

# 第三届“泰迪杯”

## 全国大学生数据挖掘竞赛

### 优秀作品

作品名称：基于数据挖掘技术的市财政收入分析预测模型

荣获奖项：一等奖

作品单位：广东金融学院

作品成员：何耀明 林凯斌 胡慧宇

指导教师：骆世广

## 基于稳定标准选取的广州财政收入预测优化模型

**摘要：**地方财政收入是区域国民经济的综合反映，也是市场经济国家的政府进行宏观调控的基础。合理地预测地方财政收入，为下一年的政策提供指导依据，是一个具有重大意义的问题。以广州财政收入为例，首先结构化财政总收入，再依次使用灰色关联度和灰色预测模型找出影响各项财政收入的关键因素。利用各项财政收入与其对应的关键因素建立 BP 神经网络模型，将各因素的灰色预测值代入模型中求得各项财政收入，汇总各项财政收入得到 2015 年财政总收入为 25533405 万元，增长率为 10.3%。最后根据实证结果进行数据分析，并结合广州近几年财政收支情况和当前经济发展现状，向广州市财政局提出建议。本文创新之处在于关键影响因素的选取，在灰色预测模型中采取不同的残差稳定标准即可得到不同的关键影响因素，尝试采用不同的关键影响因素建立 BP 神经网络模型，结合 2014 年实际值对结果进行修正，从而得出较优的预测模型。

**关键词：**灰色关联度，灰色预测，稳定标准，BP 神经网络

## The optimization model of forecast of Guangzhou revenue based on stability criteria selection

**Abstract:**The local financial revenue is the comprehensive reflection of the regional national economy, and foundation of Marco control of market economy country.It is significant to reasonably forecast the local revenue and provide the guidance for the next year. Guangzhou, for example, its finance revenue is structuralized, and it finds out the key factor of financial revenue through The Grey Relational Analysis and Gray Prediction Model.Byusing the financial income and its corresponding key factors to establish the BP neural networks model, putting the grey predictive value of each factor into model to know financial income. Itcollects financial income and know that its fiscal total revenue in 2015 is 255334.05 million yuan and the growth rate is 10.3%. Finally, according to the empirical results, data analysis, combined with the financial revenue and expenditure in Guangzhou in recent years and the current economic development situation, the Financial Bureau of Guangzhou to propose.Innovation of this paper is to choose the key factors, to take a different residual stability criteria in the gray forecast model to get a different key factors.Try using different key factors to establish BP neural network model, combined with the actual value of the result in 2014 be amended so as to arrive on better predictive models.

**Key words:**Grey Relational Analysis ,Grey prediction ,stability criteria ,Back Propagation

## 目 录

<b>1. 挖掘目标</b> .....	<b>1</b>
<b>2. 分析方法与过程</b> .....	<b>1</b>
2.1. 总体流程 .....	1
2.2. 具体步骤 .....	1
2.3. 结果分析 .....	8
<b>3. 结论</b> .....	<b>19</b>
3.1. 关键影响因素 .....	19
3.2. 财政收入预测 .....	20
3.3. 分析建议 .....	21
<b>4. 参考文献</b> .....	<b>22</b>

“泰迪杯”

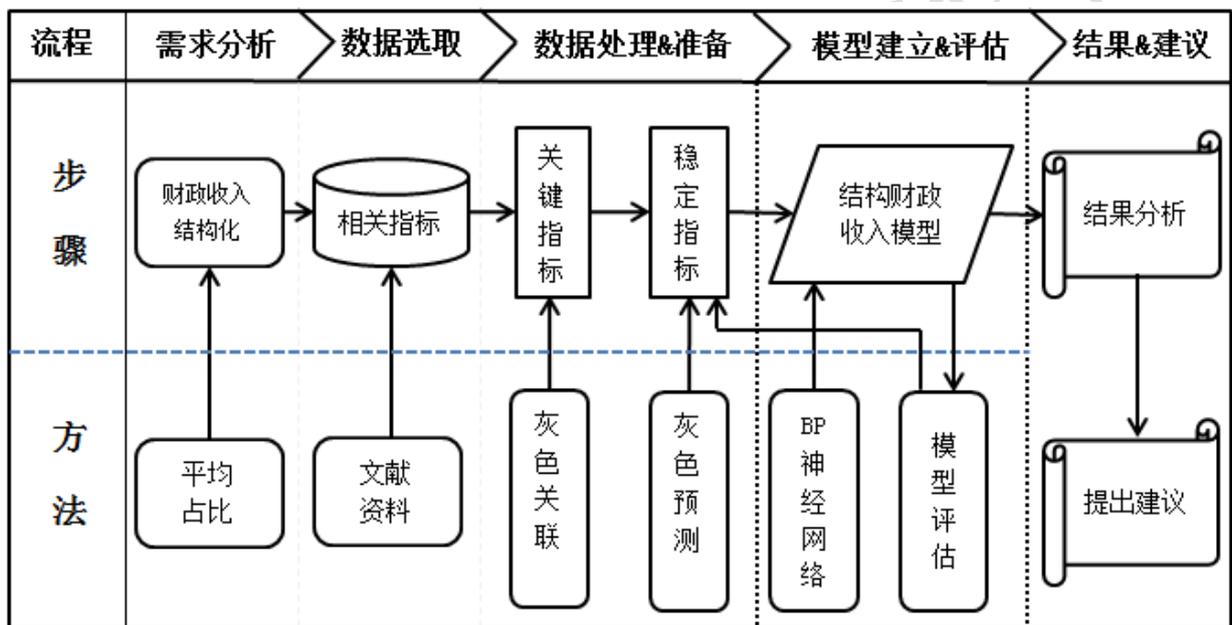
优秀作业

# 1. 挖掘目标

本次建模目标是利用往年的经济指标和非经济指标构建广州财政总收入模型。分析各项财政收入相关联指标，识别出影响各项财政收入的关键因素，根据各项财政收入和其对应的关键因素建立模型，最终得到广州财政总收入模型，为下一年有效财政收入预算做出预测，从而为下一年的政策提供指导。

## 2. 分析方法与过程

### 2.1. 总体流程



主要包括如下步骤：

步骤一：关键影响因素选取。包括财政收入结构化、相关指标选取、关键指标选取、稳定指标选取，最终使用不同稳定标准得出不同的稳定指标作为关键影响因素。

步骤二：结构化财政收入模型建立与评估。使用不同的关键影响因素建立不同的财政收入模型，并对结果进行评估比较。

步骤三：结果分析和提出建议。针对预测结果进行分析，并向广州市政局提出建议。

### 2.2. 具体步骤

#### 2.2.1 关键影响因素选取



图一 关键影响因素选取流程

### 2.2.1.1 财政收入结构化

财政总收入包括一般预算收入和政府性基金收入，若直接对财政总收入进行预测，往往造成的误差较大，所以考虑将财政总收入结构化，再对各个收入结构细化分析。

#### ■ 一般预算收入结构化

利用往年一般预算收入各类税收（个税）数据，计算个税近 15 年（1999-2013 年）来占一般预算收入平均比重，将一般预算收入分为主体税种和其它预算收入。

#### ■ 政府性基金收入结构化

查阅广州经济决算草案得知，2007 年开始政府将土地出让金列入政府性基金收入中，且占比例达八成。故将政府性基金按照组成结构分为土地出让金收入、其它政府性基金收入两类。

### 2.2.1.2 相关指标选取

由步骤一可知构成财政总收入的收入结构包括：营业税、增值税、企业所得税、个人所得税、契税、房产税、其它预算收入、其它政府性基金收入和土地出让金收入。

#### ■ 主体税种

查找相关资料，对营业税、增值税、企业所得税、个人所得税、契税、房产税的征收范围进行研究，再参考相关文献，汇总出这几种主体税收的相关影响指标。

#### ■ 其它预算收入

采用主要经济指标进行衡量，如地区生产总值，三大产业增加值等。

#### ■ 其它政府性基金收入

依据政府性基金纳入财政预算的 13 项指标（包括政府性基金：养路费、车辆购置附加费、铁路建设基金、电力建设基金、三峡工程建设基金、新菜地开发基金、公路建设基金、民航基础设施建设基金、农村教育费附加、邮电附加、港口建设费、市话初装基金、民航机场管理建设费），将其分为四项：交通建设基金、水电设施建设基金、教育

费附加基金、邮电通信建设基金，查找与这四项基金相关的指标，得出其它政府性基金收入的相关指标。

■ 土地出让金收入

土地出让金系指政府将土地使用权出让给土地使用者，按规定向受让人收取的土地出让的全部价款（指土地出让的交易总额）。由于土地出让金收入在 2007 年才开始纳入政府性基金收入预算中，且该项收入受政策影响较大，存在一定的财务杠杆效应，所以在以下的研究中将土地出让金收入和其它财务收入结构分开说明。

■ 数据来源与数据选取

材料给出了部分数据指标，对于部分不完整数据作剔除处理。从“广州统计信息网”获取其它相关指标数据，由于统计数据缺失，仅能获得 1999 年-2013 年相关指标的完整数据，利用 1999 年-2013 年相关指标数据建立模型预测 2014 与 2015 年的财政总收入，将 2014 年的预测值作为模型的实证检验。

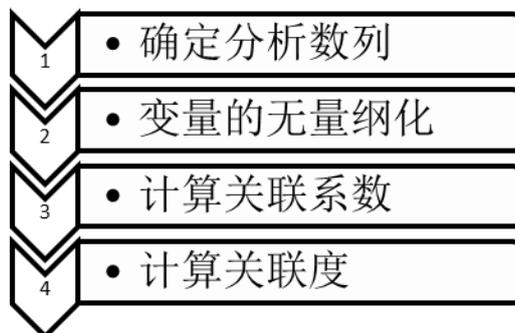
2.2.1.3 关键指标选取、稳定指标选取

(1) 关键指标选取（灰色关联度模型）

■ 问题分析

灰色关联分析方法优点在于样本容量可以少到 4 个，对数据无规律同样适用，不会出现量化结果与定性分析结果不符的情况，与需求契合。其基本思想是将评价指标原始观测数进行无量纲化处理，计算关联系数、关联度以及根据关联度的大小对待评指标进行排序。应用涉及社会科学和自然科学的各个领域，尤其在社会经济领域。

■ 灰色关联度模型



图二 灰色关联度分析步骤

目前关联度量化模型很多，如邓氏关联度、绝对关联度、T 型关联度等，这些量化模型不具备无量纲化后的保序效应；因此，本文采用改进灰色斜率关联度分析方法进行研究，基本思想：按照因素时间序列曲线的平均相对变化态势的接近程度来计算灰色关联度。

设参考数列为  $X_0$ ，因素数列为  $X_i(i=1,2,\dots,N)$  且

$$X_0 = \{x_0(1), x_0(2), \dots, x_0(n)\}, X_i = \{x_i(1), x_i(2), \dots, x_i(n)\}$$

则  $X_0$  与  $X_i$  的灰色关联度  $r(X_0, X_i)$  定义为：

$$r(X_0, X_i) = \frac{1}{n} \sum_{k=1}^n (x_0(k), x_i(k))$$

其中：

$$r(x_0, x_i) = \frac{1}{n-1} \sum_{k=1}^{n-1} \frac{1}{1 + \left| \frac{a^{(1)}(x_0(k+1))}{x_0(k+1)} - \frac{a^{(1)}(x_i(k+1))}{x_i(k+1)} \right|}$$

$$a^{(1)}(x_0(k+1)) = x_0(k+1) - x_0(k), k = 1, 2, \dots, n-1$$

$$a^{(1)}(x_i(k+1)) = x_i(k+1) - x_i(k), k = 1, 2, \dots, N$$

将所有  $N$  个序列灰色关联度  $r(x_0, x_i)$  自大而小顺序排列，得到关联度序集，以此判断数列  $X_0$  与  $X_i$  的相关大小。

采用 1999-2013 年的 15 组财政收入作为参考数列，相关指标作为因素数列进行关联度分析，利用 MATLAB 进行编程，得出各项财政收入与各个指标的关联度。

## (2) 稳定指标选取（灰色预测模型）

### ■ 问题分析

通过上步处理，得到各项财政收入的关键指标，理论上应该直接利用所有的关键指标作为因变量来构建各项财政收入模型，但考虑到不是所有的关键指标都能够稳定预测，所以在预测各项财政收入关键指标的未来值时，应该剔除不能稳定预测的关键指标，得到稳定的关键指标，再利用稳定的关键指标与各项财政收入建立模型。

灰色预测是通过原始数据的处理和灰色模型的建立，发现、掌握系统发展规律，对系统的未来状态做出科学的定量预测。灰色预测模型具有预测所需的数据量较少、预测

比较准确、精度较高、样本分布不需要有规律性、计算简便、检验方便和适用于中长期预测的优点，满足目前的背景需求。利用灰色预测的均差检验可以设定不同的稳定标准，从而可以通过修改稳定标准来不断修正最终模型达到最优化。

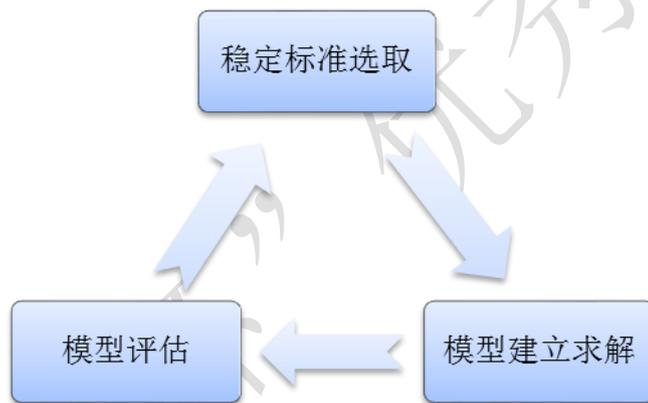
■ 灰色预测

将所有关键指标作为目标变量，采用 1 个变量 1 阶方程的灰色预测模型（GM(1,1) 模型），通过编写 MATLAB 程序实现灰色预测并得到预测值（理论见“附件 3”）。

■ 灰色模型检验

GM(1,1)模型的检验包括残差检验（误差均值检验）、关联度检验、后验差检验三种形式。残差检验数用 $\rho$ 表示、关联度检验数用 $\gamma$ 、后验差检验数用 $C$ 表示，在上述模型中需要满足 $\gamma > 0.6$ 、 $C < 0.35$ 才通过灰色模型检验， $\rho$ 可根据需求自行确定。

2.2.2 结构化财政收入模型建立与评估



图三 模型建立与优化流程

结合 BP 神经网络具有能拟合任意的非线性函数和自身强大的自学功能，使用该模型对各项财政收入与其对应关键指标进行建模求解。

■ 数据选取

本文中采用 1999-2013 年这 15 组各项财政收入及其对应稳定指标的实际数据，其中 12 组数据作为训练样本，3 组数据作为测试样本（随机抽取），输入项为 2014 年和 2015 年的稳定指标的灰色预测值，输出项为 2014 年和 2015 年各项财政收入。

■ 数据预处理

输入数据的单位不一样，有些数据的范围较大，导致的结果是神经网络收敛慢、训练时间长。为加快训练速度，我们将数据集按如下公式进行归一化处理，使得归一化后的数据分布在[-1, 1]区间中：

$$y = \frac{2 \times (x_i - \min(x_i))}{\max(x_i) - \min(x_i)} - 1$$

MATLAB 中函数“premnmx”可实现了上述归一化方式。

### ■ BP 神经网络模型建立

使用 Matlab 建立 BP 神经网络主要会使用到下面 3 个函数：

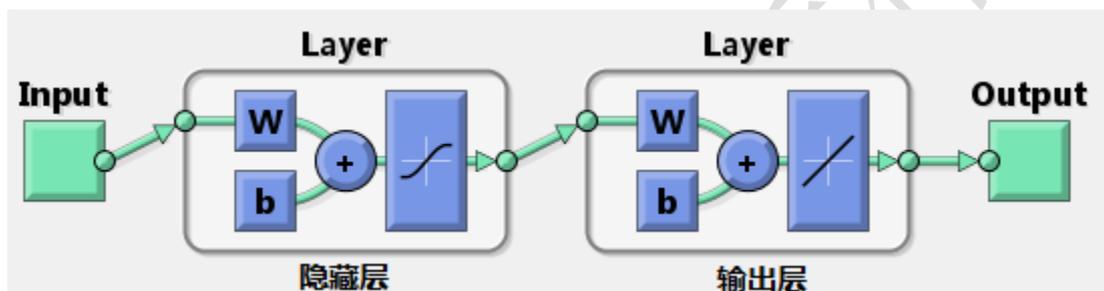
newff：前馈网络创建函数

train：训练一个神经网络

sim：使用网络进行仿真

#### ① 一般结构

建立 input—N—output 的 BP 网络结构，其中，input 表示输入项；N 为隐藏层神经元个数；output 表示输出项。一般结构图如图四：



图四 BP神经网络结构示意图

#### ② 传输函数的确定

隐藏层传输函数选择双曲正切 S 形函数：

$$f(x) = \frac{2}{1 + e^{-2x}} - 1, \quad (-1 < f(x) < 1)$$

输出层传输函数采用线性函数：

$$f(x) = x$$

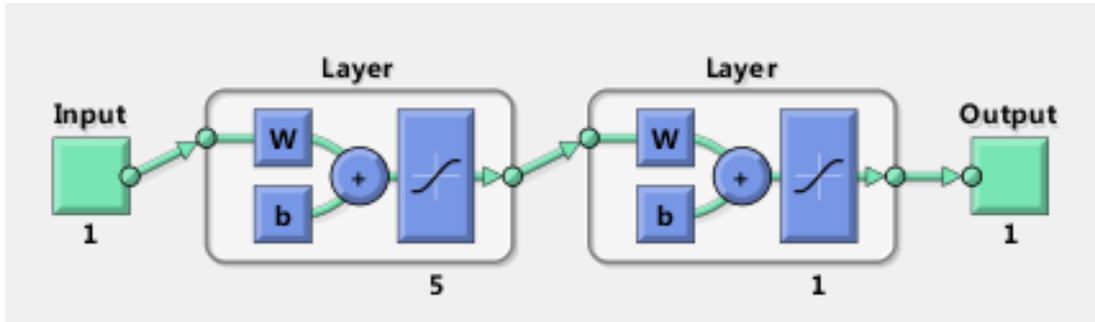
分别使用 MATLAB 神经网络工具箱中函数“tansig”和“purelin”实现传输函数的设定。

#### ③ 隐藏层的确定

隐藏层神经元个数对 BP 神经网络预测精度有显著的影响，节点数太少，网络不能很好地学习，需要增加训练次数，训练的精度也受影响；节点数太多，训练时间增加，网络容易过拟合。

有关隐含层单元的个数选取的理论尚不成熟，需要由实际问题的复杂度和规模所决定。在进一步的测试中，我们发现当两层隐含层分别设置 5 个和 1 个单元时，神经网络

训练的收敛速度比较快，于是我们把这个因果关系神经网络的各层神经元个数设为：“input-5-1-output”，最终确定神经网络结构示意图如图五：



图五 实际建立的BP神经网络结构示意图

#### ④ BP 神经网络的训练

BP 神经网络的采用梯度修正法作为权值和阈值的学习算法,从网络预测误差的负梯度方向修正权值和阈值,没有考虑以前经验的积累,学习过程收敛缓慢。对此本文采用训练算法——共轭梯度算法,优点在于将模值信赖域算法与共轭梯度算法结合起来,减少用于调整方向时搜索网络的时间。

matlab 神经网络工具箱中函数“trainscg”即实现了上述训练算法。

学习速度同样对 BP 神经网络具有重要影响作用，学习速度太小，网络学习缓慢，需要增加训练次数；学习速度太大，网络学习迅速，但是容易导致网络不收敛，影响训练的精度。我们最终确定训练的目标误差为 $10^{-3}$ ，训练步数为 500，学习率为 0.01。

#### ■ 模型求解

以营业税为例，利用BP神经网络模型对2014年和2015年营业税进行预测，数据选取如表1（加粗数据为灰色预测值）：

表 1 营业税数据表

年份	营业税	第三产业增加值	公路客运量	限额以上餐饮业主营业务收入	建筑业总产值	建筑业增加值	全社会房地产开发投资额
计量单位	万元	万元	万人次	万元	万元	万元	万元
1999	433360	11152545	18678	349730	2470523	1330484	2959027
2000	479698	13767475	19964	409725	2561326	1436406	3555816
2001	540075	16320762	21057	522331	3403870	1568267	3870207
2002	613161	18895479	23157	601220	3733922	1603966	4263898
2003	650119	21627825	23839	680647	4785787	1718007	4194806
2004	793520	25453413	27845	859141	5459314	1939100	4770315
2005	892678	29787941	30782	853315	6331382	2012633	5080846
2006	1027971	35118425	33021	1020793	6870406	2145067	5567893
2007	1235374	41646681	38949	1166412	7507109	2228495	7038031
2008	1279793	48903250	42359	1516500	8754491	2553936	7634024

2009	1516049	55607710	43921	1476434	10134050	2878166	8173449
2010	1777343	65574525	47296	1694029	12805288	3573047	9836582
2011	1625593	76419207	51107	1988868	15613171	4363837	13053605
2012	1747616	86167948	58875	2332653	17417072	4564947	13704511
2013	1623520	99643373	70815	2412402	21828895	4725256	15724289
<b>2014</b>		<b>118049333.4</b>	<b>72783.183</b>	<b>2904350.5</b>	<b>23746699</b>	<b>5319885.1</b>	<b>17081117</b>
<b>2015</b>		<b>137165274.4</b>	<b>80218.691</b>	<b>3300330.2</b>	<b>27870540</b>	<b>5919520.1</b>	<b>19476623</b>

调用 MATLAB 中“net”函数与“train”函数进行 BP 神经网络的建立与训练,为了得到最优的预测结果,以测试误差百分比小于 10%作为循环训练结束条件。

### 2.3. 结果分析

#### 2.3.1 关键影响因素选取结果

##### 2.3.1.1 财政收入结构化结果

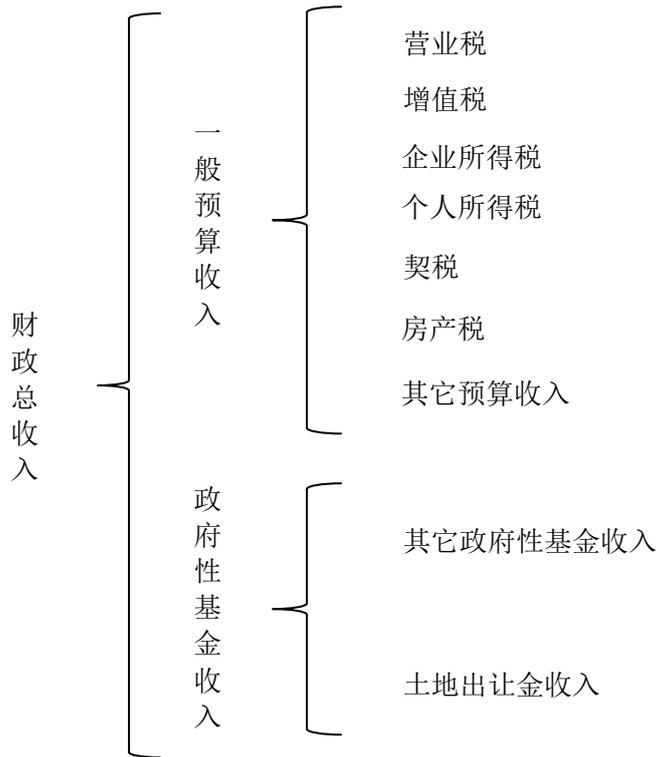
利用往年一般预算收入各类税收(个税)数据,计算个税近 15 年(1999-2013)来占一般预算收入平均比重如表 2:

表 2 税收平均比重

个税	比重	个税	比重
营业税	0.217	专项收入	0.036
增值税	0.192	罚没收入	0.035
企业所得税	0.118	印花税	0.018
城市维护建设税	0.066	国有资本经营收入	0.015
个人所得税	0.061	城镇土地使用税	0.013
契税	0.059	其它收入	0.009
房产税	0.051	车船税	0.006
行政事业性收费收入	0.049		

选取比重大于 0.05 的个税作为主体税种。由于城市维护建设税是由增值税、营业税和消费税决定,故在此不将它作为主体税种,仅将营业税和增值税作为主体税种。则可将一般预算收入分为主体税种(包括营业税、增值税、企业所得税、个人所得税、房产税、契税、房产税)、其它预算收入。

又政府性基金收入可分为占比较多的土地出让金收入和其它政府性基金收入,则可得财政总收入结构图如图六:



图六 财政总收入结构图

### 2.3.1.2 相关指标选取结果

汇总各项财政收入结构的相关指标如表 3:

表 3 各项财政收入相关指标

财政收入结构	相关指标	财政收入结构	相关指标
营业税	公路货运量	契税	商品房销售额
	公路客运量		商品房销售额#住宅
	建筑业增加值		商品房销售面积
	第三产业增加值		商品房销售面积#住宅
	全社会房地产开发投资额		房地产开发投资额
	全社会住宅投资额		房地产开发投资额#住宅
	地方财政收入		房地产开发房屋施工面积
	建筑业总产值		房地产开发房屋施工面积#住宅
	住宿和餐饮业零售额		建筑业增加值
	邮电业务收入		城镇居民储蓄存款余额
	限额以上餐饮业主营业务收入		总人口
增值税	商品进口总值	其它预算收入	建筑业企业利润总额
	地区生产总值		全社会固定资产投资额
	工业增加值		第二产业增加值
	批发零售业零售额		住宅竣工面积
	工业增加值占 GDP		建筑业总产值
	进出口总值		社会消费品零售总额
	公共财政支出		地方财政一般预算收入
	城市商品零售价格指数		地方财政一般预算支出
	城市居民家庭人均可支配收入		地区生产总值

企业所得税	社会消费品零售总额	其它基金收入	第一产业增加值
	商品销售总额		第二产业增加值
	第二产业增加值		第三产业增加值
	第三产业增加值		人均地区生产总值
	全社会固定资产投资额		社会从业人数
	城市商品零售价格指数		全社会固定资产投资额
	规模以上工业企业盈亏相抵后的利润总额		城市居民消费价格指数
	建筑业总产值		城市居民家庭人均可支配收入
	建筑业企业利润总额		社会消费品零售总额
	地方财政总收入		商品进出口总值
	规模以上工业销售产值		道路面积
	规模以上国有工业企业全员劳动生产率		全社会车辆数
	规模以上工业增加值		全社会货运量
	商品销售总额		发电量
个人所得税	城市居民年人均可支配收入	白云国际机场货邮吞吐量	
	城镇单位职工年平均工资	白云国际机场旅客吞吐量	
	城镇居民储蓄存款余额	邮电业务收入	
	地区生产总值	城市供水综合生产能力	
	第二产业增加值	普通高等学校招生数	
	城镇非私营单位从业人数	住宅竣工面积	
	地方财政收入	全社会货运量	
	城市居民消费价格指数	建筑工程投资额	
	城镇居民现金消费支出	毕业生数	
	非农业人口	全社会公路货运量	
	总人口	城市供水综合生产能力	
	城镇人口百分比	用水普及率	
	公共财政支出	地区生产总值	
	全社会住宅投资额	全社会固定资产投资额	
房产税	城市居民年人均可支配收入	第三产业增加值	
	农村居民年人均纯收入	建筑业总产值	
	平均每人居住面积	公共财政支出	
	全社会房地产开发投资额		
	地区生产总值		
	商品房销售额		
	总人口		
	城镇居民储蓄存款余额		

则可分别建立 8 项财政收入与其相关指标的模型。从“广州统计信息网”获取 1993 年-2013 年相关指标的数据，以营业税部分数据为例，如表 4（全部数据见“附件 1——初始数据”）：

表 4 营业税及其部分相关指标 1999-2013 年数据

年份	营业税	建筑业增加值	第三产业增加值	全社会房地产开发投资额	全社会住宅投资额	地方财政收入	限额以上餐饮业主营业务收入
计量单位	万元	万元	万元	万元	万元	万元	万元
1999	433360	1330484	11152545	2959027	2878473	1881388	349730
2000	479698	1436406	13767475	3555816	3250326	2199077	409725
2001	540075	1568267	16320762	3870207	3316894	2719058	522331
2002	613161	1603966	18895479	4263898	3457617	2690984	601220

2003	650119	1718007	21627825	4194806	3522168	3005475	680647
2004	793520	1939100	25453413	4770315	3712961	3384477	859141
2005	892678	2012633	29787941	5080846	3777003	4088545	853315
2006	1027971	2145067	35118425	5567893	3783416	4767231	1020793
2007	1235374	2228495	41646681	7038031	5041090	8389925	1166412
2008	1279793	2553936	48903250	7634024	5398216	8431405	1516500
2009	1516049	2878166	55607710	8173449	5246903	11076649	1476434
2010	1777343	3573047	65574525	9836582	5727122	13991612	1694029
2011	1625593	4363837	76419207	13053605	8116313	15351387	1988868
2012	1747616	4564947	86167948	13704511	8626775	15796804	2332653
2013	1623520	4725256	99643373	15724289	9969708	20881374	2412402

### 2.3.1.3 关键指标选取、稳定指标选取

#### (1) 关键指标选取结果

在 MATLAB 中编程实现各项财政收入与各个指标的关联度检测（程序见“附件 2——关键指标”）。以营业税为例进行关键指标选取，结果见表 5:

表 5 营业税相关指标的灰色关联度

营业税	第三产业增加值	公路货运量	公路客运量	全社会房地产开发投资额	建筑业增加值	住宿和餐饮业零售额
关联度	0.9392	0.93712	0.93644	0.93088	0.93063	0.92816
营业税	限额以上餐饮业主营业务收入	建筑业总产值	地方财政收入	全社会住宅投资额	邮电业务收入	
关联度	0.92025	0.9147	0.91281	0.90626	0.89577	

由表 5 可知相关指标对营业税的影响大小，从中看出营业税对邮电业务收入这一指标的依赖性明显弱于其它，因此我们只取关联度较高的前 10 个相关指标作为关键指标。

分别对各项财政收入进行相关指标的灰色斜率关联分析（详见表“附件 1——关键指标——关联度表”），可得到影响财政收入的全部关键指标，汇总结果见表 6:

表 6 关键指标表

财政收入	关键指标	财政收入	关键指标
营业税	第三产业增加值	契税	全社会固定资产投资额
	公路货运量		第二产业增加值
	公路客运量		商品房销售额
	全社会房地产开发投资额		建筑业总产值
	建筑业增加值		商品房销售额 #住宅
	住宿和餐饮业零售额		房地产开发投资额
	限额以上餐饮业主营业务收入		城镇居民储蓄存款余额
	建筑业总产值		建筑业增加值
	地方财政收入		商品房销售面积
	全社会住宅投资额		总人口
	地区生产总值		房地产开发房屋施工面积

增值税	社会消费品零售总额	其它 预算 收入	商品房销售面积 # 住宅
	批发零售业零售额		房地产开发投资额# 住宅
	公共政府支出		社会消费品零售总额
	工业增加值		社会消费品零售总额
	城市居民家庭人均可支配收入		地方财政一般预算收入
	进出口总值		地方财政一般预算支出
	商品进口总值		第三产业增加值
	商品销售总额		地区生产总值
企业 所得 税	第二产业增加值	其它 政府 性基 金收 入	商品进出口总值
	规模以上工业销售产值		城市居民家庭人均可支配收入
	第三产业增加值		第二产业增加值
	商品销售总额		人均地区生产总值
	建筑业总产值		全社会固定资产投资额
	地方财政总收入		第一产业增加值
	全社会固定资产投资额		第三产业增加值
	建筑业企业利润总额		建筑业总产值
个人 所得 税	城市商品零售价格指数	其它 政府 性基 金收 入	地区生产总值
	规模以上工业企业盈亏相抵后的利润总额		全社会货运量
	城镇人口百分比		全社会固定资产投资额
	地区生产总值		公共财政支出
	公共财政支出		普通高等学校招生数
	城镇单位职工年平均工资		城市供水综合生产能力
	城镇居民现金消费支出		白云国际机场旅客吞吐量
	城市居民年人均可支配收入		道路面积
房产 税	第二产业增加值		邮电业务收入
	地方财政收入		全社会车辆数
	城镇居民储蓄存款余额		
	地区生产总值		
	城镇居民储蓄存款余额		
	建筑业总产值		
	城市居民年人均可支配收入		
	农村居民年人均纯收入		
全社会房地产开发投资额			
平均每人居住面积			
全社会住宅投资额			

## (2) 稳定指标选取结果

### ■ 灰色预测

下面分别以地区生产总值、城市供水综合生产能力、全社会住宅投资额的预测结果作为例子进行分析。

在 MATLAB 中编写程序（程序见“附件 2：稳定指标”），分别得出结果见表 7、8、9：

表 7 地区生产总值灰色预测结果

地区生产总值			
年份	实际值	预测值	误差
1999	21391758	21391758	0.000
2000	24927434	26673356	0.070
2001	28416511	30614028	0.077
2002	32039616	35136889	0.097
2003	37586166	40327948	0.073
2004	44505503	46285925	0.040
2005	51542283	53124122	0.031
2006	60818614	60972583	0.003
2007	71403223	69980560	0.020
2008	82873816	80319359	0.031
2009	91382135	92185594	0.009
2010	107482828	105804924	0.016
2011	124234390	121436349	0.023
2012	135512072	139377133	0.029
2013	154201434	159968454	0.037

表 8 城市供水综合生产能力灰色预测结果

城市供水综合生产能力			
年份	实际值	预测值	误差
1999	350.00	350.00	0.000
2000	355.00	533.31	0.502
2001	578.00	546.72	0.054
2002	572.00	560.47	0.020
2003	552.00	574.56	0.041
2004	647.00	589.01	0.090
2005	668.74	603.82	0.097
2006	694.00	619.00	0.108
2007	666.30	634.56	0.048
2008	640.30	650.52	0.016
2009	684.30	666.88	0.025
2010	684.30	683.64	0.001
2011	702.70	700.83	0.003
2012	703.70	718.45	0.021
2013	663.50	736.52	0.110

表 9 全社会住宅投资额灰色预测结果

全社会住宅投资额			
年份	实际值	预测值	误差
1999	2878473	2878473	0.00
2000	3250326	2392308	0.26
2001	3316894	2649972	0.20
2002	3457617	2935388	0.15
2003	3522168	3251544	0.08

2004	3712961	3601752	0.03
2005	3777003	3989679	0.06
2006	3783416	4419388	0.17
2007	5041090	4895379	0.03
2008	5398216	5422637	0.00
2009	5246903	6006683	0.14
2010	5727122	6653634	0.16
2011	8116313	7370264	0.09
2012	8626775	8164080	0.05
2013	9969708	9043393	0.09

同理对其它关键指标进行灰色预测，得到的结果见“附件 1——稳定指标——预测结果”。

### ■ 灰色模型检验

得出以上三种指标的残差检验（误差均值检验） $\emptyset$ 、关联度检验 $\gamma$ 、后验差检验 $C$ 分别为：

$$\emptyset_{\text{地区生产总值}} = 0.037, \gamma_{\text{地区生产总值}} = 0.619, C_{\text{地区生产总值}} = 0.034$$

$$\emptyset_{\text{城市供水综合生产能力}} = 0.076, \gamma_{\text{城市供水综合生产能力}} = 0.756, C_{\text{城市供水综合生产能力}} = 0.404$$

$$\emptyset_{\text{全社会住宅投资额}} = 0.101, \gamma_{\text{全社会住宅投资额}} = 0.560, C_{\text{全社会住宅投资额}} = 0.151$$

在本模型中需要满足 $\gamma > 0.6$ 、 $C < 0.35$ 才能通过灰色模型检验，可知城市供水综合生产能力  $C_{\text{城市供水综合生产能力}} = 0.404 > 0.35$ 、全社会住宅投资额  $\gamma_{\text{全社会住宅投资额}} = 0.560 < 0.6$ 不满足模型检验，故将这两项指标作为不稳定指标剔除。

同理对所有关键指标进行关联度检验和后验差检验，初步设定残差检验数 $\emptyset < 0.1$ 的指标为稳定指标（记为 0.1 稳定标准），得到全部的稳定指标如表 10（详细数据见“附件 1——稳定指标——检验结果”），稳定指标即为影响各项财政收入的关键因素。

表 10 残差检验数为 0.1 下的稳定指标

财政收入	稳定指标	财务收入	稳定指标
营业税	第三产业增加值	房产税	地区生产总值
	公路客运量		城市居民年人均可支配收入
	限额以上餐饮业主营业务收入		城镇居民储蓄存款余额
	建筑业总产值		平均每人居住面积
	建筑业增加值		建筑业总产值
	全社会房地产开发投资额		全社会房地产开发投资额
增值税	批发零售业零售额	其它预算收入	农村居民年人均纯收入
	地区生产总值		第一产业增加值
	城市居民家庭人均可支配收入		第三产业增加值
	公共政府支出		地区生产总值

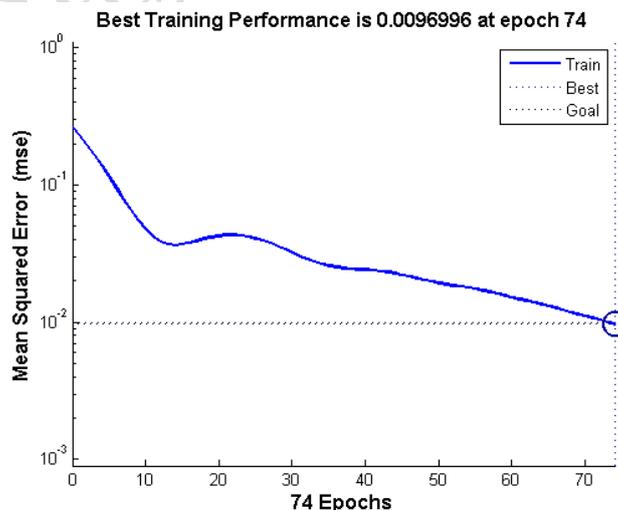
	社会消费品零售总额		城市居民家庭人均可支配收入
	商品销售总额		全社会固定资产投资额
	工业增加值		地方财政一般预算支出
企业所得税	第三产业增加值		社会消费品零售总额
	城市商品零售价格指数		社会消费品零售总额
	全社会固定资产投资额		地方财政一般预算收入
	建筑业总产值		第二产业增加值
	第二产业增加值		人均地区生产总值
个人所得税	商品销售总额	其它政府性其它基金收入	第三产业增加值
	城镇单位职工年平均工资		全社会车辆数
	地区生产总值		地区生产总值
	城市居民年人均可支配收入		全社会固定资产投资额
	城镇居民储蓄存款余额		公共财政支出
	公共财政支出		全社会货运量
	城镇居民现金消费支出		白云国际机场旅客吞吐量
第二产业增加值	建筑业总产值		
契税	全社会固定资产投资额		邮电业务收入
	总人口		
	城镇居民储蓄存款余额		
	建筑业总产值		
	第二产业增加值		
	建筑业增加值		
	房地产开发投资额		
房地产开发投资额# 住宅			

### 2.3.2 结构化财政收入模型结果

#### ■ 0.1 稳定标准下的 BP 神经网络模型

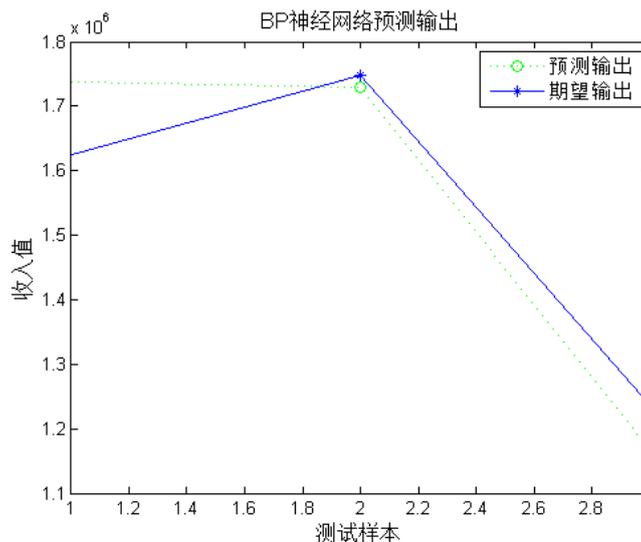
使用MATLAB编程实现BP神经网络模型的建立与求解（程序见“附件2——神经网络”），以营业税为例，得到最终结果如下：

如图七，训练在第40次迭代过程达到均方误差最小，MSE=0.000919055。此时，训练结束。

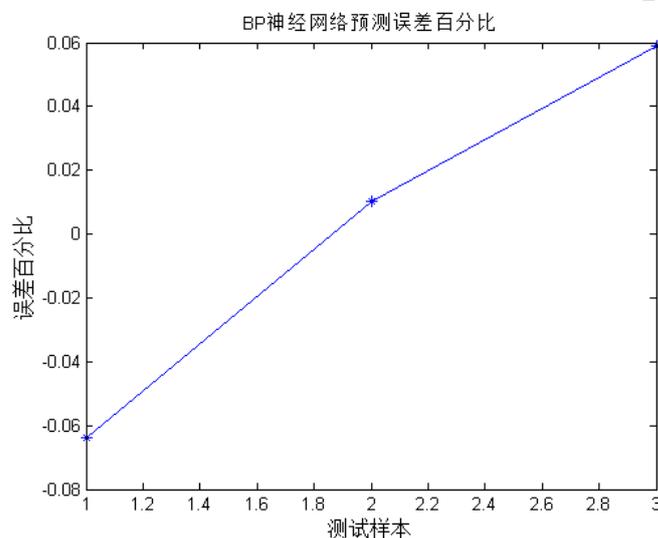


图七 神经网络迭代过程

用训练好的神经网络预测，得到预测输出值如图八，预测误差百分比如图九：



图八 BP 神经网络预测输出



图九 BP 神经网络预测误差百分比

由图八、图九可见，预测输出与期望输出相差不大，预测输出与期望输出误差百分比不超过10%，可以认为预测输出与期望输出误差较小，基本达到了我们要求的效果。

分别对各项财政收入进行BP神经网络模型训练及预测，得到结果见表11、12:

表 11 神经网络模型训练结果

各类收入	收敛步数	MSE
营业税	74	0.00096996
增值税	132	0.0009984
企业所得税	136	0.0098747
个人所得税	59	0.009453
契税	131	0.001
房产税	22	0.0008575
其它预算收入	41	0.0008385
其它政府性基金收入	197	0.0009796

表 12 神经网络模型测试结果 (单位/万元)

各类收入	实际值	预测值	误差值	相对误差
营业税	1625593	1736638	111045	0.064
	1747616	1729883	-17733	0.010
	1235374	1166380	-68994	0.059
增值税	1573830	1714628	140798	0.082
	443213	429352	-13861	0.032
	1287226	1246253	-40973	0.033
企业所得税	596693	550017	-46676	0.085
	756412	731737	-24675	0.034
	935248	937415	2167	0.002
个人所得税	472154	432987	-39167	0.090
	153080	167787	14707	0.088
	439592	481156	41564	0.086
契税	221754	225508	3754	0.017
	581872	624728	42856	0.069
	204895	188049	-16847	0.090
房产税	200221	206886	6665	0.032
	648012	685840	37828	0.055
	349616	350991	1375	0.004
其它预算收入	839073	825269	-13804	0.017
	1010447	986535	-23912	0.024
	674518	712249	37731	0.053
其它政府性基金收入	232247	260642	28395	0.109
	902997	846964	-56033	0.066
	708973	786550	77577	0.099

查找 2014 年各项财政收入实际值, 对比 BP 神经网络模型预测值, 汇总数据表如表

13:

表 13 神经网络模型预测结果 (单位/万元)

各项财政收入	2014 实际值	2014 预测值	2015 预测值	2014 年相对误差
营业税	1479906	1700547	1755803	0.149
增值税	2536757	2423528	2444431	0.045
企业所得税	1372179	1158557	123628	0.156
个人所得税	560349	492558	507204	0.121
契税	880240	895379	1034899	0.017
房产税	752931	708976	719938	0.058
其它预算收入	4832950	5661374	6386798	0.171
其它政府性基金收入	1086984	1053210	1101917	0.031

由表 13 可以看出, 增值税、契税、房产税和其它性政府性基金预测效果较好, 其它项财政收入的效果不太理想, 因此进一步缩小灰色预测检验中的残差标准, 得到更加稳定的关键指标, 并重新建立模型。

■ 关于土地出让金收入预测的特别说明

对于受政策影响较大的土地出让金收入项目, 暂且使用所需数据较少、预测精度较

高的灰色预测进行 2014、2015 年收入预测（程序见“附件 2-GM\_tdcjrj.m”），得到检验参数：

$$\emptyset_{\text{土地出让金收入}} = 0.179, Y_{\text{土地出让金收入}} = 0.640, C_{\text{土地出让金收入}} = 0.304$$

满足  $Y_{\text{土地出让金收入}} > 0.6$ 、 $C_{\text{土地出让金收入}} < 0.35$ ，即通过灰色预测检验，得到 2014 和 2015 年的预测值为 9240591、11731424（万元）。

■ 模型优化（不同稳定标准下的预测模型）

分别采用  $\emptyset < 0.08$ 、 $\emptyset < 0.06$  的残差标准，得到新的稳定关键指标（数据见“附件 1——模型优化”），在 MATLAB 中编写程序（程序见“附件 2——模型优化”），得到  $\emptyset < 0.10$ 、 $\emptyset < 0.08$ 、 $\emptyset < 0.06$  稳定标准（分别记为 0.10 稳定标准、0.08 稳定标准、0.06 稳定标准）下 2014 年财政收入预测结果和误差比较如表 14：

表 14 不同稳定标准下 2014 年财政收入预测结果（单位/万元）

各项财政收入	实际值	0.10 稳定标准		0.08 稳定标准		0.06 稳定标准	
		预测值	相对误差	预测值	相对误差	预测值	相对误差
营业税	1479906	1700547	<b>0.149</b>	1565349	0.058	1856921	<b>0.255</b>
增值税	2536757	2423528	0.045	2561844	0.010	2600628	0.025
企业所得税	1372179	1158557	<b>0.156</b>	1349780	0.016	1109898	<b>0.191</b>
个人所得税	560349	492558	<b>0.121</b>	557865	0.004	450344	<b>0.196</b>
契税	880240	895379	0.017	880584	0.000	869638	0.012
房产税	752931	708976	0.058	768396	0.021	673479	<b>0.106</b>
其它预算收入	4832950	5661374	<b>0.171</b>	5155051	0.067	4836004	0.001
其它政府性基金收入	1086984	1053210	0.031	1058599	0.026	864740	0.204
土地出让金收入	9653480	9240591	0.043	9240591	0.043	9240591	0.043

汇总各项财政收入如表 15：

表 15 不同稳定标准下 2014 年广州财政收入预测总收入（单位/万元）

各项财政收入	实际值	0.10 稳定标准		0.08 稳定标准		0.06 稳定标准	
		预测值	相对误差	预测值	相对误差	预测值	相对误差
一般预算收入	12415312	13040918	0.050	12838869	0.034	12396913	0.001
政府性基金	10740464	10293801	0.042	10299190	0.041	10105331	0.059
财政总收入	23155776	23334719	0.008	22079460	0.046	21637504	0.066

由表 14 可知在  $\emptyset < 0.10$  时模型对营业税、企业所得税、个人所得税、其它预算收入时预测误差大于 10%， $\emptyset < 0.06$  时模型对营业税、企业所得税、个人所得税、房产税预测误差大于 10%， $\emptyset < 0.08$  时对各项财政收入的预测效果较好。由表 15 可知对一般预

算收入的预测效果最好的是 0.06 稳定标准,对财政总收入和政府性基金收入的预测效果最好的是 0.08 稳定标准,0.01 稳定标准下的预测误差也较低(为 0.008),但考虑到 0.10 稳定标准和 0.06 稳定标准下模型对个别财政收入的预测效果较差,出现总的预测效果较好很可能是由于各项财政收入的误差相互抵消,从而导致了好的预测效果。相比 0.08 稳定标准下对各项财政收入的预测效果都比较理想,可以认为此时所建的各项财政收入的神经网络模型较佳,故采用 0.08 稳定标准下的预测模型作为最优模型,0.08 稳定标准下建立的模型得出广州市 2015 年各项财政收入和总收入如表 16:

表 16 广州市 2015 年财政收入预测结果(单位/万元)

各类财政收入	2015 预测值
营业税	1638245
增值税	2845002
企业所得税	1565923
个人所得税	582483
契税	924979
房产税	799934
其它预算收入	5445414
其它政府性基金收入	1100901
土地出让金收入	11731424
一般预算收入	13801980
政府性基金	12832325
财政总收入	25533405

则广州市 2015 年一般预算收入预测值为 13832325(万元)、政府性基金收入为 12832325(万元)、财政总收入为 25533405(万元)。

### 3. 结论

#### 3.1. 关键影响因素

通过以上研究,找到一种可通过变动稳定标准来选取影响财政收入的关键因素的方法,基于不同稳定标准建立起不同的 BP 神经网络财政收入预测模型,通过对不同稳定标准下 2014 年各项税收预测值与实际值进行比较,选取了误差较低的 0.08 稳定标准下的稳定指标作为关键影响因素,如表 17:

表 17 各项财政收入关键影响因素

财政收入	稳定指标	财务收入	稳定指标
营业税	第三产业增加值	房产税	地区生产总值
	公路客运量		城市居民年人均可支配收入
	限额以上餐饮业主营业务收入		城镇居民储蓄存款余额
	建筑业总产值		平均每人居住面积
	建筑业增加值		建筑业总产值
	全社会房地产开发投资额		全社会房地产开发投资额
增值税	批发零售业零售额	其它预算收入	第一产业增加值
	地区生产总值		第三产业增加值
	城市居民家庭人均可支配收入		地区生产总值
	公共政府支出		城市居民家庭人均可支配收入
	社会消费品零售总额		全社会固定资产投资额
企业所得税	第三产业增加值		地方财政一般预算支出
	城市商品零售价格指数		社会消费品零售总额
	全社会固定资产投资额		社会消费品零售总额
	建筑业总产值		地方财政一般预算收入
	第二产业增加值		第二产业增加值
个人所得税	城镇单位职工年平均工资		其它政府性其它基金收入
	地区生产总值	全社会车辆数	
	城市居民年人均可支配收入	地区生产总值	
	城镇居民储蓄存款余额	全社会固定资产投资额	
	公共财政支出	公共财政支出	
	城镇居民现金消费支出	全社会货运量	
	第二产业增加值	白云国际机场旅客吞吐量	
契 税	全社会固定资产投资额	建筑业总产值	
	总人口	邮电业务收入	
	城镇居民储蓄存款余额		
	建筑业总产值		
	第二产业增加值		
	建筑业增加值		
	房地产开发投资额		

### 3.2. 财政收入预测

在 0.08 稳定标准下, 建立各项财政收入与其对应关键影响因素的 BP 神经网络预测模型, 将灰色预测得出 2015 年关键影响因素的预测值代入所得模型, 得出 2015 年各项税收及总收入如表 18:

表 18 广州市 2015 年财政收入预测结果 (单位/万元)

各类财政收入	2014 实际值	2015 预测值	预测增长率
营业税	1479906	1638245	10.7%
增值税	2536757	2845002	12.2%
企业所得税	1372179	1565923	14.1%
个人所得税	560349	582483	4.0%
契税	880240	924979	5.1%

房产税	752931	799934	6.2%
其它预算收入	4832950	5445414	12.7%
其它政府性基金收入	1086984	1100901	1.3%
土地出让金收入	9653480	11731424	21.5%
<b>一般预算收入</b>	<b>12415312</b>	<b>13801980</b>	<b>11.1%</b>
<b>政府性基金</b>	<b>10740464</b>	<b>12832325</b>	<b>19.5%</b>
<b>财政总收入</b>	<b>23155776</b>	<b>25533405</b>	<b>10.3%</b>

广州 2014 年一般预算收入、政府性基金收入、财政总收入增长率分别为 10.9%、13.5%、10.8%，2015 年预测值分别为 11.1%、19.5%、10.3%，预测的 2015 年政府性基金收入增长率相比 2014 年较高可能是由于占政府性基金收入较大的土地出让金易受政策影响，难以较精准的预测，但是 2015 年的一般财政收入和财政总收入的预测值还是比较合理的。

### 3.3. 分析建议

近几年来广州市经济持续保持着高速健康的发展态势，但仍面临着保持领先发展与加快转型、提升质量效益的双重压力。据广州市 2015 年预算草案称，今年市本级财政收支矛盾突出，增收节支、统筹兼顾、科学安排将是财政的中心工作<sup>[13]</sup>。结合上述分析结果与广州市当前经济状况，提出以下几点建议：

1. 努力提升地区国民生产总值，带动财政收入合理增长。作为国民经济评价体系中两个最核心的指标，GDP 和财政收入存在着相互依存、相互作用的关系。从以上预测结果可知，广州 2015 年 GDP 与财政收入的预测增长率分别为 14.8% 与 10.3%，这是由于财政收入的增长有赖于 GDP 的增长的同时也会受到 GDP 增长的约束。近些年来经济增长的主要牵引力依赖于投资，消费需求始终不足，但是投资最终要转化为消费并受制于消费，因此应减少投资需求的膨胀，扩大消费需求提高消费对 GDP 的贡献率，努力提升地区生产总值，从而带动提高财政收入的增长。

2. 完善税收制度，保持财政收入稳定增长。税收在一般预算收入中占较大比重，是财政收入的主要来源。由以上预测结果可知，广州 2015 年营业税、增值税预测增长分别为 10.7% 与 12.15%，与 2014 年营业税下降 8.9%，增值说上升 14.5% 有所矛盾，这是由于广州市自 2012 年 11 月起实施“营改增”政策导致营业税收入下降，模型预测又有滞后性，但总体来说 2015 年四大税种占一般预算收入预测比例为 48.1%，与 14 年的 47.9% 相近，因此税收政策虽然短期内会导致收入的短收，但仍未对财政收入结构产生较大影

响。另外，“营改增”很大程度上可以减轻企业偷逃税的现象，甚至可以减轻增长困惑和调整转型压力，因此，在税制改革路上，不断完善税收制度，使得财政收入管理更加趋于合理，将有助于财政收入合理稳定的增长。

3.改善房地产相关政策，提高相关财政收入。由以上灰色斜率关联分析结果可知，全社会房地产开发投资额、商品房实际销售额等指标对房产税、契税、土地出让金收入等影响较大，因此房地产市场状况一定程度上影响着财政收入。2014年广州楼市低迷，虽然银行限贷政策有所松动，但受限购政策影响，广州市今年房地产开发投资和商品房销售将持续低迷，与房地产相关税收和土地出让收入形式不容乐观。因此，改善房地产相关政策，推出楼市救济计划，对提高相关财政收入有着重要意义。

4.厉行节约，保证财政支出合理化。2015年广州市财政支出重点仍是教育、文化、卫生等各项民生和社会公共事业。15年预算草案称，民生及各项公共事业支出占一般预算支出总额76.6%；民生和城市基础设施建设预算支出与政府性基金比例亦达55.9%。因此，市财政局应贯彻落实《党政机关厉行节约反对浪费条例》，努力实现公共接待费、因公出国费用、办公经费零增长，严格控制节庆、展会等一般性支出，使财政支出更加地有效与合理化，保证各项民生和社会公共事业发展得到有效保障，广州城市经济建设得以长足发展。

5.规范管理支出制度，充分发挥财政资金使用效益。合理有效的财政支出对实现收支平衡、加快社会经济发展进程有着重要意义。因此市政府应不断推进预算绩效管理改革，规范财政资金的管理，强化部门预算单位的责任意识，努力实现财政资金使用效益的最大化。

## 4. 参考文献

- [1] 刘海英.我国个人所得税收入实现的影响因素分析[J].大连理工大学学报(社会科学版),2015(01):49-54.
- [2] 辛浩,王韬.我国企业所得税税收收入能力估测研究[J].商场现代化,2009(10):342-343
- [3] 刘浩瀚.营业税影响因素计量研究[D].广州:暨南大学,2006
- [4] 郑永成.我国增值税收入的影响因素研究[J].科技广场,2014(03):137-142
- [5] 叶志锋,车文军.基于灰色关联法的房地产税收收入影响因素分析[J].2011(21):64-66
- [6] 吴笑晗.我国契税收入的影响因素及预测分析[J].税务研究,2013(12):73-75

- [7]张杰明.广州市 2007 年预算执行情况和 2008 年预算草案的报告[R].广州: 广州市财政局, 2008
- [8]张绍良, 张国良. 灰色关联度计算方法比较及其存在问题分析[J]. 系统工程, 1996, (03):45-49.
- [9]于国强, 李占斌, 张霞,等. 地下水动态的 BP 神经网络模型及改进的灰色斜率关联度分析[J]. 西安建筑科技大学学报: 自然科学版, 2009, 41(04):566-570.
- [10]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.
- [11]谢乃明, 刘思峰. 离散 GM(1,1)模型与灰色预测模型建模机理[J]. 系统工程理论与实践, 2005, 25(01):93-99.
- [12]袁静霞.关于广州市 2014 年预算执行情况和 2015 年预算草案的报告[R].广州:广州市财政局, 2015

“泰迪杯”优秀作品