

第六届“认证杯”数学中国

数学建模网络挑战赛

承 诺 书

我们仔细阅读了第六届“认证杯”数学中国数学建模网络挑战赛的竞赛规则。

我们完全明白，在竞赛开始后参赛队员不能以任何方式（包括电话、电子邮件、网上咨询等）与队外的任何人（包括指导教师）研究、讨论与赛题有关的问题。

我们知道，抄袭别人的成果是违反竞赛规则的，如果引用别人的成果或其他公开的资料（包括网上查到的资料），必须按照规定的参考文献的表述方式在正文引用处和参考文献中明确列出。

我们郑重承诺，严格遵守竞赛规则，以保证竞赛的公正、公平性。如有违反竞赛规则的行为，我们将受到严肃处理。

我们允许数学中国网站(www.madio.net)公布论文，以供网友之间学习交流，数学中国网站以非商业目的的论文交流不需要提前取得我们的同意。

我们的参赛队号为：1065

参赛队员（签名）：

队员 1：汪榕

队员 2：舒梁博

队员 3：韩旺

参赛队教练员（签名）：

参赛队伍组别：本科组

第六届“认证杯”数学中国

数学建模网络挑战赛

编号专用页

参赛队伍的参赛队号：1065

竞赛统一编号（由竞赛组委会送至评委团前编号）：

竞赛评阅编号（由竞赛评委团评阅前进行编号）：

2013 年第六届“认证杯”数学中国 数学建模网络挑战赛

题 目 公路运输业对于国内生产总值的影响分析

关 键 词 公路运输业 国内生产总值 改进加权主成分分析法模型 熵权的属性

识别模型 时间序列的单位根检验模型

摘 要

本文研究的是关于公路运输业对于国内生产总值影响的问题，通过分析题目所给的数据建立了无量纲标准化模型、指数平滑的时间序列模型、灰色系统模型、曲线拟合模型、聚类分析模型、效益综合模型、改进加权主成分分析法模型、熵权的属性识别模型、误差修正模型、时间序列的单位根检验模型、协整分析模型，利用 MATLAB 软件和 SPSS 软件对上述模型进行逐一求解，分别回答了题目所给的所有问题。

针对问题三，首先利用 SPSS 软件对原始数据填补缺失值和分析异常值，并用无量纲化处理对公路运输指标进行分类和标准化。再用指数平滑的时间序列模型、灰色系统模型、曲线拟合投资值变化趋势法分别预测公路运输投资强度，得到未来 5 年公路投资强度分别为：74.7514、74.6070、72.0358、67.2308 和 60.5528。然后用聚类城市等级量化模型对经济效果影响不同的城市进行投资力度分类（其分类效果详见 9 页），进而利用加权主成分分析法求解出各类城市对该省经济效果的贡献值。最后综合各市投资增长率、公路运输投资强度及其经济效果贡献值用熵权分析法求解出未来 5 年各城市公路建设资金投入的具体比例（详见 14 页）。

针对问题四，为了结合附件所给的数据将公路运输业对 GDP 的影响模型进行合理的修正，首先通过公路投资与国民经济发展指标的相关分析得到一般性相关关系。其次通过 Engle—Granger 两步法分析公路投资与国民经济发展指标的协整过程得出公路投资与国民经济发展之间形成了长期均衡关系。然后利用 Granger 因果关系检验 Engle—Granger 两步法所得结论得出其模型的正确性和精确度。最后得出公路投资额与国民经济发展指标 GDP 之间有着高度的正相关关系。

参赛队号： 1065

参赛密码 _____
(由组委会填写)

所选题目： C 题

英文摘要

This study is about the issues affecting the road transport industry to GDP, to establish the standardized model of the non-dimensional data subject to analysis, time series model of exponential smoothing, gray model, curve fitting model, cluster analysis model, effective integrated model, weighted principal component analysis model, entropy weight attribute recognition model, mistaking correction model, the unit root test of time series model and the cointegration analysis model,. Using the MATLAB software and the SPSS software ,we have solved the model and answered all the problems of the project.

For the problem 3. At first, we use the SPSS software to fill the missing values of the original data, classify and standard dimensionless classification of the road transport indicators. Then the exponential smoothing time series model, gray model, curve fitting the investment value trend method to predict the intensity of road transport investment, the next five years, road investment strength: 74.7514, 74.6070, 72.0358, 67.2308 and 60.5528. Then city level quantitative model clustering different impact on the economic effects of city the investment classification, and then use the weighted principal component analysis method to solve a range of urban contribution to the province's economic effect. Finally, consolidated municipalities investment growth, investment in road transport intensity value and its contribution to economic effects analysis of entropy weight method to solve a specific percentage of each city highway construction funding the next five years.

For the problem 4. At first, in order to combine the data given in the Annex to the road transport industry to GDP model is a reasonable amendment, first obtained by the correlation analysis of the indicators of road investment and the development of the national economy general relationship. On co-integration process followed by the Engle-Granger two-step analysis of highway investment and the development of the national economy indicators formed a long-term equilibrium relationship between highway investment and the development of the national economy. Then on Granger causality test Engle-Granger two-step method and concluded the correctness and accuracy of the model. Coming to the conclusion that a high degree of positive correlation between highway investment and national economic development indicators of GDP.

Key words: Highway transportation industry; GDP; Weighted principal component analysis model; Entropy of Attribute Recognition Model; Unit root test mode

一 问题的立意与背景

1.1 背景资料：

该题目是研究的公路运输业对于国内生产总值影响的问题，交通运输作为国民经济的载体，沟通生产和消费，在经济发展中扮演着极其重要的角色。纵观几百年来交通运输与经济发展的相互关系，生产水平越高，就越要求基础结构超前发展。工业化时期的基础结构，已经不允许交通运输滞后。进入现代化社会，经济社会对交通运输的要求本质上就是超前的，交通运输是国民经济的先行官，发展经济，交通先行，是经济发展的内在规律。公路运输是在公路上运送旅客和货物的运输方式，是交通运输系统的组成部分之一，主要承担中短途客货运输。发展公路运输对国内生产总值（GDP）增长的贡献产生于交通建设和客货运输两个阶段，表现为公路运输对国民经济的直接贡献、波及效果、对于相关行业的直接消费以及创造就业机会等几个方面。

某省的统计部门想通过调查研究的方法估计公路运输业对于GDP的影响，通过随机发放问卷，获得了附件1中所示的数据，该数据为真实调查得到的原始数据。通过以上数据进行数学建模，并估计了该省公路运输业对于GDP的影响（问题1）。同时，为了提高问题一的模型，考虑了所获数据的情况，因此对现有的调查项目做了调整，并设计了调查项目表（问题2）。

1.2 需要解决的问题：

- 1.如何评价深圳市这三年各主要食品领域微生物、重金属、添加剂含量等安全情况的变化趋势；
- 2.从这些数据中能否找出某些规律性的东西：如食品产地与食品质量的关系；食品销售地点（即抽检地点）与食品质量的关系；季节因素等等；

二 问题的解决思路

根据这个问题的实际背景和现有的抽检数据，首先依据2007年到2011年公路运输的统计数据对该省11个城市的投资情况进行强度分类，再结合历年公路投资的强度变化作预测分析，利用改进的加权主成分分析法模型推算出11个城市经济效益综合贡献值。从而结合熵权算法模型求解出未来5年该省每个城市公路投资比例；最后结合上述问题分析的公路运输对GDP的影响模型，利用相关分析以及循环检验更深入的讨论了公路运输业对GDP的影响效益。

问题1）通过题目所给附件的信息，结合问题的求解进行了合理的等级分类，利用将原始处理进行缺失值和异常值的处理，再根据分类标准进行数据整合。首先通过历年公路投资强度进行逐步预测分析，从而得到未来5年该省公路投资强度的变化情况，再根据前5年各城市的投资强度，利用SPSS软件进行聚类分析，从而得到重投入、较重投入、轻投入和最小投入所代表性的城市，通过改进后的加权主成分分析法模型求解出各城市公路运输对经济效益的综合贡献值。最后结合投入增长率、城市经济贡献值和城市建设强度值，利用熵权分析模型求解出未来5年各城市公路建设资金投入的具体比例，

问题2）通过对上述公路运输对GDP影响模型的分析，结合附件3中2007年到2011年公路运输统计的数据进行模型修正。首先结合公路投资强度值和GDP的相关数据初步分析出公路运输投资与GDP的相关性，再通过公路投资与国民经济协整关系分析及误差修正模型、公路投资额与GDP时间序列的单位根检验、公路投资额与GDP的协整分析和公路投资额与GDP的协整分析一系列步骤深入的分析公路运输对GDP的影响，最后结合检验环节更好的证明了上述分析的正确性。

三 基本假设

1. 假设本文附件中的缺失值和异常值是由于工作人员的忽视而造成了，而通过处理后的异常和缺失数据与原始数据差异性不大。
2. 假设该省每个城市公路建设投资的强度与城市公路经济贡献效益成正相关。
3. 假设将该省 11 个城市公路投资的强度划分为最低投入、低投入、较高投入和高投入四各等级进行分析。
4. 假设该省公路投资金额用公路投资强度来确定，其中公路里程增加的值与单位面积公路建设资金相乘表示公路投资强度。
5. 假设未来 5 年该省各城市公路投资总强度由 11 个城市公路效益经济综合值、城市建设权重和未来 5 年该省公路投资变化率来决定。

四 符号系统

u_{α} ---标准正态分布 α 分位点	p_k ---第 k 类投资所占比例
y_{ij} --- i 响应在 j 实验中的均值	Y_{ij} --- i 响应在 j 实验中标准化的均值
W_q ---第 q 个主成分的方差贡献值	λ_p ---协方差矩阵的第 p 个特征值
Δ -----预测投资强度的量	Δ_i -----该省公路第 i 年的投资变化量
g_i -----该省公路建设第 i 年的强度变化率	

五 模型的建立与求解

5.1 公路运输投资资金的分配比例预测综合模型

1 原始数据的汇总

由于题目中的数据存在缺失和异常，因此通过 SPSS 软件的处理原理对数据进行缺失值填补和异常值分析，再根据题目求解的方法，分别对该省各城市有用的信息进行筛选，其筛选原则如下所示：

I 针对公路运输投资资金方向，通过分析附件所给的信息量，结合题目所求解的方向，最终筛选出等级公路通车里程、等级公路长度和等级公路面积作为公路投资资金的集中体现形式。

II 针对公路运输对经济效益的影响方向，通过前面两问的求解结果进行分析，最终筛选出全社会旅客运输总量、全社会货运总量、道路运输从业人员和汽车维修服务网点数作为影响效益的体现形式。

通过如上筛选原则，将其筛选的信息量数据进行整合处理，最终得到汇总表格见（25 页）附录 1。

2 公路运输指标的分类

结合题目所给的附件，为了更好的求解题目所设定的题目，需要首先对表格中所给的信息指标量进行分类处理，根据查阅相关文献进行分类，其效果图如下所示：

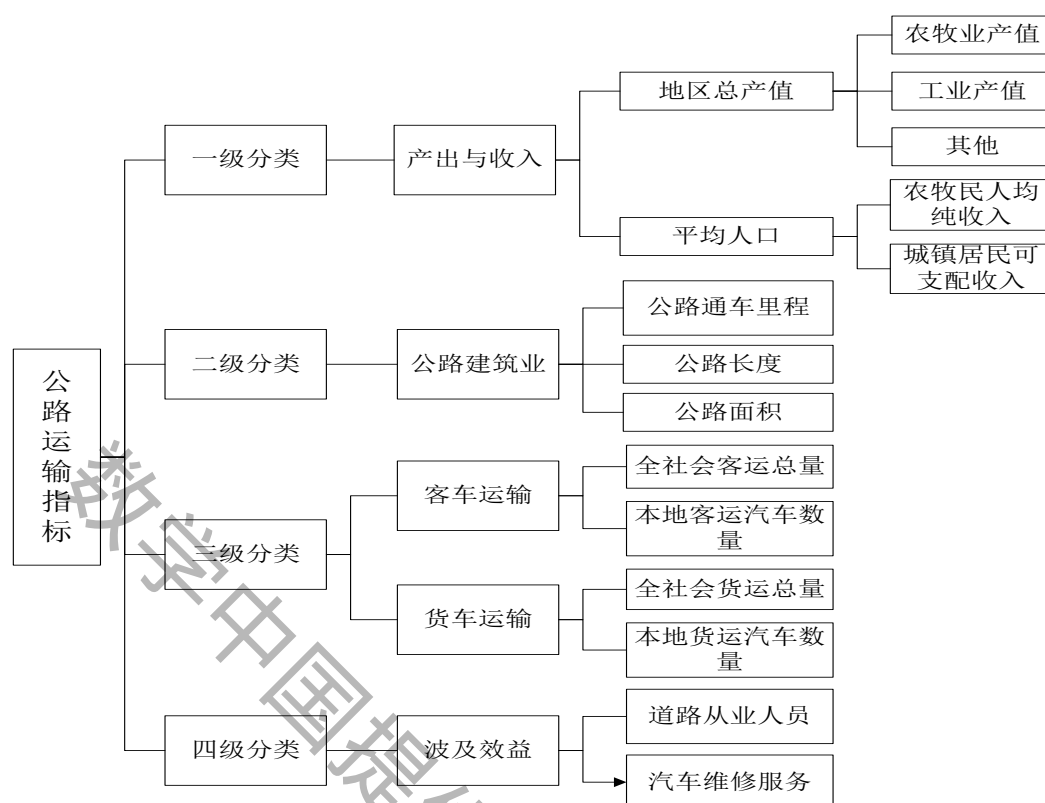


图1 公路运输分类图

如上图1所示，需要说明一点，运输指标是公路发展水平的一个重要标志，它表现在严格而高效的数据统计方法和该地区的发展情况，最终体现为该城市公路资金的投入比例，鉴于很多指标难以量化，从可操作性考虑，这里需要对汇总数据进行数据标准化处理。

3 分类数据的无量纲化处理

1) 模型前期的数据处理^{[1][2]}

通过对附件的数据观测，发现存在缺失值数据以及异常数据，因此我们首先对异常数据进行筛选，再将得到筛选表进行缺失值处理，其理论依据如下所示：

① 缺失值的表现形式：

- i 完全随机缺失，表示缺失和变量的取值无关。
- ii 随机缺失，缺失分布中调查变量只依赖于数据组数中有的记录变量。
- iii 非随机缺失，数据的缺失不仅和去他变量的取值有关，也和自身有关。

SPSS 主要对 MCAR 和 MAR 两种缺失值情况进行分析。区别 MCAR 和 MAR 的含义在于：由于 MCAR 通常实际上很难遇到，应该在进行调查之前就去考虑那些重要变量可能会有非无效的未回答，还要尽量在调查中包含工变量，以使用这些变量来估算缺失值。

② 针对不同情况的缺失值，SPSS 操作给出了三种处理方法：

i 删除缺失值，这种方法适用于缺失值非常少的时候，它不需要专门的步骤，通常在相应的分析对话框中的“选项”子对话框中进行设置。

ii 替换缺失值，利于“转化”菜单中的“替换缺失值”命令过程将所各有的记录看成一个序列，然后采用某种指标对缺失值进行填充。

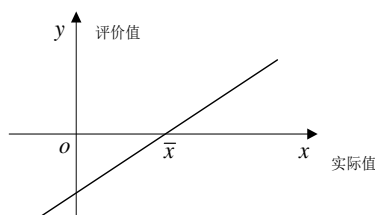
iii 缺失值分析过程，缺失值分析过程是 SPSS 专门针对缺失值分析而提供的模块。

2) 公路投资数据的标准化方法

统计学理论告诉我们，要对多组不同量纲的数据进行比较，可以先将它们分别标准化，转化成无量纲的标准化数据。综合评价就是要将多组不同的数据进行综合，因而可以借助于标准化方法来消除数据量纲的影响。标准化公式为：

$$y_i = \frac{x_i - \bar{x}}{s} \quad (1)$$

上式中 $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ ， $s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$ ，指标实际值与评价值的关系如下所示：



可以看出，无论指标的实际值如何，指标的评价值总是分布在零的两侧。指标实际值比平均值大的，其评价值为正，反之为负，实际不同在于：第一，它利用了原始数据的所有信息；第二，它要求样本数据较多；第三，它的评价值结果超出 $[0,1]$ 区间，有正有负。为了更符合习惯，我们可以将其转化为百分数形式，比如用公式：

$$\begin{aligned} y_i &= 60 + \frac{x_i - \bar{x}}{10s} \times 100 \\ &= 60 + \frac{x_i - \bar{x}}{s} \times 10 \end{aligned} \quad (2)$$

均值转化为 60. 超过均值的转化为 60 以上，反之在 60 一下。这种“百分数”还不同于一般的百分数，因为个别极端数值的转化只可能超过 $[0,100]$ 区间。另外，也可以将均值转化为 50。

4. 公路运输投资强度的综合预测模型

模型一：指数平滑的时间序列模型

① 模型的建立

步骤一：确定预测方程如下：

$$\hat{L}(k) = a \times b^k \quad (a, b \text{ 为待定系数})$$

步骤二：对测量值取对数可得

$$\lg \hat{L}(k) = \lg a + k \cdot \lg b$$

令 $\hat{a} = \lg a$ ， $\hat{b} = \lg b$ ，则

$$\lg \hat{L}(k) = \hat{a} + \hat{b} \times k$$

步骤三：采用三项加权平均值求解指数方程的待定系数，令

$$\begin{cases} R = \frac{1}{6} [\lg L(1) + 2 \lg L(2) + 3 \lg L(3)] \\ T = \frac{1}{6} [\lg L(n-2) + 2 \lg L(n-1) + 3 \lg L(n)] \end{cases}$$

根据 R , T 计算出 \hat{a} , \hat{b} , 得

$$\begin{cases} \hat{b} = \frac{T-R}{N-3} \\ \hat{a} = R - \frac{7}{3} \hat{b} \end{cases}$$

步骤四：将 \hat{a} , \hat{b} 代入 $\lg \hat{L}^{(i)}(k) = \hat{a} + \hat{b} \times k$ 即可得预测方程。

② 模型求解

对蛋制品中所给的抗生素指标数据进行曲线拟合，考虑到数据的正负号，因此首先通过公式(2)进行负向指标正向化处理，最后再利用 MATLAB 软件求解结果如下所示：

$$R = 2.040379, T = 2.391256; \quad \hat{b} = 0.175439, \hat{a} = 1.631022;$$

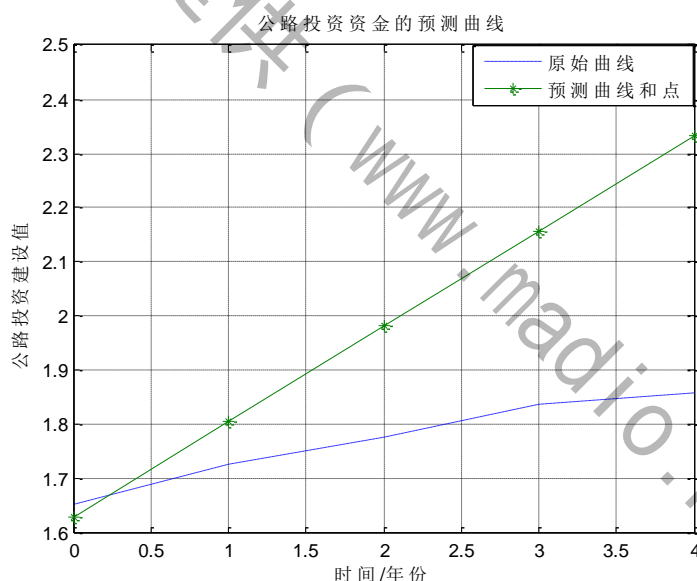


图 2 公路投资强度的预测

预测结果分析：预测点的值与原始值相差太大，由此认为用平滑指数的时间序列法对投资资金进行预测不可行。

模型二：灰色系统模型

① 模型的建立

设时间序列 $X^{(0)}(k) = \{x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n)\}$ ，共有 n 个观察值，其中

$$x^{(0)}(k) \geq 0 \quad k = 1, 2, \dots, N,$$

步骤一：对 $X^{(0)}$ 做一次累加生成列 $X^{(1)}$ ，即

$$X^{(1)}(k) = \sum_{i=1}^k X^{(0)}(i) = X^{(1)}(k-1) + X^{(0)}(k)$$

步骤二：定义矩阵 $\mathbf{B} = \begin{bmatrix} -\frac{1}{2}[X^{(1)}(1) + X^{(1)}(2)] & 1 \\ -\frac{1}{2}[X^{(1)}(2) + X^{(1)}(3)] & 1 \\ \vdots & \vdots \\ -\frac{1}{2}[X^{(1)}(n-1) + X^{(1)}(n)] & 1 \end{bmatrix}$ $\mathbf{Y}_n = \begin{bmatrix} X^{(0)}(2) \\ X^{(0)}(3) \\ \vdots \\ X^{(0)}(n) \end{bmatrix}$

步骤三：利用最小二乘法求解可得

$$\hat{\mathbf{a}} = (a, u)^T = (\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T \mathbf{Y}_n$$

求解微分方程

$$\frac{dX^{(1)}}{dt} + aX^{(1)} = u$$

可得预测模型

$$\hat{X}^{(1)}(k+1) = \left[X^{(1)}(1) - \frac{u}{a} \right] e^{-ak} + \frac{u}{a} \quad k=1, 2, \dots, n$$

则预测值为

$$\hat{X}^{(0)}(k+1) = \hat{X}^{(1)}(k+1) - \hat{X}^{(1)}(k) \quad k=1, 2, \dots, n$$

当 $1 \leq k \leq N$ 时，上式反映了原始数据列的变化情况；当 $k > N$ ，上次为预测值。

② 模型求解

步骤一：由表经过筛选后的数据得原始序列：

$$X^{(0)} = \{44.48557, 54.53204, 59.79179, 69.03974, 72.15087\}$$

步骤二：作一次累加后得序列：

$$X^{(1)}(k) = \{44.48557, 99.01761, 158.8094, 227.8491, 300\}$$

步骤三：由矩阵 \mathbf{B} 的值计算得

$$\hat{\mathbf{a}} = (a, u)^T = \begin{bmatrix} -0.0241 \\ 48.0294 \end{bmatrix}$$

$$\begin{aligned} \hat{X}^{(1)}(k+1) &= \left[X^{(1)}(1) - \frac{u}{a} \right] e^{-ak} + \frac{u}{a} \\ &= 2037.407e^{0.0241k} - 1992.92 \end{aligned}$$

步骤四：进而可求得

$$\hat{X}^{(0)}(k+1) = \hat{X}^{(1)}(k+1) - \hat{X}^{(1)}(k) = 2037.407[e^{0.0241k} - e^{0.0241(k-1)}] \quad k=1, 2, \dots, n$$

为了形象的表示出来，作出图 3

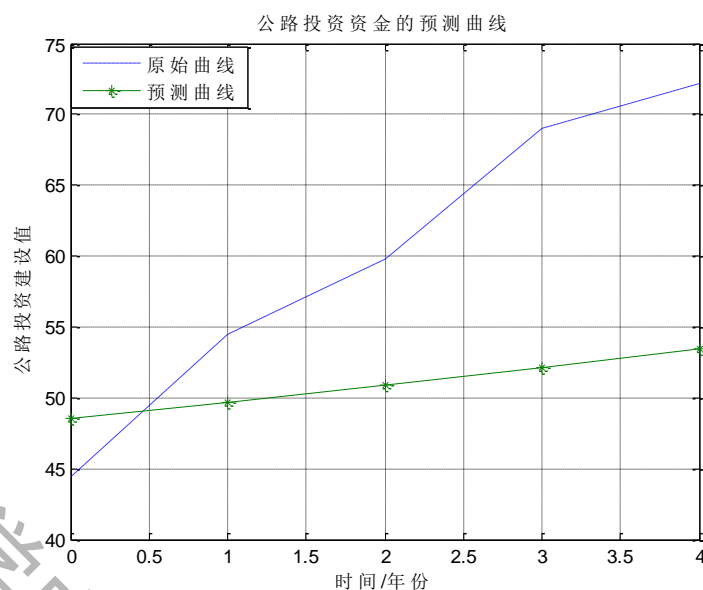


图3 原始值与预测值对比图

结果分析：由上图3可知，预测的公路投资值变化趋势呈现缓慢增长，即该省公路运输的投资情况可以无限制的增加，这显然与实际情况不吻合。

模型三：曲线拟合投资值变化趋势法

步骤一：对上述原始数据进行曲线拟合如下所示。

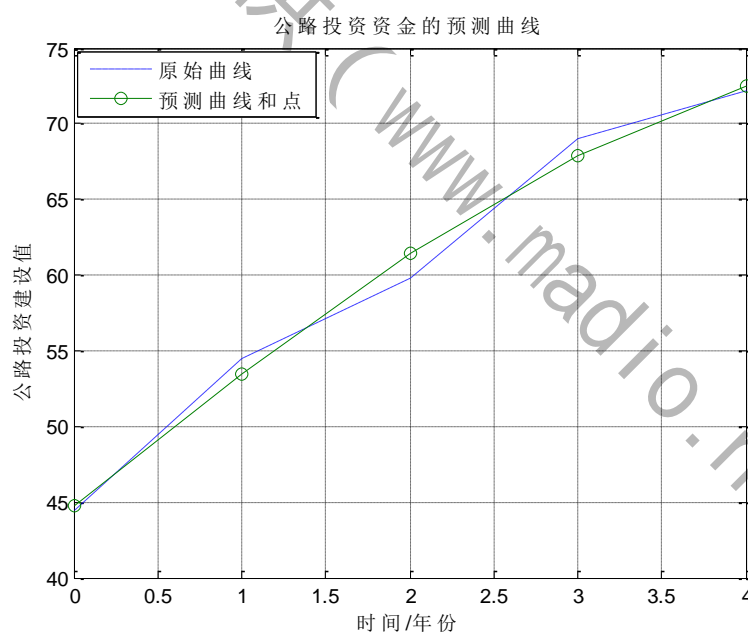


图4 公路运输投资强度曲线拟合

步骤二：为了清晰地观察投资的情况变化率，对图4得到的拟合方程求导，得到投资金额的变化率曲线，如图5所示：

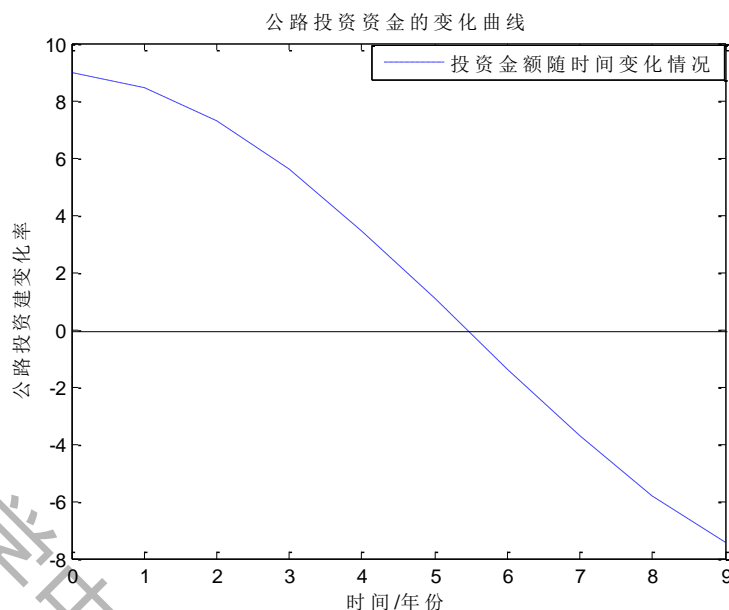


图 5 投资强度随时间变化情况

定义最佳投资时间和公路饱和的判断标准：

① 由于该省的公路发展存在一个饱和情况，通过上图可以观察到公路投资在什么时候达到饱和情况。

② 随着公路的投资加大，该省所需要修的公路越来越少，因此公路运输投资的情况会愈来愈小。

通过上述求解，预测出未来 5 年该省公路建筑的投资统计表如下所示：

表 1 未来 5 年公路运输投资强度的变化表

年份	2012	2013	2014	2015	2016
公路投资强度	74.7514	74.6070	72.0358	67.2308	60.5528

通过表 1 所求解的数据，就可以得到未来 5 年该省公路投资强度的具体变化数据值。

5.不同城市公路投资强度的聚类效果^[3]

为了更好地分析今年来公路建设资金来源结构的变化，将运用多远容积分析中的聚类分析方法来分析自 1991 年以来 公路投资资金来源结构的变动状况及变动原因。

聚类分析的基本思想是首先将每一个样品当做一类，然后根据样品之间的相似程度并类，并计算新类与其他类之间的距离，在选择相似者并类，每合并一次减少一类，继续这一过程，知道所有的样品都并成一类为止。在聚类过程中，本文选用欧氏距离

$$d_{ij}(2) = \left[\sum_{h=1}^m (x_{ik} - x_{jh})^2 \right]^{\frac{1}{2}} \quad (m \text{ 表示每个样品的变量数})$$

来度量类与类之间的相似程度，聚类方法采用最短距离法

$$D_s(p, q) = \min \{ d_{ij} \mid j \in G_p, k \in G_q \},$$

它等于类 G_p 和类 G_q 中最邻近的两个样品的距离。

考虑到城市之间发展的特殊性，因此各个城市的投资比例中，我们结合附件 2 整合后的数据，利用 SPSS 软件进行聚类分析，最终得到该省公路投资情况在 11 个城市比例聚类效果如下图所示：

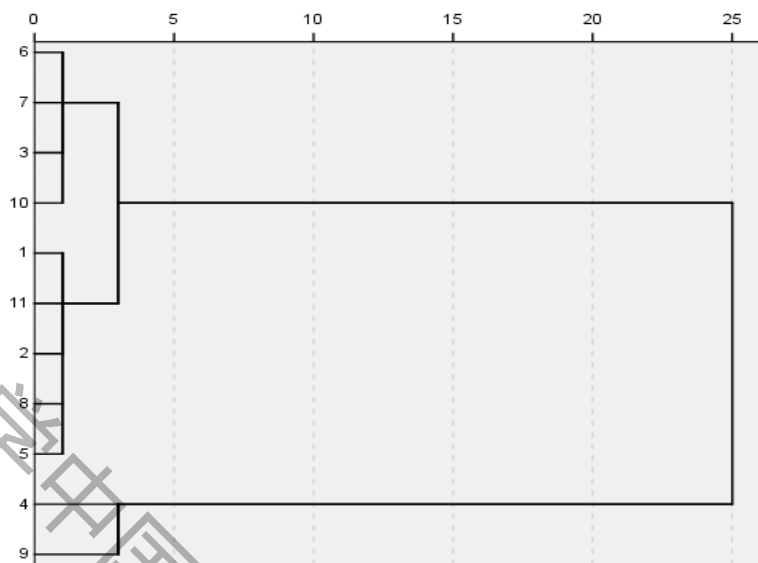


图 6 不同城市投资聚类图

6. 聚类城市等级量化指标模型

1) 定性指标的分类标准

根据安全指标处理后的数据特点，我们结合 Likert 等级分类原理进行等级划分，其区间划分效果如下：

表 2 不同城市公路运输投资的强度类别

类别编号	城市运输类别	城市名
1	低投入	城市 1
		城市 2
		城市 5
		城市 8
		城市 11
2	最低投入	城市 3
		城市 6
		城市 7
		城市 10
3	较高投入	城市 9
4	高投入	城市 4

通过表 2 所分类后的城市投资情况，了解到城市 4 为重点投资的城市，其次为城市 9，接着为城市 1、2、5、8 和 11，而最低投入为城市 3、6、7 和 10。

2) 定性指标的量化方法

在综合评价时，会遇到上述定性的指标，这些指标必须经过处理才能与其他量化指标一起运用，若用统计的术语来叙述，使公式更易于理解和表示。设 u_{α} 使

$$\int_{-\infty}^{u_{\alpha}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \alpha$$

则称 u_{α} 是标准正态分布 $N(0,1)$ 的 α 分位点，因此上面的计算中的 x_i ，可用 α 分位点

表示： $x_1 = u_{0.02}$ ， $x_1 = u_{0.02}$ ， $x_1 = u_{0.02}$ ， $x_1 = u_{0.02}$

一般来说，若有 k 类 a_1, a_2, \dots, a_k ， a_1 最差， a_2 比 a_1 好，……，依次递升， a_k 最好， a_i 类各占的比例和累计比例如下述：

类	a_1	a_2	a_3	$\dots a_k$
各类比例	p_1	p_2	p_3	$\dots p_k$
累计（由 a_1 起始）的比例	p_1	$p_1 + p_2$	$p_1 + p_2 + p_3$	$\dots 1$

于是 a_i 对应的 x_i 应有性质：

$$P(x < x_i) = \sum_{j=1}^{i-1} p_j + \frac{1}{2} p_i, i=1, 2, \dots, k$$

用 $N(0,1)$ 的 α 分位点 u_{α} 来表示，就得

$$x_1 = u_{\frac{1}{2}p_1}, x_2 = u_{\frac{p_1 + \frac{1}{2}p_2}{2}}, \dots, x_i = u_{\frac{\sum_{j=1}^{i-1} p_j + \frac{1}{2}p_i}{2}}, \dots, x_k = u_{\frac{\sum_{j=1}^{k-1} p_j + \frac{1}{2}p_k}{2}} \quad (2)$$

这就给出了一般化的表达式。

最终通过公式（2）求解得到每一类开发强度的城市所量化后的数据见下表所示：

表 3 城市分类指标的量化处理

城市开发强度	高投入	较高投入	低投入	最低投入
量化数据	0.8301	0.8061	0.7227	0.5721

通过上表所量化后的数据，就可以得到不同发展城市公路建设的开发值的数据，通过此数据，再结合经济效益数据进行问题分析。

7. 每年内的公路投资对该省各个城市的经济效益综合模型

1) 原始数据的预处理

I 对原始数据标准化处理。公式为 x_i （正指标）， $x'_{ij} = \sum \frac{x - x_{ij}}{s_i}$ （逆指标）。

II 求出各指标两两间的相关系数。公式为

$$r_{ij} = \frac{1}{n} \sum x'_{ik} x'_{jk}$$

III 定义有 p 个响应变量，共进行 m 次试验，以 y_{ij} 表示第 i 个响应在第 j 个实验中的均值。为消除量纲的影响，对各响应均值进行标准化，得到标准化均值，如下：

$$Y_{ij} = \frac{y_{ij} - \mu}{\sigma_i} \quad (3)$$

通过上述的数据处理方法，最终得到的整合后的多重输入与响应表，详细见附录 1。

2) 模型的前期理论分析^{[4][5]}

用 Y_j 表示第 j 个响应的标准化信噪比(望大特性)，对 Y_j 进行主成分分析，可将 p (假设多响应问题有 p 个输出响应) 个响应转化为 $k(k \leq p)$ 个无关的主成分，即：

$$Z_q = e_{q1} Y_1 + e_{q2} Y_2 + \dots + e_{qp} Y_p \quad (4)$$

其中， Z_q 表示第 q 个主成分， $q=1,2,\dots,k$ 。 $Y=(Y_1, Y_2, \dots, Y_p)$ ，第 q 主成分的系数 $e_{q1}, e_{q2}, \dots, e_{qp}$ 是 Y 的协方差矩阵的第 q 个特征值向量的元素，且满足 $e_{q1}^2 + e_{q2}^2 + \dots + e_{qp}^2 = 1$ 。

主成分的方差贡献率为该主成分所解释的方差占总方差的比例，反映各主成分的信息量含量多少，其值为：

$$W_q = \lambda_q / \sum_{q=1}^k \lambda_q$$

其中， W_q 为第 q 个主成分的方差贡献率， λ_q 为 Y 的协方差矩阵的第 q 个特征值。

3) 综合指标显著性影响与城市公路贡献效益的加权主成分分析法模型

各相应的标准化均由前文中得到，利用相应均值可得到 11 个相应的回归模型及 R^2 值，如下所示：

$$y_1 = 0.091x_1 - 0.033x_2 - 0.143x_3 + 0.071x_4 + 0.03, R^2 = 88.08\%;$$

$$y_2 = -0.038x_1 - 0.021x_2 - 0.067x_3 - 0.061x_4 + 0.079, R^2 = 78.58\%;$$

$$y_3 = 0.027x_1 + 0.135x_2 - 0.025x_3 - 0.088x_4 + 0.076, R^2 = 89.06\%;$$

$$y_4 = 0.029x_1 - 0.106x_2 - 0.209x_3 + 0.068x_4 + 0.12, R^2 = 90.02\%;$$

$$y_5 = 0.062x_1 + 0.038x_2 - 0.001x_3 - 0.041x_4 + 0.327, R^2 = 93.25\%;$$

$$y_6 = -0.055x_1 - 0.044x_2 + 0.033x_3 - 0.005x_4 + 0.1, R^2 = 79.65\%;$$

$$y_7 = 0.06x_1 - 0.053x_2 - 0.058x_3 + 0.066x_4 + 0.036, R^2 = 68.86\%;$$

$$y_8 = 0.008x_1 - 0.039x_2 + 0.005x_3 - 0.007x_4 + 0.146, R^2 = 84.98\%;$$

$$y_9 = -0.029x_1 - 0.004x_2 - 0.004x_3 - 0.007x_4 + 0.15, R^2 = 69.86\%;$$

$$y_{10} = -0.339x_1 + 0.082x_2 + 0.738x_3 - 0.009x_4 - 0.328, R^2 = 98.06\%;$$

$$y_{11} = 0.014x_1 - 0.002x_2 + 0.013x_3 - 0.009x_4 - 0.001, R^2 = 87.46\%;$$

上述表达式中： x_1 x_2 x_3 x_4 代表全社会旅客运输总量（万个）、全社会货运总量（万吨）、道路运输从业人员（个）和汽车维修服务网点数（个）。

由此得到标准化的预测能力如下表：

表 4 标准化预测数据表

城市	r_1	r_2	r_3	r_4	r_5	r_6	r_7	r_8	r_9	r_{10}	r_{11}
数据	0.094	0.084	0.096	0.097	0.100	0.085	0.074	0.091	0.075	0.105	0.094
	9	7	0	0	5	8	2	6	3	6	2

上述所求数据的方差协方差矩阵为 $\begin{pmatrix} 1, 0.000 \\ 0.000, 1 \end{pmatrix}$ ，可见两主成分相关性不存在，适合于

上述提出的加权主成分分析法。

利用 SPSS 软件进行主成分分析，可得到主成分的特征值和特征向量矩阵，如下表所示为城市 1 的加权分析表（其余城市的见附录 2）。

表 5 城市 1 加权主成分分析表

主成分	初始特征值			特征值向量			
	合计	方差%	累计%	X_1	X_2	X_3	X_4
1	3.966	99.143	99.143	0.4962	-0.5275	0.4296	0.4263
2	0.033	0.821	99.963	0.1728	-0.1438	0.3982	-0.1839

由式（4），得到相应的两个主成分：

$$Z_1 = 0.4962 * 0.0949X_1 - 0.5275 * 0.0949X_2 + 0.4296 * 0.0949X_3 + 0.4263 * 0.0949X_4$$

$$Z_2 = 0.1728 * 0.0949X_1 - 0.1438 * 0.0949X_2 + 0.3982 * 0.0949X_3 - 0.1839 * 0.0949X_4$$

由两个主成分，可得到改进后的多响应城市公路运输经济效益值：

$$Z = 0.1873X_1 - 0.1990X_2 + 0.1629X_3 + 0.1599X_4$$

通过上述公式（2）的分析，本文将经济效益中的变化进行百分数转化，再根据上式可计算出城市 1 的公路运输效益综合值为，利用这个方法，分别对其他 10 个城市进行加权主成分分析法，其方差贡献率见附录 2，通过计算最终贡献效益如下表所示：

表 6 城市综合效益数据表

城市类别	城市 1	城市 2	城市 3	城市 4	城市 5	城市 6
经济效益得分	7.78	7.49	6.98	8.84	7.85	6.70
城市类别	城市 7	城市 8	城市 9	城市 10	城市 11	城市效益 综合得分 表
经济效益得分	6.49	7.88	9.08	6.80	7.06	

通过表 6 可看出各城市公路运输经济效益的综合值，其中综合值越大，说明该城市效益情况较为显著，应该加大投资力度，确保该省经济的高速发展。

7.公路投资资金在各市的分配比例模型

1) 投资强度变化率的分析模型

城市建设投资对 GDP 增长的贡献是多方面的,但最根本的是满足经济发展对运输日益变化的需求,正因为投资增长(可用公路里程增加强度来表示)是经济发展的必须结果,因此,我们可以把该省公路建设投资的变化量与其对应年份的公路投资的强度之比看做对该省公路建设增长的建设强度变化率。可用公式表示为:

$$g_i = \frac{\Delta_i}{\Delta} \quad (5)$$

公式中的 g_i 代表该省公路建设这年的强度变化率, Δ_i 代表该省公路建设投资的变化量, Δ 代表预测投资强度的量。

利用上述模型求解得到未来 6 年内公路投资强度 Q 的预测数据为:

$$Q = [74.7514, 74.6070, 72.0358, 67.2308, 60.5528, 64.3582]$$

根据数据求得到未来五年的投资强度增长率分别为:

$$q = [-0.0019, -0.0345, -0.0667, -0.0993, 0.0628]$$

结合上述问题求解得到的公路运输对经济贡献值,通过分析了解到公路投资对产生的经济效益是一个互相联系的关系,投入决定产出效果,因此假设经济效益变化率与投入增长率相同,最后利用前 5 年的经济贡献矩阵,求解得到未来 5 年公路运输经济贡献值为矩阵:

2) 各城市未来 5 年公路建设投入比例模型

2.1) 未来 5 年各省公路运输经济效益矩阵

结合表 8 所求得该省各城市公路运输贡献经济值和未来 5 年投资强度增长率,可以得到未来 5 年的各城市公路运输效益预测值如下表示所示:

表 7 未来 5 年各城市公路运输贡献效益综合值

综合值	城市 1	城市 2	城市 3	城市 4	城市 5	城市 6	城市 7	城市 8	城市 9	城市 10	城市 11
2012 年	7.77	7.48	6.97	8.82	7.84	6.69	6.48	7.87	9.06	6.79	7.05
2013 年	7.50	7.22	6.73	8.52	7.56	6.46	6.25	7.59	8.75	6.55	6.80
2014 年	7.00	6.74	6.28	7.95	7.06	6.03	5.84	7.09	8.17	6.12	6.35
2015 年	6.30	6.07	5.65	7.16	6.36	5.43	5.26	6.38	7.36	5.51	5.72
2016 年	6.70	6.45	6.01	7.61	6.76	5.77	5.59	6.78	7.82	5.85	6.08

结合表 7 所求解的数据,利用本文下面模型进行各城市投资比例分析。

2.2) 基于熵权的属性识别模型^[7]

① 熵值法确定权重系数

在信息论中,信息熵反映了信息无序化程度,信息熵越小,系统无序度越小,信息效用值越大,反之亦然。对于所讨论的 n 个方案的 m 个评价指标的初始矩阵,决策矩阵显然是一种信息的载体,故可用信息熵评价所获系统信息的有序度及效用,由于信息熵可用来度量 m 个指标的信息效用值,因此,利用熵值法计算个指标的权重,其本质就是利用该指标信息的效用值来计算的,效用值越高,其对评价的重要性越大。步骤如下:

i 构建 n 个方案的 m 个评价指标的判断矩阵：

$$\mathbf{R} = (x_{ij})_{nm} \quad (i = 1, 2, \dots, n, j = 1, 2, \dots, m)$$

ii 把 \mathbf{R} 归一化处理，得到 \mathbf{B} 矩阵：

$$b_{ij} = \frac{x_{ij} - x_{\min}}{x_{\max} - x_{\min}}$$

上式中， x_{\max} ， x_{\min} 分别表示同指标下不同方案中最满意者或最不满意者（根据不同的情况）。

iii 根据熵的定义， n 个方案的 m 个评价指标，可以确定评价指标的熵为

$$H_j = -\frac{1}{\ln n} \left(\sum_{i=1}^n f_{ij} \ln f_{ij} \right) \quad j = 1, 2, \dots, m \quad (6)$$

式中

$$f_{ij} = \frac{1 + b_{ij}}{\sum_{i=1}^n (1 + b_{ij})} \quad (7)$$

iv 计算评价指标的熵权 W ：

$$w_j = \frac{1 - H_j}{m - \sum_{j=1}^m H_j} \quad W = (w_j)_{j=1}^m \quad (8)$$

结合附录 5 数据归一化处理，利用熵权确定模型，通过公式（6）~（8）式进行求解，最终得到具体指标间的熵权如下所示：

表 8 未来 5 年各城市公路运输投资比例值

综合值	城市 1	城市 2	城市 3	城市 4	城市 5	城市 6	城市 7	城市 8	城市 9	城市 10	城市 11
2012 年	0.0945	0.0901	0.0824	0.1106	0.0955	0.0782	0.0751	0.0960	0.1143	0.0797	0.0836
2013 年	0.0874	0.0891	0.0769	0.1240	0.0880	0.0670	0.0712	0.0840	0.1524	0.0788	0.0812
2014 年	0.0869	0.0802	0.0816	0.1280	0.0920	0.0670	0.0751	0.0960	0.1323	0.0780	0.0829
2015 年	0.0938	0.0930	0.0841	0.1065	0.0946	0.0808	0.0783	0.0950	0.1095	0.0820	0.0851
2016 年	0.0964	0.0923	0.0852	0.1129	0.0950	0.0820	0.0762	0.0988	0.0945	0.0798	0.0869

通过上述表格所给的信息，就可以确定未来 5 年该省各城市投入比例的分布情况，针对每个城市内部县城的投资情况，再结合差异性进行分析。

5.2 公路投资与国民经济关系分析

投资与国民经济的增长，两者是辨证的关系。一方面投资能够促进国民经济的增长，国民经济的增长有众多至关重要的因素。另一方面，投资的来源又离不开国民经济的增长，投资规模的大小及其在国民收入中所占的比重，都要受国民经济发展水平的制约。因为从较为长远的角度来考察，一定时期内投资的规模最终要由国民经济的增长状况所决定。

1.公路投资与国民经济发展指标的相关分析^{[8][9][10]}

根据题目所提供的资料，得到了表 2 及图 3，公路投资强度是指实际完成的数额，国民经济发展指标选用国内生产总值即 GDP。

表 9 公路投资强度和 GDP 汇总表

年份	公路投资强度	GDP/亿元	年份	公路投资强度	GDP/亿元	年份	公路投资强度	GDP/亿元
2007	44.4856	628.386	2009	59.7918	1033.729	2011	72.1509	1543.426
2008	54.5320	829.59	2010	69.040	1236.320	2012	74.7514	1756.254

通过表中所给的数据，绘制出如下所示的数据随时间的变化图：

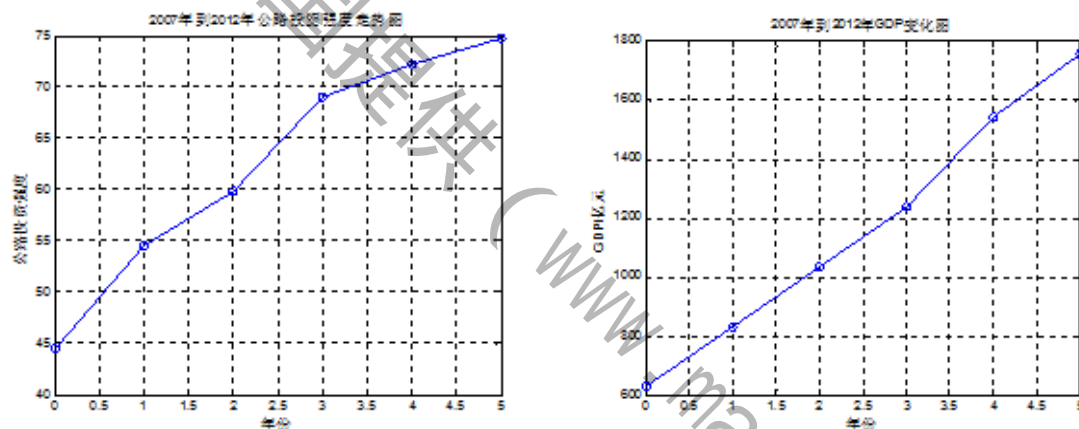


图 7 公路投资和 GDP 随时间变化走势图

由图 7 显而易见，2007 年以来，我国国民经济发展迅速，公路建设投资力度也大大加强了，二者之间是否存在相关关系，需进行相关分析才能确定。

2.公路投资与 GDP 的相关分析

相关关系是相关分析的研究对象。具体地说，相关分析是研究一个变量(y)与另一个变量(x)或另一组变量(x_1, x_2, \dots, x_n)之间相关方向和相关密切程度的一种统计分析方法。

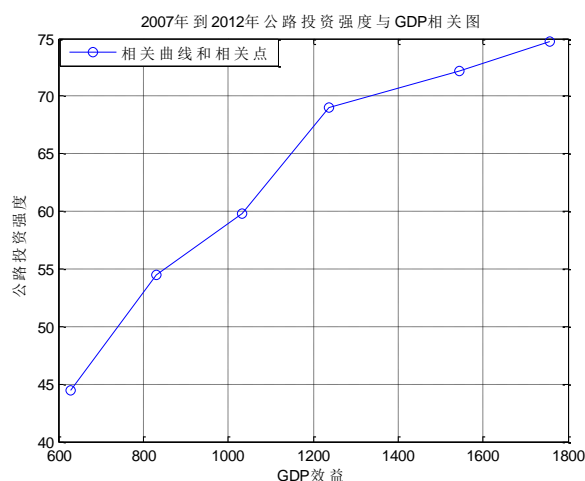


图8 公路路投资额与GDP的相关图

相关系数的基本公式为：

$$R = \frac{n \cdot \sum xy - \sum x \cdot \sum y}{\sqrt{\{n \cdot \sum x^2 - (\sum x)^2\} \cdot \{n \cdot \sum y^2 - (\sum y)^2\}}}$$

式中 n 指样本数， x 、 y 分别指相关因素。

当 $|R|=1$ 时，表示 X 与 Y 变量为完全线性相关，两者之间存在着确定的函数关系。

当 $R=0$ 时，表示 Y 的变化与 X 无关，即两者完全没有线性相关。

当 $0 < |R| < 1$ 时，表示两者之间存在着一定程度的线性相关。 $|R|$ 的数值越大，愈接近于 1，表示两者相关程度愈高。反之， $|R|$ 的数值越小，愈接近于 0，表示两者相关程度愈低。

由图 8 可看出，随着经济的增长，公路投资额也呈增长趋势，这说明公路投资额与 GDP 呈现正相关关系。进一步运用公式(3—1)来分析二者的相关程度。根据表 3—1 的数据可计算得 $R=0.960$ ，由此可知公路投资额与国民经济发展指标 GDP 之间有着高度的正相关关系。

3.公路投资与国民经济协整关系分析及误差修正模型

进一步运用协整检验来分析公路投资与国民经济发展指标的相关关系，并在此基础上建立误差修正模型。协整分析方法包括三个步骤：①在进行协整检验前，先进行单位根检验；②协整检验；③如果存在协整向量，就可以利用 ECM 来表述。本节就是根据以上三个步骤来进行研究的。

① 非平稳过程

如果随机过程 $\{y_t\}$ 的均值、方差等统计特征随时间而改变，则称此随机过程为非平稳的。一个比较常见的非平稳过程是随机游走。

对于随机过程 $\{Y_t, t=1, 2, \dots\}$ ，如果

$$Y_t = \mu + Y_{t-1} + \varepsilon_t$$

其中， $\{\varepsilon_t\}$ 为白噪声过程即独立同分布，且 $E(\varepsilon_t) = 0$ ， $Var(\varepsilon_t) = E(\varepsilon_t^2) = \sigma^2 < \infty$ ，则称其为随机游走。

显然，随机游走是非平稳过程的一个特例。通过直接迭代

$$Y_t = Y_0 + \sum_{i=1}^t (\mu + \varepsilon_i) = Y_0 + \mu t + \sum_{i=1}^t \varepsilon_i$$

即， Y_t 是许多随机变量的一个积累，具有一个明显的趋势。 Y_t 的期望

$$E(Y_t) = E\left[Y_0 + \sum_{i=1}^t \mu + \sum_{i=1}^t \varepsilon_i\right] = Y_0 + \mu t$$

和它的方差

$$\text{Var}(Y_t) = E(\varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_t)^2 = t\sigma^2$$

都是时间的函数，而且随时间分散到无穷大。序列 Y_t 并不是以大致相等的振幅围绕其趋势值波动。

当 $\mu=0$ 时，称 $Y_t = Y_{t-1} + \varepsilon_t$ 为随机游走。显然纯随机游走也是非平稳过程。当 $\mu \neq 0$ 时，称 $Y_t = \mu + Y_{t-1} + \varepsilon_t$ 为带漂移的随机游走。其一阶差分 $Z_t = Y_t - Y_{t-1} = \mu + \varepsilon_t$ 是一白噪声过程加上 Z_t ，的均值 μ ，显然这是一平稳过程。

② 单位根过程

单位根过程是较随机游走更为一般的非平稳过程。随机过程称 $\{Y_t, t=1, 2, \dots\}$ 为单位根过程，如果

$$Y_t = \rho Y_{t-1} + \varepsilon_t, \quad t=1, 2, \dots$$

其中， $\rho=1$ ， $\{\varepsilon_t\}$ 为一稳定过程，且 $E(\varepsilon_t)=0$ ， $\text{Cov}(\varepsilon_t, \varepsilon_{t-s}) = \eta_s < \infty$ 这里 $s=0, 1, 2, \dots$ 随机过程 $\{Y_t, t=1, 2, \dots\}$ 称为带常数项的单位根过程，如果

$$Y_t = \mu + \rho Y_{t-1} + \varepsilon_t$$

其中， $\mu \neq 0$ ， $\rho=1$ ， $\{\varepsilon_t\}$ 为一稳定过程。同样，它是对带漂移的随机游走的推广，

该过程反复迭代可得 $Y_t = \mu t + \sum_{i=1}^t \varepsilon_i$ 。显然，该过程也具有一个明显的趋势。

随机过程 $\{Y_t, t=1, 2, \dots\}$ 称为带趋势的平稳过程，如果 $Y_t = c + rt + \varepsilon_t$ 。其中， $\{\varepsilon_t\}$ 为一平稳过程。显然，随机变量 Y_t 的期望值是时间的函数，即 $E(Y_t) = c + rt$ 。

由此可见，带趋势的平稳过程也是非平稳过程，以上情况，其数据生成过程都可写成：

$$(1-L)Y_t = \alpha + \varepsilon_t$$

其中， L 为滞后算子， α 分别为 0 、 μ 和 Y ， $\{\varepsilon_t\}$ 为稳定过程。它们的特征方程 $1-\omega=0$ 有一个单位根 $\omega=1$ ，因此可将以上三种随机过程称为单位根过程。

③ 单位根检验 “迪基-福勒检验”

对每个时间序列，可以采用迪基-福勒检验法(记为 DF)和增广的迪基-福勒检验法(检验其是否服从单位根过程。介绍这种检验方法。

假设我们要检验时间序列 Y_t 是否为 $Y_t = \rho Y_{t-1} + \varepsilon_t$ 形式的单位根过程，则可以对模型 $Y_t = \alpha + \rho Y_{t-1} + \varepsilon_t$ 进行估计，然后检验 $\rho=1$ 或 $\alpha=1$ 和 $\rho=1$ 。

但是，当 $\rho=1$ 为真时，最小二乘估计 $\hat{\rho}$ ， t_{ρ} ，统计量和 F_t 统计量有非标准的极限分布。所以在检验时我们能直接采用传统的 t 检验和 F 检验方法，这种检验方法称为迪基-福勒检验，即 DF 检验。 t_{ρ} 和 F_t 的分布决定于估计的回归模型中是否包括常数项或时间趋势以及真实过程是否具有非零漂移特征。

4.公路投资额与 GDP 时间序列的单位根检验。

为了研究的方便，并考虑到对各时间序列的数据取对数之后不会改变各时间序列之间的协整关系，且所得到的数据容易得到平稳的序列，在这里对时间序列进行对数处理，在以下的分析中，在字母前加上 Ln 的变量表示是由相应数据经过取对数转换而得到的新变量。本文采用的是增强的迪基-福勒检验(ADF 检验)来对变量进行单位检验。检验 $\{Y_t\}$ 是否存在单位根的回归方程如下：

$$Y_t = c + \rho Y_{t-1} + \alpha t + \sum_{i=1}^p Y_{t-i} \Delta Y_{t-i} + \varepsilon_t$$

其中， c 表示常数项， t 表示时间趋势， ε_t 为残差项。原假设为： $H_0: \rho=1$ ，备择假设为 $H_1: \rho<1$ 。根据(3—4)式回归方程中系数 ρ 的 t 检验值进行判断，若 t 检验值小于 ADF 分布的临界值，则拒绝原假设，接受备择假设，说明序列 $\{Y_t\}$ 是平稳过程；如果 t 检验值大于临界值，则接受原假设，说明序列(r)存在单位根。方程 (3-4)中加入 p 个滞后项是为了使残差项为白噪声。

表 11 各时间序列 ADF 检验结果

变量	情形 (c, t, n)	ADF 检验值	ADF 分布临界值
$LnGGU$	$(c, t, 3)$	-2.186	-3.576
$\Delta LnGGU$	$(c, 0, 3)$	-3.933	-3.241
$LnGDP$	$(c, t, 1)$	-2.096	-3.2253
$\Delta LnGDP$	$(c, 0, 3)$	-2.608	-2.652

通过上表的检验结果，发现用时间序列检测的效果吻合度较好。

5.公路投资额与 GDP 的协整分析

1) 协整过程

协整的经济意义在于：两个变量，虽然它们具有各自的长期波动规律，但是如果它们是协整的则它们之间存在着一个长期稳定的比例关系；反之，如果两个变量具有各自的长期波动规律，但是它们不是协整的，则它们之间就不存在着一个长期稳定的比例关系。

模型一：Engle—Granger 两步法

若 $I(1)$ 的随机变量 Y_t 和 Z_t ，是协整的，它们之间的协整关系可表示为

$$Y_t = \alpha + \beta Z_t + v_t$$

由此得到的最小二乘估计 $\hat{\alpha}$ 和 $\hat{\beta}$ 是一致估计量，估计残差 $\hat{v}_t = Y_t - \hat{\alpha} - \hat{\beta} Z_t$ 互构成 $I(0)$ 过程。但是，若 Y_t 和 Z_t 非协整，(3—5)给出的是 Y_t 和 Z_t 之间的伪回归关系，由此得到的估计残差 \hat{v}_t 仍为 $I(1)$ 过程。所以 \hat{v}_t 是否为单位根过程反映了 Y_t 和 Z_t 之间是否协整。于是，对 Y_t 和 Z_t 协整性的检验便可以通过对 \hat{v}_t 的单位根检验进行。首先做形如(3—5)式

的回归，称为协整回归，然后对估计残差 \hat{v}_t 做单位根检验，因为整个过程经过了两个步骤，所以此方法称为 Engle—Granger 两步法。

对 \hat{v}_t 的单位根检验可用 DF 法或 ADF 法进行，但是由于 \hat{v}_t 是最小二乘估计的残差，对它做单位根检验与一般情况有所不同。对 \hat{v}_t 的单位根检验，总是在模型

$$\hat{v}_t = \beta_1 \Delta \hat{v}_{t-1} + \beta_2 \Delta \hat{v}_{t-2} + \cdots + \beta_p \Delta \hat{v}_{t-p} + \rho \hat{v}_{t-1} + \varepsilon_t$$

中进行，但临界值表必须根据所检验的协整关系，即协整回归模型的形式进行选择。具体地，协整关系的检验可归结为以下三种情况：

情况一： Y_t 和 Z_t ，都不含有常数项时，检验协整关系 $Y_t = Y \cdot Z_t + v_t$

其中，不包括常数项和时间趋势。

情况二： Y_t 和 Z_t ，都不含有常数项时，检验协整关系 $Y_t = \alpha + Y \cdot Z_t + v_t$

情况三： Y_t 和 Z_t 中至少有一个常数项或时间趋势时，检验协整关系 $Y_t = \alpha + Y \cdot Z_t + v_t$

至于，如何确定所检验的协整关系是三种情况中的哪一种，可以根据经济变量的实际经济意义和统计性质判断。

2) 公路投资与国民经济发展指标的协整分析

本文对公路投资额序列与 GDP 序列之间的协整关系的检验采用的是 Johansen 协整检验，结果见表 12。

表 12 $\ln GGU$ 与 $\ln GDP$ 的协整关系检验

Eigen 值	似然比	5%临界值	1%临界值	无协整假设
0.547902	16.81359	15.41	20.04	None
0.087046	1.730316	3.76	6.65	At most 1

检验结果表明，公路投资额与 GDP 存在着协整关系(在 5%的水平下，检验值大于临界值，即 16.81359>15.41)，也就是说公路投资与国民经济发展之间形成了长期均衡关系，它们之间有着某种必然的联系。

6.模型的检验与修正

1) 公路投资额与 GDP 的误差修正模型

通过检验发现了公路投资额与 GDP 之间的协整关系以后，就可以求出协整向量，以这种关系一期滞后构成误差修正项；然后将误差修正项看作一个解释变量，连同其他反映短期波动的解释变量一起，建立短期模型，这种误差修正模型既区分了变量之间的长期均衡关系和短期动态关系，又避免了谬误回归，具有较强的经济意义，是含有单位根的协整过程的有效表示形式。通过给出的 $\Delta \ln GGU$ 的实际值、拟合值与残差图，可以观察结果的理想程度。模型显示(对数的)我国实际完成的公路投资额 GGU 和我国的 GDP 序列存在着稳定的紧密的关系，这也就是说从增长率角度来看，GDP 的增长对我国公路投资额具有很强的促进作用。

2) 公路投资与国民经济因果关系分析

通过检验我们知道公路投资额与 GDP 之间存在协整关系即它们有着一种长期均衡关系，那么，到底是国民经济发展决定公路投资规模还是公路投资规模影响国民经济的发展？抑或二者是一种互为因果的关系？

① Granger 因果关系

所谓因果关系是指变量之间的依赖性，作为结果的变量是由作为原因的变量所决定的，原因变量的变化引起结果变量的变化。我们已经知道因果关系不同于相关关系，从一个回归关系式中，我们无法确定变量之间是否具有因果关系，虽然有时我们说回归方程中解释变量是被解释变量的因，但是，这一因果关系实际上是先验设定的，或者是在回归之前就已确定。

② Granger 因果检验

检验 X 是否为引起 Y 变化的原因的 Granger 因果检验过程如下：

首先检验原假设 $H_0: X$ 不是引起 Y 变化的 Granger 原因。估计下列两个回归模型：

$$\text{无约束回归模型}(u): Y_t = \sum_{i=1}^p \alpha_i Y_{t-i} + \sum_{i=1}^q \beta_i X_{t-i} + \varepsilon_t$$

$$\text{有约束回归模型}(r): Y_t = \sum_{i=1}^p \alpha_i Y_{t-i} + \varepsilon_t$$

然后，用它们的回归残差平方和 RSS_u 和 RSS_r 构造 F 统计量

$$F = \frac{RSS_r - RSS_u}{RSS_u / (n - p - q - 1)} \sim F(q, n - p - q - 1)$$

检验 $\beta_1, \beta_2, \dots, \beta_q$ 是否显著不为零，其中 n 为样本容量。如果显著不为零，则拒绝原

假设 $H_0: X$ 不是引起 Y 变化的 Granger 原因。

其次，将 Y, X 的位置交换，按同样的方法检验原假设 $H_0: Y$ 不是引起 X 的 Granger 原因。

最后，要得到 X 是 Y 的 Granger 原因的结论，必须同时拒绝原假设“ X 不是引起 Y 变化的 Granger 原因”和接受原假设“ Y 不是引起 X 的 Granger 原因”。Granger 因果关系检验可以借助于软件 Eviews3.1 轻松自动完成。

③ 公路投资额与 GDP 的 Granger 因果关系检验

运用软件对公路投资额与 GDP 进行因果检验，检验结果见表 3-4。由表 3—4 可知，检验统计量的值分别为 0.77560 和 0.02147，表明经济增长(LnGDP)没有引起公路投资额(LnGGU)波动的零假设在 5% 的显著性水平下被拒绝，因为其 F 统计量的概率很小，为 0.02147，这说明经济增长具有推断和预测公路投资额的能力，公路投资额增长趋势是由经济景气变化导致的。这符合“投资规模受国力制约规律”的基本原理。改革开放以来，我国经济发展取得了很大的成就，为了使我国经济保持持续、稳定的发展，为了消除公路运输业的“瓶颈”制约，我国加大了公路投资的力度。

表 13 公路投资额与 GDP 的 Granger 因果检验

原假设	自由度	F 统计量	P 值
LnGGU 不是 LnGDP 的 Granger 原因	20	0.37053	0.77560
LnGDP 不是 LnGGU 的 Granger 原因		4.56552	0.02147

通过上面投资额与 GDP 的因果检验结果，就可以认为本文中求解结果的准确性，

也对公路建筑业对国内生产总值的影响进行了模型修正。

六 模型的评价与改进

在本文中，我们没有使用单一的模型对问题进行分析求解，而是综合运用了多种数学模型，并且增加了一定的检验环节，这样就增加了我们所求得数据的合理性，同时也增加了论文的实际参考价值。在运用归一化法综合预测公路投资强度情况的变化趋势时，我们很巧妙的将多个指标转化为单一指标，这样就很好的解决了多指标难以解决的难题。在分析 11 个城市公路效益贡献值时，通过改进后的加权主成分分析法和熵权法的有机结合，解决了公路效益随机性和波动性较大从而难以准确预测的难题。而在解决问题二时，合理的变量相关性分析将 GDP 与公路运输之间的联系解释的很完善，最后的结果检测更好的反应了模型的可信性。

在改进方面我们认为：在问题一中的熵权模型确定的显著性指标可以改进为用模糊神经网络进行判断分析，可以更好的反应指标权重，以及不同指标的关系。在分析变化趋势时，我们用以预测的一次累加拟合模型虽然在预测点内误差比较小，但在数据点外，进行预测时，误差将是不可知的，可以改进为使用时间序列法、灰色系统理论、BP 神经网络进行预测，或者将三者有机的结合在一起，通过优化组合形成最优化权值组合法，这样既可以达到更高的精度要求，又可以避免使用大量参数和数据，达到最小误差预测的目的，在问题一最后，可以各城市中每个县投资的差异性，为了定量分析近年来公路投资在各县地区之间的差异性，现分别采用数理统计的方差分析与非参数检验中的秩和检验，对公路投资在各个城市每个县区之间的差异性进行统计检验。通过设 X 为一总体，将一容量为 N 的样本观察值按自小到大的次序编号排列成 $x_{(1)} < x_{(2)} < \cdots < x_{(n)}$ ，称 i 为 $x_{(i)}$

的秩， i 为 $1, 2, 3, \cdots, n$ ，用秩号代替原始数据后，所得某些秩号之和，称为秩和。用其进行秩和检验即可。

参考文献

- [1] 杨虎, 钟震, 程代杰, 基于不完全数据的异常挖掘算法研究, 计算机研究与发展, 第 41 卷第 9 期: 第 1532 页到 1539 页, 2004 年。
- [2] 聂庆华, 杨利华, 缺失或不完备数据的填补方法分析, 科技创新与应用, 2012 年 5 月 (上): 第 46 页, 2012 年。
- [3] 陈胜可, SPSS 统计分析从入门到精通, 北京: 清华大学出版社, 2010 年。
- [4] 张迎冬, 何桢, 改进的主成分分析法在多响应优化中的应用, 组合机床与自动化加工技术, 2012 年第 11 期: 第 97 页至 100 页, 2012 年。
- [5] 胡永宏, 贺思辉, 综合评价方法, 北京: 科学出版社, 2000 年。
- [6] 庄宇等, 应用数据可信度分析重构科技创新指标体系, 情报杂志, 2006 年第 1 期: 第 12 页至 16 页, 2006 年。
- [7] 吕跃进等, 基于判断可信度的层次分析排序方法, 控制与决策, 27 卷 5 期: 第 787 页至 796 页, 2012 年。
- [8] HAN Min, LIN Yun, SUN Yannan, QI Donghai, Study of construction bidding system based on neural networks, JOURNAL OF SYSTEMS ENGINEERING, Vol. 18 No. 4: Page.348-354, Aug. 2003。
- [9] 王恺, 发挥公路建设作用推动和谐社会建设进程, 学术纵横, 2011 年第 07 期: 第 105 页, 2011 年。
- [10] 戴添发, 赵鸿, 随机变异-优化选择神经网络在铁货运量预测方面的应用, 新疆师范大学学报, 第 30 卷第 2 期: 第 70 页至 74 页, 2011 年。
- [11] 杨璐, 高自友, 用神经网络进行变量选择, 北方交通大学学报, 第 23 卷第 3 期: 第 53 至 56 页, 1999 年。
- [12] 梅长林, 周家良, 实用统计方法, 北京: 科学出版社, 2002 年。
- [13] 胡永宏, 贺思辉, 综合评价方法, 北京: 科学出版社, 2000 年。
- [14] 姜启源, 谢金星, 叶俊, 数学模型, 北京: 高等教育出版社, 2003 年。
- [15] 曹戈, MATLAB 教程及实训, 北京: 机械工业出版社, 2008 年。

附件

附录一：多元回归处理数据

城市	年份	全社会旅客 运输总量（万 人）	全社会货运 总量（万吨）	道路运输从 业人员（个）	汽车维修服 务网点数 （个）	地区生产 总值（万 元）
1	2007	0.4555	0.4155	0.4892	0.4170	1091714
	2008	0.5415	0.5833	0.5157	0.4956	1781800
	2009	0.5955	0.5815	0.5836	0.5891	2424550
	2010	0.6827	0.6973	0.6616	0.6976	2990722
	2011	0.7248	0.7974	0.7500	0.8755	3824928
2	2007	0.2374	0.1170	0.1548	0.1714	3416200
	2008	0.2488	0.1148	0.1663	0.1887	4390600
	2009	0.1638	0.2282	0.2274	0.2116	5098600
	2010	0.1743	0.2682	0.2157	0.2156	6033300
	2011	0.1830	0.2719	0.2415	0.2232	7184500
3	2007	0.1999	0.3268	0.1132	0.0000	1277
	2008	0.1999	0.4575	0.1641	0.0000	1760
	2009	0.2000	0.6723	0.1959	0.0000	2169
	2010	0.2996	0.7367	0.2362	0.0000	2461
	2011	0.2199	0.7497	0.2958	0.0000	3005
4	2007	2.1326	2.6769	2.2278	2.7540	12499357
	2008	2.4490	2.2416	2.2314	2.2759	15814579
	2009	2.2787	2.3783	2.3800	2.5512	18257721
	2010	2.8990	2.5219	2.7715	2.8716	21649213
	2011	3.2472	2.7914	3.3894	3.2264	26943462
5	2007	1.1213	0.9636	1.9124	1.1553	24125737
	2008	1.3955	1.4518	1.1355	1.3833	34113016
	2009	1.6542	1.6632	1.4618	1.5787	44128540
	2010	1.8278	1.9152	1.6585	1.8155	54063652
	2011	2.2422	2.2283	1.6856	2.6718	67878486
6	2007	0.8393	0.2840	0.9966	0.1400	4622613
	2008	0.5742	0.1687	0.1747	0.1714	5492021
	2009	0.2319	0.5726	0.0000	0.0000	6427645
	2010	0.2366	0.8670	0.4348	0.3486	7559056
	2011	0.1265	1.6232	1.3581	1.3400	9001039
7	2007	2.3951	2.2694	2.1320	2.1946	5010684
	2008	2.5754	2.5169	2.4273	2.3919	6081389
	2009	2.7257	2.8150	2.4862	2.5951	7901317
	2010	3.5554	3.3976	2.9182	2.6914	9471923
	2011	3.2483	3.3589	3.1553	3.1275	11609108
8	2007	1.3394	1.6757	1.1947	0.9719	5624332
	2008	1.9452	1.6383	1.4143	1.1493	7919598

	2009	1.1477	1.5437	1.7641	1.7856	9315589
	2010	1.2783	1.7534	2.6597	2.2245	10403491
	2011	1.2896	1.5612	1.5615	1.8689	13728628
9	2007	2.1979	1.9736	1.8642	1.6679	3644752
	2008	2.2142	1.9676	2.1927	1.7522	4617138
	2009	2.2815	1.8456	2.1856	2.2256	5276411
	2010	2.1370	2.6949	2.4227	2.5334	6019840
	2011	2.1694	2.1437	2.5816	2.8222	7328326
10	2007	0.5759	0.2278	0.7214	0.6185	1370000
	2008	0.6329	0.3352	0.7221	0.6477	958807
	2009	0.7455	0.7638	0.7449	0.8824	2388648
	2010	1.1449	1.7177	0.8260	0.9267	3120767
	2011	1.3121	0.9562	0.9856	0.9990	3601376
11	2007	1.2229	1.1283	1.1423	0.6643	1431895
	2008	1.1733	1.1463	1.1399	0.7227	1788287
	2009	1.2265	1.2836	1.2257	0.9975	2151749
	2010	1.2648	1.3712	1.2549	1.3299	2317596
	2011	1.3952	1.1483	1.2372	1.2886	3239787

附录 2：解释的总方差结果

1. 城市 1

成份	初始特征值			提取平方和载入		
	合计	方差的 %	累积 %	合计	方差的 %	累积 %
1	3.966	99.143	99.143	3.966	99.143	99.143
2	0.033	0.821	99.963			
3	0.001	0.028	99.991			
4	0	0.009	100.000			

2. 城市 2

成份	初始特征值			提取平方和载入		
	合计	方差的 %	累积 %	合计	方差的 %	累积 %
1	3.777	94.425	94.425	3.777	94.425	94.425
2	.152	3.808	98.232			
3	.052	1.311	99.544			
4	.018	.456	100.000			

3. 城市 3

成份	初始特征值			提取平方和载入		
	合计	方差的 %	累积 %	合计	方差的 %	累积 %
1	2.306	76.867	76.867	2.306	76.867	76.867
2	.621	20.703	97.570			
3	.073	2.430	100.000			

4. 城市 4

成份	初始特征值	提取平方和载入
----	-------	---------

	合计	方差的 %	累积 %	合计	方差的 %	累积 %
1	3.314	82.858	82.858	3.314	82.858	82.858
2	.645	16.131	98.990			
3	.026	.642	99.632			
4	.015	.368	100.000			

5. 城市 5

成份	初始特征值			提取平方和载入		
	合计	方差的 %	累积 %	合计	方差的 %	累积 %
1	2.915	72.872	72.872	2.915	72.872	72.872
2	1.029	25.723	98.595	1.029	25.723	98.595
3	.055	1.363	99.958			
4	.002	.042	100.000			

6. 城市 6

成份	初始特征值			提取平方和载入		
	合计	方差的 %	累积 %	合计	方差的 %	累积 %
1	2.860	71.509	71.509	2.860	71.509	71.509
2	1.075	26.870	98.379	1.075	26.870	98.379
3	.065	1.619	99.998			
4	.000	.002	100.000			

7. 城市 7

成份	初始特征值			提取平方和载入		
	合计	方差的 %	累积 %	合计	方差的 %	累积 %
1	3.715	92.883	92.883	3.715	92.883	92.883
2	.265	6.614	99.497			
3	.019	.469	99.965			
4	.001	.035	100.000			

8. 城市 8

成份	初始特征值			提取平方和载入		
	合计	方差的 %	累积 %	合计	方差的 %	累积 %
1	2.259	56.483	56.483	2.259	56.483	56.483
2	1.243	31.064	87.546	1.243	31.064	87.546
3	.473	11.817	99.363			
4	.025	.637	100.000			

9. 城市 9

成份	初始特征值			提取平方和载入		
	合计	方差的 %	累积 %	合计	方差的 %	累积 %
1	2.823	70.578	70.578	2.823	70.578	70.578
2	.940	23.489	94.068			
3	.158	3.943	98.011			
4	.080	1.989	100.000			

10. 城市 10

成份	初始特征值			提取平方和载入		
	合计	方差的 %	累积 %	合计	方差的 %	累积 %

	合计	方差的 %	累积 %	合计	方差的 %	累积 %
1	3.367	84.166	84.166	3.367	84.166	84.166
2	.506	12.654	96.820			
3	.126	3.162	99.982			
4	.001	.018	100.000			

11. 城市 11

成份	初始特征值			提取平方和载入		
	合计	方差的 %	累积 %	合计	方差的 %	累积 %
1	2.906	72.650	72.650	2.906	72.650	72.650
2	1.050	26.261	98.911	1.050	26.261	98.911
3	.031	.780	99.691			
4	.012	.309	100.000			

附录 3 城市投资强度聚类数据

年份	城市 1	城市 2	城市 3	城市 4	城市 5	城市 6	城市 7	城市 8	城市 9	城市 10	城市 11
2007	56.8159	57.9994	54.0394	86.1642	61.9903	50.5945	51.0023	59.0052	70.9035	51.4852	60.0002
2008	56.3089	59.5350	53.6209	82.9016	60.9596	50.2970	50.4498	58.7882	76.3221	51.8967	58.9203
2009	56.1921	59.3480	53.4332	80.9629	61.6394	50.3823	50.3924	59.1935	78.6521	52.2109	57.5931
2010	55.6874	62.1237	52.9663	78.3698	61.9032	50.0744	49.8590	59.6938	80.3732	52.3810	56.5682
2011	55.5588	61.7882	52.8066	78.7751	62.7479	50.0721	49.6711	59.8689	79.7887	52.5815	56.3412

附录4 预测方程程序

1.

```
x=[0:1:4];
y=[1.654277124 1.728771002 1.776948578 1.839460144 1.859779993];
y1=1.631022+0.175439*x;
plot(x,y,'--',x,y1,'-*');
title('预测曲线');
xlabel('时间');
ylabel('预测值');
grid on;
```

2.

```
B=[-287.0063459 1
-515.6540044 1
-773.3170607 1
-1055.69827 1];
B1=[-287.0063459 -515.6540044 -773.3170607 -1055.69827
1 1 1 1];
Y=[54.5320405
59.79178874
69.03973939
72.15086514];
q=inv(B1*B)*B1*Y
```

3. 曲线方程

```
x=[0:1:4];
y=[44.48556623 54.5320405 59.79178874 69.03973939 72.15086514];
y1=2037.407*(exp(0.0241*x)-exp(0.0241*(x-1)));
plot(x,y,'--',x,y1,'-*');
title('预测曲线图');
xlabel('时间');
ylabel('预测值');
grid on;
```

4.

```
x=[0:1:4];
y=[44.48556623 54.5320405 59.79178874 69.03973939 72.15086514];
cftool
```

5.

```
x=[0:1:4];
y=[44.48556623 54.5320405 59.79178874 69.03973939 72.15086514];
y3=2.064e+005*exp(-0.07529*x) -2.063e+005*exp(-0.07536 *x);
plot(x,y,'--',x,y3,'-o');
grid on;
```

5.

```
x=[0:1:4];
y=[44.48556623 54.5320405 59.79178874 69.03973939 72.15086514];
```

```
y2=42.29 + 2.449 *cos(x*0.2749) + 32.61*sin(x*0.2749);
plot(x,y,'--',x,y2,'-o');
grid on;
6.
x=[0:1:9];
y7=42.29 + 2.449 *cos(x*0.2749) + 32.61*sin(x*0.2749);
y7
1.
x1=[0.455544275 0.541542453 0.595533854 0.682681875 0.724812977 ];
x2=[0.415468779 0.583256115 0.581482314 0.697288343 0.797434953 ];
x3=[0.489219269 0.515736684 0.583646355 0.661576338 0.749965397];
x4=[0.416962175 0.495571922 0.589116229 0.697595173 0.875468267];
y=[44.48557 54.53204 59.79179 69.03974 72.15087];
X=[ones(length(y),1),x1',x2',x3',x4'];
Y=y';
[b,bint,r,rint,stats]=regress(Y,X);
b,bint,r,rint,stats
```