

# 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：二等奖

作品单位：华南师范大学

作品成员：陈凯东 刘依依 李梓欣

指导教师：

## 广州市财政收入预测模型研究

**摘要：**广州市在实现经济快速发展，地区生产总值飞跃的同时，财政收入也在增加。对广州市未来的财政收入做出合理、有效的预算，能够克服年度地方预算收支规模确定的随意性和盲目性，为财政政策的制定提供指导依据。为了对广州市 2015 年的财政收入作出分析和预测，文章基于 1999 年至 2014 年广州市的相关数据，考虑了各种财政因素对财政收入的影响，按照定量与定性相结合的原则，应用逐步回归思想，最小二乘原理，BP 神经网络原理分别建立逐步回归模型、多元线性回归模型和 BP 神经网络模型，再利用三种模型的有效组合预测模型，建立了广州市财政收入的预测模型，并在此基础上为下一年的政策提供指导依据。

**关键词：**广州，财政收入，多元线性回归，BP 神经网络，组合预测，建议

# Study on the forecasting model of Guangzhou's financial revenue

**Abstract:** As Guangzhou achieving rapid economic development and great leap of regional GDP, its revenue is growing as well. To make reasonable and effective prediction of Guangzhou's revenue can help government avoid randomness and blindness when it intends to figure out the scale of annual regional revenues and expenditures. Also, it provides guidance for formulating financial policy. In order to make analysis and prediction of Guangzhou's revenue of 2015, this paper, being based on the data of Guangzhou from 1999 to 2014, according to the principle of combining qualitative and quantitative analysis, and considering various influential factors of financial revenue, make use of stepwise regression, the least squares principle and the principle of BP neural network to establish stepwise regression model, Multiple Linear Regression model and BP neural network model respectively. It combines the above three models effectively to establish the forecasting model of Guangzhou's financial revenue which can provide guidance for policy-making next year.

**Key words:** Guangzhou, financial revenue, Multiple Linear Regression, BP neural network, Combined Forecast, guidance

## 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. 单项预测模型的构建和预测 .....	2
2.3. 组合预测模型的构建和预测 .....	11
2.4. 模型评价 .....	12
<b>3. 对策与建议</b> .....	<b>12</b>
<b>4. 参考文献</b> .....	<b>13</b>
<b>5.附录</b> .....	<b>13</b>

“泰迪杯” 优秀作品

# 1. 挖掘目标

财政收入是政府实现其公共职能的重要基础。中央财政收入是我国整体经济社会长期稳定发展的重要保证，而地方财政收入作为我国地方性政府部门的公共收入，也同样是保证我国区域经济社会长期稳定发展的重要基础，是区域国民经济的综合反映以及市场经济国家的政府进行宏观调控的基础。

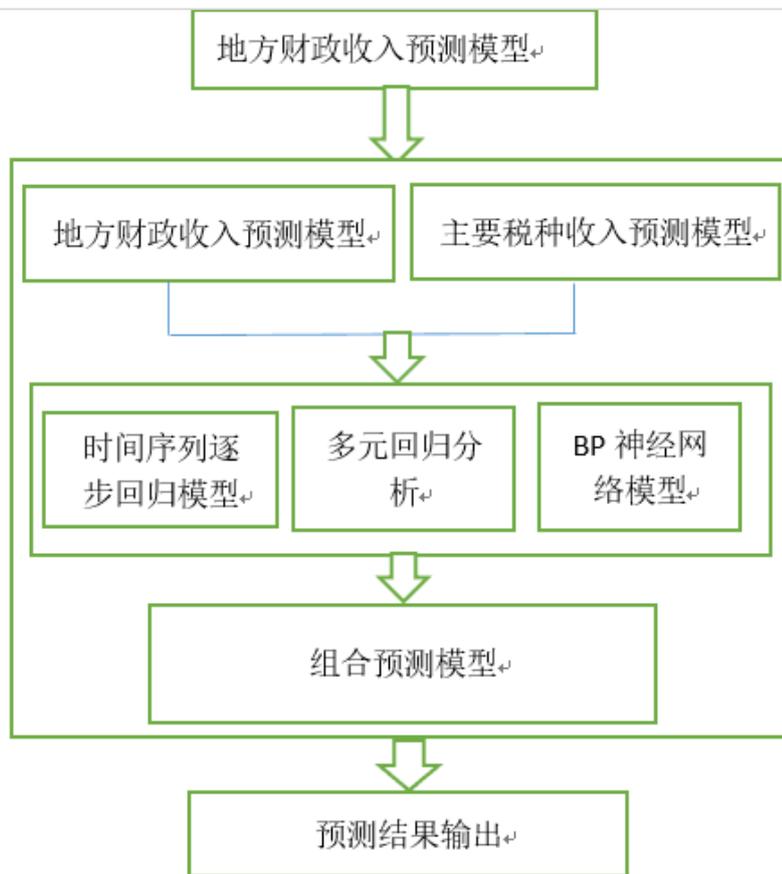
从地方公共财政收入总额来看，国内各省收入差距较大，广东夺魁，而广州市作为广东省的省会，改革开放的前沿城市，在实现经济快速发展，地区生产总值飞跃的同时，财政收入也在增加。对广州市未来的财政收入作出合理、有效的预算，能够克服年度地方预算收支规模确定的随意性和盲目性，为财政政策的制定提供指导依据。

本文基于1999年至2013年广州市的相关数据，首先应用最小二乘原理，建立多元线性回归模型再利用逐步回归模型（多项式拟合模型），以及BP神经网络模型探究、分析影响广州市财政收入的关键影响因素；然后立足于对影响因素的分析，对广州市2015年的财政收入及相关类别收入作出分析和预测；最后结合社会经济发展状况及广州市的财政情况，对广州市财政局提出建议。

## 2. 分析方法与过程

### 2.1. 总体流程

本文先确定了影响广州市财政收入的关键影响因素，接着进行建模，主要思路是将财政收入预测模型分为两个主要的模块，其中一个模块是将广州市财政收入和四种主要税种作为预测对象，通过运用时间序列方法、多元回归分析和BP神经网络原理三种预测方法分别对其进行建模预测，最后利用组合预测的方法将三种方法融合在一起，以减少各单项预测方法由于各自的缺陷没有考虑到的信息，改进单项预测的精确度。



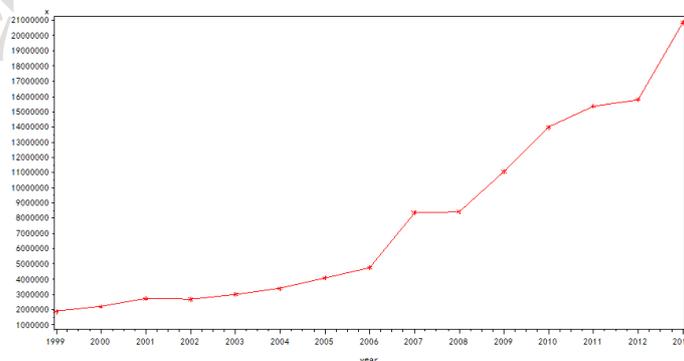
另外一个模块是基于模型确定的关键影响因素和预测结果，以及考虑广州近几年的财政收入及支出等情况，从财政收入和支出预算的角度，给广州市财政局提出的建议。

## 2.2. 单项预测模型的构建和预测

### 2.2.1. 逐步回归模型的构建和预测

首先，从时间序列图（如图 1 所示）中可以看出，财政收入具有明显的增长趋势；

图 1 广州市财政收入时间序列图



再进行纯随机性检验可判断其为非白噪声序列。

简单观察时序图后，先用趋势拟合对序列进行拟合，

图 2 线性拟合

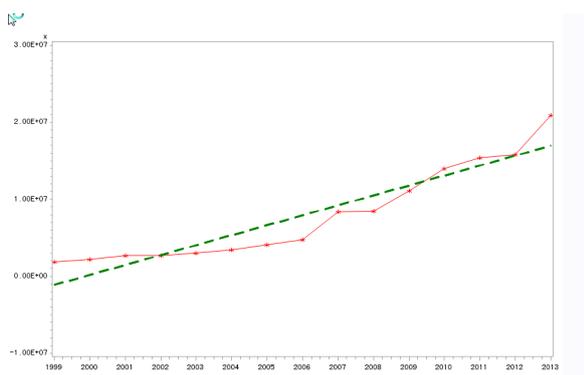
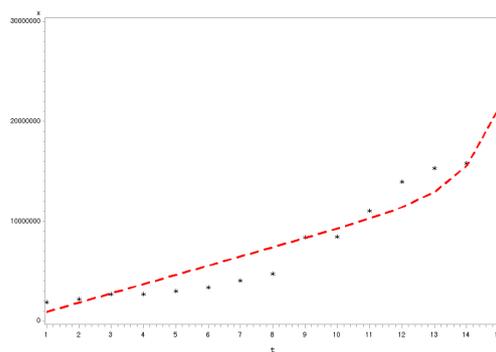


图 3 曲线拟合



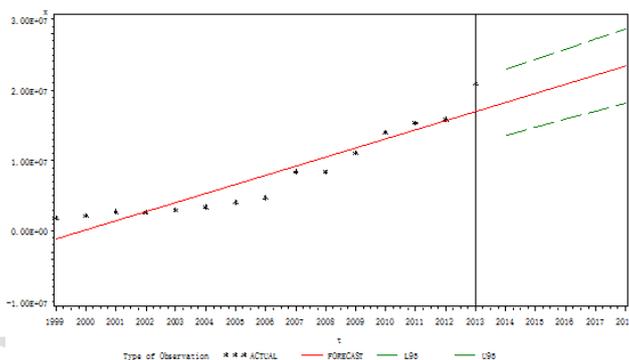
可见线性拟合和曲线拟合效果都不够理想，下面进行逐步回归；

逐步回归的原理是先使用多项式拟合趋势拟合好之后，再对残差序列进行自回归拟合，SAS 会默认一个高阶自回归阶数，然后通过逐步回归的方法，筛选合适的延迟阶数，得到最优 AR 模型，该方法也称为逐步自回归方法。

分别用 2 次、3 次 进行拟合，得到的拟合效果图如下：

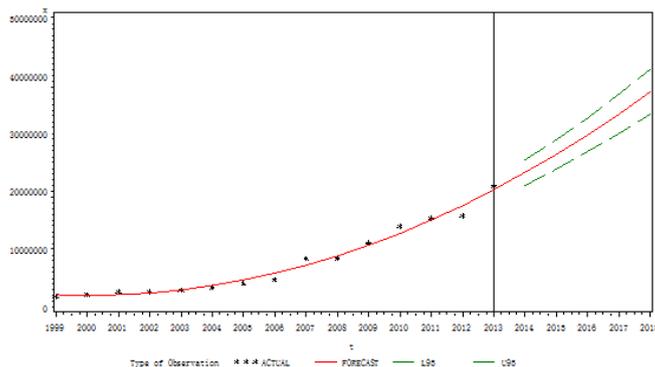
(1) trend=2: 即要求拟合线性趋势  $x_t = a + bt$

图 4 拟合效果图



(2) trend=3: 即要求为曲线拟合

图 5 拟合效果图



显然曲线拟合效果更好，更符合财政收入的增长趋势。模型的预测结果如下：（单位：万元）

图 6 模型的预测结果

t	观测类型	预测期间数	x
2013	ACTUAL	0	20881374
2013	FORECAST	0	20283825.62
2014	FORECAST	1	23224030.42
2014	L95	1	20978900.04
2014	U95	1	25469160.79
2015	FORECAST	2	26384091.48
2015	L95	2	23842776.45
2015	U95	2	28925406.51

预测 2014 年财政收入为 23224030 万元，区间为 20978900-25469160；由 2014 年广州市统计局的报告，2014 年的实际财政收入为 24539000 万元，在模型预测范围之内，可见预测的精准度在可接受的范围内。2015 年地方财政收入预测值为 **26384091.482 万元**。

以同样的方法，我们预测出了个人所得税、营业税、增值税和企业所得税四大税种的预测值：

预测 2014 年个人所得税值为 512540.16，2015 年为 536824.77；

营业税 2014 年值为 1939984.45，2015 年为 2047753.45；

增值税 2014 年值为 2058348.52，2015 年为 2187122.41；

企业所得税 2014 年值为 1188464.21，2015 年为 1236212.28。

### 2.2.2. 多元回归分析模型的构建和预测

#### (1) 指标变量的选择：

根据宏观经济学的相关理论，可以将影响财政收入的因素归纳为以下几个方面，分别为土地、资本、劳动力、政策，除此之外，还有众多其它的影响因素。根据对参考文献的总结，有前人认为财政收入的规模受到经济规模等诸多因素的影响，如第一产业总产值、第二产业总产值、第三产业总产值、社会消费品零售额、国土受灾面积、新增固定资产投资等；也有研究人员认为影响因素主要有总税收、国内生产总值、其他收入和就业人数等。因此，在考虑分析宏观经济理论以及前人有

关研究的基础之上，笔者决定从国民经济水平因素、资本因素、劳动力因素、消费因素及贸易因素这几个方面选取影响因素，作为指标变量。

GDP 代表国民经济总体水平，主要反映的是财政收入构成中的新创造的归社会支配的剩余产品价值部分，即社会生活中的总产值，国民生产总值是第一产业、第二产业以及第三产业生产总值的加和，全面反映了国民经济的发展情况，是财政收入的主要影响因素，当国民生产总值增加时，财政收入也会随之增加，一般来说财政收入将与国民经济生产总值保持相对稳定的增长态势。同样地，地区财政收入也应与地区生产总值有着密切联系，因此选取广州市主要年份的第一产业生产总值、第二产业生产总值及第三产业生产总值作为国民经济水平因素的变量。

全社会固定资产投资代表社会投资水平，反映的是国民经济的未来增长趋势，在经济发展态势良好的情况下，将会引致投资的增加，进而引起国民经济生产总值的持续增加，形成良性循环；而当国民经济生活发展态势消极时，将会导致居民消费需求降低，经济循环能力不足，进而降低闲置资金再投资的积极性。因此，固定资产投资也对地区财政收入有一定程度的影响，应选取作为资本因素的变量。

全社会消费品零售总额代表社会整体消费情况，是可支配收入在经济生活中的实现。当社会消费品零售总额增长时，表明社会消费意愿强烈，这一部分程度上会导致财政收入中增值税的增长，同时，当消费增长时，也会引起经济系统中其他方面发生变动，最终引致财政收入的增长，因此可选作消费因素的变量。

城镇居民储蓄存款余额、城镇单位职工年平均工资反映的是人民生活水平以、社会分配情况，是社会总产值中供个人支配的部分，还反映了社会劳动力状况，将主要影响财政收入中个人所得税、房产税、以及潜在消费能力，作为劳动力因素的变量。

进、出口总值会间接通过作用于关税的征收从而影响财政收入。就现行的出口退税政策说，作为一种间接性财政支出手段，能够最直接地起到扩大出口的作用，并且通过扩大出口，促进经济增长，最终带来财政收入的增加。可见进、出口总值与财政收入也有联系，作为贸易因素选作变量。综上，确定引入的影响因素变量有：第一产业生产总值( $x_2$ )、第二产业生产总值( $x_3$ )、第三产业生产总值( $x_4$ )、商品进口总值( $x_5$ )、商品出口总值( $x_6$ )、全社会固定资产投资额( $x_7$ )、社会消费品零售总额( $x_8$ )、城镇居民储蓄存款余额( $x_9$ )以及城镇单位职工年平均工资( $x_{10}$ )。(数据来源：广州市统计信息网)。

## (2) 多元线性回归原理：

### 1、多元线性回归模型概述

当有多个自变量可以同时影响因变量的变化时，可以考虑建立多元线性回归模型。设随机变量  $y$  与一般变量  $x_1, x_2, \dots, x_p$  的线性回归模型为

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \varepsilon$$

其中,  $\beta_0, \beta_1, \dots, \beta_p$  是  $p+1$  个未知参数,  $\beta_0$  称为回归常数,  $\beta_1, \dots, \beta_p$  称为回归系数。

$y$  称为被解释变量 (因变量),  $x_1, x_2, \dots, x_p$  是  $p$  个可以精确测量并控制的一般变量, 称为解释变量 (自变量)。

## 2、回归参数的普通最小二乘估计

对于回归模型  $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \varepsilon$ , 所谓最小二乘法, 就是寻找参数  $\beta_0, \beta_1, \dots, \beta_p$  的估计值  $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p$ , 使离差平方和  $Q(\beta_0, \beta_1, \dots, \beta_p) = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \dots - \beta_p x_{ip})^2$  达到极小, 求出的  $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p$  就称为回归参数  $\beta_0, \beta_1, \dots, \beta_p$  的最小二乘估计。

## 3、回归方程的显著性检验

### (1) F 检验

对多元线性回归方程的显著性检验就是要看自变量  $x_1, x_2, \dots, x_p$  从整体上对随机变量  $y$  是否有明显的影响。为此提出原假设

$$H_0: \beta_1 = \beta_2 = \dots = \beta_p = 0$$

构造 F 检验统计量如下:

$$F = \frac{SSR / p}{SSE / (n - p - 1)}$$

其中,  $SSR = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$ ,  $SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2$ 。

当  $F > F_\alpha(p, n - p - 1)$  时, 拒绝原假设  $H_0$ , 认为在显著性水平  $\alpha$  下,  $y$  与  $x_1, x_2, \dots, x_p$  有显著的线性关系, 即回归方程是显著的, 也即“自变量全体对因变量  $y$  产生线性影响”这一结论犯错误的概率不超过  $\alpha$ ; 反之, 当  $F \leq F_\alpha(p, n - p - 1)$  时, 则认为回归方程不显著。

### (2) t 检验

从回归方程中剔除那些次要的、可有可无的变量, 建立更为简单的回归方程, 需要对每个自变量进行显著性检验。如果某个自变量  $x_j$  对  $y$  的作用不显著, 那么在回归模型中, 它的系数  $\beta_j$  就取值为零。因此, 检验变量  $x_j$  是否显著等价于检验假设

$$H_{0j}: \beta_j = 0, \quad j = (1, 2, \dots, p)$$

构造 t 统计量

$$t_j = \frac{\hat{\beta}_j}{\sqrt{c_{jj} \hat{\sigma}}}$$

其中  $\hat{\sigma} = \sqrt{\frac{1}{n-p-1} \sum_{i=1}^n e_i^2} = \sqrt{\frac{1}{n-p-1} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$ ,  $(\mathbf{X}\mathbf{X})^{-1} = (c_{ij})$ ,  $\mathbf{X}$  是资料矩阵。

当  $|t_j| \geq t_{\alpha/2}$  时, 拒绝原假设  $H_{0j}: \beta_j = 0$ , 认为  $\beta_j$  显著不为零, 自变量  $x_j$  对因变量  $y$  的线性效果显著; 当  $|t_j| < t_{\alpha/2}$  时, 接受原假设  $H_{0j}: \beta_j = 0$ , 认为  $\beta_j$  为零, 自变量  $x_j$  对因变量  $y$  的线性效果不显著。

回归自变量的选择是建立回归模型的重要问题。有些自变量对问题的研究可能并不重要, 有些自变量数据的质量可能很差, 这都将影响回归方程的应用。因此, 可以用逐步回归法进行自变量的选择。

逐步回归的基本思想是有进有出, 引入一个变量或从回归方程中剔除一个变量, 为逐步回归的一步, 每一步都要进行 F 检验, 以确保每次引入新的变量之前回归方程中只包含显著的变量。这个过程反复进行, 直到既无显著的自变量选入回归方程, 也无不显著自变量从回归方程中剔除为止, 保证了最后所得的回归子集是最优回归子集。

#### 4、拟合优度

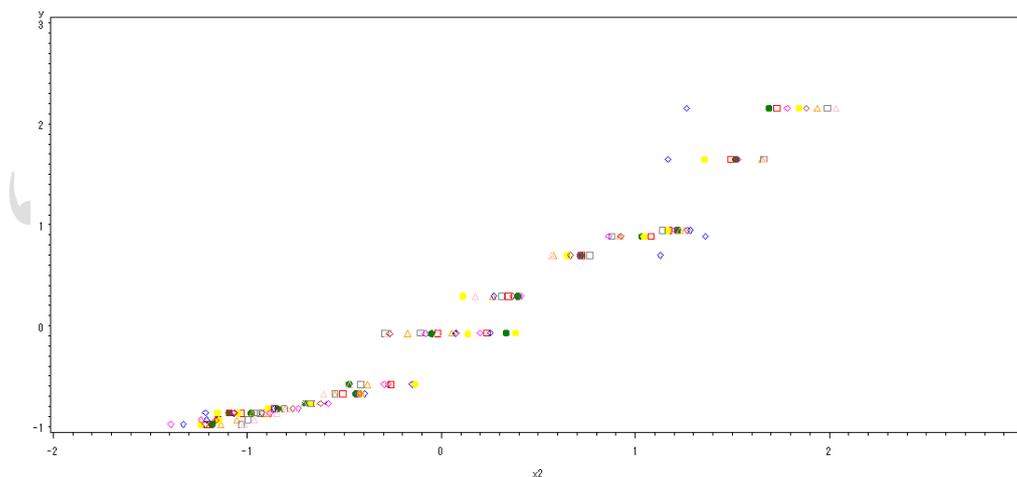
拟合优度用于检验回归方程对样本观测值的拟合程度。在多元线性回归中, 定义样本决定系数为

$R^2 = \frac{SSR}{SST}$ , 其中  $SST = SSR + SSE$ 。样本决定系数  $R^2$  的取值在  $[0, 1]$  区间内,  $R^2$  越接近 1, 表明回归拟合的效果越好;  $R^2$  越接近 0, 表明回归拟合的效果越差。

#### (3) 构建多元线性回归模型

通过对数据做散点图 (结果如图 7 所示), 可以看到财政收入与所选变量之间大体成线性关系, 因此用多元线性回归模型进行拟合及分析。

图 7 财政收入与 10 个变量散点图



基于最小二乘原理、回归方程的显著性检验原理以及逐步回归的分析方法, 运用现代统计软件 SAS 对数据进行多元线性回归 (SAS 程序见附录 1), 其结果如图 8 所示。

图 8 逐步回归结果

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	-1300646	718184	1.287536E12	3.28	0.0952
x4	0.21670	0.08401	2.811867E12	6.65	0.0241
x7	0.52660	0.14506	5.17328E12	13.18	0.0034
x9	-1793.28845	502.30916	5.003459E12	12.75	0.0039

得到模型： $y = 0.2167x_4 + 0.5266x_7 - 1793.2885x_9 - 1300646$

从回归方程可以看到，第三产业生产总值( $x_4$ )、全社会固定资产投资额( $x_7$ )、城镇居民储蓄存款余额( $x_9$ )对广州市财政收入影响显著，可见它们是影响广州市财政收入的关键影响因素；且根据回归系数可以看到，广州市财政收入与第三产业生产总值和全社会固定资产投资额是正相关的，而与城镇居民储蓄存款余额为负相关。

当社会固定资产投资额增加一亿元时，会带来财政收入 0.5266 亿元的增加，这与固定资产投资规模的加大会使财政收入增加的经济意义相符合；同理，第三产业生产总值增加一亿元时，相应的财政收入增加 0.2167 亿元，这也符合当前经济意义；而当城镇居民储蓄存款余额增加一亿元时，财政收入会相应地减少 1793.288 亿元，存款增加说明消费减少进而影响财政收入的减少这一实际。地方财政收入绝大部分由各种税收收入构成，故影响因素可能通过影响各种税收收入从而影响到财政收入。鉴于希望通过细化分析税收收入来研究地方财政收入，在此拟合四种主要税种的回归模型，个人所得税、企业所得税、增值税、营业税的回归模型分别如下：

$$y_1 = 0.27151x_2 - 104560$$

$$y_2 = 0.01647x_8 + 118492$$

$$y_3 = 0.01095x_4 + 18.07015x_{10} - 147539$$

$$y_4 = 3861.607x_5 - 1268.612x_6 + 232357$$

上述回归模型体现了个人所得税、企业所得税分别主要受第一产业生产总值和社会消费品零售总额的影响；增值税主要受第三产业生产总值和城镇单位职工年平均工资的影响；营业税主要受进出口总值的影响。根据模型及 2015 年各指标变量的数据得到 2015 年广州市地方财政收入及四种主要税种的预测值，如下表所示：

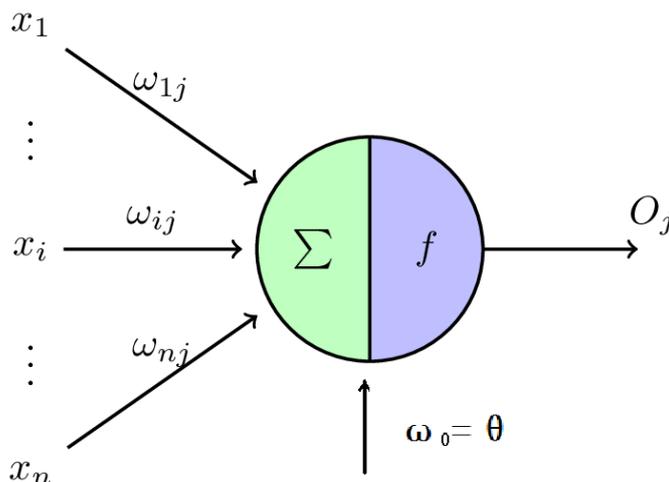
财政收入	个人所得税	企业所得税	增值税	营业税
108293.75	604608.5657	1556227.523	2638411.741	1660656.347

### 2.2.3. BP 神经网络模型的构建和预测

#### (1) BP 神经网络模型基本原理

BP (Back Propagation) 网络是 1986 年由 Rinehart 和 McClelland 为首的科学家小组提出，是一种按误差进行逆传播算法训练的多层前馈网络，是目前应用最广泛的神经网络模型之一。BP 网络

能学习和存贮大量的输入-输出模式映射关系，而无需事前揭示描述这种映射关系的数学方程。它的学习规则是使用最速下降法，通过反向传播来不断调整网络的权值和偏置，使网络的误差平方和最小。BP 神经网络模型拓扑结构包括输入层 (input)、隐层 (hide layer) 和输出层 (output layer)。



图中  $x_1 \square x_n$  是从其他神经元传来的输入信号， $w_{ij}$  表示表示从神经元  $j$  到神经元  $i$  的连接权值， $\theta$  表示一个偏置 ( bias )。则神经元  $i$  的输出与输入的关系表示为：

$$net_i = \sum_{j=1}^n w_{ij}x_j - \theta$$

$$y_i = f(net_i)$$

图中  $y_i$  表示神经元  $i$  的输出，函数  $f$  称为激活函数 ( Activation Function ) 或转移函数 ( Transfer Function ) ，  $net$  称为净激活 (net activation)。若将偏置看成是神经元  $i$  的一个输入  $x_0$  的权重  $w_{i0}$ ，则上面的式子可以简化为：

$$net_i = \sum_{j=0}^n w_{ij}x_j$$

$$y_i = f(net_i)$$

(1) 具体操作

1) 选取 2.2 中的关键指标作为输入变量，首先，对数据进行归一化处理，利用函数：

$$y = 2 \times (x - min) / (max - min) - 1$$

2) 此处隐层和输出层的传输函数采用 [0, 1] 型 sigmoid 函数

$$f(x) = \frac{1}{1 + e^{-x}} (0 < f(x) < 1)$$

和 purelin 线性函数

$$y = ax + b$$

网络的训练函数为 Quasi-Newton Algorithms - BFGS 算法。BFGS 是一种准牛顿算法，基本思想如下。首先构造目标函数在当前迭代  $x_k$  的二次模型：

$$m_k(p) = f(x_k) + \nabla f(x_k)^T p + \frac{p^T B_k p}{2}$$

$$p_k = -B_k \nabla f(x_k)$$

这里  $B_k$  是一个对称正定矩阵，于是我们取这个二次模型的最优解作为搜索方向，并且得到新的迭代点  $x_{k+1} = x_k + a_k p_k$ ，其中我们要求步长  $a_k$  满足 Wolfe 条件。这样的迭代与牛顿法类似，区别就在于用近似矩阵  $B_k$  代替真实的 Hessian 矩阵。所以拟牛顿法最关键的地方就是每一步迭代中矩阵  $B_k$  的更新。现在假设得到一个新的迭代  $x_{k+1}$ ，并得到一个新的二次模型：

$$m_{k+1}(p) = f(x_{k+1}) + \nabla f(x_{k+1})^T p + \frac{p^T B_{k+1} p}{2}$$

我们尽可能地利用上一步的信息来选取  $B_k$ 。具体地，我们要求

$$\nabla f(x_{k+1}) - \nabla f(x_k) = a_k B_{k+1} p_k$$

从而得到

$$B_{k+1}(x_{k+1} - x_k) = \nabla f(x_{k+1}) - \nabla f(x_k)$$

(2) 模型建立实现步骤：

1、将 1999-2013 年数据设为训练集，2014、2015 年为测试集（其中 14 年为实际数据，15 年为预算收入数据）。

2、参数设置：

最大迭代次数 (epochs)：40

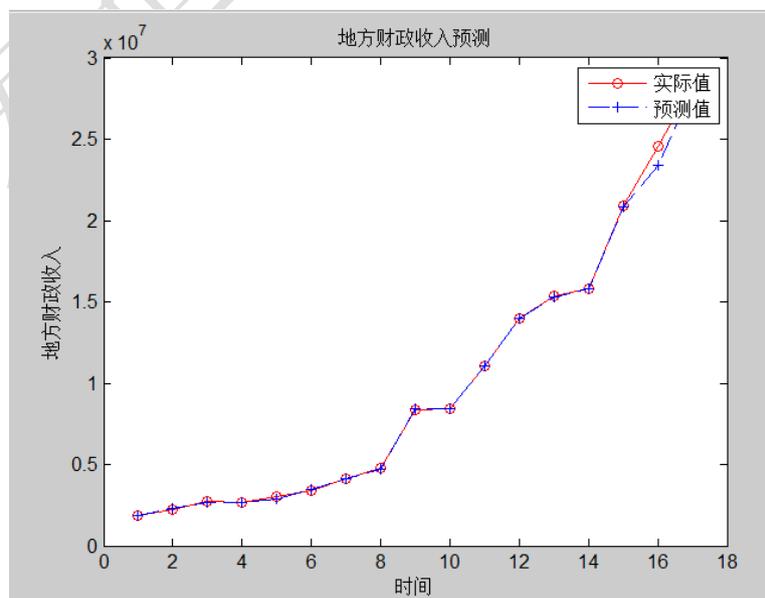
神经网络训练的目标误差 (goals)：0.0000001

显示中间结果的周期 (show)：2000

学习率 (lr)：0.15

3、输入训练集和测试集，使模型充分学习，使预测值逼近实际值。

4、输出预测值，由于数据进行过归一化处理，现进行反归一化，并画图。



预测值为 2014 年 23394443.98 万元、2015 年 28135630.27 万元（其中，2014 年报实际值为 24539000 万元，2015 预算值为 28436000 万元）

### 2.3. 组合预测模型的构建和预测

由于单因素的逐步回归模型、多因素的多元回归分析以及 BP 神经网络模型各自具有的不同的特性，我们考虑用组合模型的方法，来提取各个模型的优点，缩小误差范围。

我们采用拟合优度法

取

$$w_i = \frac{\sum_{i=1}^n \sigma - \sigma_i}{\sum_{i=1}^n \sigma} - \frac{1}{n-1}$$

式中  $\sigma_i$  ——第 i 个预测模型标准误差，标准误差为预测值与真实值的差的平方和。

当各种预测结果较分散，该模型能予以预测标准误差最小的模型以最大的权重，使预测结果保证拟合优度。

实际操作如下：

年份	地方财政收入	逐步回归预测值	多元回归分析预测值	神经网络预测值
1999	1881388	2205866.67	2082389.977	1870751.862
2000	2199077	2068083.7	2530085.354	2301053.42
2001	2719058	2150157.01	2723993.385	2663786.766
2002	2690984	2452086.58	2490660.552	2694938.464
2003	3005475	2973872.42	2890355.885	2869228.429
2004	3384477	3715514.53	3684860.801	3502648.814
2005	4088545	4677012.91	4143571.615	4120924.654
2006	4767231	5858367.56	5267752.976	4712944.914
2007	8389925	7259578.47	7512956.488	8415968.037
2008	8431405	8880645.66	8069418.947	8430230.089
2009	11076649	10721569.11	10492116.01	11080274.27
2010	13991612	12782348.84	13831534.39	13981064.14
2011	15351387	15062984.83	15236666.94	15308424.99
2012	15796804	17563477.09	16880288.57	15830212.17
2013	20881374	20283825.62	21773871.06	20864972.88
2014	24539000	23224030.42	23583714.6	23394443.98
2015	28436000	26384091.48	27108293.75	28135630.27

已知：逐步回归模型和 BP 神经网络模型的标准误差分别为  $1.47194E+13$ 、 $1.45437E+12$ 。先计算权重得：

$$w_1 = \frac{(1.47194E+13) + (1.45437E+12) + (6.47359E+12) - (1.47194E+13)}{(1.47194E+13) + (1.45437E+12) + (6.47359E+12)} \times \frac{1}{3-1} = 0.175031$$

同理可得： $w_2 = 0.357078$ ， $w_3 = 0.467891$

所以计算出来的 2014 年和 2015 年预测值为：

$$\hat{y}_{2014} = 23394443.98 \times 0.175031 + 23583714.6 \times 0.357078 + 23224030 \times 0.467891 = 23382293.2 \text{ 万元}$$

$$\hat{y}_{2015} = 28436000 \times 0.175031 + 27108293.75 \times 0.357078 + 26384091.482 \times 0.467891 = 27001835.6 \text{ 万元}$$

组合模型得到的 2014 年财政收入值为 23382293.2 万元，与广州市统计局给出的预算值 24539000 万元相比，误差在 5% 以内，视为合理，则说明组合模型的拟合效果比较好。

同理，根据组合预测模型，求得四大税种的预测值如下。

企业所得税 2015 年预测值：1515712.791 万元

个人所得税 2015 年预测值：608611.4696 万元

增值税 2015 年预测值：2707989.582 万元

营业税 2015 年预测值：1668158.517 万元

## 2.4. 模型评价

本文使用了三种单项指标模型和一个组合模型对财政收入和四大主要税种进行预测，分别是单指标的逐步回归模型、多指标的多元回归分析模型、以及 BP 神经网络模型。逐步回归模型注重于历史数据对未来预测的显著影响，多元回归分析侧重于各指标对财政收入的结构影响，而 BP 神经网络则采用了不同于传统时间序列预测模型的新方式，对财政收入及四大主要税种进行预测。组合模型则综合了上述各项模型，通过给误差较小的模型赋予较大的权重，来得到更优于单项预测模型的预测结果。

## 3. 对策与建议

### (1) 全社会固定资产投资额对财政收入的影响

由模型结果可知，全社会固定资产投资额对广州地区财政收入的影响很大，而且随着全社会固定资产投资额的增加，财政收入也增加。固定资产投资额的增加会使得各项相应税费额增加，而税收正是构成财政收入的主要部分。2014 年，广州全年完成固定资产投资 4889.50 亿元，比上年增长 14.5%。分投资主体看，国有经济投资 1391.84 亿元，增长 23.7%；民间投资 1814.37 亿元，增长 24.9%；港澳台、外商经济投资 864.92 亿元，下降 0.2%。从上述统计数据可以看到，固定资产投资额处于上升的趋势，而且还有提高的空间。因此，笔者认为，广州政府可以先加强投资立法，为投

资主体创造良好的投资环境，保护投资主体的财产所有权和收益权，然后积极进行招商引资工作，大力引进外资，通过增加固定资产投资额促进广州地方政府财政收入的增加。

### (2) 第三产业生产总值对财政收入的影响

通过回归分析可以发现，第三产业生产总值的增加能引起广州财政收入的增加，而且它对广州财政收入的影响贡献也是相对较大的。因此，广州市政府应重视第三产业的发展，应重视金融、旅游、交通运输等第三产业的发展，加大对第三产业的扶持力度，制定优惠政策，提供完善的法律保障制度，注重第三产业的结构调整，发挥广州地区优势，积极引进先进生产设备及科学技术，吸引高科技人才，促进第三产业结构优化升级，进而促进广州财政收入的增加。

### (3) 城镇居民储蓄存款余额对财政收入的影响

城镇居民储蓄存款余额存款的增加可能是由于居民收入增加，也可能是因为消费的减少。对于收入增加的情况，无疑会促使广州财政收入的增加。而由回归模型可以看到，城镇居民储蓄存款余额的增加会导致财政收入额的减少，这在一定程度上能够说明居民消费对财政收入的影响是大于居民收入的影响的。因此，政府可以积极推出刺激居民消费的政策。

## 4. 参考文献

- [1]何晓群,刘文卿. 应用回归分析[M].中国人民大学出版社.2012
- [2]刘心竹,李昊,刘青青,周慧昕. 我国地方财政收入影响因素的实证分析[J].中国集体经济,2012(3月):84-85
- [3]吕宁.地方财政一般预算收入预测模型研究—以浙江省为例[M].浙江大学
- [4]龚三乐.地区财政收入水平与质量评价及影响因素分析——以广西为例[J].区域金融研究.2011(461期):78-83
- [5]孙元,吕宁.地方财政一般预算收入预测模型及实证分析[J].数量经济技术研究.2007(1 期)
- [6]广州统计信息网: <http://www.gzstats.gov.cn/>

## 5.附录

### 5.1. SAS 逐步回归

```
data example4_4;  
input x@@;  
t=1999+_n_-1;
```

```

cards;
1881388 2199077 2719058 2690984 3005475 3384477 4088545 4767231 8389925 8431405
  11076649  13991612  15351387  15796804  20881374
;
proc gplot;
plot x*t=1;
symbol1 i=join v=none c=black;
run;
proc forecast data=example4_4 method=stepar trend=3 lead=5
              out=out outfull outest=est;
  id t;
  var x;
proc gplot data=out;
plot x*t= _type_ / href=2013;
  symbol1 i=none v=star c=black;
  symbol2 i=join v=none c=red;
  symbol3 i=join v=none c=green l=2;
  symbol4 i=join v=none c=green l=2;
run;

```

## 5.2. SAS 多元回归程序:

```

data a1;
input y y1-y4 x1-x10;
cards;
1881388 133621 277375 288972 433360 21391758 928522 9310691 11152545
  93.18 98.67 8782586 10006848 2040.18 16202
2199077 185625 309764 350495 479698 24927434 943718 10216241 13767475
  115.6 117.91 9236676 11211340 2239.86 19091
2719058 254892 483421 443213 540075 28416511 972806 11122943 16320762
  114.13 116.24 9782093 12482848 2600.43 22141
2690984 159684 236416 526377 613161 32039616 1030721 12113416
  18895479 141.49 137.78 10092421 13706815 3132.8 25583
3005475 153080 268360 581898 650119 37586166 1099080 14859261
  21627825 180.52 168.89 11751668 14942742 3727.33 28237
3384477 167379 326556 528365 793520 44505503 1171452 17880638
  25453413 233.14 214.74 13489283 16777731 4256.82 31025
4088545 198017 373397 816119 892678 51542283 1302159 20452183
  29787941 268.07 266.68 15191582 19058398 5024.69 33853
4767231 231794 455820 967265 1027971 60818614 1285029 24415160
  35118425 313.85 323.77 16963824 21991379 5562.36 36321
8389925 295316 596693 1115007 1235374 71403223 1498737 28257805
  41646681 355.91 379.03 18633437 26242399 5589.51 40187
8431405 353372 756412 1287226 1279793 82873816 1691849 32278717
  48903250 389.47 429.26 21055373 31873862 6867.29 45365

```

11076649	389824	732282	1375085	1516049	91382135	1722837	34051588
55607710	392.82	374.03	26598516	36157655	7954.22	49215	
13991612	472154	935248	1594182	1777343	107482828	1885645	40022658
65574525	553.89	483.79	32635731	44763780	9069.26	54807	
15351387	462098	1061594	1573830	1625593	124234390	2045420	
45769763	76419207	596.94	564.74	34122005	52430246	10032.62	57473
15796804	439592	1075045	1758311	1747616	135512072	2137620	
47206504	86167948	582.52	589.15	37583868	59772666	11310.69	63752
20881374	489777	1155923	2216017	1623520	154201434	2288680	
52273770	99638984	560.82	628.06	44545508	68828473	12253.98	69692
24539000	560000	1372000	2537000	1480000	167068700	2375200	
56064100	108629400	578.85	727.15	48895000	76978500	13608.376	74301

```

;
run;
proc gplot data=a1;
symbol1 c=green i=none w=2 v=dot;
symbol2 c=red i=none w=2 v=square;
symbol3 c=orange i=none w=2 v=triangle;
symbol4 c=blue i=none w=2 v=diamond;
symbol5 c=yellow i=none w=2 v=dot;
symbol6 c=gray i=none w=2 v=square;
symbol7 c=pink i=none w=2 v=triangle;
symbol8 c=brown i=none w=2 v=diamond;
symbol9 c=magenta i=none w=2 v=diamond;
plot y*x2=1 y*x3=2 y*x4=3 y*x5=4 y*x6=5 y*x7=6 y*x8=7 y*x9=8 y*x10=9/overlay;
run;
proc reg data=a1;
model y=x2-x10/selection=stepwise;
run;
proc reg data=a1;
model y1=x2-x10/selection=stepwise;
run;
proc reg data=a1;
model y2=x2-x10/selection=stepwise;
run;
proc reg data=a1;
model y3=x2-x10/selection=stepwise;
run;
proc reg data=a1;
model y4=x2-x10/selection=stepwise;

```

run;

### 5.3. MATLAB 神经网络程序

```

P=[21391758 928522 9310691 11152545 93.18 98.67 8782586 10006848 2040.18
16202
24927434943718 10216241 13767475115.6 117.91 9236676 11211340 2239.86 19091
28416511 972806 11122943 16320762 114.13 116.24 9782093 12482848
2600.43 22141
32039616 1030721 12113416 18895479 141.49 137.78 10092421 13706815
3132.8 25583
37586166 1099080 14859261 21627825 180.52 168.89 11751668 14942742
3727.33 28237
44505503 1171452 17880638 25453413 233.14 214.74 13489283 16777731
4256.82 31025
51542283 1302159 20452183 29787941 268.07 266.68 15191582 19058398
5024.69 33853
60818614 1285029 24415160 35118425 313.85 323.77 16963824 21991379
5562.36 36321
71403223 1498737 28257805 41646681 355.91 379.03 18633437 26242399
5589.51 40187
82873816 1691849 32278717 48903250 389.47 429.26 21055373 31873862
6867.29 45365
91382135 1722837 34051588 55607710 392.82 374.03 26598516 36157655
7954.22 49215
107482828 1885645 40022658 65574525553.89 483.79 32635731 447637809069.26
54807
124234390 2045420 4576976376419207 596.94 564.74 34122005 52430246
10032.62 57473
135512072 2137620 47206504 86167948 582.52 589.15 37583868 59772666
11310.69 63752
154201434 2288680 5227377099638984 560.82 628.06 44545508 68828473
12253.98 69692
167068700 2375200 56064100 108629400 578.85 727.15 48895000 76978500
13608.376 74301
    
```

187790838.2 2612038.916 62700824.24 122477975.1 624.98 776.54 54470240 87294203  
14953.57457 79956

]; 用了 10 个关键指标

T=[1881388

2199077

2719058

2690984

3005475

3384477

4088545

4767231

8389925

8431405

11076649

13991612

15351387

15796804

20881374

24539000

28436000

];%地方财政收入的数据，2014 是有的，2015 的是预算收入

S=T

%数据预处理，归一化到[-1,1],函数 mapminmax 是  $y = 2 * (x - \min) / (\max - \min) - 1$ ;

[P,Ps]=mapminmax(P);

[T,Ts]=mapminmax(T);

%P1 是 2015 年的输入数据，各个指标的数据都是用回归拟合得到的，准备作为模型预测的输入，T1 是预算收入

P1=P(:,16:17);

%P 是归一化后 1999-2014 年的输入数据，T 是 1999-2014 年地方财政收入

P=P(:,1:15);

T=T(:,1:15);

%BP 神经网络，隐层和输出层的传输函数为[0,1]型 sigmoid 函数和 purelin 线性函数， $y=a*x+b$

%网络的训练函数为 Quasi-Newton Algorithms - BFGS Algorithm/BFGS 算法

[n1,m1]=size(P);

net1=newff(minmax(P),[15 1],{'tansig','purelin'},'trainbfg');

```
net1.trainParam.epochs=40;
net1.trainParam.show=2000;
net1.trainParam.lr=0.15;
net1.trainParam.mc=0.7;
net1.trainParam.goal=0.0000001;
net1=train(net1,P,T);
%输出预测值，放在 Y
net_out=sim(net1,P);
net_out1=sim(net1,P1);
Y=[net_out,net_out1];
%反归一化，并画图
Y = mapminmax('reverse',Y,Ts)
plot([1:size(Y,2)],S,'r-o',[1:size(Y,2)],Y,'b--+');
hold off;
title('地方财政收入预测')
xlabel('时间')
ylabel('地方财政收入')
legend('实际值','预测值')
```