次报价价格} = \text{第二次报价价格} \times \text{第二次调价比例}$$

7. 对附件 2 所给定任务的定价进行评价

问题二中本文建立了关于评价定价的模糊综合评价模型，对于给定任务的定价进行评价，由于数据量过多，无法一一展示，同样随机挑选了六组数据，通过隶属度函数得到的评价矩阵如下所示：

任务 17281

$$R_1 = \begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ r_4 \end{bmatrix} = \begin{bmatrix} 0 & 0.104 & 1 \\ 0 & 1 & 0.81 \\ 0.96 & 1 & 0.038 \\ 0 & 0 & 1 \end{bmatrix}$$

任务 15629

$$R_2 = \begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ r_4 \end{bmatrix} = \begin{bmatrix} 0 & 0.18 & 1 \\ 0 & 1 & 0.38 \\ 0.94 & 1 & 0.058 \\ 0 & 0 & 1 \end{bmatrix}$$

任务 3117

$$R_3 = \begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ r_4 \end{bmatrix} = \begin{bmatrix} 0 & 0.028 & 1 \\ 0 & 0.58 & 1 \\ 1 & 0.57 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

任务 11140

$$R_4 = \begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ r_4 \end{bmatrix} = \begin{bmatrix} 0 & 0.016 & 1 \\ 0 & 0.35 & 1 \\ 1 & 0.57 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

任务 3117

$$R_5 = \begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ r_4 \end{bmatrix} = \begin{bmatrix} 0 & 0.011 & 1 \\ 0 & 0.21 & 1 \\ 1 & 0.64 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

任务 3110

$$R_6 = \begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ r_4 \end{bmatrix} = \begin{bmatrix} 0 & 0.074 & 1 \\ 0 & 1 & 0.93 \\ 1 & 0.9 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

最终得到的六组数据的模糊综合评判向量，以及最终评分结果如表 30，31 所示，由最终评分可以看出最终评分结果都比较高，说明本文建立的定价模型对报价的预测具有较大参考意义。

表 30 六组数据模糊综合评价向量

向量名称	大小	大小	大小
B_1	0.04320	1.5451	2.3619
B_2	0.4320	1.7244	1.9065

B_3	0.4500	0.9131	2.5500
B_4	0.4500	1.0418	2.1600
B_5	0.8400	0.5267	2.1600
B_6	0.4500	1.5649	2.4744

表 31 六组数据最终评分

名称	评分
S_1	7.2689
S_2	6.5374
S_3	6.0131
S_4	5.3618
S_5	5.8467
S_6	6.5137

5.4 问题四的建议

通过上述三个模型的结果进行分析，并结合货运线路历史定价和预测定价的相关数据数据，我们对于无车承运人平台的定价提出以下四点建议：

(1) 平台需要在既有的模型上优化计划时间差，以增强对实际时间差预测。通过对历史数据的主成分分析得到的相关性可知，线路定价与计划时间差的相关性有 0.989，与实际时间差的相关性，却只有 0.253，这导致了当实际时间差是计划时间差的一倍以上时，无车承运人就会遭受到一定的损失，同时也会出现两个不同任务的路线定价一致，但实际时间差相差过大的情况，这将使平台的可信度下降。故平台需要选用更合理的模型以增强平台的可信度并保障委托人和无车承运人的利益，促进商品的快速成交。

(2) 建立责任保险制度，无车承运可以承担所有类型的道路货物运输，一旦发生运输事故，很难保证委托人的权益。保险制度能有效防范风险，因此，可以参考国外无车承运保险制度建立我国相应的无车承运保险制度，达到防范风险的目的。

(3) 增加报价次数，从中选择最准确线路报价价格。相对于以往的一次报价，多次报价可以减小误差，避免出现在不同任务的因素相差较大的情况下，线路价格一致。因此，在多处报价的情况下，可以令线路价格趋向合理。同时使平台，无车承运人和委托人三方都满意此次交易，从而提高运单完成率。

(4)* 加强科技人才队伍培养，提高企业核心竞争内力。无车承运平台的经营发展，涉及信息系统设计、大数据分析、市场营销、金融投资、运输组织 and 安全管理等各个方面，属于技术密集、知识密集型企业，需要强大的技术和管理团队。互联网平台型企业不仅要加强市场开拓，稳定货源，更要加快科技人才储备，特别是高端管理人才和技术人才的引进，以提高企业的核心内力和竞争实力。

综上所述，平台需要提高时间预测的准确性；建立相应责任保险制度；多处报价令线路价格趋向合理；加强科技人才队伍培养，提高企业核心竞争内力；最后，综合以上四项措施实施，相信定能为无车承运平台的提高带来新的机遇。

六、模型的评价与推广

6.1 模型的优点

对于问题一，从线路总成本、线路指导价、线路价格的关系着手，结合采用主成分分析和相关性矩阵方法，分析出线路总成本、线路指导价、调价比例的主要影响因素，以此研究出影响无车承运人平台进行货运线路定价的主要因素，用少数几个集中了原始变量的大部分信息的综合变量可代替原始多个变量，可减少无关数据干扰，全方位、科学地分析出影响无车承运人平台进行货运线路定价的主要因素。

对于问题二，采用模糊综合评价来对已经成交货运线路历史交易数据中的定价进行评价，能够通过精确的数字手段处理模糊的评价对象，以矢量的方式呈现评价结果，使得评价结果更加的准确和具有参考价值。另外发掘出定价不合理之处，对模型进行了优化。

对于问题三，采用多元回归模型来确定总成本定价和三次报价，建立多个自变量间的线性因果关系，从多个自变量的最优组合进行预测，这种方法使预测更加简单，让预测结果贴合现实、更加准确，具有现实意义。

6.2 模型的缺点

问题一的主成分分析模型在降维的过程中使综合变量的含义有点模糊，没有原始变量的含义那么清晰，同时得到的特征值带有负号使得综合评价函数意义就不明确。但由于仅需寻找主要因素，所以无需分析各主成分的内涵。

问题二的模糊综合评价的计算过程偏复杂，对指标的权重的确定的主观性较强。

问题三的多元回归模型在回归分析中对于选用的因子和该因子采用的表达式都只是一种推测，影响了用电因子的多样性和某些因子的不可测性，使回归分析在某些情况下受到限制。

6.3 模型的推广

主成分分析能够将众多的因素通过它们的相关性分析进行一个整合，整合成新的一个或者多个相对独立的综合因素，使分析更加全面，能够运用于测量一个平台的客户的满意程度。

模糊综合评价能够通过精确的数字手段处理模糊的评价对象，可以应用于市场分析评价、风险评价、选址问题等考虑多因素的问题。

多元回归模型利用多个变量间的线性关系对结果进行一个预测，使预测更加贴合现实，能够用于解决变量之间的关系问题和受多变量影响的结果预测问题。

七、参考文献

- [1] 董娜. 无车承运人的优势分析和发发展建议[J]. 交通标准化, 2011 (24): 87-90.
- [2] 黄少波, 李挥剑. 论无车承运在我国发展的必要性和可行性[J]. 交通运输部管理干部学院学报, 2016, 26 (1): 17-20. 张文彤.SPSS 统计分析高级教程[M].北京:高等教育出版社,2004.9.
- [3] 李艳双, 曾珍香. 主成分分析法在多指标综合评价方法中的应用[J]. 河北工业大学学报, 1999(1):94-97.
- [4] 陈述云, 张崇甫. 对多指标综合评价的主成分分析方法的改进[J]. 统计研究, 1995.
- [5] 李萍, 毛琼, 王新颖,等. 基于多元线性回归分析的操作成本预测模型建立与应用
- [6]. 石油规划设计, 2018, v.29; No.161(03):37-41.
- [7] 林彬. 多元线性回归分析及其应用[J]. 中国科技信息, 2010, 000(009):60-61.
- [8] 吴刚, 张敬信等, 数学建模与数学实验, 中国商业出版社, 2017

八、附录

问题二 *MATLAB* 图代码

(计划时间)

对 2 隶属度

```
a1=1;
a2=16;
x=0:0.1:60;
for i=1:length(x)
    y(i)=x(i)/(a2-a1)-a1/(a2-a1);
    if y(i)>1
        y(i)=1;
    end
    if y(i)<0
        y(i)=0;
    end
end
plot(x,y)
```

对 1 隶属度

```
x=0:0.1:60;
y=trapmf(x,[1 16 32 50]);
plot(x,y);
xlabel('输出入');
```

```
ylabel('输入值');  
grid on
```

对 0 隶属度

```
a1=32;  
a2=50;  
x=0:0.1:60;  
for i=1:length(x)  
    y(i)=a2/(a2-a1)-x(i)/(a2-a1);  
    if y(i)>1  
        y(i)=1;  
    end  
    if y(i)<0  
        y(i)=0;  
    end  
end  
plot(x,y)
```

(需求紧急程度)

对 0 隶属度

```
a1=0;  
a2=0.5;  
x=0:0.1:5;  
for i=1:length(x)  
    y(i)=x(i)/(a2-a1)-a1/(a2-a1);  
    if y(i)>1  
        y(i)=1;  
    end  
    if y(i)<0  
        y(i)=0;  
    end  
end  
plot(x,y)
```

对 1 隶属度

```
x=0:0.1:5;  
y=trimf(x,[0.5 1.5 3]);  
plot(x,y);
```

```
xlabel('输出');  
  
ylabel('输入值');  
  
grid on
```

对 2 隶属度

```
a1=1.5;  
a2=3;  
x=0:0.1:5;  
for i=1:length(x)  
    y(i)=a2/(a2-a1)-x(i)/(a2-a1);  
    if y(i)>1  
        y(i)=1;  
    end  
    if y(i)<0  
        y(i)=0;  
    end  
end  
plot(x,y)
```

(调价比例)

```
a1=1.04;  
a2=1.56;  
x=0:0.1:3;  
for i=1:length(x)  
    y(i)=a2/(a2-a1)-x(i)/(a2-a1);  
    if y(i)>1  
        y(i)=1;  
    end  
    if y(i)<0  
        y(i)=0;  
    end  
end  
plot(x,y)  
x=0:0.1:3;  
y=trapmf(x,[0.62 1.04 1.56 2.03]);  
plot(x,y);  
  
xlabel('输出');  
  
ylabel('输入值');  
  
grid on
```

```

a1=1.56;
a2=2.03;
x=0:0.1:3;
for i=1:length(x)
    y(i)=x(i)/(a2-a1)-a1/(a2-a1);
    if y(i)>1
        y(i)=1;
    end
    if y(i)<0
        y(i)=0;
    end
end
plot(x,y)

```

问题三

实际时间差的多元回归分析结果

模型摘要

模型	R	R 方	调整后 R 方	标准估算的错误
1	.661 ^a	.068	.068	38.25436

a. 预测变量: (常量), 线路总成本, 线路价格 (不含税), 计划发车

ANOVA

模型		平方和	自由度	均方	F	显著性
1	回归	1689681.614	4	422420.403	288.658	.000 ^b
	残差	23190433.965	15847	1463.396		
	总计	24880115.578	15851			

a. 因变量: 实际时间差

b. 预测变量: (常量), 线路总成本, 线路价格 (不含税), 计划发车和到达时间差, 总里程

系数^a

模型		未标准化系数		标准化系数	t	显著性
		B	标准错误	Beta		
1	(常量)	-1.085	.374		14.931	.000
	总里程	-.00235	.009	-.412	-2.656	.008
	线路价格 (不含税)	.002	.001	.184	3.533	.000
	计划发车和到达时间差	.942	.380	.281	2.477	.013
	线路总成本	.003	.001	.208	2.029	.043