

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：暨南大学

作品成员：陈赓 卢丹丹 万浩文

指导教师：王国长

基于数据挖掘技术的市财政收入分析预测模型

摘要： 在分税制条件下，地方财政收入不仅是国家财政收入的重要组成部分，而且具有相对独立和固定的构成内容。文章选取 1994 年-2013 年广州市的财政收入及相关经济指标的数据，以及 1999 年-2013 年广州市的增值税、营业税、企业所得税、个人所得税及相关指标的数据，运用 Adaptive-Lasso 方法梳理并识别出影响市财政收入及各个类别收入的关键影响因素，并以此为解释变量，用灰色预测与 BP 神经网络的组合模型预测广州市 2015 年的财政收入及各个类别的收入。结果表明，广州市 2015 年财政收入达到 2879.3 亿元，与其预算值(2843.6 亿元)相差 35.7 亿元，预测效果良好。并且文章从财政收支与税源角度给出了相应的对策建议，给广州市财政局在实现公共财政过程中以参考。

关键词： 财政收入 Adaptive-Lasso 灰色预测 BP 神经网络

Analysis and forecast model of municipal finance revenue based on Data Mining Technology

Abstract: Under the conditions of the tax system, local revenues is not only an important part of state revenue, and also has a relatively independent and fixed content. The paper selects the data of the financial revenue and the relevant economic indicators of Guangzhou from 1994 to 2013, as well as the 1999 -2013 Guangzhou's VAT, business tax, corporate income tax, personal income tax and related indicators. Then using Adaptive-Lasso method to card and identify key factors that influence the municipal finance revenue and the income of each category, and as explanatory variables, with a combination model of gray forecasting and BP neural network to forecast Guangzhou 2015 fiscal revenue and the various categories of income. The results show that the Guangzhou 2015 revenue reached 287.93 billion RMB, with its budget value (284.36 billion RMB) differ 3.57 billion RMB, so the forecast effect was good. And the article also gives the corresponding countermeasures from the perspective of the revenue and expenditure and tax sources, in the hope that these are helpful to the Guangzhou Municipal Finance Bureau

Key words: Finance revenue Adaptive-Lasso Grey prediction
BP neural network

目 录

摘要.....	1
英文摘要.....	2
第 1 章 引言.....	1
1.1 研究背景.....	1
1.2 文献综述.....	1
第 2 章 挖掘目标.....	2
第 3 章 方法介绍.....	2
3.1 Adaptive-Lasso 正则方法.....	2
3.2 灰色预测与 BP 神经网络的组合模型.....	3
第 4 章 市财政收入影响因素分析.....	3
4.1 数据说明.....	3
4.1.1 财政收入数据.....	4
4.1.2 影响因素数据.....	4
4.2 描述分析.....	5
4.2.1 数据概括性度量.....	5
4.2.2 相关分析.....	6
4.3 Adaptive-Lasso 模型建立.....	6
第 5 章 收入预测.....	7
5.1 财政收入预测.....	7
5.2 增值税预测.....	8
5.3 营业税预测.....	10
5.4 企业所得税预测.....	12
5.5 个人所得税预测.....	13
5.6 政府性基金收入预测.....	15
第 6 章 对策建议.....	16
6.1 财政收支角度.....	16
6.2 税源角度.....	17
参考文献.....	19

第 1 章 引言

1.1 研究背景

在我国现行的分税制财政管理体制下，地方财政收人不仅是国家财政收入的重要组成部分，而且具有其相对独立的构成内容。因此，对地方财政收人进行预测，不仅是必要的，而且也是可能的。科学、合理地预测地方财政收人，对于克服年度地方预算收支规模确定的随意性和盲目性，正确处理地方财政与经济的相互关系具有十分重要的意义。

广州作为改革开放的前沿城市，其经济发展在全国经济中的地位举足轻重。目前，广州市在财政收入规模、结构等方面与北京、上海、深圳等城市仍有一定差距，存在不断完善的空间。本文旨在通过研究，发现影响广州市目前以及未来地方财源建设的因素，并对其进行深入分析，提出对广州市地方财源优化的具体建议，供政府决策参考，同时为学界和其他经济发展较快的城市提供借鉴。

1.2 文献综述

国内学者关于财政收入影响因素的分析，主要有以下方面：

尚元君和殷瑞锋^[1]选用 1993—2005 年中国 31 个省、市、自治区的财政、经济指标的面板数据，使用固定效应模型，对影响地方政府财政收入能力的因素进行了实证分析，研究发现：产业发展、城市经济水平指标与财政收入能力高度正相关。

刘心竹等^[2]选用 2009 年中国各省市财政以及经济数据，使用多元线性回归模型，并且加入地区虚拟变量，对影响地方财政收入的因素进行了实证分析，结果发现：固定资产投资、第一、第三产业生产总值对财政收入的影响非常显著。

成军^[3]选用河北省 1995—2000 年河北省财政以及经济数据，使用 BP 神经网络、多元线性回归组合模型预测 2001 年地方财政收入。结果显示：预测误差为 0.36 %，达到了较高的预测精度。

何邓娇^[4]选用广州市 1978—2011 年财政以及经济数据，使用 VAR 模型来分析影响广州市财政收入的因素。结果发现：国内生产总值、税收收入对财政收入的影响最大，社会从业人数对财政收入增加的贡献最大。

邓宏乾^[5](2007)提出分税制改革下，中央对地方财力“剥夺”过于严重，认为地方财源建设要完善和健全分级财政体制，确立房地产税收与城市土地地租在城市财政中的主

体地位。

刘晓^[6](2008)从经济增长、财政收入、财源建设的内在关联机制出发,研究湖南省财源建设问题。

总的来看,我国很多学者已经对财政收入的影响因素进行了很多研究,但是他们大多先建立财政收入与各待定的影响因素之间的多元线性回归模型,运用最小二乘估计方法来估计回归模型的系数,通过系数能否通过检验来检验他们之间的关系,这样的结果对数据的依赖程度很大,并且普通最小二乘估计求得的解往往是局部最优解,后续的检验可能就会失去应有的意义。

近几十年来,现代统计技术不断完善和发展,对新的数据运用新的方法来考察地方财政收入的影响因素是有必要的。本文在已有研究的基础上运用 Adaptive-lasso 变量选择方法来研究影响地方财政收入的因素。其优点为:新的现代统计方法的运用,能更好的度量各因素对股价的影响程度,从而全面的反映各指标与地方财政收入的相关程度。

第 2 章 挖掘目标

本次建模目标是利用广州市统计信息网的真实数据,采用 Adaptive-Lasso 方法,梳理影响广州市财政收入关联指标的有关数据,分析并识别影响财政收入的关键影响因素,构建广州市财政收入预测模型。进而利用 BP 神经网络与 ARIMA 组合模型对广州市 2015 年的财政总收入及各个类别收入进行预测。最后结合社会经济发展和广州市近几年的财政收支等情况,从财政收入和支出预算的角度,向广州市财政局提出对策建议。

第 3 章 方法介绍

3.1 Adaptive-Lasso 正则方法

在以往的文献中,对影响财政收入的因素的分析中大多使用普通最小二乘法来对回归模型的系数进行估计,预测变量的选取则采用的是逐步回归。然而,无论是最小二乘法还是逐步回归,都有其不足之处。它们一般都局限于局部最优解而不是全局最优解。如果预测变量过多,子集选择的计算过程具有不可实行性,且子集选择具有内在的不连

续性，从而导致子集选择极度多变。Lasso 是近年来被广泛应用于参数估计和变量选择的方法之一，并且 Lasso 进行变量选择在确定的条件下已经被证明是一致的。本文选用了 Adaptive-Lasso 方法来探究地方财政收入与各因素之间的关系。

Lasso 是由 Tibshirani^[7] (1996)提出的将参数估计与变量选择同时进行的一种正则化方法。Lasso 参数估计被定义如下：

$$\hat{\beta}(\text{lasso}) = \underset{\beta}{\operatorname{argmin}} \left\| y - \sum_{j=1}^p x_j \beta_j \right\|^2 + \lambda \sum_{j=1}^p |\beta_j| \quad (3-1)$$

其中， λ 为非负正则参数， $\lambda \sum_{j=1}^p |\beta_j|$ 称为惩罚项。

Lasso 方法虽然可以解决最小二乘法和逐步回归局部最优估计的不足，但是其自身需要满足一定的苛刻条件。Hui ZOU^[8](2006)提出了一种改进的 Lasso 方法，其改进之处在于给不同的系数加上了不同的权重，被称为 Adaptive-Lasso 方法，定义如下：

$$\hat{\beta}^{*(n)} = \underset{\beta}{\operatorname{argmin}} \left\| y - \sum_{j=1}^p x_j \beta_j \right\|^2 + \lambda_n \sum_{j=1}^p \hat{\omega}_j |\beta_j| \quad (3-2)$$

式中，权重 $\hat{\omega}_j = \frac{1}{|\hat{\beta}_j|^\gamma} (\gamma > 0)$ ， $j=1,2,\dots,p$ ， $\hat{\beta}_j$ 为由普通最小二乘法得出的系数。

3.2 灰色预测与 BP 神经网络的组合模型

在 Adaptive-Lasso 变量选择的基础上，鉴于灰色预测对小数据量数据预测的优良性能，对单个选定的影响因素建立灰色预测模型，得到它们在 2014 年及 2015 年的预测值。由于神经网络较强的适用性和容错能力，对历史数据建立训练模型，把灰色预测的数据结果代入训练好的模型中，就得到了充分考虑历史信息的预测结果，即 2015 年广州市财政收入及各个类别的收入。

第 4 章 市财政收入影响因素分析

4.1 数据说明

考虑到数据的可得性，本文所用的财政收入分为地方一般预算收入和政府性基金收入。地方一般预算收入包括：(1)税收收入，主要包括企业所得税和地方所得税中中央和地方共享的 40%，地方享有的 25%的增值税、营业税、印花税等；(2)非税收入，包括

专项收入、行政事业性收费、罚没收入、国有资本经营收入和其他收入等。政府性基金收入是国家通过向社会征收以及出让土地、发行彩票等方式取得收入，并专项用于支持特定基础设施建设和社会事业发展的收入^[9]。

4.1.1 财政收入数据

由于 1994 年我国对财政体制进行了重大改革，开始实行分税制财政体制，影响了财政收入相关数据的连续性，在 1994 年前后不具有可比性。由于没有合适的数学手段来调整这种数据的跃变，仅对 1994 年及其以后的数据进行分析，本文所用数据均来自《广州市统计年鉴》(1995-2014)。

4.1.2 影响因素数据

影响财政收入的因素有很多，在查阅大量文献的基础上，通过经济理论对财政收入的解释以及对实践的观察，考虑一些与能源消耗关系密切并且直观上有线性关系的因素，初步选取以下因素为自变量，分析它们之间的关系。

X1. 社会从业人数：就业人数的上升伴随着居民消费水平的提高，从而间接影响财政收入的增加。

X2. 在岗职工工资总额：反映的是社会分配情况，主要影响财政收入中的个人所得税、房产税以及潜在消费能力。

X3. 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，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，最终导致财政收入的增长。

X4. 城镇居民人均可支配收入：居民收入越高消费能力越强，同时意味着其工作积极性越高，创造出的财富越多，从而能带来财政收入的更快和持续增长。

X5. 城镇居民人均消费性支出：居民在消费商品的过程中会产生各种税费，税费又是调节生产规模的手段之一。在商品经济发达的如今，居民消费的越多，对财政收入的贡献就越大。

X6. 年末总人口：在地方经济发展水平既定的条件下，人均地方财政收入与地方人口数呈反比例变化。

X7. 全社会固定资产投资额：是建造和购置固定资产的经济活动，即固定资产再生产活动。主要通过投资来促进经济增长，扩大税源，进而拉动财政税收收入整体增长。

X8. 地区生产总值：表示地方经济发展水平。一般来讲，政府财政收入来源于即期的地区生产总值。在国家经济政策不变、社会秩序稳定的情况下，地方经济发展水平与地方财政收入之间存在着密切的相关性，越是经济发达的地区，其财政收入的规模就越大。

X9. 第一产业产值：取消农业税、实施三农政策，第一产业对财政收入的影响更小。

X10. 税收：由于其具有征收的强制性、无偿性和固定性特点，可以为政府履行其职能提供充足的资金来源。因此，各国都将其作为政府财政收入的最重要的收入形式和来源。

X11. 居民消费价格指数：反映居民家庭购买的消费品及服务价格水平的变动情况，影响城乡居民的生活支出和国家的财政收入。

X12. 第三产业与第二产业产值比：表示产业结构。三次产业生产总值代表国民经济水平，是财政收入的主要影响因素，当产业结构逐步优化时，财政收入也会随之增加。

X13. 居民消费水平：在很大程度上受整体经济状况 GDP 的影响，从而间接影响地方财政收入。

4.2 描述分析

首先对已有数据进行描述性统计分析，获得对数据的整体性认识。

4.2.1 数据概括性度量

表 4-1 各变量统计量

	Min	Max	Mean	SD
x1	3831732.000	7599295.000	5579519.9500	1.262195e+06
x2	181.540	2110.780	765.0350	5.956983e+02
x3	448.190	6882.850	2370.8250	1.919167e+03
x4	7571.000	42049.140	19644.6850	1.020302e+04
x5	6212.700	33156.830	15870.9480	8.199771e+03
x6	6370241.000	8323096.000	7350513.6000	6.213419e+05
x7	525.710	4454.550	1712.2390	1.184714e+03
x8	985.310	15420.140	5705.7990	4.478400e+03
x9	60.620	228.460	129.4935	5.050983e+01
x10	65.660	852.560	340.2165	2.515779e+02
x11	97.500	120.000	103.3050	5.513283e+00
x12	1.029	1.906	1.4222	2.532348e-01
x13	5321.000	41972.000	17273.8000	1.110919e+04
y	64.870	2088.140	618.0840	6.092545e+02

4.2.2 相关分析

相关系数可以用来描述定量变量之间的关系，初步判断因变量与解释变量之间是否具有线性相关性。

表 4-2 变量 Pearson 相关系数表

	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	x13	y
x1	1.00	0.95	0.95	0.97	0.97	0.99	0.95	0.97	0.98	0.98	-0.29	0.94	0.96	0.94
x2	0.95	1.00	1.00	0.99	0.99	0.92	0.99	0.99	0.98	0.98	-0.13	0.89	1.00	0.98
x3	0.95	1.00	1.00	0.99	0.99	0.92	1.00	0.99	0.98	0.99	-0.15	0.89	1.00	0.99
x4	0.97	0.99	0.99	1.00	1.00	0.95	0.99	1.00	0.99	1.00	-0.19	0.91	1.00	0.99
x5	0.97	0.99	0.99	1.00	1.00	0.95	0.99	1.00	0.99	1.00	-0.18	0.90	0.99	0.99
x6	0.99	0.92	0.92	0.95	0.95	1.00	0.93	0.95	0.97	0.96	-0.34	0.95	0.94	0.91
x7	0.95	0.99	1.00	0.99	0.99	0.93	1.00	0.99	0.98	0.99	-0.15	0.89	1.00	0.99
x8	0.97	0.99	0.99	1.00	1.00	0.95	0.99	1.00	0.99	1.00	-0.15	0.90	1.00	0.99
x9	0.98	0.98	0.98	0.99	0.99	0.97	0.98	0.99	1.00	0.99	-0.23	0.91	0.99	0.98
x10	0.98	0.98	0.99	1.00	1.00	0.96	0.99	1.00	0.99	1.00	-0.17	0.90	0.99	0.99
x11	-0.29	-0.13	-0.15	-0.19	-0.18	-0.34	-0.15	-0.15	-0.23	-0.17	1.00	-0.43	-0.16	-0.12
x12	0.94	0.89	0.89	0.91	0.90	0.95	0.89	0.90	0.91	0.90	-0.43	1.00	0.90	0.87
x13	0.96	1.00	1.00	1.00	0.99	0.94	1.00	1.00	0.99	0.99	-0.16	0.90	1.00	0.99
y	0.94	0.98	0.99	0.99	0.99	0.91	0.99	0.99	0.98	0.99	-0.12	0.87	0.99	1.00

从相关分析结果可以看到，居民消费价格指数(x11) 与财政收入的线性关系不显著，而且呈现负相关。其余变量均与财政收入呈现高度的正相关关系。

4.3 Adaptive-Lasso 模型建立

运用 LARS 算法来解决式(3-2)的 Adaptive-Lasso 估计，对于每给一个 γ ，该算法会寻找一个最优的 λ_n 。此处取 $\gamma=1$ ，用 R 语言编制相应的程序后运行得到如下结果，见表 4-3。

表 4-3 系数表

x1	x2	x3	x4	x5	x6	x7
-0.0001	-0.2309	0.1375	-0.0401	0.0760	0.0000	0.3069
x8	x9	x10	x11	x12	x13	
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	

由表 4-3 可看出，年末总人口、地区生产总值、第一产业产值、税收、居民消费价格指数、第三产业与第二产业产值比以及居民消费水平等因素的系数为 0，即在模型建立的过程中这几个变量被剔除了。这是因为居民消费水平与城镇居民人均消费性支出存在明显的共线性，Adaptive-Lasso 方法在构建模型的过程中剔除了这个变量；由于广州存在流动人口与外来打工人口多的特性，年末总人口并不显著影响广州市财政收入；居

民消费价格指数与财政收入的相关性太小以致可以忽略；由于农牧业各税在各项税收总额中所占比重过小，而且广州于 2005 年取消了农业税，因而第一产业对地方财政收入的贡献率极低；其他变量被剔除均有类似于上述的原因。这说明 Adaptive-Lasso 方法在构建模型时，能够剔除存在共线性关系的变量，同时体现了 Adaptive-Lasso 方法对多指标进行建模的优势。

综上所述，利用 Adaptive-Lasso 方法识别影响财政收入的关键影响因素是社会从业人数、在岗职工工资总额、社会消费品零售总额、城镇居民人均可支配收入、城镇居民人均消费性支出以及全社会固定资产投资额。

第 5 章 收入预测

5.1 财政收入预测

对Adaptive-Lasso 变量选择方法识别的影响财政收入的因素建立灰色预测与BP神经网络的组合预测模型，其参数设置为误差精度 10^{-7} ，学习次数10000次，神经元个数为Lasso变量选择方法选择的变量个数 6。社会从业人数(x1)、在岗职工工资总额(x2)、社会消费品零售总额(x3)、城镇居民人均可支配收入(x4)、城镇居民人均消费性支出(x5)、固定资产投资额(x7)指标的2014年及2015年数值通过R语言建立灰色预测模型得出，预测精度等级见表5-1，灰色预测模型有很好的效果。代入地方财政收入所建立的BP神经网络预测模型，得到广州市财政收入2015年的预测值为**2879.26**亿元，相关数据见表5-2，其中红色字体的数据为预测数据。图5.1为BP神经网络地方财政收入真实值与预测值对比图。

表5-1 灰色预测模型地方财政收入相关因素精度表

	X1	X2	X3	X4	X5	X7
2014 预测值	8142148	2239.295	7042.313	43611.84	35046.63	4600.405
预测精度等级	好	好	好	好	好	好

表5-2 地方财政收入及其相关因素历史数据和预测表

	x1	x2	x3	x4	x5	x7	y
1994	3831732	181.54	448.19	7571	6212.7	525.71	64.87
1995	3913824	214.63	549.97	9038.16	7601.73	618.25	99.75
1996	3928907	239.56	686.44	9905.31	8092.82	638.94	88.11
1997	4282130	261.58	802.59	10444.6	8767.98	656.58	106.07
1998	4453911	283.14	904.57	11255.7	9422.33	758.83	137.32

1999	4548852	308.58	1000.69	12018.52	9751.44	878.26	188.14
2000	4962579	348.09	1121.13	13966.53	11349.47	923.67	219.91
2001	5029338	387.81	1248.29	14694	11467.35	978.21	271.91
2002	5070216	453.49	1370.68	13380.47	10671.78	1009.24	269.1
2003	5210706	533.55	1494.27	15002.59	11570.58	1175.17	300.55
2004	5407087	598.33	1677.77	16884.16	13120.83	1348.93	338.45
2005	5744550	665.32	1905.84	18287.24	14468.24	1519.16	408.86
2006	5994973	738.97	2199.14	19850.66	15444.93	1696.38	476.72
2007	6236312	877.07	2624.24	22469.22	18951.32	1863.34	838.99
2008	6529045	1005.37	3187.39	25316.72	20835.95	2105.54	843.14
2009	6791495	1118.03	3615.77	27609.59	22820.89	2659.85	1107.67
2010	7110695	1304.48	4476.38	30658.49	25011.61	3263.57	1399.16
2011	7431755	1700.87	5243.03	34438.08	28209.74	3412.21	1535.14
2012	7512997	1969.51	5977.27	38053.52	30490.44	3758.39	1579.68
2013	7599295	2110.78	6882.85	42049.14	33156.83	4454.55	2088.14
2014	8142148	2239.295	7042.313	43611.84	35046.63	4600.405	2453.9
2015	8460489	2581.142	8166.916	47792.22	38384.22	5214.78	2879.26

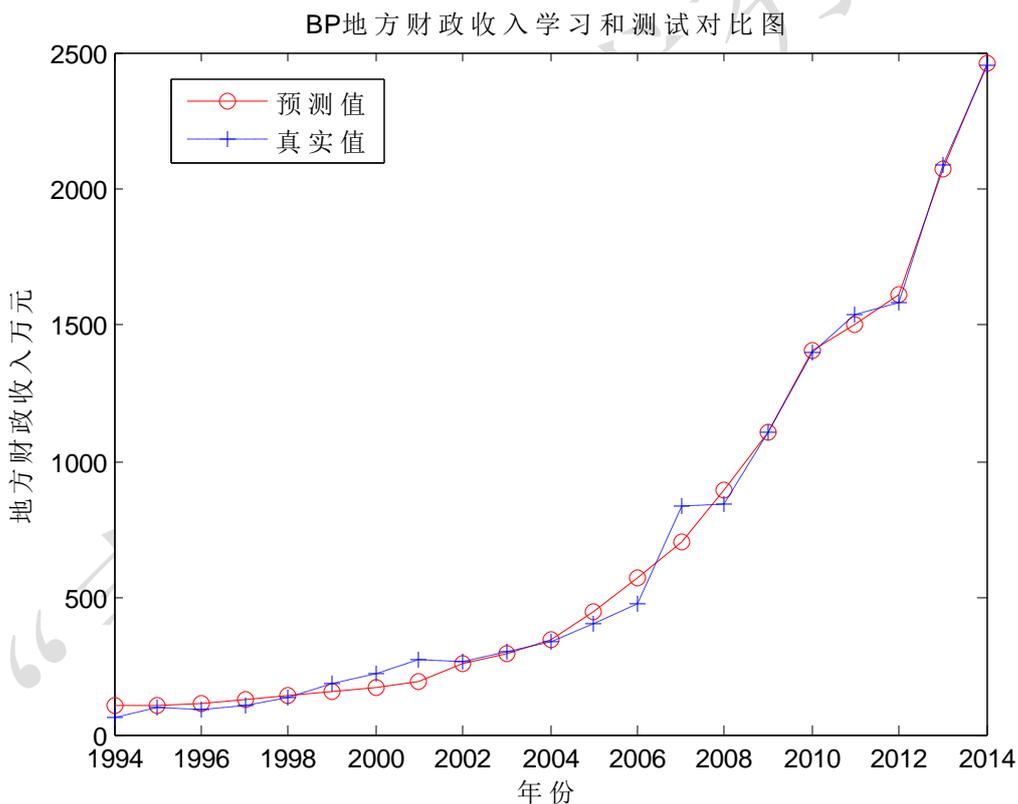


图5.1 BP神经网络地方财政收入真实值与预测值对比图

5.2 增值税预测

利用Adaptive-Lasso 方法进行增值税影响因素的变量选择, 通过表5-3可以看出, 商

品进口总值(x1)、工业增加值(x3) 和工业增加值占GDP (x5) 比重这三个因素进入选择，其他因素的系数为0。这会因为根据工业增加值及其占GDP 比重可以算出地区生产总值，所以Adaptive-Lasso 方法在构建模型的过程中剔除了地区生产总值这个变量；由于批发零售业对增值税的贡献率较低，所以该因素也被剔除。

表5-3 系数表

x1	x2	x3	x4	x5
-1365.17346	0.00000	0.06098	0.00000	-447747.88928

对Adaptive-Lasso 变量选择方法识别的影响增值税的因素建立BP神经网络预测模型，其参数设置为误差精度 10^{-7} ，学习次数10000次，神经元个数为Lasso变量选择方法选择的变量个数 3。商品进口总值(x1)、工业增加值(x3) 和工业增加值占GDP 比重(x5) 指标的2014年及2015年数值通过R语言建立灰色预测模型得出，后验差比值、预测精度等级见表5-4，灰色预测模型有较好的效果。代入增值税所建立的BP神经网络预测模型，得到**增值税的2015年预测值为1835178万元**，相关数据见表5-5，其中红色字体的数据为预测数据。图5.2为BP神经网络增值税真实值与预测值对比图。

表5-4 灰色预测模型增值税相关因素精度表

	X1	X3	X5
后验差比值	0.1853(<0.35)	0.0807	0.5067([0.5,0.65])
预测精度等级	好	好	勉强合格

表5-5 增值税及其相关因素历史数据和预测表

	x1	x3	x5	增值税
1999	93.18	7980207	0.373051	288972
2000	115.6	8779835	0.352216	350495
2001	114.13	9554676	0.336237	443213
2002	141.49	10509450	0.328014	526377
2003	180.52	13141254	0.34963	581898
2004	233.14	15941538	0.358193	528365
2005	268.07	18439550	0.357756	816119
2006	313.85	22270093	0.366172	967265
2007	355.91	26029310	0.36454	1115007
2008	389.47	29724781	0.358675	1287226
2009	392.82	31173422	0.341133	1375085
2010	553.89	36449611	0.33912	1594182
2011	596.94	41405926	0.333289	1573830
2012	582.52	42641557	0.31467	1758311
2013	560.89	47548175	0.308351	2216017
2014	767.59	58163230	0.3290099	1752521
2015	862.30	65803730	0.3271446	1835178

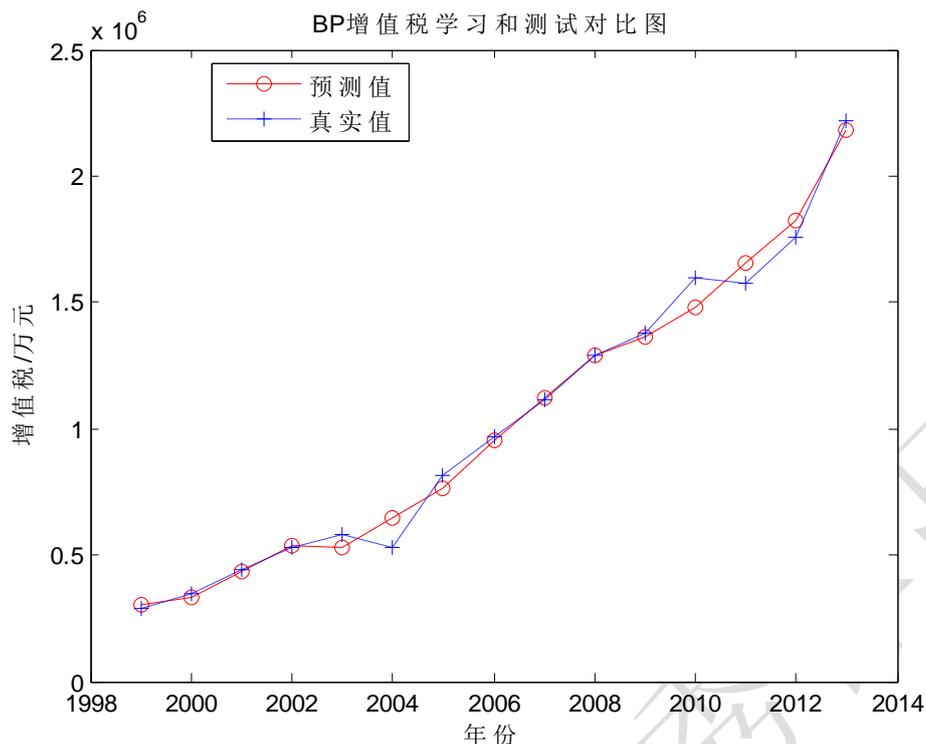


图5.2 BP神经网络增值稅真实值与预测值对比图

5.3 营业税预测

利用Adaptive-Lasso 方法进行企业所得税影响因素的变量选择，通过表5-6可以看出，全社会固定资产投资额(x3)、城市商品零售价格指数(1978=100) (x4)、规模以上国有及国有控股工业企业亏损面(x6) 和建筑业企业利润总额 (x8) 这四个因素进入选择，其他因素的系数为0。

表5-6 系数表

x1	x2	x3	x4	x5
0.00000	-0.00000	0.14197	0.05383	0.00000
x6	x7	x8	x9	x10
-0.18523	0.00000	-0.14107	0.00000	0.00000

对Adaptive-Lasso 变量选择方法识别的影响营业税的因素建立BP神经网络预测模型，其参数设置为误差精度 10^{-7} ，学习次数10000次，神经元个数为Lasso变量选择方法选择的变量个数 3。变量选择的指标的2014年及2015年数值通过R语言建立灰色预测模型得出，后验差比值、预测精度等级见表5-7，灰色预测模型有很好的效果。代入营业税所建立的BP神经网络预测模型，得到营业税的2015年预测值为**1431472**万元，相关数据见表5-8，其中红色字体的数据为预测数据。图5.3为BP神经网络营业税真实值与预测值对比图。

表5-7 灰色预测模型营业税相关因素精度表

	X3	X4	X6	X8
后验差比值	0.1153	0.0179	0.1566	0.0719
预测精度等级	好	好	好	好

表5-8 营业税及其相关因素历史数据和预测表

	x3	x4	x6	x8	营业税
1999	1330484	11152545	2878473	2470523	433360
2000	1436406	13767475	3250326	2561326	479698
2001	1568267	16320762	3316894	3403870	540075
2002	1603966	18895479	3457617	3733922	613161
2003	1718007	21627825	3522168	4785787	650119
2004	1939100	25453413	3712961	5459314	793520
2005	2012633	29787941	3777003	6331382	892678
2006	2145067	35118425	3783416	6870406	1027971
2007	2228495	41646681	5041090	7507109	1235374
2008	2553936	48903250	5398216	8754491	1279793
2009	2878166	55607710	5246903	10134050	1516049
2010	3573047	65574525	5727122	12805288	1777343
2011	4363837	76419207	8116313	15613171	1625593
2012	4564947	86167948	8626775	17417072	1747616
2013	4725256	99643373	9969708	21828895	1623520
2014	5319885	118049300	10017410	23746700	1520054
2015	5919520	137165300	11096340	27870540	1431472

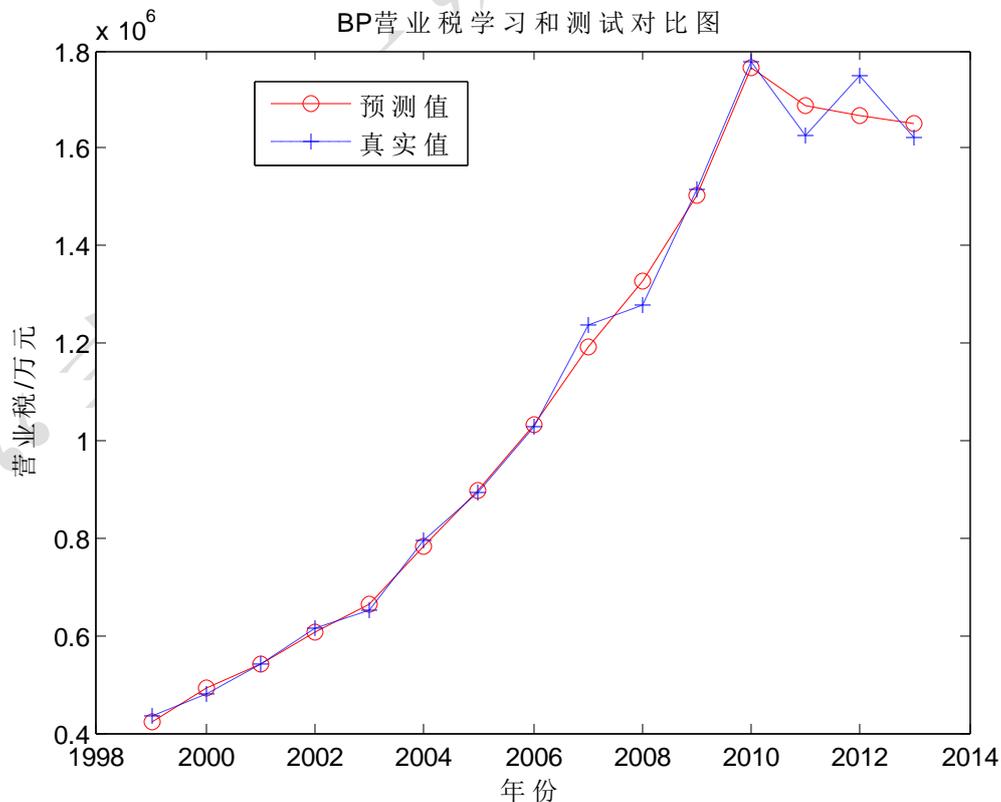


图 5.3 BP 神经网络营业税真实值与预测值对比图

5.4 企业所得税预测

利用Adaptive-Lasso 方法进行企业所得税影响因素的变量选择，通过表5-9可以看出，规模以上工业企业盈亏相抵后的利润总额和建筑业企业利润总额这两个因素的系数为0。这会因为建筑业企业利润总额与建筑业总产值存在线性关系，所以该变量被剔除。第二产业增加值(x1)、第三产业增加值(x2)、全社会固定资产投资额(x3)、城市商品零售价格指数(1978=100)(x4)、规模以上国有及国有控股工业企业亏损面(x6)、建筑业总产值(x7)、限额以上连锁店(公司)零售额(x9)、地方财政总收入(x10) 8个变量被选入影响企业所得税(y)的因素中。

表5-9 系数表

x1	x2	x3	x4	x5
0.01474	-0.00753	-0.00607	3507.63391	0.00000
x6	x7	x8	x9	x10
-8893.78600	0.02010	0.00000	0.00712	0.00526

对Adaptive-Lasso 变量选择方法识别的影响企业所得税的因素建立BP神经网络预测模型，其参数设置为误差精度 10^{-7} ，学习次数10000次，神经元个数为Lasso变量选择方法选择的变量个数 8。变量选择的指标的2014年及2015年数值通过R语言建立灰色预测模型得出，后验差比值、预测精度等级见表5-10，灰色预测模型有很好的效果。代入企业所得税所建立的BP神经网络预测模型，得到**企业所得税的2015年预测值为1292470万元**，相关数据见表5-11，其中红色字体的数据为预测数据。图5.4为BP神经网络企业所得税真实值与预测值对比图。

表5-10 灰色预测模型企业所得税相关因素精度表

	x1	x2	x3	x4	x6	x7	x9	x10
后验差比值	0.0696	0.0179	0.0743	0.2294	0.3182	0.0719	0.2904	0.1038
预测精度等级	好	好	好	好	好	好	好	好

表5-11 企业所得税及其相关因素历史数据和预测表

	x1	x2	x3	x4	x6	x7	x9	x10	y
1999	9310691	11152545	8782586	593.4		2470523		1881388	277375
2000	10216241	13767475	9236676	589.8		2561326		2199077	309764
2001	11122943	16320762	9782093	574.5	31.25	3403870		2719058	483421
2002	12113416	18895479	10092421	559.6	31.99	3733922	1053156	2690984	236416
2003	14859261	21627825	11751668	554.5	29.87	4785787	1154425	3005475	268360
2004	17880638	25453413	13489283	566.1	30.69	5459314	1434440	3384477	326556
2005	20452183	29787941	15191582	575.2	31.63	6331382	3621757	4088545	373397
2006	24415160	35118425	16963824	582.1	28.95	6870406	4196301	4767231	455820

2007	28257805	41646681	18633437	599	24.88	7507109	7068265	8389925	596693
2008	32278717	48903250	21055373	633.1	30.85	8754491	17829885	8431405	756412
2009	34051588	55607710	26598516	612.8	23.16	10134050	17019222	11076649	732282
2010	40022658	65574525	32635731	632.4	20.42	12805288	26192835	13991612	935248
2011	45769763	76419207	34122005	664.7	22.55	15613171	21639131	15351387	1061594
2012	47206504	86167948	37583868	677.3	20.9	17417072	21396742	15796804	1075045
2013	52273431	99643373	44545508	680.7	19.7	21828895	22659148	20881374	1155923
2014	63377150	118049300	49881480	681.3904	19.50026	23746700	38486970	25899680	1293134
2015	71600540	137165300	56932590	692.1273	18.65565	27870540	45686090	30994690	1292470

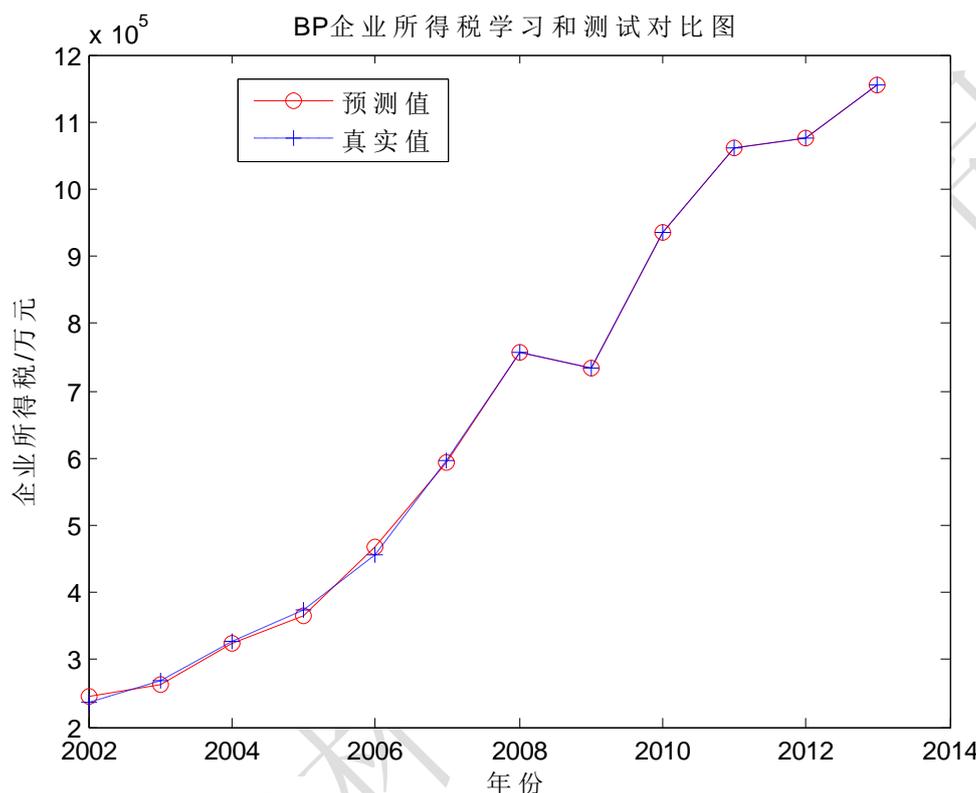


图 5.4 BP 神经网络企业所得税真实值与预测值对比图

5.5 个人所得税预测

利用Adaptive-Lasso 方法进行个人所得税(y) 影响因素的变量选择, 通过表5-12可以看出, 城市居民年人均可支配收入(x1)、地区生产总值(x4)、第二产业增加值(x5)和地方财政收入(x7)这四个因素进入选择, 其他因素的系数为0。

表5-12 系数表

x1	x2	x3	x4	x5	x6	x7
16.98345	0.00000	0.00000	-0.01035	0.01759	0.00000	0.02417

对Adaptive-Lasso 变量选择方法识别的影响个人所得税的因素建立BP神经网络预测模型, 其参数设置为误差精度 10^{-7} , 学习次数10000次, 神经元个数为Lasso变量选择

方法选择的变量个数 4。变量选择的指标的2014年及2015年数值通过R语言建立灰色预测模型得出，后验差比值、预测精度等级见表5-13，灰色预测模型有很好的效果。代入个人所得税所建立的BP神经网络预测模型，得到个人所得税的**2015年预测值为525004万元**，相关数据见表5-14，其中红色字体的数据为预测数据。图5.5为BP神经网络个人所得税真实值与预测值对比图。

表5-13 灰色预测模型个人所得税相关因素精度表

	X1	X4	X5	X7
后验差比值	0.0747	0.0349	0.0696	0.1038
预测精度等级	好	好	好	好

表5-14 个人所得税及其相关因素历史数据和预测表

	x1	x4	x5	x7	y
1999	12019	21391758	9310691	1881388	133621
2000	13967	24927434	10216241	2199077	185625
2001	14694	28416511	11122943	2719058	254892
2002	13380	32039616	12113416	2690984	159684
2003	15003	37586166	14859261	3005475	153080
2004	16884	44505503	17880638	3384477	167379
2005	18287	51542283	20452183	4088545	198017
2006	19851	60818614	24415160	4767231	231794
2007	22469	71403223	28257805	8389925	295316
2008	25317	82873816	32278717	8431405	353372
2009	27610	91382135	34051588	11076649	389824
2010	30658	107482828	40022658	13991612	472154
2011	34438	124234390	45769763	15351387	462098
2012	38054	135512072	47206504	15796804	439592
2013	42049	154201434	52273431	20881374	489777
2014	45375	183601900	63377150	24539000	498049
2015	50041	210726900	71600540	28792638	525004

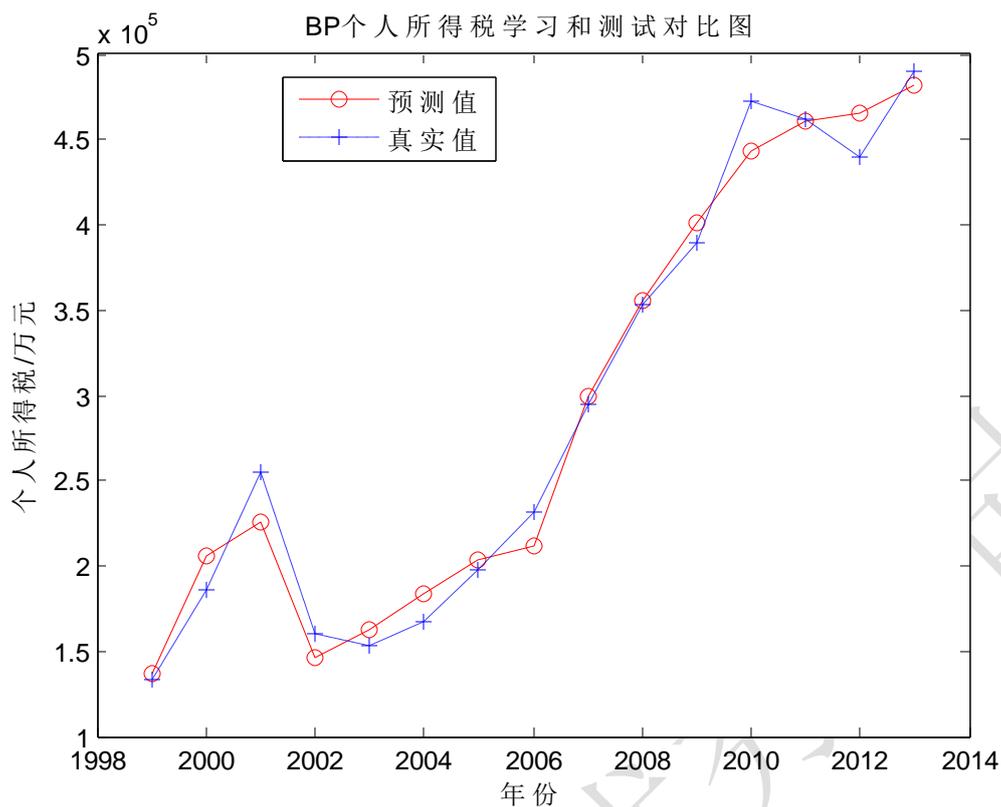


图 5.5 BP 神经网络个人所得税真实值与预测值对比图

5.6 政府性基金收入预测

相比于2006年及以往年份，2007年的广州土地出让金大幅上涨，而土地出让金收入的大幅上涨直接影响了政府性基金收入。所以为了数据的连续性，文章利用灰色预测法对2007年至2013年的政府性基金收入进行预测，得到2014年政府性基金收入为10595746万元，2015年政府性基金收入为13366207万元。灰色预测的后验差比值为0.3052，小于0.35，预测精度为好。

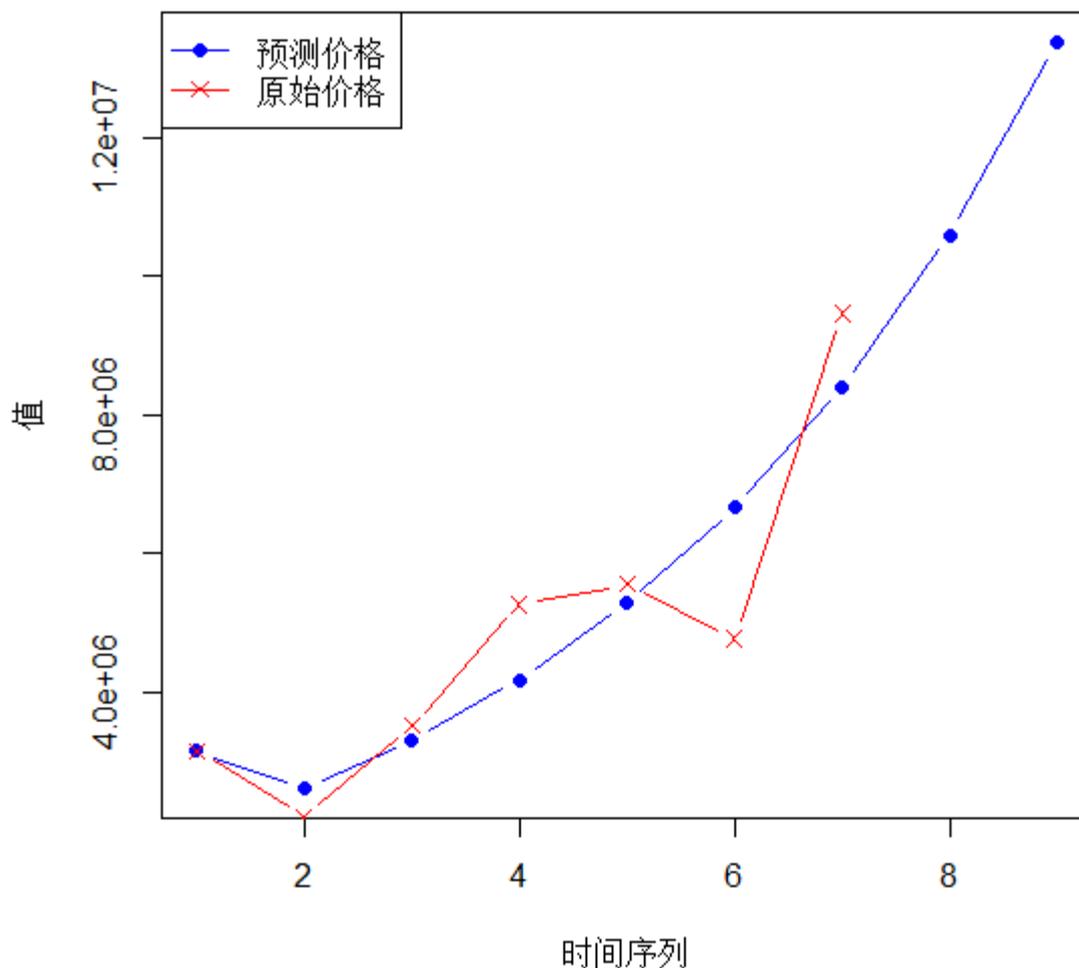


图 5.6 灰色预测政府性基金收入真实值与预测值对比图

第 6 章 对策建议

6.1 财政收支角度

(一) 经济增速放缓，楼市调控风暴下导致房地产业的税收减收，税收政策调整(个人所得税、增值税、营业税起征点提高等)使得相关税种减收，在这些环境下，广州财政收入增长面临考验。而且，随着其他省市一系列优惠措施出台，广州市税源流失情况日益严重，一些重点企业外迁到珠三角其他地市、长三角中心城市或将原在广州市业务全部或部分转移至外地总部。

财政收入增速放缓应成为改革和转型的号角。从短期来看，这可能意味着地方需要适应财政收入进入个位数增长的现实，需要决心和魄力克服既得利益的阻碍，主动推进改革，才能实现经济的转型和长远增长。

(二) 财政收入“缩水”的同时，广州的财政支出却出现较快增长，两者形成了鲜明的对比。

在广州市这样的财政收支情况下，广州可以适当削减三公支出，但是在公共民生支出必须保障，广州应该在开源上做文章，培育新的税源和经济增长点，特别是在战略性新兴产业方面，可以挖掘环保、节能、创意产业、金融服务业等。

6.2 税源角度

制约广州市财源的主要因素：现行中央、省对广州市财政体制的制约，广州市对财政收入组织和支配自由度不大；工业企业自主创新能力不足；第三产业税收结构有待优化，金融、物流、咨询、通讯等现代服务业的发展潜力未得到充分发挥；总部经济竞争力有待提升，总部经济税收聚集效应未能有效发挥；市场投资动力不足，全市投资增长更多地依赖于基础设施投资和服务业投资，尤其依赖政府的投资刺激计划。

针对广州市地方财源现状，文章提出如下财源优化的对策：

1、向中央和广东省争取有利的财政政策，力争地方参与优质税源税收的共享分成，提高财政收入组织与支配的自由度。

现行广东省对广州市的财政管理体制从 1996 年实行以来，一直没有大的改变，其中存在的一些不合理、不完善问题已对广州市财力造成较大的负面影响。广州市可利用《珠江三角洲地区改革发展规划纲要》实施的契机，争取广东省为广州市的财政发展营造更好的政策环境，进一步壮大广州市财力。尤其需要支持广州市引进世界顶尖企业，引进重大项目，促进金融业的发展，争取广东省在金融保险业营业税上增加对广州市的税收返还，形成产业发展—加大财政激励—产业加快发展的良性循环机制，达到省、市“水涨船高”的财政收入态势。

2、把第三产业发展成骨干财源，拓展第三产业发展空间

鉴于第三产业对税收收入尤其是地方税收收入的贡献率较高，其在产业结构中比重的提高必将对广州市税收产生积极作用。因此，广州市应积极培育新的财源增长点，尤其是现代物流、金融保险、邮电通信和管理咨询等税收量高的现代服务业，要逐步改变现行对第三产业的财税优惠政策仅限于金融、软件和动漫、连锁经营等少数几个行业、某些优惠政策则仅限于特定企业的状况，制定针对物流、社会中介、会展等行业具有普惠性的财税优惠政策，全面推动第三产业的发展。

3、扶持民营经济发展，培育新的财政收入增长点

民营经济从事的行业大多属于第二、第三产业，且其带来的税收收入多为地方固定收入，无须与中央、省分成，是地方财力增长的重要“引擎”。近年来，广州市民营经济迅速发展，显示出巨大的生命力。但从总体上看，民营企业虽然有其市场化程度高、经营灵活、社会负担较轻等优势，但发展水平却与所应具有法定身份不相适应。广州市应进一步放宽市场准入条件，加大对民营企业的金融支持力度，优化民营企业发展环境等。

4、力促总部经济发展，强化总部经济税收聚集效应

据统计，2014年广州市共认定总部企业323家，不到企业总数的1%，但贡献了全市税收的三分之一。鉴于广州的物流业、金融业、会展业、中介服务业等日益完善，发展总部经济的条件向好，大力发展总部经济，使之成为税源的重要增长点。整合珠三角资源以拓展广州发展腹地。任何一个经济体的可持续发展都有赖于与周边地区的良好互动合作。为此，广州可从市内各区、广佛经济圈和穗港澳三个层面实现资源整合。

5、加快高新技术产业发展，支持自主创新，培育优质财源

高新技术产业利大税高，对财源建设的作用日趋明显。加快高新技术产业发展，引领支撑广州市实施提升产业竞争力与提升自主创新能力。

6、提升招商引资层次，坚持引进优质项目，引资与增税并重

在利用外资上，引入技术含量高、环境保护、资源消耗、新增就业和国内配套比例等综合指标，建立起科学的吸引外资评价体系，逐步形成内外资企业公平竞争、政策一致的市场经济环境。

7、加强财政收入征管，完善财税征管联动机制

收入征管部门要密切关注组织收入中的增减收现象，从经济、体制、政策等角度分析原因，测算各种因素对收入的影响程度，及时提出加强和改进征管的建议和措施。

参考文献

- [1] 尚元君,殷瑞锋.对地方政府财政收入能力影响因素的实证分析[J].中央财经大学学报.2009,5:12-15.
- [2] 刘心竹,李昊等.我国地方财政收入影响因素的实证分析[J].中国集体经济.2012,9:84-85.
- [3] 成军.地方财政收入预测模型设计及实证分析[J].经济研究参考.2003,88:27-34.
- [4] 何邓娇.广州市财政收入影响因素的实证研究[J].经济研究导刊.2014,15:104:106.
- [5] 邓宏乾.国城市主体财源问题研究--房地产税与城市土地地租[M],2007.
- [6] 刘晓.湖南财政收入及财源建设实证分析[M],2008.
- [7] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso[J]. J. Royal. Statist. Soc B., Vol. 58, No. 1, pages 267-288.
- [8] Zou H. The Adaptive Lasso and its oracle properties [J]. Journal of the American Statistical Association, 2006, 101(476): 1418-1429
- [9] 陈共.财政学.北京:中国人民大学出版社,2012:141 页.