

## 摘要

财政收入，是政府为履行其职能，实施公共政策和提供公共物品与服务需要而筹集的一切资金的总和，是衡量一国政府财力的重要指标，为国家职能的履行提供保障，对于国家的发展至关重要。准确预测财政收入，有益于政府财政预算的编制，合理进行赤字安排，最大可能地实现长期的财政收支平衡。

本文对辽宁省 1984-2016 年财政预算公开数据进行抽样调查，并将用样本估计的总体指标与实际财政收入情况进行对比，发现由于 1994 年以后地方财政收入与以前实行的总额分成财政体制下的地方财政收入在内容和范围上有一定差别，2002 年所得税收入分配方式又有重大改革，2008 年金融危机爆发之后，我国实行了较大强度的积极财政政策，导致历史数据纵向不完全可比，财政收入预算不够准确。为了解决这一问题，本文采用实际财政收入数据，引入人工智能优化算法，建立有较强时效性与解释力的财政收入预测模型。

首先，对辽宁省 1985-2016 年财政收入数据进行时间序列分析，发现辽宁省计算得财政收入在 1985-2016 年平均数高于其预算水平与全国平均水平，总的来说，辽宁省在 1985-2016 年间财政收入呈 8.8% 增长的趋势。然后对 1985-2016 年财政收入数据进行移动平均后，将财政收入公开的四项组成运用多元线性回归方法，拟合财政收入预测曲线。结果表明：国有资本经营收入、国有企业计划亏补、预算各项收入与财政收入回归不显著相关。

为了探索影响辽宁省财政收入的影响因素，接着，采用因子分析方法，对财政收入各项细分数据进行主成分分析后得到，公共因子 1 主要构成有个人所得税等 12 项收入，方差贡献率为 30.947%；公共因子 2 主要构成有营业税等 8 项收入，方差贡献率为 30.924%；公共因子 3 的主要构成有消费税、罚没收入，方差贡献率为 17.892%；构成公共因子 4 的主要有其他税收等 3 项收入，方差贡献率为 15.061%。并根据相关得分矩阵对数据进行降维。

以降维后的四个公共因子作为输入，各年实际财政收入作为输出。训练得到由 1 个输入层，2 个隐藏层与 1 个输出层，其中隐藏层 1、2 神经元个数分别为 2、1，拟合优度为 99.46%，训练集拟合优度达到 100%，具有非常好的预测能力的人工神经网络。

最后，讨论财政收入预测的影响因素，提出相关政策建议，以助力辽宁经济平稳健康运行。

## 关键词

财政收入；人工神经网络；时间序列分析；因子分析

## Abstract

Fiscal revenue is the sum of all funds raised by the government to fulfill its functions, implement public policies, and provide public goods and services. It is a significant indicator to measure the financial power of a country's government and provides guarantees for the performance of state functions, which is essential for the development of the country.

This article conducted a sample survey concerning the public data of 1984-2016 fiscal revenue budget in Liaoning Province, comparing the population indicators estimated by samples with the actual fiscal income. It was found that since 1994, after the local fiscal revenue which had been previously divided was reformed, there exists a certain difference in the content and scope of local fiscal revenue; in 2002, there was a major reform in the distribution of income tax revenues; after the outbreak of the financial crisis in 2008, China implemented a relatively strong proactive fiscal policy. With these concerns, the historical data are not completely comparable vertically and the fiscal revenue budget is not accurate enough. To solve this problem, using the actual fiscal revenue data and artificial intelligence optimization algorithm, a fiscal revenue forecasting model with strong timeliness and explanatory power was established.

First, the time series analysis of Liaoning Province's 1985-2016 fiscal revenue data was conducted. It was found that the calculated average fiscal revenue of Liaoning Province in 1985-2016 is higher than its budget level and the national average level. Overall, Liaoning's fiscal revenue grew by 8.8% between 1985 and 2016. After the moving average smoothing of the 1985-2016 fiscal revenue data, four components of fiscal revenue are applied to fit the multivariate linear regression model of fiscal revenue. The results show that there is no significant correlation between state-owned capital operating income, state-owned enterprises' plans to make up for losses, and budget revenues and fiscal revenue.

In order to explore the influencing factors affecting the fiscal revenue of Liaoning Province, then, using the factor analysis method, the principal component analysis of the subdivided data of the fiscal revenue was obtained: Component 1 mainly consists of 12 income such as personal income tax, and the variance contribution rate is 30.947%; Component 2 mainly consists of 8 items such as business tax, and the variance contribution rate is 30.924%; Component 3 consists of consumption tax and confiscation income, and variance contribution rate is 17.892%; Component 4 mainly consists of 3 items such as other taxes, and the variance contribution rate is 15.061%. And the data size was reduced according to the relevant score matrix.

Taking the four public factors after dimension reduction as input and the actual fiscal revenue for each year as output, an artificial neural network with 1 input layer, 2 hidden layers, and 1 output layer was established, among which, the number of neurons in the hidden layer 1 and 2 is 2 respectively, 1. The goodness of fit is 99.46% and the goodness of fit in the training set reaches 100%, which means the artificial neural

network has a sound predictive ability.

Finally, the influencing factors of fiscal revenue forecast was discussed, and relevant policy recommendations was put forward to help the Liaoning economy to run steadily and healthily.

## **KEY WORDS**

Fiscal revenue; Artificial neural network; Time series analysis; Factor analysis

# 目录

摘要.....	I
Abstract.....	II
1 绪论.....	1
1.1 研究背景及意义.....	1
1.2 财政收入预测研究现状.....	2
1.3 研究思路及行文框架.....	3
2 辽宁省财政收入预算公开数据抽样调查.....	4
2.1 调查背景.....	4
2.2 抽样方法.....	4
2.3 抽样结果与分析.....	4
3 财政收入的时间序列分析.....	6
3.1 辽宁省财政收入的水平分析.....	6
3.1.1 发展水平与平均发展水平.....	6
3.1.2 增长量与平均增长量.....	6
3.2 辽宁省财政收入的速度分析.....	8
3.2.1 发展速度与增长速度.....	8
3.2.2 平均发展速度与平均增长速度.....	9
3.3 时间序列的构成分析和预测方法.....	10
3.3.1 辽宁省财政收入的构成与分解.....	10
3.3.2 辽宁省财政收入的预测方法.....	10
4 财政收入的因子分析模型.....	14
4.1 因子分析简介.....	14
4.2 因子分析模型建立.....	15
4.3 模型求解与分析.....	16
5 基于神经网络的财政收入预测模型.....	20
5.1 BP 神经网络简介.....	20
5.2 BP 神经网络模型建立.....	21
5.3 模型求解与分析.....	22
总结与展望.....	25
参考文献.....	26
附录.....	27
1 BP 神经网络模型源码.....	27
2 BP 神经网络输出结果.....	29
致谢.....	32

# 1 绪论

## 1.1 研究背景及意义

财政收入，是政府为履行其职能，实施公共政策和提供公共物品与服务需要而筹集的一切资金的总和，其表现为政府部门在一定时期内（一般为一个财政年度）所获得的货币收入。近日，财政部召开新闻发布会介绍 2017 年财政收支情况，去年财政运行总体平稳，财政收入增速加快。据统计，2017 年全国一般公共预算收入 172567 亿元，同比增长 7.4%，其中税收收入 144360 亿元，同比增长 10.7%；全国一般公共预算支出首次突破 20 万亿元。

财政收入是衡量一国政府财力的重要指标，为国家职能的履行提供保障，对于国家的发展至关重要。其一，财政收入是财政支出的前提。财政分配是收入与支出的统一过程，财政支出是财政收入的目的，财政收入则是财政支出的保证，一般的，收入的数量决定着财政支出的规模，只有在发展生产的基础上积极筹集资金，才能为财政支出创造前提。其二，财政收入是实现国家职能的重要保证。国家为了实现其职能，必须掌握一定数量的社会产品，财政收入正是家资金的重要手段。对实现国家职能有重要意义。其三，财政收入是正确处理各方面物质利益关系的重要方式。财政收入的取得不仅仅是个聚集资金的问题，在具体操作过程中，取得多少、采取何种方式，关系到党的方针政策的贯彻落实，涉及到各方面的物质利益关系的处理。只有在组织财政收入的过程中正确处理各种物质利益关系，才能达到充分调动各方面的积极性，达到优化资源配置，协调分配关系的目的。

我国财政管理体系不断发展完善，目前采用的是分税制管理。要了解国家财政收入则需先了解地方财政收入。地方财政收入包括地方财政预算收入和预算外收入。具体包括地方所属企业收入和各项税收收入；各项税收收入包括营业税、地方企业所得税、个人所得税、城镇土地使用税、固定资产投资方向调节税、土地增值税、城镇维护建设税、房产税、车船使用税、印花税、农牧业税、农业特产税、耕地占用税、契税，增值税、证券交易税（印花税）的 25%部分和海洋石油资源税以外的其他资源税；中央财政的调剂收入，补贴拨款收入及其它收入；地方财政预算外收入的内容主要有各项税收附加，城市公用事业收入，文化、体育、卫生及农、林、牧、水等事业单位的事业收入，市场管理收入及物资变价收入等。

财政收入的目的是为了能够更好地分配与支出，准确预测财政收入，有益于政府财政预算的编制，合理进行赤字安排，最大可能地实现长期的财政收支平衡。对财政收入进行定量分析并对其做出较为准确的预测则可以为相关部门或者企业定制发展规划、实施相关措施提供可靠的理论预测参考。

## 1.2 财政收入预测研究现状

随经济的发展,统计制度的不断完善。统计指标的数目日趋庞大,形成了巨大的信息资源,使得利用时间序列分析经济规律、预测经济走势成为可能。由于时间序列预测法所需要的只是序列本身的历史数据,因此,这一类方法应用较为广泛,具体方法有时间序列分析法、移动平均法、指数平滑法、趋势外推法、自适应滤波法、平稳时间序列预测法、灰色预测法、状态空间模型和卡尔曼滤波等。

财政系统既具有非线性、关联性和系统性,又具有一定的随机性,甚至某种混沌性。目前在经济学实证研究中主要以计量方法为主,大部分采用线性模型进行研究,但线性模型在发挥巨大作用的同时暴露了许多不可避免的固有缺陷,即很难把控宏观经济系统中的非线性现象,导致经济预测上的误差较大,使得财政收入预测不准,在此基础上指定的财政政策可能失去效力。因此,经济学家做了大量改进工作,如建立分段线性模型,参数时变性模型等,但结果往往并不理想,这迫使学者寻求一些非线性工具进行宏观经济建模,神经网络模型等在这一情况下被一些学者引入经济预测领域。

冯润(2008)[1]观察到之前的财政收入预测多运用一种方法直接预测总收入,而未考虑构成总收入的各细分收入在稳定程度及适用预测方法上的差异,从而导致准确性不够。他提出一套全新分解预测法,将总收入分成多分项收入,对于波动较大的采用一系列方法进行调整后与稳定变化的一通进行预测,从而达到更好的预测准确度。屈丽萍等(2008)[2]通过建立回归模型,并采用  $H_\infty$  滤波算法对国家财政收入中各分项进行估计预测。郑鹏辉等(2008)[3]将 ARIMA 模型应用于我国财政收入分析与预测,结果表明 ARIMA 是一种短期预测精度较高的预测模型,但随时间延长其预测误差会逐渐加大。舒成(2010)[4]利用灰色模型,通过原始数据的整理寻求其变化规律的特点,对小样本、时间序列的县级财政收支预测进行了分析。

范敏等(2008)[5]根据地方财政收入预测收到许多因素影响和经济系统的非线性本质的特点提出了一种基于灰色预测模型和神经网络组合预测方法。首先通过灰色关联分析确定财政收入的主要指标,然后用灰色预测模型对各指标进行预测,最后将个指标的预测值作为输入,相应的地方财政收入实际值作为输出,训练并建立神经网络模型。江澜(2008)[6]通过对历史数据的数学处理,采用人工神经网络方法建立财政收入预测模型。运用前向神经网络 BP(Back Propagation)算法,利用福建省 1990-2004 年财政收入数据建立神经网络预测,结果表明神经网络预测模型精度高,且具有较大的实际应用价值。张育军(2011)[7]运用 GRNN 神经网络和 BP 神经网络对财政收入预测的结果进行比较,结果表面两者预测能力并没有太大区别。赵海华(2016)[8]针对财政收入及其影响因素原始数据呈随机性、非线性变化,两者之间具有非线性相关关系的特点,在分析 RBF 神经网络和无偏 GM(1,1)模型两者建模优点的基础上,结合回归分析思想,提出了基于灰色 RBF 神经网络的多因素财政收入预测模型。

总的来说,我国财政收入预测的研究成果相当可观,但研究体系尚不成熟,具有较大的发展空间。神经网络自从开创以来就备受关注,它是人工智能领域的一项重要研究,因其具有较好的非线性处理特点,能从多维度的角度出发进行预测,弥补了时间序列等预测方法预测周期短、单线预测的缺陷,故广泛被应用于

非线性预测。且多数预测结果显示，神经网络的预测精确度较其他预测方法要高的多。故本文选取典例，通过构建人工神经网络建立财政收入预测模型。

### 1.3 研究思路及行文框架

为了准确预测财政收入，并提出相关的政策建议，促进财政收支平衡。首先由辽宁省统计局获得历年财政预算，并使用随机抽样的方法，抽取若干年了解财政收入预算大致情况。并对辽宁省历年财政收入进行时间序列分析，通过各指标了解其发展水平、增长水平、速度水平。然后通过移动平均法和多元线性回归拟合，验证现在公开的财政收入各组成成分与财政收入的关系。

接着，对财政收入各细分部分进行探索性因子分析，从而发现财政收入的各个影响因素。同时对数据实现降维。以提高神经网络的训练速度，同时提高模型拟合优度。

本文第一章为绪论，从财政收入和神经网络的研究意义及现状出发说明本文研究的必要性，并给出本文研究思路及框架。第二章主要为辽宁省财政收入情况抽样调查，并给出相关结果分析，论证本研究的必要性。第三章为财政收入的时间序列分析，对比其各指标，说明辽宁省财政收入历年状况，同时分析现行预测模型可靠性。第四章通过建立因子分析模型，以得到影响财政收入的关键因素，同时实现数据在信息量损失较少的情况降维。第五章通过神经网络建立辽宁省财政收入预测模型，分析其性能。最后总结全文，同时给出相关政策建议。

## 2 辽宁省财政收入预算公开数据抽样调查

### 2.1 调查背景

财政预算是政府调节经济和社会发展的重要工具。在市场经济条件下，当市场难以保持自身均衡发展时，政府可以根据市场经济运行状况，选择适当的预算总量或结构政策，用预算手段去弥补市场缺陷，谋求经济的稳定增长。

根据我国审计署公布的相关统计数据，虽然目前我国地方政府的债务整体处于安全范围，但是部分地区已出现了严重过量负债或债务管理不规范等问题。辽宁在全国财政收入的排名正在迅速下降。2010 年，辽宁在 31 省市中排名第 7 位，紧随在北京之后，排在四川之前。到了 2015 年，排名下降到第 17 位，是所有省市中排名下降最高的省份。财政收入增速的降低，对地方政府的偿债也形成了一定的压力，已经披露的预算报告显示，2015 年末辽宁省政府债务余额为 8718.5 亿，债务率为 157%，相比 2012 年末上升了 88 个百分点，已经超过 100% 的债务率线。

准确进行财政预算，准确预测财政收入，做好政府财政预算的编制，合理进行赤字安排，最大可能地实现长期的财政收支平衡，对于辽宁经济振兴发展显得尤为重要。

### 2.2 抽样方法

为了了解辽宁省财政收入预算的情况，从 1985-2016 这 32 年间公开的辽宁省地区财政收入预算收入中，采用随机抽样方法，不重复地随机抽取 5 个年份进行研究。

### 2.3 抽样结果与分析

运用 Python 实现重复随机抽样，结果如图 2-1 所示。

```
import random

list = [1985,1986,1987,1988,1989,1990,1991,1992,1993,1994,
1995,1996,1997,1998,1999,2000,2001,2002,2003,2004,
2005,2006,2007,2008,2009,2010,2011,2012,2013,2014,
2015,2016]

sample_year = random.sample(list,5)

print(sample_year)

[2003, 2001, 2005, 2015, 2006]
```

图 2-1 随机抽样结果图

通过查阅辽宁省统计局统计年鉴获得样本相关财政收入数据如表 2-1 所示。



表 2-1 随机抽样结果表

年份	真实值
1985	85.2
1986	98.9
1994	153.7
1997	228.2
1999	279.6

平均数 $\bar{x}$ 重复抽样误差 $\mu_{\bar{x}}$ 计算公式:

$$\mu_{\bar{x}} = \sqrt{\frac{\sigma^2}{n} \left(1 - \frac{n}{N}\right)} \quad (2-1)$$

其中 $\sigma^2$ 为总体平均方差, 这里用样本平均方差替代。

解得 $\bar{x}=887.56$ ,  $\sigma^2=2048656.52$ ,  $\mu_{\bar{x}}=415.76$ , 令置信度为 1.96, 则 1985-2016 年的财政收入有 95%的概率落在 72.67-1702.44 亿元之间。

抽出的样本中, 有 $p=20\%$ 的概率年财政收入预算在 2000 亿元以上。

$$\mu_{\bar{x}} = \sqrt{\frac{p(1-p)}{n} \left(1 - \frac{n}{N}\right)} = 0.1643$$

令置信度为 1.65, 则在 1985-2016 年的财政收入在 2000 亿元以上有的 90% 概率占比在 0%-52.21%之间。

抽样调查时间短、收效快、费用省且易推广, 但由于抽样调查所调查的对象中的一部分, 抽样调查的结果是从抽取样本中获取的信息推断出来的, 所以抽样调查存在着抽样误差。抽样误差是客观存在的, 在一定范围内也是允许的。

但是, 由于 1994 年以后地方财政收入与以前实行的总额分成财政体制下的地方财政收入在内容和范围上有一定差别, 从 2002 年 1 月 1 日起, 国所得税收入分配方式又有重大改革, 2002 年所得税收入中央分享 50%, 地方分享 50%。因此, 历年数据不完全可比。且 2008 年金融危机爆发之后, 我国实行了较大强度的积极财政政策, 直接导致了地方政府债务的急剧扩张, 导致之前的数据纵向可比性较差。事实上, 1985-2016 年财政收入平均数为 1613.90 亿万元, 财政收入在 2000 亿元以上的年份占比 28.125%, 抽样调查情况与实际出入较大, 说明现有的预测模型不够理想。

因此, 在下面的研究中, 有探索具有时效性的财政收入预算模型, 以促进财政收支平衡, 助力辽宁省经济平稳运行的必要。

### 3 财政收入的时间序列分析

#### 3.1 辽宁省财政收入的水平分析

##### 3.1.1 发展水平与平均发展水平

在财政收入时间序列中，每一年的财政收入叫做发展水平，它是时间序列分析的基础。用 $a_i$ 表示，代表在这一年上达到的一种数量状态，辽宁省 1985-2016 年对应的财政收入值表示为 $a_1, a_2, \dots, a_{32}$ ，如图 3-1 所示。

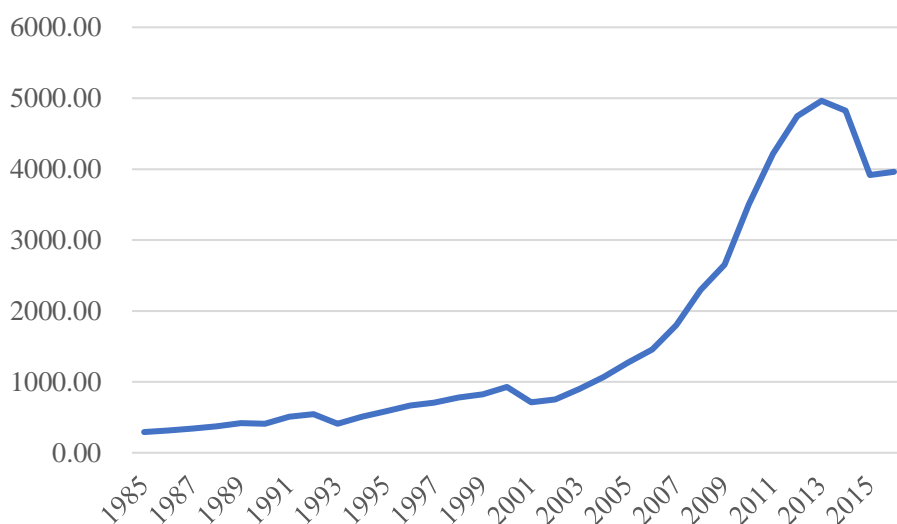


图 3-2 辽宁省 1985-2016 财政收入发展水平图

其中 $a_1=292.2$  亿元为最初水平， $a_{32}=3961.078922$  为最末水平， $a_2, \dots, a_{31}$ 为中间水平。因其为时期序列，平均数 $\bar{a}$ 计算公式为式(3-1)：

$$\bar{a} = \frac{\sum_{i=1}^{32} a_i}{32} \quad (3-1)$$

计算得财政收入在 1985-2016 年平均数 $\bar{a}$ 为 1613.90 亿元。高于其平均预算收入 897.00 亿元，且高于全国平均财政收入预算 1289.68 亿元。

从 1985-2013 年的财政收入预算和实际收入可以看出，辽宁省在该阶段财政收入运行持续向好，但在 2014 年开始出现负增长情况，2015 年财政收入出现大幅下降。这与财政收入“挤水分”相关，辽宁省委副书记、省长陈求发代表省政府作政府工作报告时，首次对外确认，辽宁省所辖市、县，在 2011 年至 2014 年存在财政数据造假的问题，陈求发在政府工作报告中解释，“我们顶着面子上难看的压力，认真地挤压水分，2015 年夯实了财政收据，2016 年以来努力夯实其他经济数据。”

##### 3.1.2 增长量与平均增长量

增长量又称增减量是在一定时期内所增减的绝对量，即报告期水平与基期水平之差。它说明某种社会经济现象报告期水平比基期水平增加(或减少)了多少。计算结果为正值，表示增加量；计算结果为负值，表示减少量。由于比较的基期不

同，增长量分为累计增长量和逐期增长量两种。

逐年增长量

$$\Delta_{1i} = a_i - a_{i-1} \quad (3-2)$$

累计增长量

$$\Delta_{2i} = a_i - a_1 \quad (3-3)$$

其中  $i = 2, 3, \dots, 9$ 。平均增长量为逐期增长量之和除以逐期增长量个数，即累计增长量除以观察值个数减 1，由式(3-2)、(3-3)计算得：

表 3-1 辽宁省 2008-2016 财政收入增长量表

年份	地区财政收入/亿元	逐年增长量/亿元	累计增长量/亿元	年份	地区财政收入/亿元	逐年增长量/亿元	累计年增长量/亿元
1985	292.20	—	—	2001	710.86	-219.2	418.7
1986	312.80	20.6	20.6	2002	751.32	40.5	459.1
1987	341.90	29.1	49.7	2003	899.62	148.3	607.4
1988	371.70	29.8	79.5	2004	1069.45	169.8	777.3
1989	416.30	44.6	124.1	2005	1270.34	200.9	978.1
1990	409.70	-6.6	117.5	2006	1456.21	185.9	1164.0
1991	508.60	98.9	216.4	2007	1799.87	343.7	1507.7
1992	542.60	34.0	250.4	2008	2295.04	495.2	2002.8
1993	411.00	-131.6	118.8	2009	2651.89	356.9	2359.7
1994	507.00	96.0	214.8	2010	3504.61	852.7	3212.4
1995	585.10	78.1	292.9	2011	4216.10	711.5	3923.9
1996	668.90	83.8	376.7	2012	4749.42	533.3	4457.2
1997	706.00	37.1	413.8	2013	4963.68	214.3	4671.5
1998	779.00	73.0	486.8	2014	4822.58	-141.1	4530.4
1999	825.20	46.2	533.0	2015	3914.74	-907.8	3622.5
2000	930.10	104.9	637.9	2016	3961.08	46.3	3668.9

年平均增长量为 114.65 亿元。由表 3-1 可知，总体看来，辽宁省 1985-2016 年财政收入呈增长趋势。

## 3.2 辽宁省财政收入的速度分析

### 3.2.1 发展速度与增长速度

发展速度是反应社会经济发展程度的相对指标，它是现象的报告期水平与基期水平之商。由于采用的基期不同，发展速度可分为环比发展速度和定基发展速度。在本文，环比发展速度 $R_{1i}$ 说明辽宁省财政收入在观察期内逐年变化的程度，定基发展速度 $R_{2i}$ 说明辽宁省 1985-2016 年总的发展变化程度，计算分别如式(3-4)和式(3-5)所示。

$$R_{1i} = \frac{a_i}{a_{i-1}} \quad (3-4)$$

$$R_{2i} = \frac{a_i}{a_1} \quad (3-5)$$

其中 $i = 2, 3, \dots, 32$ 。

另一个速度指标，增长速度也称增长率，增长速度可以是正数，也可以是负数。正数表示增长，负数表示降低。增长速度由于采用的基期不同，可分为环比增长速度 $G_{1i}$ 和定基增长速度 $G_{2i}$ ，其计算公式如式(3-6)和(3-7)所示。

$$G_{1i} = \frac{a_i - a_{i-1}}{a_{i-1}} \quad (3-6)$$

$$G_{2i} = \frac{a_i - a_1}{a_1} \quad (3-7)$$

其中 $i = 2, 3, \dots, 32$ 。

由式(3-4)至(3-7)可得：

表 3-4 辽宁省 2008-2016 财政收入发展水平一览表

年份	财政收入/亿元	环比发展速度	定基发展速度	环比增长速度	定基增长速度
1985	292.20	—	—	—	—
1986	312.80	1.0705	1.0705	0.0705	0.0705
1987	341.90	1.093031	1.170089	0.093031	0.170089
1988	371.70	1.08716	1.272074	0.08716	0.272074
1989	416.30	1.119989	1.424709	0.119989	0.424709
1990	409.70	0.984146	1.402122	-0.01585	0.402122
1991	508.60	1.241396	1.740589	0.241396	0.740589
1992	542.60	1.06685	1.856947	0.06685	0.856947
1993	411.00	0.757464	1.406571	-0.24254	0.406571
1994	507.00	1.233577	1.735113	0.233577	0.735113

1995	585.10	1.154043	2.002396	0.154043	1.002396
1996	668.90	1.143223	2.289185	0.143223	1.289185
1997	706.00	1.055464	2.416153	0.055464	1.416153
1998	779.00	1.103399	2.665982	0.103399	1.665982
1999	825.20	1.059307	2.824093	0.059307	1.824093
2000	930.10	1.127121	3.183094	0.127121	2.183094
2001	710.86	0.764283	2.432786	-0.23572	1.432786
2002	751.32	1.056917	2.571253	0.056917	1.571253
2003	899.62	1.197386	3.078782	0.197386	2.078782
2004	1069.45	1.18878	3.659993	0.18878	2.659993
2005	1270.34	1.187844	4.347502	0.187844	3.347502
2006	1456.21	1.146315	4.983607	0.146315	3.983607
2007	1799.87	1.235996	6.159719	0.235996	5.159719
2008	2295.04	1.275114	7.854346	0.275114	6.854346
2009	2651.89	1.155487	9.075599	0.155487	8.075599
2010	3504.61	1.321552	11.99387	0.321552	10.99387
2011	4216.10	1.203015	14.42881	0.203015	13.42881
2012	4749.42	1.126496	16.25401	0.126496	15.25401
2013	4963.68	1.045113	16.98727	0.045113	15.98727
2014	4822.58	0.971573	16.50437	-0.02843	15.50437
2015	3914.74	0.811752	13.39746	-0.18825	12.39746
2016	3961.08	1.011838	13.55606	0.011838	12.55606

### 3.2.2 平均发展速度与平均增长速度

根据各期的环比发展速度采用集合平均法计算公式为式(3-8):

$$\bar{R} = \sqrt[31]{\frac{a_2}{a_1} \times \frac{a_3}{a_2} \times \cdots \times \frac{a_{32}}{a_{31}}} = \sqrt[31]{\prod_{i=2}^{32} \frac{a_i}{a_{i-1}}} = \sqrt[31]{\frac{a_{32}}{a_1}} \quad (3-8)$$

计算得 $\bar{R}=1.088$ ，则平均增长速度 $\bar{G}=0.088$ 。

总的来说，辽宁省在 1985-2016 年间财政收入呈 8.8%增长的趋势，1991、1994、2007、2010、2011 年增长较快，环比增长速度均在 20%以上，在 1993、2001、2014、2015 年，财政收入出现负增长情况。其中 2008 年前后的大幅增长与次贷危机相关，而 2014 与 2015 年出现较大幅度的负增长与辽宁省财政收入数据挤出水分有关。

### 3.3 时间序列的构成分析和预测方法

#### 3.3.1 辽宁省财政收入的构成与分解

目前，公开财政收入预算范围主要包括各项税收、国有资本经营收入、国有企业计划亏补和预算其他各项收入构成。

#### 3.3.2 辽宁省财政收入的预测方法

##### (1) 移动平均法

辽宁省财政收入由 32 期，以 4 为移动平均项数，可以计算出：

$$\begin{aligned}\bar{Y}_{1.5}^{(1)} &= \frac{1}{2}(Y_1 + Y_2) \\ \bar{Y}_{2.5}^{(1)} &= \frac{1}{2}(Y_2 + Y_3) \\ \bar{Y}_{3.5}^{(1)} &= \frac{1}{2}(Y_3 + Y_4) \\ \bar{Y}_{4.5}^{(1)} &= \frac{1}{2}(Y_4 + Y_5) \\ &\vdots\end{aligned}\tag{3-9}$$

再进行一次移动项数为 2 的移动平均，即：

$$\begin{aligned}\bar{Y}_2^{(2)} &= \frac{1}{2}(\bar{Y}_{1.5}^{(1)} + \bar{Y}_{2.5}^{(1)}) \\ \bar{Y}_3^{(2)} &= \frac{1}{2}(\bar{Y}_{2.5}^{(1)} + \bar{Y}_{3.5}^{(1)}) \\ \bar{Y}_4^{(2)} &= \frac{1}{2}(\bar{Y}_{3.5}^{(1)} + \bar{Y}_{4.5}^{(1)}) \\ \bar{Y}_5^{(2)} &= \frac{1}{2}(\bar{Y}_{4.5}^{(1)} + \bar{Y}_{5.5}^{(1)}) \\ &\vdots\end{aligned}\tag{3-10}$$

根据式(3-9)和(3-10)，可以计算出 32 年的移动平均值，如表 3-5 所示。

表 3-5 移动平均结果一览表

年份	$\bar{Y}_{t+1}^{(1)}$	$\bar{Y}_t^{(2)}$	年份	$\bar{Y}_{t+1}^{(1)}$	$\bar{Y}_t^{(2)}$
1985	—	—	2001	822.98	813.67
1986	329.65	—	2002	857.81	840.39
1987	360.68	345.16	2003	997.68	927.75
1988	384.90	372.79	2004	1173.91	1085.79
1989	426.58	405.74	2005	1398.97	1286.44
1990	469.30	447.94	2006	1705.37	1552.17

1991	467.98	468.64	2007	2050.75	1878.06
1992	492.30	480.14	2008	2562.85	2306.80
1993	511.43	501.86	2009	3166.91	2864.88
1994	543.00	527.21	2010	3780.51	3473.71
1995	616.75	579.88	2011	4358.45	4069.48
1996	684.75	650.75	2012	4687.94	4523.20
1997	744.78	714.76	2013	4612.60	4650.27
1998	810.08	777.43	2014	4415.52	4514.06
1999	811.29	810.68	2015	—	—
2000	804.37	807.83	2016	—	—

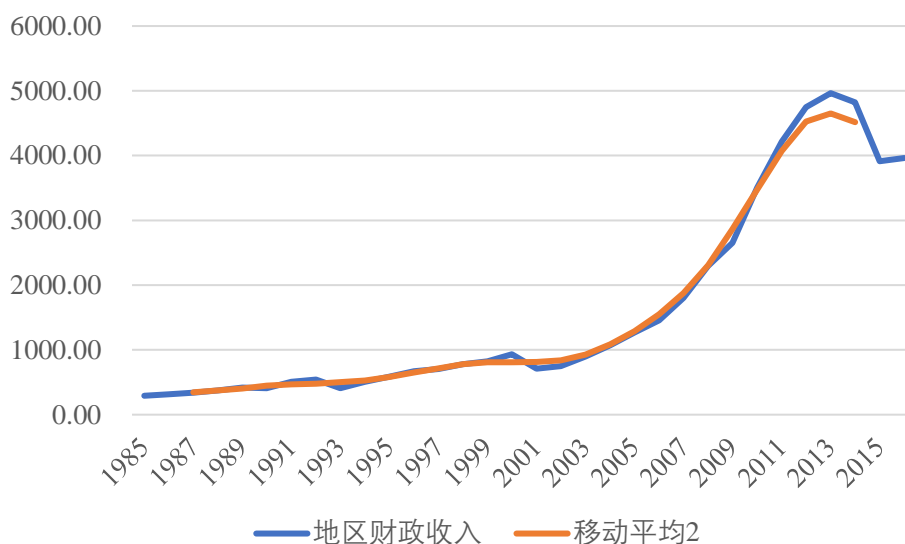


图 3-5 移动平均结果图

从图 3-5 可以看出，经移动平均后，数据变得平滑。

## （2）趋势方程拟合法

考虑到财政收入与各项税收、国有资本经营收入、国有企业计划亏补和预算其他各项收入呈线性关系，因此采用多元线性回归进行拟合。

对于多元线性回归模型

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \varepsilon_i \quad (3-12)$$

多元线性回归模型的参数估计比一元线性回归模型要复杂得多，为了便于计算和分析，便于将结果由四变量总体推广到一般的多变量总体，引入矩阵这一工具简化计算和分析。

设 $(y_i, x_{1i}, x_{2i}), i = 1, 2, \dots, n$ 是取自总体的一组随机样本。在该组样本下，总体回归模型可以写成方程组的形式：

$$\begin{aligned}
y_1 &= \beta_0 + \beta_1 x_{11} + \beta_2 x_{21} + \beta_3 x_{31} + \beta_4 x_{41} + \varepsilon_1 \\
y_2 &= \beta_0 + \beta_1 x_{12} + \beta_2 x_{22} + \beta_3 x_{32} + \beta_4 x_{42} + \varepsilon_2 \\
&\vdots \\
y_n &= \beta_0 + \beta_1 x_{1n} + \beta_2 x_{2n} + \beta_3 x_{3n} + \beta_4 x_{4n} + \varepsilon_n
\end{aligned} \tag{3-13}$$

利用矩阵运算，可表示为：

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{21} & x_{31} & x_{41} \\ 1 & x_{12} & x_{22} & x_{32} & x_{42} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1n} & x_{2n} & x_{3n} & x_{4n} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

记  $y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$ ,  $X = \begin{bmatrix} 1 & x_{11} & x_{21} & x_{31} & x_{41} \\ 1 & x_{12} & x_{22} & x_{32} & x_{42} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1n} & x_{2n} & x_{3n} & x_{4n} \end{bmatrix}$ ,  $\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix}$ ,  $\varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$ 。

则在该组样本下，总体回归模型的矩阵表示为：

$$y = \beta X + \varepsilon \tag{3-14}$$

该模型满足以下假定：

假定 1 回归模型是参数线性的，并且是设定正确的。

假定 2 随机误差项与解释变量不相关。即  $Cov(y, \varepsilon) = 0$ 。如果解释变量是非随机的，则该假设自动满足。

假定 3 零均值假定。即  $E(\varepsilon) = 0$ 。

假定 4 同方差假定。即  $Var(\varepsilon) = \sigma^2$ 。

假定 5 无自相关假定。即  $Cov(\varepsilon_i, \varepsilon_j) = 0$ 。

假定 6 解释变量之间不存在完全共线性，即两个解释变量之间无确切的线性关系。

假定 7 正态性假定。即  $\varepsilon \sim N(0, \sigma^2)$ 。

系数向量  $\beta$  的 OLS 估计为：

$$\hat{\beta} = (X^T X)^{-1} X^T y \tag{3-15}$$

在随机误差项服从正态分布的假定下，系数向量的估计量也服从正态分布：

$$\hat{\beta}_j \sim N(\beta_j, \sigma^2)$$

记  $C = (X^T X)^{-1}$  第  $j$  个主对角元素为  $c_{jj}$ ，则

$$\hat{\beta}_j \sim N(\beta_j, \sigma^2 c_{jj})$$



有了系数估计量的分布,就可以对总体参数做假设检验。与双变量总体相同,总体误差 $\varepsilon$ 是不可观察的,因而其方差 $\sigma^2$ 是未知的。若用 $\sigma^2$ 的无偏估计量 $\widehat{\sigma^2}$ 代替 $\sigma^2$ ,则 OLS 估计量服从自由度 $n - 5$ 的 $t$ 分布,而不是正态分布,即:

$$\frac{\widehat{\beta}_j - \beta_j}{se(\widehat{\beta}_j)} \sim t(n - 5)$$

$$\text{其中, } se(\widehat{\beta}_j) = \sqrt{\widehat{\sigma^2} c_{jj}}, \quad \widehat{\sigma^2} = \frac{\sum \varepsilon_i^2}{n-5}。$$

点估计 $\widehat{y}$ 的标准差为 $se(\widehat{y}_i) = \sqrt{\widehat{\sigma^2}(1 + x_i^T X^T X x_i)}$ ,  $\widehat{y}_i$ 置信度为 $1 - \alpha$ 的置信区间为 $\left[\widehat{y}_i - t_{\frac{\alpha}{2}}(n - 5)se(\widehat{y}_i), \widehat{y}_i + t_{\frac{\alpha}{2}}(n - 5)se(\widehat{y}_i)\right]$ 。其中, 每一个 $se(\widehat{y}_i)$ 与第 2 章中的 $\mu_{\bar{x}}$ 的区别为 $\mu_{\bar{x}}$ 为估计一个点, 而 $se(\widehat{y}_i)$ 为估计一条线。

运用 SPSS, 求解得到拟合优度 $R^2 = 1 - \frac{\sum \varepsilon_i^2}{\sum (\widehat{y}_i - \bar{y})^2} = 99.2\%$ , 显著性水平为 0.000 的多元线性回归模型, 令各项税收、国有资本经营收入、国有企业计划亏损、预算各项收入分别为 $x_1, x_2, x_3, x_4$ , 得到 $se = 135.34$ , 显著性水平为 0.000 的多元线性回归方程:

$$y = 260 + 2.075x_1 - 2.276x_2 + 4.897x_3 + 0.276x_4$$

但是,  $x_2, x_3, x_4$ 的显著性水平分别为 0.037, 0.136, 0.863 均高于 0.005, 故拒绝其相关性假设。因此, 为了进一步探索影响税收的因子, 运用更具有细化数据建立更具有时效性及准确性的模型, 对辽宁省近 9 年财政收入各项目进行因子分析。

## 4 财政收入的因子分析模型

### 4.1 因子分析简介

因子分析 (Factor analysis): 就是用少数几个因子来描述许多指标或因素之间的联系, 以较少几个因子来反映原资料的大部分信息的统计学分析方法。从数学角度来看, 主成分分析是一种化繁为简的降维处理技术。主成分分析 (Principal component analysis): 是因子分析的一个特例, 是使用最多的因子提取方法。它通过坐标变换手段, 将原有的多个相关变量, 做线性变化, 转换为另外一组不相关的变量。选取前面几个方差最大的主成分, 这样达到了因子分析较少变量个数的目的, 同时又能与较少的变量反映原有变量的绝大部分的信息。两者关系: 主成分分析 (PCA) 和因子分析 (FA) 是两种把变量维数降低以便于描述、理解和分析的方法, 而实际上主成分分析可以说是因子分析的一个特例。

其具有以下几大特点:

- (1) 因子变量的数量远少于原有的指标变量的数量, 因而对因子变量的分析能够减少分析中的工作量。
- (2) 因子变量不是对原始变量的取舍, 而是根据原始变量的信息进行重新组构, 它能够反映原有变量大部分的信息。
- (3) 因子变量之间不存在显著的线性相关关系, 对变量的分析比较方便, 但原始部分变量之间多存在较显著的相关关系。
- (4) 因子变量具有命名解释性, 即该变量是对某些原始变量信息的综合和反映。

在保证数据信息丢失最少的原则下, 对高维变量空间进行降维处理 (即通过因子分析或主成分分析)。显然, 在一个低维空间解释系统要比在高维系统容易的多。其分析原理为有  $n$  个样本, 每个样本共有  $p$  个变量, 构成一个  $n \times p$  阶的数据矩阵:

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}$$

当  $p$  较大时, 在  $p$  维空间中考察问题比较麻烦。这就需要进行降维处理, 即用较少几个综合指标代替原来指标, 而且使这些综合指标既能尽量多地反映原来指标所反映的信息, 同时它们之间又是彼此独立的。记  $x_1, x_2, \dots, x_p$  为原变量指标,  $z_1, z_2, \dots, z_m$  ( $m \leq p$ ) 为新变量指标 (主成分), 则其线性组合为:

$$\begin{cases} z_1 = l_{11}x_1 + l_{12}x_2 + \cdots + l_{1p}x_p \\ z_2 = l_{21}x_1 + l_{22}x_2 + \cdots + l_{2p}x_p \\ \vdots \\ z_m = l_{m1}x_1 + l_{m2}x_2 + \cdots + l_{mp}x_p \end{cases} \quad (4-1)$$

$L_{ij}$  是原变量在各主成分上的载荷。  $z_i$  与  $z_j$  无关,  $z_1$  是  $x_1, x_2, \dots, x_p$  的一切线性组合中方差最大者,  $z_2$  是与  $z_1$  不相关的  $x_1, x_2, \dots, x_p$  的所有线性组合中方差最大者。则新变量指标  $z_1, z_2, \dots, z_m$  分别称为原变量指标的第一, 第二,  $\dots$ , 第  $m$  个主成分。  $Z$  为因子变量或公共因子, 可以理解为在高维空间中互相垂直的  $m$  个坐标轴。

主成分分析实质就是确定原来变量 $x_j$  ( $j=1, 2, \dots, p$ ) 在各主成分 $z_i$  ( $i=1, 2, \dots, m$ ) 上的荷载 $l_{ij}$ 。从数学上容易知道, 从数学上也可以证明, 它们分别是相关矩阵的 $m$ 个较大的特征值所对应的特征向量。无论是哪一种因子分析方法, 其相应的因子解都不是唯一的, 主因子解仅仅是无数因子解中之一。

## 4.2 因子分析模型建立

本文采用主成分分析法, 通过坐标变换, 转化为另一组不相关的变量 $z_i$ , 求相关系数矩阵的特征根 $\lambda_i$  ( $\lambda_1, \lambda_2, \dots, \lambda_{25} > 0$ ) 和对应的标准正交的特征向量 $l_1$ , 根据相关系数矩阵的特征根即 $z_j$ 的方差贡献, 即因子载荷矩阵 $L$ 中第 $j$ 各元素的平方和), 计算公共因子 $z_j$ 的方差贡献率 $c_1$ 与累积贡献率 $c_2$ 。

$$c_1 = \frac{\lambda_i}{\sum_{k=1}^{25} \lambda_k} \quad (4-2)$$

$$c_2 = \frac{\sum_{k=1}^i \lambda_k}{\sum_{k=1}^{25} \lambda_k} \quad (4-3)$$

其中,  $i=1, 2, \dots, 25$ 。

主成分分析是在一个多维坐标轴中, 将原始变量组成的坐标系进行平移变换, 使得新的坐标原点和数据群点的重心重合。新坐标第一轴与数据变化最大方向对应。通过计算特征根 (方差贡献) 和方差贡献率与累积方差贡献率等指标, 来判断选取公共因子的数量和公共因子 (主成分) 所能代表的原始变量信息。

公共因子个数的确定准则: 1) 根据特征值的大小来确定, 一般取大于 1 的特征值对应的几个公共因子/主成分。2) 根据因子的累积方差贡献率来确定, 一般取累计贡献率达 85-95% 的特征值所对应的第一、第二、 $\dots$ 、第  $m$  ( $m \leq p$ ) 个主成分。也有学者认为累积方差贡献率应在 80% 以上。

因子变量的命名解释是因子分析的另一个核心问题。经过主成分分析得到的公共因子 $z_1, z_2, \dots, z_m$ 是对原有变量的综合。在实际的应用分析中, 主要通过对载荷矩阵进行分析, 得到因子变量和原有变量之间的关系, 从而对新的因子变量进行命名。利用因子旋转方法能使因子变量更具有可解释性。构建负载矩阵 $A$ :

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{251} & a_{252} & \cdots & x_{25m} \end{bmatrix} = \begin{bmatrix} l_{11}\sqrt{\lambda_1} & l_{12}\sqrt{\lambda_2} & \cdots & l_{1m}\sqrt{\lambda_m} \\ l_{21}\sqrt{\lambda_1} & l_{22}\sqrt{\lambda_2} & \cdots & l_{2m}\sqrt{\lambda_m} \\ \vdots & \vdots & \ddots & \vdots \\ l_{251}\sqrt{\lambda_1} & l_{252}\sqrt{\lambda_2} & \cdots & l_{25m}\sqrt{\lambda_m} \end{bmatrix}$$

正交旋转和斜交旋转是因子旋转的两类方法。前者由于保持了坐标轴的正交性, 因此使用最多。正交旋转的方法很多, 其中以方差最大化法最为常用。

方差最大正交旋转 (varimax orthogonal rotation) 基本思想是, 使公共因子的相对负荷的方差之和最大, 且保持原公共因子的正交性和公共方差总和不变。可使每个因子上的具有最大载荷的变量数最小, 因此可以简化对因子的解释。

因子旋转的目的是使因子负荷两极分化, 要么接近于 0, 要么接近于 1。从而使原有因子变量更具有可解释性。

因子变量确定以后, 对于每一个样本数据, 希望得到它们在不同因子上的具体数据值, 即因子得分。估计因子得分的方法主要有: 回归法、Bartlette 法等。

计算因子得分应首先将因子变量表示为原始变量的线性组合。

回归法得分是由贝叶斯思想导出的，得到的因子得分是有偏的，但计算结果误差较小。贝叶斯判别思想是根据先验概率求出后验概率，并依据后验概率分布作出统计推断。**Bartlett 法**：Bartlett 因子得分是极大似然估计，得到的因子得分是无偏的，但计算结果误差较大。本文选取回归法计算得分。

### 4.3 模型求解与分析

采用 SPSS 22 统计软件中的 FACTOR 模块，运用因子分析法对各指标进行处理，得出主因子对应的特征值和贡献率，如表 4-1 所示，按累计方差贡献率达到 90%，解释了指标大部分的信息，因此选取前 4 个因子代表 25 个变量。

表 4-1 相关矩阵特征方程特征根表

成分	初始特征值			旋转提取因子的负载平方和		
	特征值	方差贡献率/%	累计方差贡献率/%	特征值	方差贡献率/%	累计方差贡献率 %
1	15.753	63.011	63.011	7.737	30.947	30.947
2	4.605	18.419	81.430	7.731	30.924	61.871
3	1.980	7.922	89.352	4.473	17.892	79.763
4	1.368	5.472	94.824	3.765	15.061	94.824
5	.673	2.690	97.514	—	—	—
6	.375	1.500	99.014	—	—	—
7	.198	.792	99.806	—	—	—
8	.049	.194	100.000	—	—	—
9	1.260E-15	5.039E-15	100.000	—	—	—
10	6.616E-16	2.646E-15	100.000	—	—	—
11	4.522E-16	1.809E-15	100.000	—	—	—
12	4.090E-16	1.636E-15	100.000	—	—	—
13	3.284E-16	1.313E-15	100.000	—	—	—
14	1.973E-16	7.892E-16	100.000	—	—	—
15	1.386E-16	5.543E-16	100.000	—	—	—
16	7.477E-17	2.991E-16	100.000	—	—	—
17	2.810E-17	1.124E-16	100.000	—	—	—
18	-7.777E-18	-3.111E-17	100.000	—	—	—
19	-1.154E-16	-4.614E-16	100.000	—	—	—
20	-2.425E-16	-9.700E-16	100.000	—	—	—
21	-2.943E-16	-1.177E-15	100.000	—	—	—
22	-3.985E-16	-1.594E-15	100.000	—	—	—
23	-4.769E-16	-1.908E-15	100.000	—	—	—
24	-7.185E-16	-2.874E-15	100.000	—	—	—
25	-1.425E-15	-5.701E-15	100.000	—	—	—

接着，用方差最大正交旋转法求解旋转后的成分矩阵如表 4-2 所示。

表 4-2 旋转后的成分矩阵表

组成	主成分			
	1	2	3	4
个人所得税	.851	-.028	.432	-.042
其他税收	-.838	.153	-.149	.219
关税	.778	.464	.087	.370
企业所得税	.777	.507	-.199	.262
增值税	.719	.415	.341	.298
车辆购置税	.690	.452	.510	.229
车船使用税	.690	-.019	.371	.601
印花税	.689	.554	.417	.194
房产税	.671	.103	.480	.547
外商外国企业所得税	.665	.314	.608	.292
城市维护建设税	.654	.391	.540	.340
教育费附加	.654	.414	.532	.329
地方教育费收入	.649	.396	.401	.488
耕地占用税	.032	.979	.108	.012
资源税	.160	.952	.093	.164
土地增值税	.223	.940	.175	.104
契税	.099	.927	.246	-.174
营业税	.174	.886	.260	-.099
城镇土地使用税	.342	.832	.231	.306
文化事业建设费	-.419	.576	-.209	-.542
投资方向调节税	-.186	-.251	-.941	.004
消费税	.386	.209	.868	.115
罚没收入	.039	.109	.084	-.946
其他收入	-.559	-.306	-.400	-.617
烟叶税	.283	.573	.330	.607

可以看出公共因子 1 主要构成有个人所得税、关税、企业所得税、增值税、车辆购置税、车船使用税、印花税、房产税、外商外国企业所得税、城市维护建设税、教育费及附加、地方教育费收入，公共因子 2 主要构成有耕地占用税、资源税、土地增值税、契税、营业税、城镇土地使用税、文化事业建设费、其他收入，公共因子 3 的主要构成有消费税、罚没收入，构成公共因子 4 的主要有其他税收、投资方向调节税、烟叶税。可以看出个人及企业是财政收入的主要来源。

根据其成分得分为权重，如表 4-3 所示，加总得到降维结果见表 4-4。

表 4-3 成分得分相关矩阵表

变量名	组成	主成分			
		1	2	3	4
$x_1$	增值税	.109	.016	-.038	-.005
$x_2$	消费税	-.103	-.045	.345	-.046
$x_3$	营业税	-.026	.132	.033	-.081
$x_4$	个人所得税	.215	-.078	.022	-.181
$x_5$	外商外国企业所得税	.029	-.019	.124	-.008
$x_6$	企业所得税	.265	.070	-.343	-.011
$x_7$	资源税	-.038	.158	-.067	.044
$x_8$	投资方向调节税	.175	.033	-.433	.066
$x_9$	城市维护建设税	.029	.000	.084	.018
$x_{10}$	房产税	.028	-.048	.062	.107
$x_{11}$	印花税	.090	.037	.003	-.048
$x_{12}$	城镇土地使用税	-.024	.118	-.034	.069
$x_{13}$	土地增值税	-.024	.146	-.036	.003
$x_{14}$	车船使用税	.059	-.062	.006	.131
$x_{15}$	耕地占用税	-.064	.167	-.023	-.002
$x_{16}$	契税	-.039	.144	.044	-.100
$x_{17}$	车辆购置税	.066	.011	.058	-.036
$x_{18}$	烟叶税	-.110	.067	.041	.207
$x_{19}$	关税	.178	.041	-.193	.026
$x_{20}$	其他税收	-.317	.079	.109	.247
$x_{21}$	教育费附加	.032	.005	.078	.013
$x_{22}$	文化事业建设费	-.036	.138	-.031	-.156
$x_{23}$	地方教育费收入	.036	.012	.006	.089
$x_{24}$	罚没收入	.152	.009	.060	-.425
$x_{25}$	其他收入	.015	.000	-.030	-.161

表 4-4 降维后数据表

年份	公共因子 1	公共因子 2	公共因子 3	公共因子 4
2008	218.8364859	98.0712151	-110.3269796	-56.2301568
2009	132.71944	100.47983	52.41545	-72.85787
2010	184.70689	133.83052	56.11221	-95.73748
2011	230.2240972	185.4011636	20.4499613	-109.7332492
2012	223.9152404	235.4202445	22.5226104	-100.7011863
2013	232.1145137	255.8840088	12.16610071	-102.3925829
2014	252.6672073	217.6286098	12.14986681	-91.26324602
2015	218.0618766	106.8717683	95.3769832	-94.0470943
2016	278.8086773	90.58318471	29.60143371	-74.15949582

主成分分析是将多项指标转化为少数几项综合指标,用综合指标来解释多变量的方差—协方差结构。综合指标即为主成分。所得出的少数几个主成分,要尽可能多地保留原始变量的信息,且彼此不相关。因子分析是研究如何以最少的信息丢失,将众多原始变量浓缩成少数几个因子变量,以及如何使因子变量具有较强的可解释性的一种多元统计分析方法。

因子分析是寻找潜在的起支配作用的因子模型的方法。因子分析是根据相关性大小把变量分组,使得同组内的变量之间相关性较高,但不同的组的变量相关性较低,每组变量代表一个基本结构,这个基本结构称为公共因子。对于所研究的问题就可试图用最少个数的不可测的所谓公共因子的线性函数与特殊因子之和来描述原来观测的每一分量。通过因子分析得来的新变量是对每个原始变量进行内部剖析。

因后文神经网络需要的是隐藏了潜在信息,且互不相关特征作为输入,故此处使用成分得分作为权重,对数据进行降维。

由表 4-4 可以看出,公共因子 1、2 涉及的财政收入来源均对财政收入产生正面影响,因此可以主要通过增进就业、增加进出口贸易、刺激消费等措施,同时抑制居民对烟叶的消费,综合扩大内需与稳定外需提高政府财政预算收入。

## 5 基于神经网络的财政收入预测模型

### 5.1 BP 神经网络简介

在人工神经网络的发展历史上，感知机（Multilayer Perceptron, MLP）网络曾对人工神经网络的发展发挥了极大的作用，也被认为是一种真正能够使用的人工神经网络模型，它的出现曾掀起了人们研究人工神经网络的热潮。单层感知网络（M-P 模型）做为最初的神经网络，具有模型清晰、结构简单、计算量小等优点。但是，随着研究工作的深入，人们发现它还存在不足，例如无法处理非线性问题，即使计算单元的作用函数不用阀函数而用其他较复杂的非线性函数，仍然只能解决线性可分问题。不能实现某些基本功能，从而限制了它的应用。增强网络的分类和识别能力、解决非线性问题的途径是采用多层前馈网络，即在输入层和输出层之间加上隐含层。构成多层前馈感知器网络。

20 世纪 80 年代中期，David Rumelhart、Geoffrey Hinton 和 Ronald Williams、David Parker 等人[9]发现了误差反向传播算法（Error Back Propagation Training），简称 BP，系统解决了多层神经网络隐含层连接权学习问题，并在数学上给出了完整推导。人们把采用这种算法进行误差校正的多层前馈网络称为 BP 网。

BP 神经网络具有任意复杂的模式分类能力和优良的多维函数映射能力，解决了简单感知器不能解决的异或（Exclusive OR, XOR）和一些其他问题。从结构上讲，BP 网络具有输入层、隐藏层和输出层；从本质上讲，BP 算法就是以网络误差平方为目标函数、采用梯度下降法来计算目标函数的最小值。

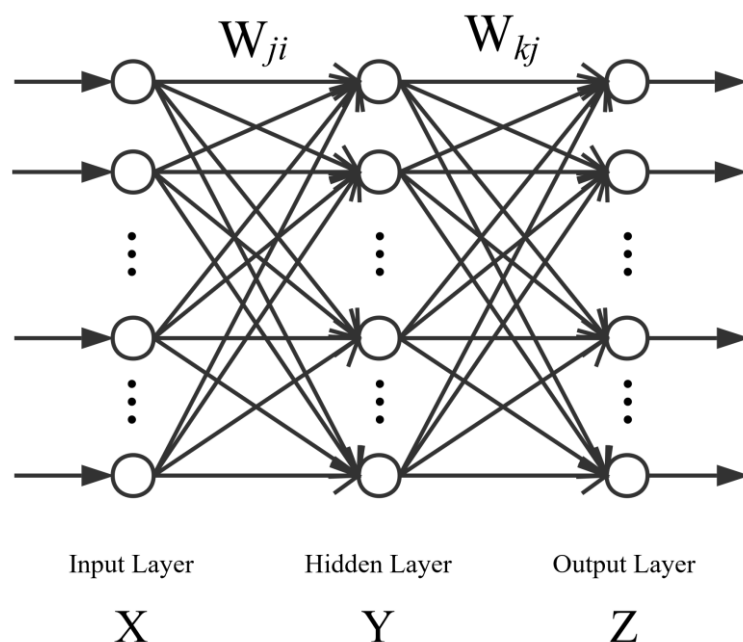


图 5-1 神经网络拓扑结构图



其具有以下几个特点：

(1) 非线性映射性

已有数理理论证明[10]，三层的神经网络就能够以任意精度逼近任何非线性连续函数，具有很强的非线性映射能力。因此，适用于解决内部机制复杂的经济问题。

(2) 自学习、自适应性

在训练过程中，BP 神经网络能通过学习自动提取输入、输出数据间的映射规则，通过梯度下降等方法不断更新权值，将其储存于网络的知识库中，具有高度的自学习与自适应能力。

(3) 泛化性

泛化性指的是通过设计网络能够保证对所需预测对象正确对策及对未见过的模式、噪声污染的模式进行正确的预测的性能。BP 神经网络拥有将训练成果应用于新的、未知的环境的能力。[11]市场经济条件下，由于经济变量的微小变化，都可能衍生出一种全新的预测模型，因此，泛化性之于经济预测模型显得尤为关键。

(4) 容错能力

BP 神经网络的局部神经元受损后，对于全局的训练效果不会造成太大影响。

## 5.2 BP 神经网络模型建立

首先，对数据进行归一化处理， $X$  代表变量  $x_i$  所在的特征数据集，此处共有四个特征，即表 4-4 中的公共因子 1、2、3、4。

$$x_i := \frac{x_i - \min(X)}{\max(X) - \min(X)} \quad (5-1)$$

采用非线性激活函数为 Sigmoid 函数， $W$  代表各特征的权值集合，权重初始值为 0。

$$f(x) = \frac{1}{1 + \exp(-W^T X)} \quad (5-2)$$

本文采用最小二乘代价函数如式(5-3)所示。

$$E_s = \sum_{k=1}^{\alpha_{u+1}} \frac{1}{2} (y_k - t_k)^2 \quad (5-3)$$

其中， $E_s$  为网络在第  $s$  个样本输入下的误差， $s=1,2,3,4$ ； $\alpha_u$  为第  $u$  层神经元的个数； $y_k$  为输出层第  $k$  个神经元的输出； $t_k$  为输出层第  $k$  个神经元的目标值， $k=1,2,\dots,\alpha_{u+1}$ 。

BP 神经网络的核心为误差从输出层逐层反向传播，各层权值通过梯度下降法 (gradient descent algorithm) 进行更新，即：

$$w := w - \eta \nabla E_s(w) \quad (5-4)$$

上式中， $\eta$  是每次更新的步长， $\nabla E_s(w)$  是第  $s$  个样本输入下的输出偏差对某一层权值的偏导数，表示每输入一个样本更新一次参数。

下面以  $w_{ji}^{(u)}$  为例推导梯度项：

$$\frac{\partial E_s}{\partial w_{ji}^{(u)}} = \frac{\partial E_s}{\partial z_j^{(u)}} \frac{\partial z_j^{(u)}}{\partial w_{ji}^{(u)}} = \frac{\partial E_s}{\partial z_j^{(u)}} b_i^{(u-1)} \quad (5-5)$$

其中,  $b_i^{(u-1)}$  为第  $u-1$  层的偏执项, 定义  $\delta_j^{(u)} = \frac{\partial E_s}{\partial z_j^{(u)}}$ , 对于输出层, 可以得

出  $\delta_j^{(u+1)} = y_k - t_k = b_i^{(u+1)} - t_k$ , 则式(5-5)可表示为:

$$\frac{\partial E_s}{\partial w_{ji}^{(u)}} = \frac{\partial E_s}{\partial z_j^{(u)}} \frac{\partial z_j^{(u)}}{\partial w_{ji}^{(u)}} = \delta_j^{(u)} b_i^{(u-1)} \quad (5-6)$$

现在求解  $\delta_j^{(u+1)}$ :

$$\delta_j^{(u)} = \frac{\partial E_s}{\partial z_j^{(u)}} = \sum_{k=1}^{\alpha_{u+1}} \frac{\partial E_s}{\partial z_k^{(u+1)}} \frac{\partial z_k^{(u+1)}}{\partial z_j^{(u)}} \quad (5-7)$$

根据  $\delta$  的定义可知  $\frac{\partial E_s}{\partial z_k^{(u+1)}} = \delta_k^{(u+1)}$ , 带入上式, 则有:

$$\delta_j^{(u)} = \sum_{k=1}^{\alpha_{u+1}} \delta_k^{(u+1)} \frac{\partial z_k^{(u+1)}}{\partial z_j^{(u)}} = \sum_{k=1}^{\alpha_{u+1}} \delta_k^{(u+1)} \frac{\partial z_k^{(u+1)}}{\partial b_j^{(u)}} \frac{\partial b_j^{(u)}}{\partial z_j^{(u)}} \quad (5-8)$$

由  $z_k^{(u+1)}$  和  $b_j^{(u)}$  定义:

$$\frac{\partial z_k^{(u+1)}}{\partial b_j^{(u)}} = w_{kj}^{(u+1)} \quad (5-9)$$

$$\frac{\partial b_j^{(u)}}{\partial z_j^{(u)}} = f'(z_j^{(u)}) \quad (5-10)$$

将式(5-9)和(5-10)代入式(5-8)得:

$$\delta_j^{(u)} = \sum_{k=1}^{\alpha_{u+1}} \delta_k^{(u+1)} w_{kj}^{(u+1)} f'(z_j^{(u)}) = f'(z_j^{(u)}) \sum_{k=1}^{\alpha_{u+1}} w_{kj}^{(u+1)} \delta_k^{(u+1)} \quad (5-11)$$

由此得到了误差从输出层向低层反向传播的递推公式, 进而可以求出误差对于每一层权值的梯度  $\nabla E_s(w)$

### 5.3 模型求解与分析

以表 4-4 中公共因子 1、2、3、4 分别作为四个特征输入, 每年的财政收入作为对应的输出, 将数据集划分为 70%训练集、15%信度测试集, 15%测试集进行训练。经多次调参, 迭代过程如图 5-2 所示。

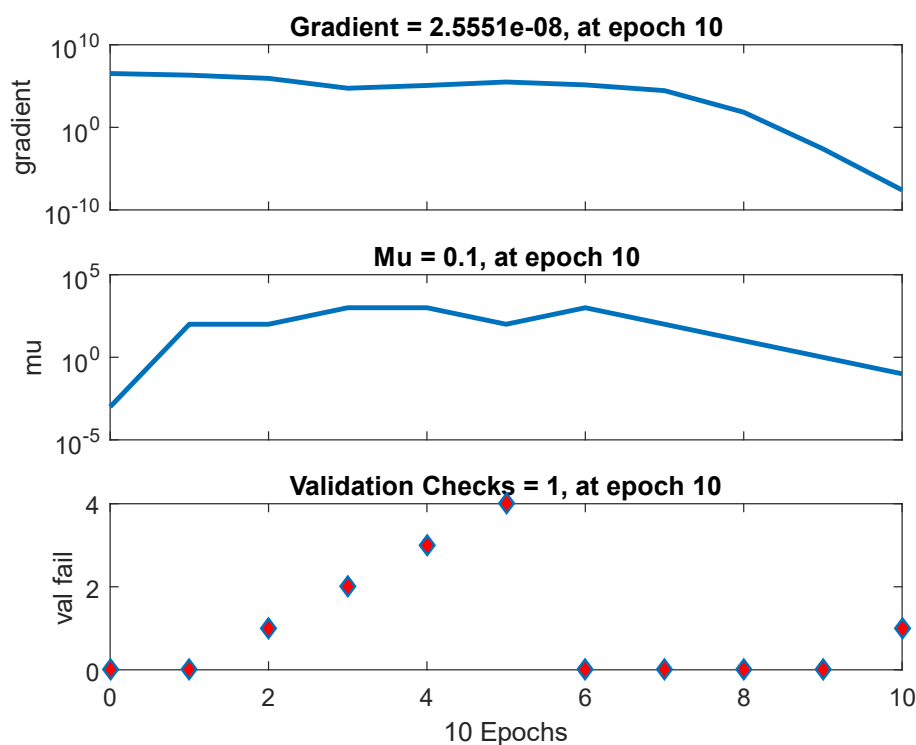


图 5-2 神经网络迭代过程图

可以看出，在迭代 10 次后达到收敛，梯度下降至最优。拟合出的神经网络示意图如图 5-2 所示。

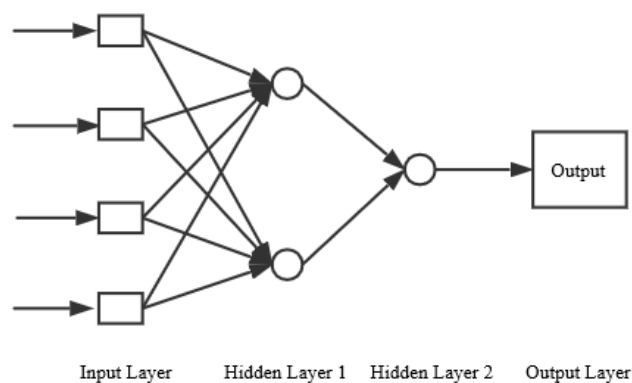


图 5-3 神经网络模型图

该神经网络具有 1 个输入层，2 个隐藏层与 1 个输出层。其中隐藏层 1、2 神经元个数分别为 2、1，其对应的偏置与权重向量分别为：

$$B_1 = [-0.46; 1.92], W_1 = [-2.20, -1.97, -0.24, 1.67; 1.44, 0.84, 3.73, 3.07]$$

$$B_2 = [-0.68736866457646162], W_2 = [-0.80, 0.76]$$

模型拟合优度为 99.46%，其中训练集拟合优度达到 100%，具有非常好的预测能力。

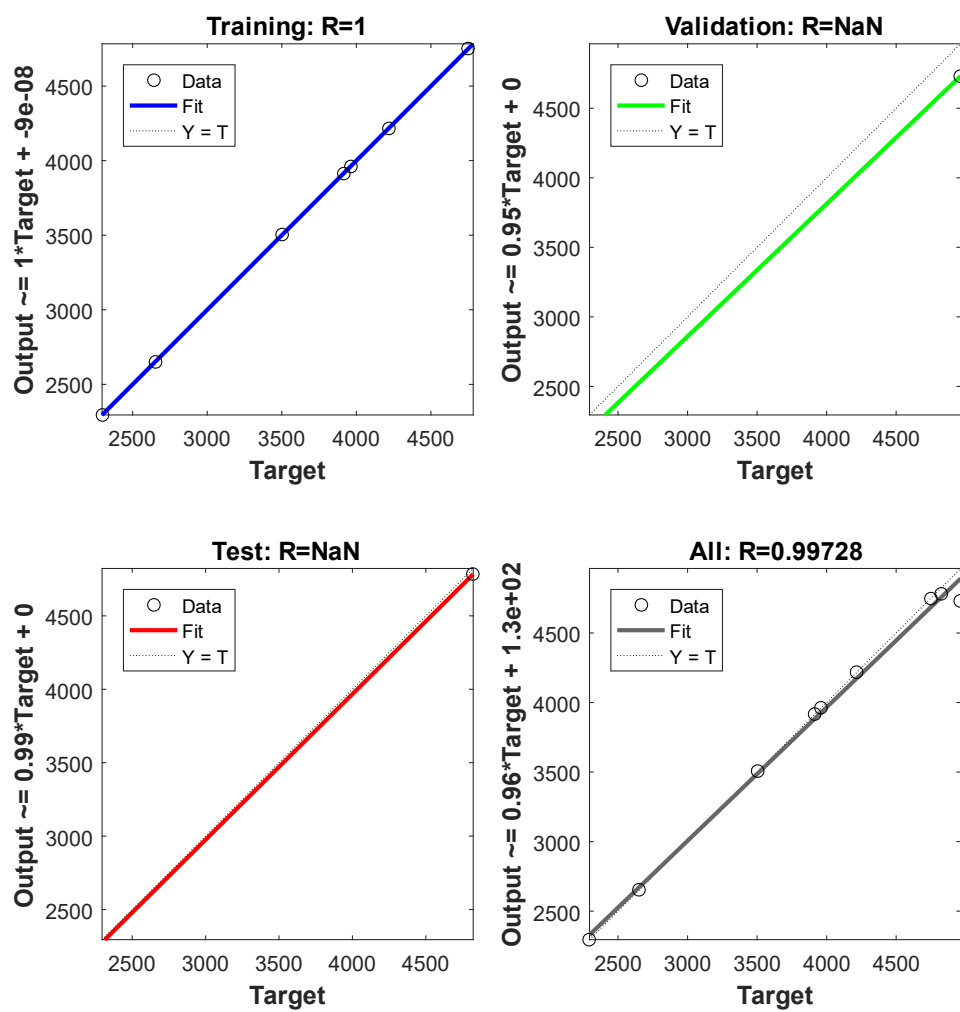


图 5-4 神经网络拟合优度结果图

结果表明：在数据预处理较好的情况下，神经网络有较好的预测财政收入的效果。

## 总结与展望

本文通过随机抽样,发现辽宁省财政收入预算与实际存在收入存在较大差异,挖掘差异原因,通过时间序列分析、因子分析及人工神经网络,建立具有较好时效性与较高准确性的辽宁省财政收入人工神经网络预测模型。

首先,通过时间序列分析,从 1985-2013 年的财政收入预算和实际收入可以看出,辽宁省在该阶段财政收入运行持续向好,但在 2014 年开始出现负增长情况,2015 年财政收入出现大幅下降。总的来说,辽宁省在 1985-2016 年间财政收入呈 8.8%增长的趋势,1991、1994、2007、2010、2011 年增长较快,环比增长速度均在 20%以上,在 1993、2001、2014、2015 年,财政收入出现负增长情况。其中 2008 年前后的大幅增长与次贷危机相关,而 2014 与 2015 年出现较大幅度的负增长与辽宁省财政收入数据挤出水分有关。然后通过多元回归分析,发现财政收入的四项构成中,国有资本经营收入、国有企业计划亏补、预算各项三项与财政收入回归不显著相关。

为了进一步探索影响税收的因子,运用更具有细化数据建立更具有时效性及准确性的模型,接着,采用因子分析方法,对财政收入各项细分数据进行主成分分析,公共因子 1 主要构成有个人所得税等 12 项收入,方差贡献率为 30.947%;公共因子 2 主要构成有营业税等 8 项收入,方差贡献率为 30.924%;公共因子 3 的主要构成有消费税、罚没收入,方差贡献率为 17.892%;构成公共因子 4 的主要有其他税收等 3 项收入,方差贡献率为 15.061%。由因子分析结果可知,公共因子 1、2 涉及的财政收入来源均对财政收入产生正面影响,因此可以主要通过增进就业、增加进出口贸易、刺激消费等措施,同时抑制居民对烟叶的消费,综合扩大内需与稳定外需提高政府财政预算收入。

最后,根据相关得分矩阵对数据进行降维,以降维后的四个公共因子作为输入,各年实际财政收入作为输出。训练得到由 1 个输入层,2 个隐藏层与 1 个输出层。其中隐藏层 1、2 神经元个数分别为 2、1,拟合优度为 99.46%,训练集拟合优度达到 100%,具有非常好的预测能力的人工神经网络。探索出一种具有时效性的财政收入预算模型,以期政府做好合理财政预算,实现财政收支平衡做出贡献,助力辽宁省经济平稳运行。

## 参考文献

- [1] 冯润. 分解预测法在财政收入预测中的应用[J]. 中国财政, 2008(13):51-53.
- [2] 屈丽萍, 毛加强.  $H_{\infty}$ 滤波算法在国家财政收入预测中的应用[J]. 统计与决策, 2008(21):157-159.
- [3] 郑鹏辉, 单锐, 陈静. 时间序列分析在我国财政收入预测中的应用[J]. 重庆高教研究, 2008, 27(2):15-18.
- [4] 舒成. 基层财政收支灰色预测模型及应用[J]. 统计与决策, 2010(3):49-50.
- [5] 范敏, 石为人, 梁勇林,等. 组合预测模型在地方财政收入预测中的应用[J]. 重庆大学学报, 2008, 31(5):536-540.
- [6] 江澜. 免疫进化算法和神经网络在建立财政收入预测模型中的应用[J]. 内蒙古农业大学学报(社会科学版), 2008, 10(5):95-97.
- [7] 张育军. 基于 GRNN 和 BP 神经网络的财政收入预测模型研究[D]. 厦门大学, 2011.
- [8] 赵海华. 基于灰色 RBF 神经网络的多因素财政收入预测模型[J]. 统计与决策, 2016(13):79-81.
- [9] Rumelhart D E, Hinton G E, Williams R J. Learning representations by back-propagating errors.[J]. Parallel Distributed Processing Explorations in the Microstructure of Cognition, 1986, 323(6088):399-421.
- [10] Zhang Y, Wang J. Recurrent neural networks for nonlinear output regulation[J]. Automatica, 2001, 37(8):1161-1173.
- [11] Cohn D, Atlas L, Ladner R. Improving generalization with active learning[J]. Machine Learning, 1994, 15(2):201-221.
- [12] 金秀, 于春海. 统计学 : Statistics[M]. 清华大学出版社, 2014.

## 附录

### 1 BP 神经网络模型源码

```
function [Y,Xf,Af] = myNeuralNetworkFunction(X,~,~)
%MYNEURALNETWORKFUNCTION neural network simulation function.
%
% Generated by Neural Network Toolbox function genFunction, 06-May-2018
18:27:20.
%
% [Y] = myNeuralNetworkFunction(X,~,~) takes these arguments:
%
%   X = 1xTS cell, 1 inputs over TS timesteps
%   Each X{1,ts} = 4xQ matrix, input #1 at timestep ts.
%
% and returns:
%
%   Y = 1xTS cell of 1 outputs over TS timesteps.
%   Each Y{1,ts} = 1xQ matrix, output #1 at timestep ts.
%
% where Q is number of samples (or series) and TS is the number of timesteps.

%#ok<*RPMT0>

% ===== NEURAL NETWORK CONSTANTS =====

% Input 1
x1_step1.xoffset = [132.71944;90.58318471;-110.3269796;-109.7332492];
x1_step1.gain =
[0.0136902624516611;0.0120991532317533;0.0097227101159181;0.037381016877
4469];
x1_step1.ymin = -1;

% Layer 1
b1 = [-0.45652560416222432;1.9184970012572138];
IW1_1 = [-2.2003177724187979 -1.9652855852363482 -0.242046434187805
1.6721531195885651;1.438104130045575 0.84291677959734046
3.7266550701943899 3.0692278656672785];

% Layer 2
b2 = -0.68736866457646162;
```

```
LW2_1 = [-0.79979791215417539 0.75728867606731742];
```

```
% Output 1
```

```
y1_step1.ymin = -1;  
y1_step1.gain = 0.000749445319967701;  
y1_step1.xoffset = 2295.04;
```

```
% ===== SIMULATION =====
```

```
% Format Input Arguments
```

```
isCellX = iscell(X);  
if ~isCellX, X = {X}; end;
```

```
% Dimensions
```

```
TS = size(X,2); % timesteps  
if ~isempty(X)  
    Q = size(X{1},2); % samples/series  
else  
    Q = 0;  
end
```

```
% Allocate Outputs
```

```
Y = cell(1,TS);
```

```
% Time loop
```

```
for ts=1:TS
```

```
    % Input 1
```

```
    Xp1 = mapminmax_apply(X{1,ts},x1_step1);
```

```
    % Layer 1
```

```
    a1 = tansig_apply(repmat(b1,1,Q) + IW1_1*Xp1);
```

```
    % Layer 2
```

```
    a2 = repmat(b2,1,Q) + LW2_1*a1;
```

```
    % Output 1
```

```
    Y{1,ts} = mapminmax_reverse(a2,y1_step1);
```

```
end
```

```
% Final Delay States
```

```
Xf = cell(1,0);
```



```

Af = cell(2,0);

% Format Output Arguments
if ~isCellX, Y = cell2mat(Y); end
end

% ===== MODULE FUNCTIONS =====

% Map Minimum and Maximum Input Processing Function
function y = mapminmax_apply(x,settings)
y = bsxfun(@minus,x,settings.xoffset);
y = bsxfun(@times,y,settings.gain);
y = bsxfun(@plus,y,settings.ymin);
end

% Sigmoid Symmetric Transfer Function
function a = tansig_apply(n,~)
a = 2 ./ (1 + exp(-2*n)) - 1;
end

% Map Minimum and Maximum Output Reverse-Processing Function
function x = mapminmax_reverse(y,settings)
x = bsxfun(@minus,y,settings.ymin);
x = bsxfun(@rdivide,x,settings.gain);
x = bsxfun(@plus,x,settings.xoffset);
end

```

## 2 BP 神经网络输出结果

```

function [y1] = myNeuralNetworkFunction(x1)
%MYNEURALNETWORKFUNCTION neural network simulation function.
%
% Generated by Neural Network Toolbox function genFunction, 06-May-2018
18:30:31.
%
% [y1] = myNeuralNetworkFunction(x1) takes these arguments:
%     x = 4xQ matrix, input #1
% and returns:
%     y = 1xQ matrix, output #1
% where Q is the number of samples.

%#ok<*RPMT0>

```

```

% ===== NEURAL NETWORK CONSTANTS =====

% Input 1
x1_step1.xoffset = [132.71944;90.58318471;-110.3269796;-109.7332492];
x1_step1.gain =
[0.0136902624516611;0.0120991532317533;0.0097227101159181;0.037381016877
4469];
x1_step1.ymin = -1;

% Layer 1
b1 = [-0.45652560416222432;1.9184970012572138];
IW1_1 = [-2.2003177724187979 -1.9652855852363482 -0.242046434187805
1.6721531195885651;1.438104130045575 0.84291677959734046
3.7266550701943899 3.0692278656672785];

% Layer 2
b2 = -0.68736866457646162;
LW2_1 = [-0.79979791215417539 0.75728867606731742];

% Output 1
y1_step1.ymin = -1;
y1_step1.gain = 0.000749445319967701;
y1_step1.xoffset = 2295.04;

% ===== SIMULATION =====

% Dimensions
Q = size(x1,2); % samples

% Input 1
xp1 = mapminmax_apply(x1,x1_step1);

% Layer 1
a1 = tansig_apply(repmat(b1,1,Q) + IW1_1*xp1);

% Layer 2
a2 = repmat(b2,1,Q) + LW2_1*a1;

% Output 1
y1 = mapminmax_reverse(a2,y1_step1);
end

```

```
% ===== MODULE FUNCTIONS =====
```

```
% Map Minimum and Maximum Input Processing Function
```

```
function y = mapminmax_apply(x,settings)
y = bsxfun(@minus,x,settings.xoffset);
y = bsxfun(@times,y,settings.gain);
y = bsxfun(@plus,y,settings.ymin);
end
```

```
% Sigmoid Symmetric Transfer Function
```

```
function a = tansig_apply(n,~)
a = 2 ./ (1 + exp(-2*n)) - 1;
end
```

```
% Map Minimum and Maximum Output Reverse-Processing Function
```

```
function x = mapminmax_reverse(y,settings)
x = bsxfun(@minus,y,settings.ymin);
x = bsxfun(@rdivide,x,settings.gain);
x = bsxfun(@plus,x,settings.xoffset);
end
```

## 致谢

正如著名统计学家 C.R.Rao 所说：在理性的世界里，所有的判断都是统计学。经过十周统计学课程的学习与近一月的统计学实践，当报告进行到这一步的时候，意味着文章进入了尾声，我们的学习生涯又将再一次到达一个新的节点。

回首这短暂而充实的十周，从理论学习到实证研究，每一步都因有老师和同学，特别是队友的陪伴而弥足珍贵。

首先，对任课教师金秀老师道声感谢。从课程伊始，您就带领我们领略了统计学的魅力。通过课程学习，我们了解到统计学是研究总体在一定条件下的数量特征及其规律性的方法论的学科，统计学也是检验经济学的理论是否符合客观事物的发展规律的重要工具。学习统计学，我们不能零散的学习，而是需要从系统的，比较的角度来学习。

其次，感谢与我一同学习成长的同学们。在案例讨论中，我感受到了基于不同想法的多样性，在思维的碰撞中寻找正确解决方案；从课程互动问答中，我触摸到一个个真实的、奋发向上的灵魂……课程的学习不仅仅是向老师学习，更是向身边的同学们学习。感谢同学们一起营造的令人满意的学习环境，我为自己有幸能在这样一个集体中学习而感到荣幸。

最后，感谢东北大学开设诸如统计学等课程，并安排金秀等对待学术严谨、对待教育认真、对待学生负责的老师教授，为我们提供了一个资源丰富、全面发展的平台。

这里的句点不是终止，而是刚刚开始，愿一路上，常怀一颗感恩、向上、钻研、进取的心，以更加昂扬的斗志，去迎接更美好的未来，让青春之花绽放在祖国最需要的地方！

2018 年 5 月 9 日  
于东北大学书