

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：广东金融大学

作品成员：周婉倩 侯晓佳 陈国权

指导教师：骆世广

广州市财政收入的实证分析与预测

摘要：科学、合理地预测地方财政收入，对于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。为了探究广州市财政收入的影响因素和预测 2015 年的财政总收入及各个类别收入，选取地区生产总值、第二三产业产值比、税收收入、社会从业人员、固定资产投资额，收集 1985-2013 年广州市的相关数据为样本。

首先，建立 VAR 模型，通过 Johansen 协整检验，判断出影响广州财政总收入的关键因素是社会从业人员，地区生产总值和税收收入。

接着，运用 GM(1, 1)白化预测模型，预测出 2015 年五个影响因素的预测值，再代入 VAR 模型，求得 2015 年广州财政总收入预测值为 23978544.28 万元；将财政收入分为个人所得税、企业所得税、营业税、增值税、其他税收收入、非税收收入和政府性基金七大类，利用时间序列和神经网络模型，个人所得税、企业所得税、营业税、增值税、其他税收收入、非税收收入和政府性基金的其预测值分别为 346249.60、982475.60、1708531.43、2353663.07、4209813.22、2929400.73、15986526.18 万元。

最后，在 VAR 模型的基础上，通过 VEC 模型、脉冲响应函数和方差分解，对财政收入与影响因素的关系进行进一步分析，并提出实现充分就业，以经济增长带动财政收入增长，提高税收征管执行力度和优化产业结构等相关建议。

关键词：财政收入；VAR 模型；Johansen 检验；灰色预测；时间序列；神经网络

The empirical analysis and forecast of Guangzhou fiscal revenue

Abstract: Scientific and reasonable prediction of local fiscal income, to correctly handle the relationship between local finance and economy has very important significance. In order to explore the key factors influencing the Guangzhou revenue, and forecast 2015 total revenue and income in each category, this article selects GDP, the secondary industry and the tertiary industry output value, tax revenue, social workers, fixed assets investment and collects in the city of Guangzhou in 1985-2013 phase closed data as the sample.

First, through the establishment of VAR model and Johansen cointegration test, the key factors affecting the total fiscal revenue of Guangzhou are social workers, GDP and tax revenues.

Then, using GM (1,1) gray prediction model to predict the predicted values for each variable in 2015. And then substituted into the VAR model, obtained in 2015 the total revenue of Guangzhou predictive value of 239785.4428 million. For the 2015 income projections for each category, this article divides into personal income tax revenue, corporate income tax, sales tax, VAT, other tax revenues, non-tax revenue and government funds received seven categories, time series and neural network models, seeking obtaining predictive values were 3462.4960, 9824.7560, 17085.3143, 23536.6307, 42098.1322, 29294.0073, 159865.2618 million.

Finally, based on VAR model, through the VEC model, impulse response function and variance decomposition, relationship of fiscal income and the influencing factors were further analysed, and we make some related suggestions like realizing full employment, promoting economic growth to drive revenue growth, improving tax collection and management implementation efforts, optimizing the industrial structure and so on.

Key words: Fiscal revenue, VAR model, Johansen test, gray prediction, time series, neural network

目 录

1.挖掘目标	1
2.分析方法与过程	1
2.1.总体流程.....	1
2.2.具体步骤.....	2
3.结论与建议	16
3.1.结论.....	16
3.2.建议.....	17
4.参考文献	18

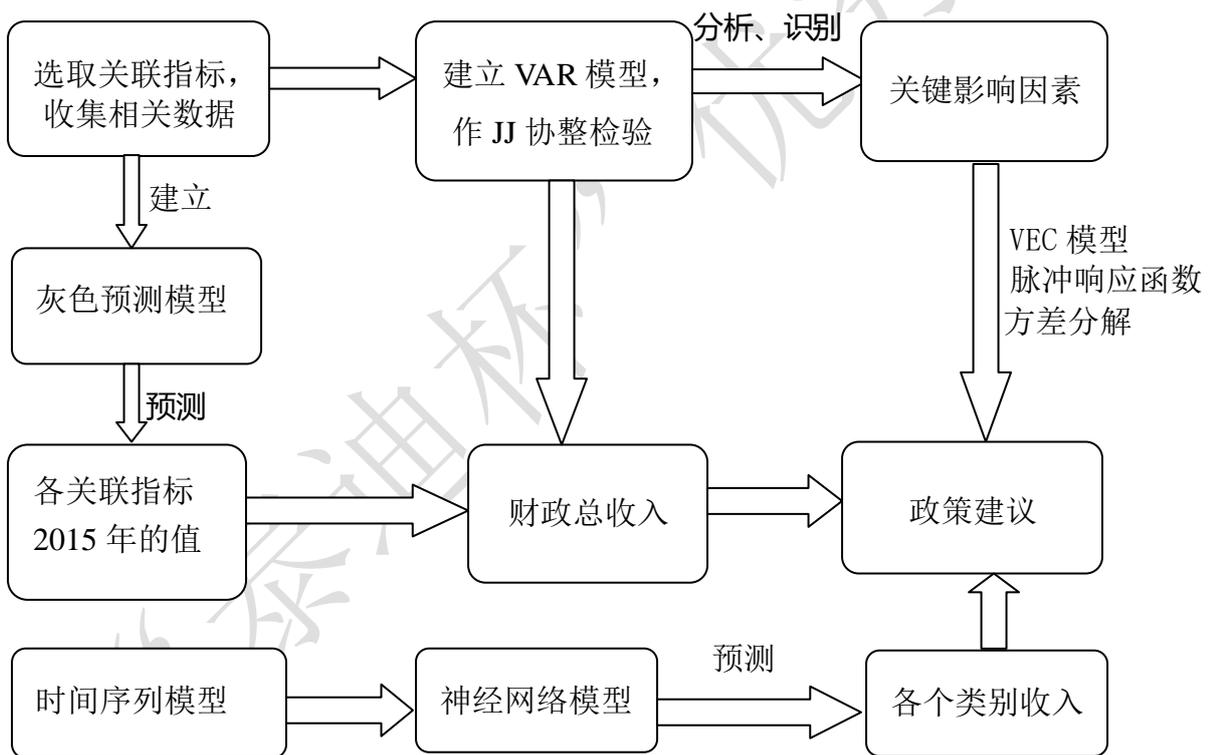
“泰迪杯” 优秀作品

1.挖掘目标

本次建模的目的是对广州市财政收入进行实证分析与预测。首先，利用 1985-2013 年广州市财政收入与地区生产总值、第二第三产业产值比、税收收入、社会从业人员、固定资产投资额的相关数据，分析、识别影响财政收入的关键因素；接着，基于数据挖掘技术，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入及支出等情况，向广州市财政局提出促进就业等几点建议。

2.分析方法与过程

2.1.总体流程



主要包括以下步骤：

步骤一：选取影响财政收入的因素和收集相关数据

步骤二：建立 VAR 模型，通过 JJ 协整检验分析关键影响因素

步骤三：结合灰色预测模型和 VAR 模型，预测 2015 年广州的财政收入

步骤四：利用时间序列预测法、神经网络模型对财政收入进行细化预测

步骤五：建立 VEC 模型，利用脉冲响应函数和方差分解进行进一步分析

2.2.具体步骤

步骤一：财政收入影响因素的选取和相关数据的收集处理

(1)因素的选取：影响财政收入的因素有很多，结合广州市实际情况，本文选取广州市地区生产总值 X_1 、第二第三产业产值比 X_2 、税收收入 X_3 、社会从业人员 X_4 、固定资产投资额 X_5 作为考虑因素进行分析研究。

(2)数据的收集：利用大赛给出的附件和广州市统计局等网站收集了 1985-2013 年财政总收入 Y /万元、地区生产总值 X_1 /万元、第二第三产业产值比 X_2 、税收收入 X_3 /万元、社会从业人员 X_4 /人、固定资产投资额 X_5 /万元的数据。

(3)数据预处理：

利用 EViews 6.0 软件，导入 1985-2013 年序列 Y 与 X_i 的数据，分别做出时序折线图，进行初步的观察与分析，结果如图 1-6 所示。（Eviews 操作过程和结果详见附录一）

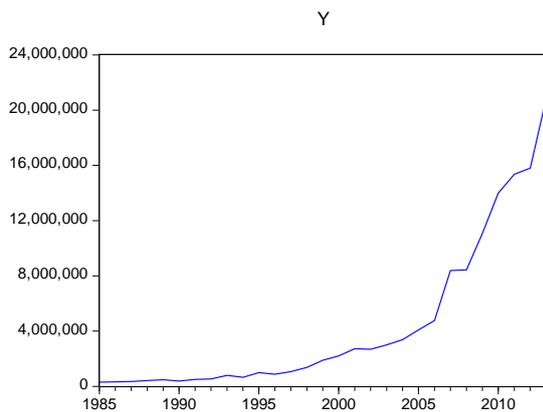


图 1 Y 的时序折线图

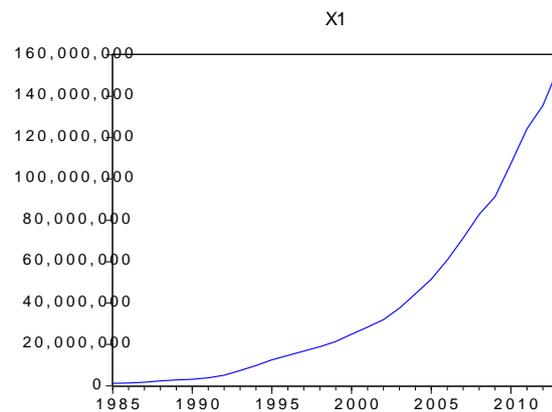


图 2 X_1 的时序折线图

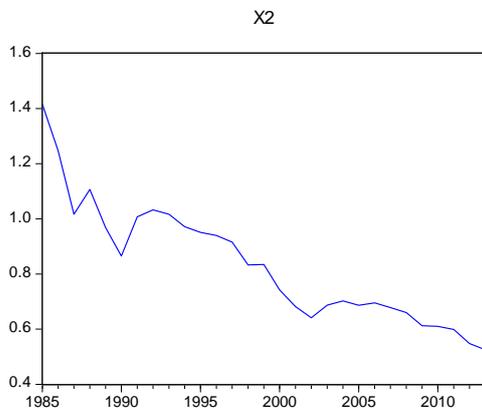


图 3 X_2 的时序折线图

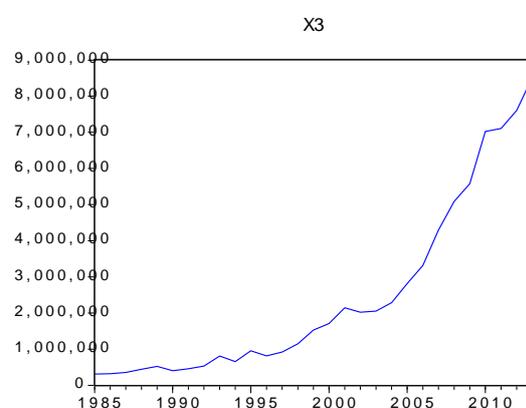


图 4 X_3 的时序折线图

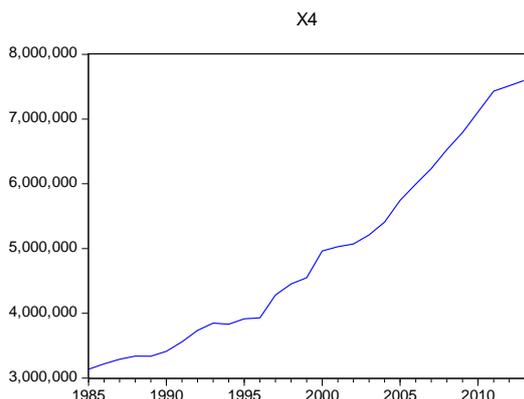


图 5 X_4 的时序折线图

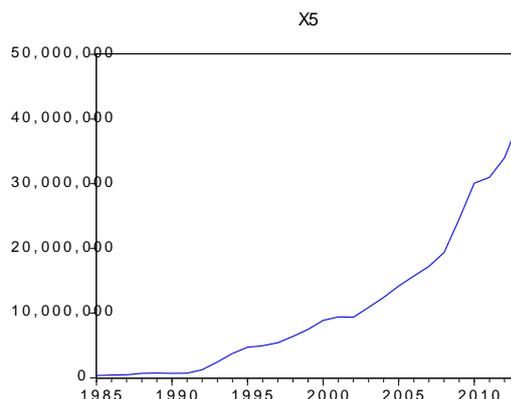


图 6 X_5 的时序折线图

从图 1-6 可知，七个指标的时间序列数据均呈明显的指数上升趋势，直观来看，显著非平稳。根据序列的上升趋势，为了减小上升趋势中的波动，同时消除 Y 与 X_i 的异方差，对序列 Y 与 X_i 进行对数化处理（对数化的数据并不影响原序列之间的协整关系），相应地得到对数序列 YL 与 X_iL 。

为了证实对数序列是否平稳，进一步对对数序列 YL 与 X_iL 做 ADF 检验，结果如表 1。

表 1 对数序列 YL 与 X_iL 的 ADF 检验结果

	t 统计量				P 值
	ADF 检验值的 t 统计量	1% level	5% level	10% level	
YL	1.270052	-3.69987	-2.97626	-2.62742	0.9978
X1L	-1.47889	-3.72407	-2.98623	-2.6326	0.5275
X2L	-1.52636	-3.68919	-2.97185	-2.62512	0.5059
X3L	0.169464	-3.69987	-2.97626	-2.62742	0.9652
X4L	0.753017	-3.68919	-2.97185	-2.62512	0.9912
X5L	-1.82293	-3.71146	-2.98104	-2.62991	0.3616

由表 1 可知，对数序列 YL 与 X_iL 计算出来的 ADF 检验值的 t 统计量都比 1%、5%和 10%的检验水平下的 t 统计量大，P 值显著较大，也就是说，在 1%、5%和 10%的检验水平下不能拒绝原假设，可以认为对数序列 YL 与 X_iL 分别至少有一个单位根，即在 1%、5%和 10%的检验水平下，对数序列 YL 与 X_iL 不平稳

为得到平稳序列，我们对对数序列 YL 与 X_iL 进行一阶差分，再对其进行 ADF 检验，结果如表 2。

表 2 对数序列 YL 与 X_iL 一阶差分的 ADF 检验结果

	t 统计量				P 值
	ADF 检验值的 t 统计量	1% level	5% level	10% level	

ΔYL	-8.45539	-3.69987	-2.97626	-2.62742	0.0000
$\Delta X1L$	-2.62082	-3.69987	-2.97626	-2.62742	0.1000
$\Delta X2L$	-5.44056	-3.69987	-2.97626	-2.62742	0.0001
$\Delta X3L$	-7.36399	-3.69987	-2.97626	-2.62742	0.0000
$\Delta X4L$	-4.63559	-3.69987	-2.97626	-2.62742	0.0010
$\Delta X5L$	-4.26158	-3.71146	-2.98104	-2.62991	0.0027

由表 2 可知,对数序列 YL 与 X_iL 的一阶差分序列 ADF 检验值的 t 统计量都比 1%、5% 和 10%的检验水平下的 t 统计量小, P 值显著偏小 (小于 0.1), 也就是说, 在 1%、5% 和 10%的检验水平下可以拒绝原假设, 即可以认为对数序列 YL 与 X_iL 的一阶差分没有单位根。

因此, 对数序列 YL 与 X_iL 是一阶差分平稳 $I(1)$ 的。

步骤二：建立 VAR 模型，通过 JJ 协整检验分析关键影响因素

(1)单方程协整检验(E_G 两步法)

由于对数序列 YL 与 X_iL 是一阶差分平稳 $I(1)$ 型的时间序列, 需要对这些变量进行进一步的协整检验, 以确定变量之间是否存在长期稳定的均衡关系。目前, 常用的协整检验有 E_G 两步法和 Johansen 协整检验法两种, 本文先从 E_G 两步法入手, 寻找对数序列 YL 与 X_iL 之间的稳定性关系, 再通过 Johansen 协整检验法得出多个经济变量之间的关系。

第一步：运用普通最小二乘法估计参数建立对数序列 YL 与 X_iL 之间的关系, 得到估计方程为:

$$YL = -22.7477 - 0.0560X_1L + 0.1093X_2L + 0.9904X_3L + 1.6158X_4L - 0.0468X_5L \tag{1}$$

该方程相应的检验结果如表 3。

表 3 方程 (1) 的检验结果

Dependent Variable: YL				
Method: Least Squares				
Date: 05/21/15 Time: 10:37				
Sample: 1985 2013				
Included observations: 29				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-22.74770	5.756931	-3.951359	0.0006
X1L	-0.055956	0.135540	-0.412839	0.6835
X2L	0.109331	0.171303	0.638231	0.5296
X3L	0.990412	0.134332	7.372870	0.0000
X4L	1.615814	0.501313	3.223165	0.0038
X5L	-0.046752	0.102387	-0.456620	0.6522

R-squared	0.997878	Mean dependent var	14.46396
Adjusted R-squared	0.997416	S.D. dependent var	1.354704
S.E. of regression	0.068862	Akaike info criterion	-2.331438
Sum squared resid	0.109065	Schwarz criterion	-2.048549
Log likelihood	39.80585	Hannan-Quinn criter.	-2.242840
F-statistic	2162.702	Durbin-Watson stat	2.131984
Prob(F-statistic)	0.000000		

由表 3 分析可得，方程调整后的 F 值对应的 P 值为 0，修正后的拟合度 R^2 较大接近于 1，DW 统计量在 2 附近，表明模型的总体效果显著，拟合度较高，但有部分系数不是非常显著，总体上模型估计效果比较理想。

第二步：利用模型估计结果生成残差序列并对其进行单位根检验，检验结果如表 4。

表4 残差序列的单位根检验结果

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-5.598590	0.0001
Test critical values:		
1% level	-3.689194	
5% level	-2.971853	
10% level	-2.625121	

由表 4 分析可得，该残差序列平稳，这说明，对数序列 YL 与 X_tL 之间存在长期的均衡关系。

(2) JJ 协整检验

JJ 协整检验是一种以 VAR (p) 模型为基础，因此，在进行 JJ 检验之前，先要建立初步的 VAR 模型，并由此考察模型的滞后长度。

◇ 确定 VAR 模型的最大滞后长度

对对数序列 YL 与 X_tL 尝试建立起 VAR 模型，并查看 LR、AIC、SC 等多种检验准则，结果如表 5。

表 5 不同滞后值下的 LogL 值、LR 值、AIC 值和 SC 值

Lag	LogL	LR	FPE	AIC	SC	HQ
0	106.0728	NA	2.43e-11	-7.412797	-7.124833	-7.327170
1	269.4503	242.0408*	2.09e-15	-16.84817	-14.83242*	-16.24878
2	313.8625	46.05708	1.75e-15*	-17.47129*	-13.72776	-16.35815*

由表 5 分析可知，AIC 最小值对应的 p 值为 2，SC 最小值对应的 P 值为 1，相互矛盾，所以根据这两个信息准则不能确定滞后值 P，只能用似然比确定 P 值，直观上看便是取*数多的滞后值，即滞后值 P 为 2。所以可以考虑建立 VAR (2) 模型。

◇ VAR 模型的建立

对对数序列 YL 与 X_iL 建立 VAR (2) 模型, 方程式如 (2)。

$$YL_t = + \begin{bmatrix} -0.3011 \\ -0.7036 \\ 0.5898 \\ -0.0320 \\ 3.7706 \\ 0.1234 \end{bmatrix}^T \begin{bmatrix} YL \\ X_1L \\ X_2L \\ X_3L \\ X_4L \\ X_5L \end{bmatrix}_{t-1} + \begin{bmatrix} 0.0479 \\ 0.8279 \\ -0.0105 \\ 0.0567 \\ 2.0008 \\ -0.1541 \end{bmatrix}^T \begin{bmatrix} YL \\ X_1L \\ X_2L \\ X_3L \\ X_4L \\ X_5L \end{bmatrix}_{t-2} - 72.0464 \quad (2)$$

根据 VAR 模型的分析, 对数序列 YL 与 X_iL 服从 I(1)过程, 由协整的定义可知, 对数序列 YL 与 X_iL 的某个线性组合可能是平稳的 I(0)过程。因此, 可以利用 JJ 检验来研究货对数序列 YL 与 X_iL 之间的关系, 建立对数序列 YL 与 X_iL 的经济模型。

根据 Johanson 协整检验的理论, 先确定的最佳滞后阶数为 P=2, 检验假设是含截距项、不含时间项的线性趋势, 其检验结果如表 6。

表 6 Johanson 非约束协整关系检验结果 (协整阶数为 2)

至少存在协整关系个数r	特征值	迹统计量	5%显著水平下的临界值	P值	最大特征值统计量	5%显著水平下的临界值	P值
0	0.856930	143.2200	95.75366	0.0000	52.49945	40.07757	0.0012
1	0.800285	90.72055	69.81889	0.0005	43.49331	33.87687	0.0027
2	0.541533	47.22724	47.85613	0.0572	21.05641	27.58434	0.2728

从表 6 可知, 在 5% 的显著水平下, 迹统计量 (90.72055>69.81889) 和最大特征值统计量 (43.49331>33.87687) 都表明应该拒绝没有协整关系的原假设 (r=0), 而且应该接受至少有 2 个协整关系, 一般而言, 第一个协整向量具有较强的经济解释能力, 对该协整向量进行正规化, 得到的具体方程为:

$$YL = -1.8518X_1L - 0.7033X_2L - 1.4932X_3L + 13.1696X_4L + 1.0158X_5L + 151.6015 \quad (3)$$

(0.39531) (0.52529) (0.47731) (1.63655) (0.28236)

注: 括号中为对应系数标准差的值

根据协整方程 (3), 其他因素指标保持不变时, 当 X_1L 每变化1.8518个单位, 或者当 X_2L 每变化0.7033个单位, 或者当 X_3L 每变化1.4932个单位, 或者当 X_4L 每变化13.1696个单位, 或者当 X_5L 每变化1.0158个单位时, YL 变化一个单位。也就是说, 从各变量的系数绝对值来看, 对广州财政总收入影响最大的因素是社会从业人员即社会就业, 其次是地区生产总值即经济发展水平, 税收收入, 较小是固定资产投资额, 最小是第二三产业产值比即产业结构。

因此, 针对问题一, 考虑地区生产总值、第二三产业产值比、税收收入、社会从业

人员和固定资产投资额这五个影响因素的基础上，社会就业，经济发展水平和税收收入是影响广州财政总收入的关键因素。

步骤三：结合灰色预测模型和 VAR 模型，预测 2015 年广州的财政收入

针对问题二，针对本文研究的时间数据序列特点，样本数据较少，理论上可采用灰色预测法。灰色预测法主要是针对小样本，贫信息根据客观事物的过去和现在的发展规律，对其未来的发展趋势和状况进行描述和分析。

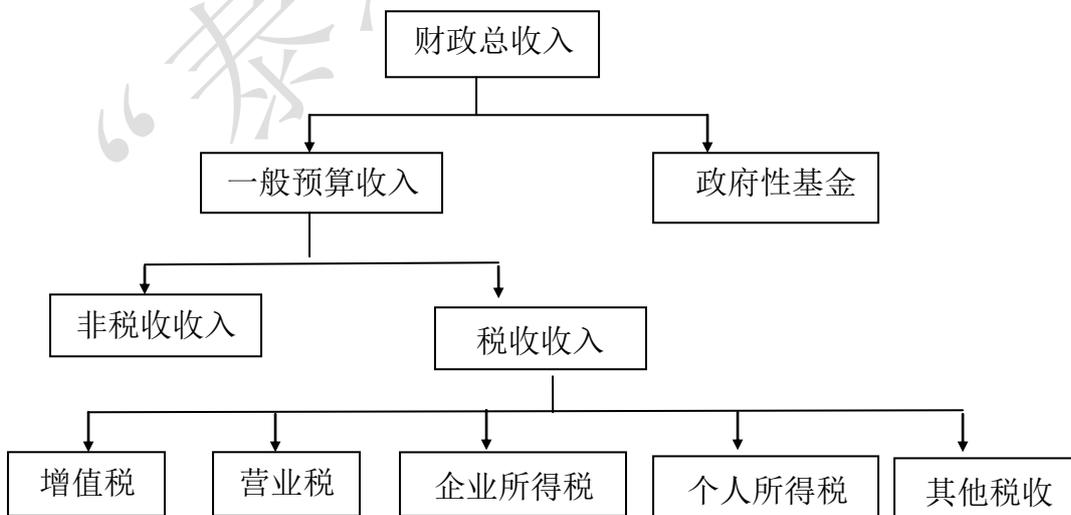
对于预测2015年广州财政总收入，本文运用GM(1, 1)白化预测模型 $\hat{x}^{(1)}(t+1) = [x^{(0)}(1) - \frac{u}{a}]e^{-at} + \frac{u}{a}$ ，预测出2015年的 X_iL 的预测值（具体相关数据见表7），再代入VAR(2)模型即方程(2)，求得2015年广州财政总收入预测值为23978544.28万元。（相关Matlab程序代码和结果见附录二）

表7 预测 X_iL 的相关数据

	α	μ	2015 年预测值
地区生产总值 X1	-0.1378	2.19E+07	210730000
第二三产业产值比 X2	0.019	0.7537	0.54963
税收收入 X3	-0.1306	1.36E+06	11817000
社会从业人员	-0.0377	4.53E+06	8426000
固定资产投资额 X5	-0.1278	6.06E+06	50897000

步骤四：利用时间序列预测法、神经网络模型对财政收入进行细化预测

对于预测 2015 年各类别收入，本文将财政收入划分为七个类别收入如下：



采用时间序列和神经网络模型，对七大类收入进行预测求解。因为收入和时间有很强关联性，时间序列预测方法更能反映其变化规律。神经网络是一种模拟人类大脑神经网络结构和功能，进行分布式并行信息处理的算法数学模型，实际上是由大量简单处理单元以某种方式互相连接而成，通过调整内部这些大量处理单元之间互相连接的关系，从而达到处理信息的目的。

第一步：对数据进行预处理，然后在 SPSS Clementine 数据挖掘平台上建立时间序列模型，分析 1999 年至 2013 年的收入数据，并对在 2015 年的收入进行预测，使用的数据流如图 7 所示：

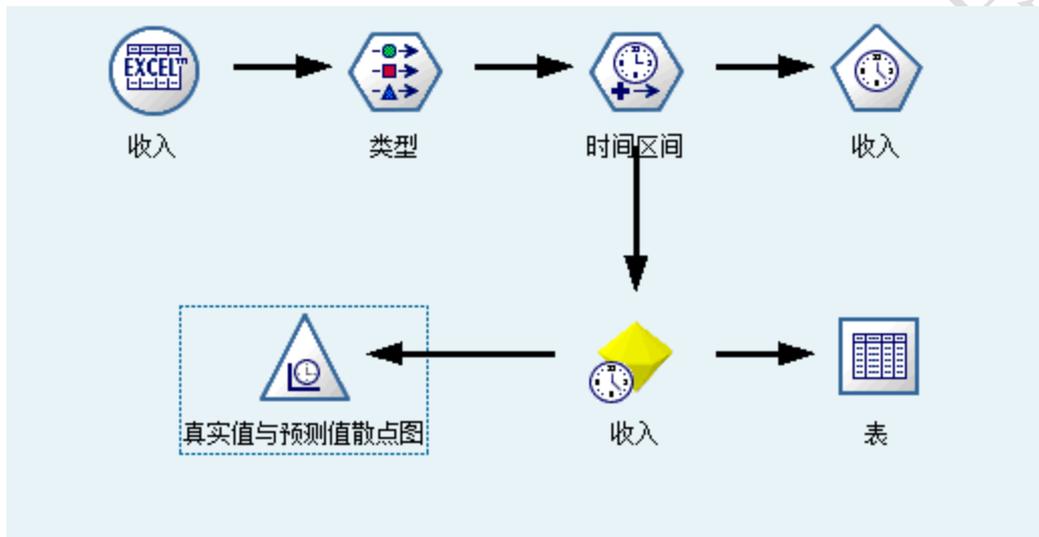


图 7 收入预测

第二步：画出各类收入真实值与预测值之间的时间散点图如图 8：

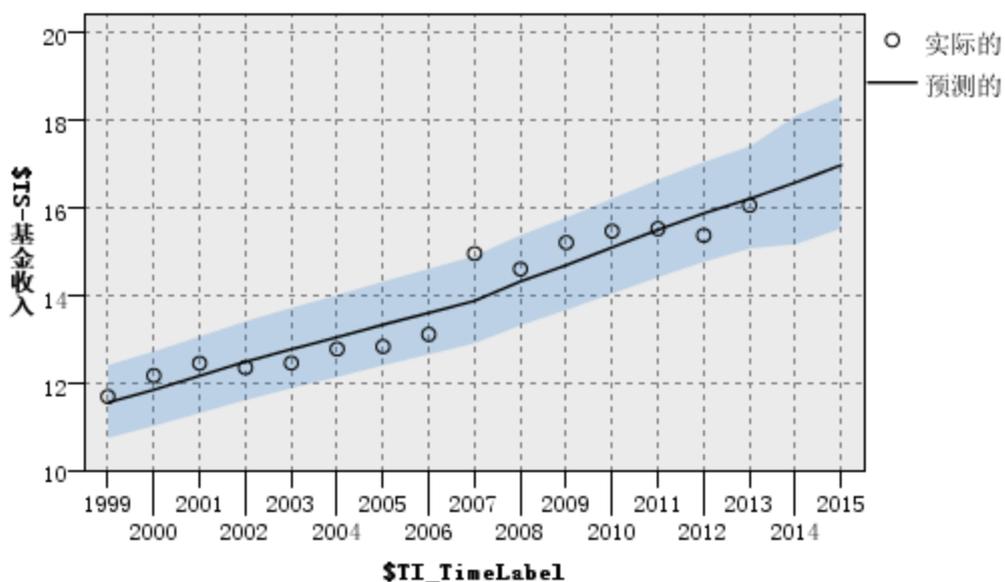


图 8 基金收入拟合时间散点图

由图 8 可知，基金收入的实际值基本落在拟合曲线上，所以模型预测效果较好。（其余六类收入拟合时间散点图详见附录三）

第三步：验证模型的准确度： $1 - \frac{|y - \hat{y}|}{y}$ ，其中 y 表示实际收入， \hat{y} 表示预测收入，营业税、增值税、非税收收入，基金预算收入，其他税收收入的准确度如表 8。

表 8 各类收入的准确度

	时间	营业税	增值税	其他税收	非税收	政府性基金
拟合准确度	1999			99.79%	99.10%	98.82%
	2000	91.94%	78.28%	98.79%	99.88%	97.26%
	2001	95.44%	89.86%	98.61%	95.78%	97.62%
	2002	98.06%	89.65%	98.68%	99.65%	98.93%
	2003	92.61%	85.89%	98.92%	84.73%	97.52%
	2004	92.64%	63.82%	99.14%	89.58%	97.93%
	2005	98.42%	81.61%	90.65%	93.95%	96.24%
	2006	95.11%	98.60%	91.57%	92.33%	96.38%
	2007	90.09%	99.09%	97.23%	83.90%	92.22%
	2008	96.83%	97.31%	98.51%	93.03%	98.01%
	2009	90.02%	96.38%	98.26%	88.47%	96.50%
	2010	90.08%	94.89%	98.52%	80.78%	97.53%
	2011	85.44%	89.96%	99.25%	87.83%	99.84%
	2012	97.88%	97.34%	98.75%	95.39%	96.83%
2013	87.12%	85.56%	98.84%	74.29%	99.08%	
预测值	2014	1708531.43	2353663.07	4209813.22	2929400.73	15986526.18
	2015	1793542.86	2491309.14	4995527.99	3549795.99	23815494.59

由表 8 可知，利用时间序列模型，求解营业税，增值税，其他税收，非税收，政府性基金的预测值得拟合准确度都比较好，基本处于 85% 以上，所以可看出该模型的拟合情况较好，对于 2015 年的预测值是可信的。

第四步：对税收收入进行分类别预测时，发现个人所得税和企业所得税由于样本量过少，用时间序列模型进行拟合时存在较大误差。所以采用时间序列模型预测出来的各变量值代入神经网络模型对个人所得税和企业所得税未来两年的值进行预测，如图 9、图 10。

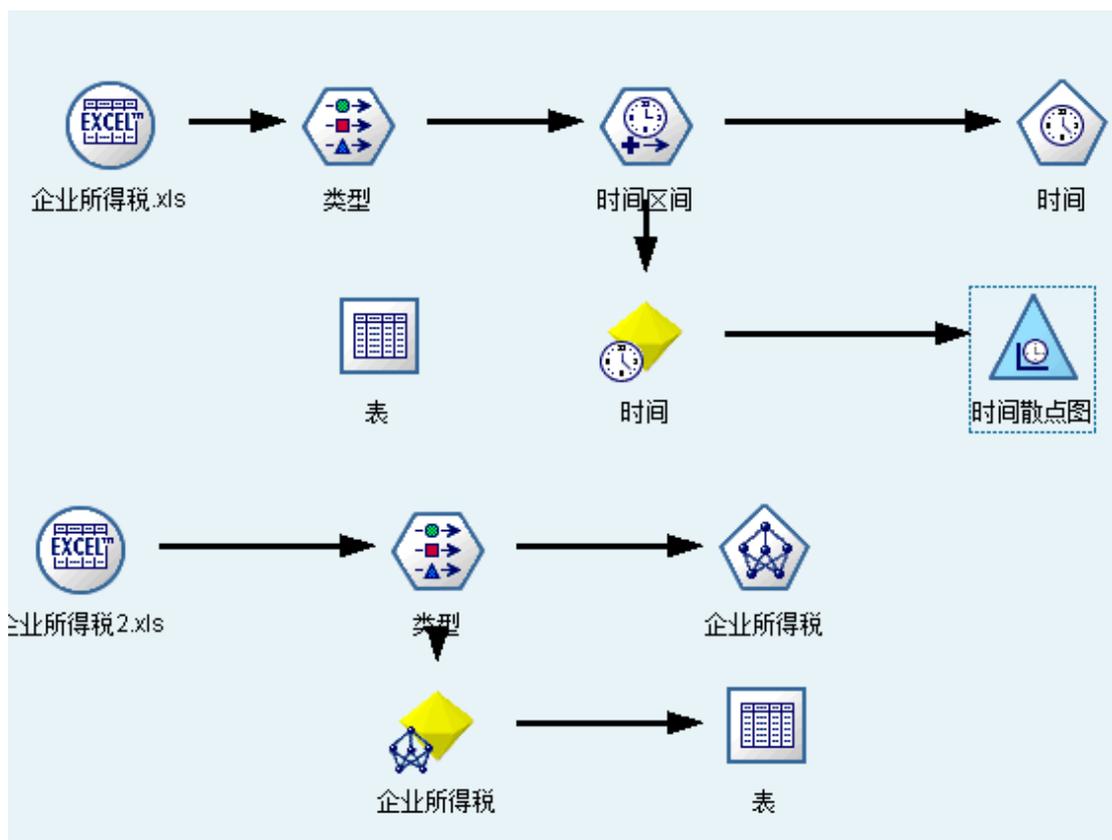
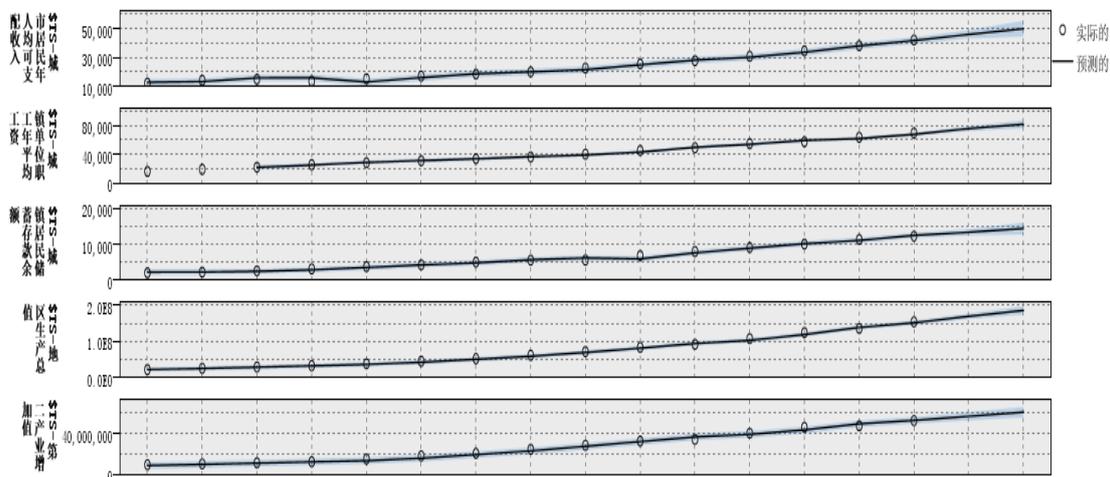


图 9 时间序列模型与神经网络模型数据流



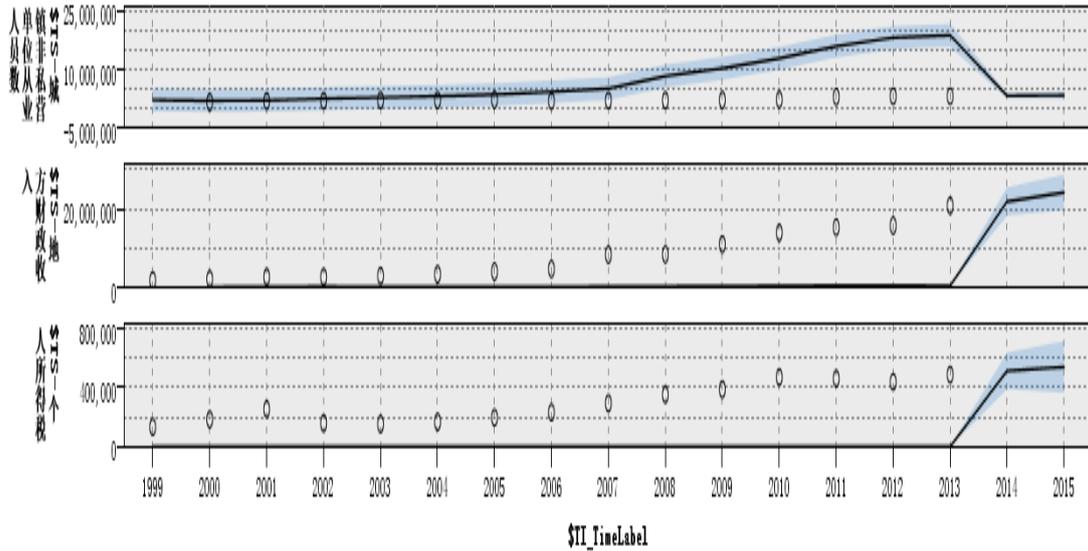


图 10 个人所得税拟合时间散点图

由图 10 可看出，个人所得税的城镇非私营单位从业人员数与地方财政收入这两个影响因素的拟合偏差较大，所以将其剔除，再进行神经网络模型网络。

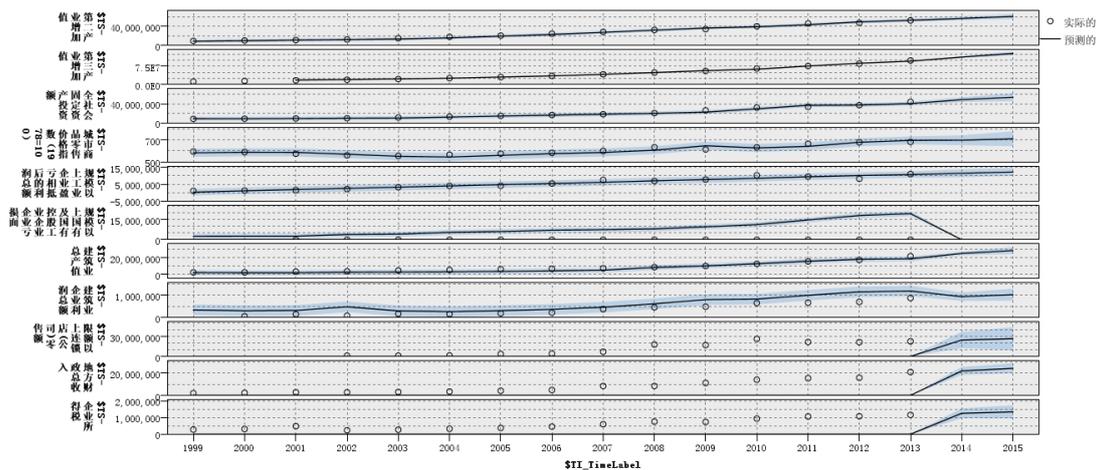


图 11 个人所得税拟合时间散点图

由图 11 可以看出，企业所得税的限额以上连锁店(公司)零售额、建筑业企业利润总额、规模以上国有及国有控股工业企业企业亏损面与地方财政收入这四个影响因素的拟合偏差较大，所以将其剔除，再进行神经网络模型预测。模型拟合准确度如表 9：

表 9 神经网络模型拟合准确度

	时间	个人所得税	企业所得税
	1999	89.89%	97.56%
	2000	81.54%	88.44%
	2001	61.38%	56.84%

拟合 准确度	2002	98.67%	82.12%
	2003	86.59%	87.80%
	2004	85.31%	94.38%
	2005	89.99%	96.72%
	2006	91.48%	93.00%
	2007	99.32%	90.91%
	2008	99.41%	97.19%
	2009	99.45%	92.05%
	2010	89.79%	96.11%
	2011	96.16%	87.06%
	2012	96.17%	86.48%
	2013	94.97%	83.32%
预测值 /万元	2014	325768.5	974273.6
	2015	346249.6	982475.6

第五步：从以上模型预测得出各个类别收入的预测值如表 10。

表 10 2015 年各类别收入的预测值

时间	个人所得税	企业所得税	营业税	增值税
2015	346249.6	982475.6	1708531.43	2353663.07
时间	其他税收收入	非税收收入	政府性基金	
2015	4209813.22	2929400.73	15986526.18	

步骤五：建立 VEC 模型，利用脉冲响应函数和方差分解进行进一步分析

(1) 建立 VEC 模型

Johansen 协整检验结果表明对数序列 Y_L 与 X_L 之间存在协整关系，即存在长期均衡关系，但从短期来看这种关系可能会失衡，根据恩格尔与格兰杰提出的格兰杰表述定理，如果变量之间是协整的，则它们间的短期非均衡关系总能由一个误差修正模型表述。

接下来我们用上面已经正规化的协整方程（即误差修正项），估计向量误差修正（VEC）模型，即对各个变量施加协整关系约束条件的向量自回归，由此建立起 VEC 模型为：

$$\Delta YL_t = -0.0632VECM_{t-1} + \begin{bmatrix} -0.5846 \\ -1.3488 \\ -0.6138 \\ 0.0093 \\ 2.0303 \\ 0.4822 \end{bmatrix}^T \begin{bmatrix} \Delta YL \\ \Delta X_1L \\ \Delta X_2L \\ \Delta X_3L \\ \Delta X_4L \\ \Delta X_5L \end{bmatrix}_{t-1} + 0.3479 \quad (4)$$

由方程 (4) 分析可知，方程误差修正项系数为负，表明存在反向修正机制，从短期来看，对广州财政总收入影响最大的因素仍然是社会从业人员，其次是地区生产总值，第二第三产业产值比，较小是固定资产投资额，最小是税收收入。

(2) 脉冲响应函数分析

脉冲响应函数描述的是内生变量对残差冲击的反应，用于分析衡量随机扰动项的一个标准差新息冲击时被解释变量的响应程度和持续时间。

✧ Var 的稳定性判定

对于 4.3.1 中建立的 VAR(2) 即方程 (4-2)，AR 特征多项式有 12 个根，其平稳性检验结果如图 4-7。

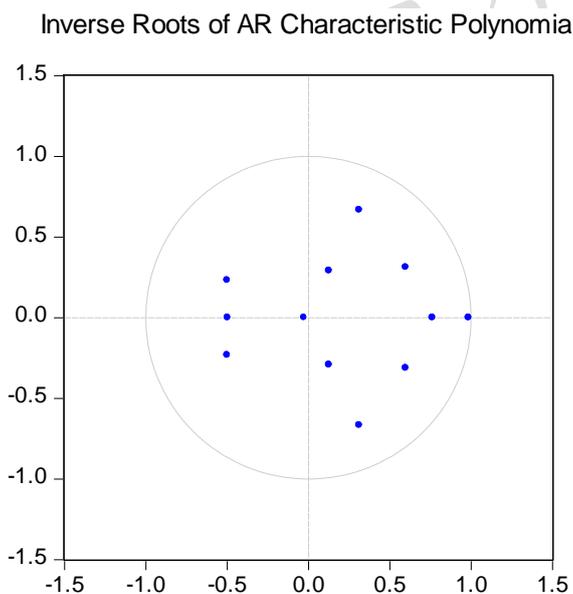


图 4-7 VAR(2)的平稳性检验结果图

由图 4-7 可以直观看出，VAR(2)的所有根的倒数的模小于 1，即位于单位圆内，也就是说，VAR(2)模型是稳定的。因此，进行脉冲响应函数和方差分解的结果将是有效的。

✧ 脉冲响应函数分析结果

在基于 VAR (2) 模型的基础上，通过脉冲响应函数来进一步分析对数序列 X_tL 对 YL 的效应，绘制出脉冲响应函数图如图 4-8。

Response to Cholesky One S.D. Innovations ± 2 S.E.

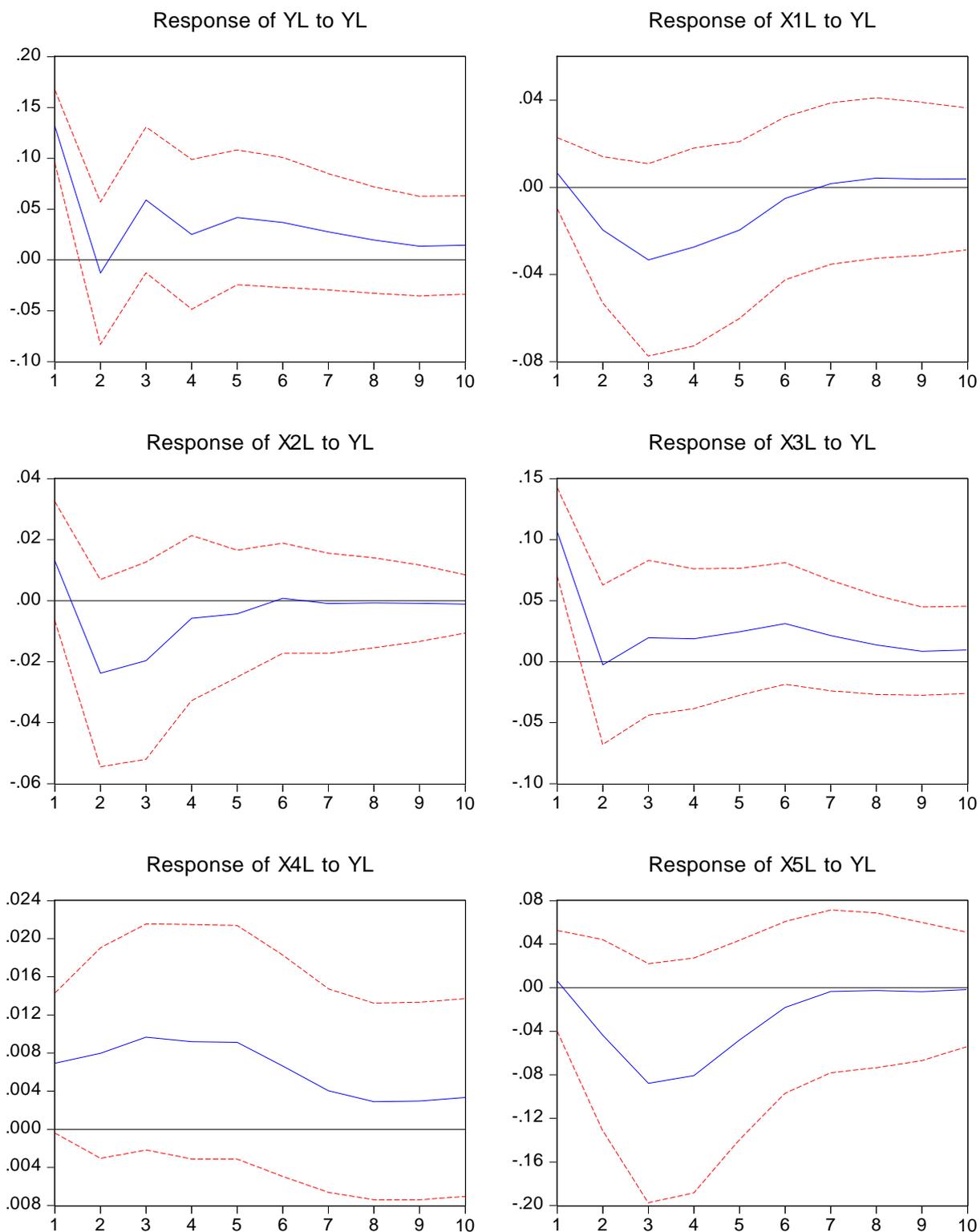


图 4-8 脉冲响应函数分析图

由图 4-8 分析可得， YL 对 X_1L 一个标准差新息冲击的响应在整个分析期内呈现出波动性的影响。 X_1L 的冲击在前 7 期内基本上对 YL 带来负面响应，第 7 期后才带来较小的正面响应； X_2L 的冲击在前 6 期内基本上对 YL 带来负面响应，第 6 期后响应几近为 0； X_3L

的冲击对YL始终带来正面响应，在第一期达到最大，在第2期响应几近为0； X_4L 的冲击对YL始终带来正面响应，在第3期达到最大； X_5L 的冲击在前7期内基本上对YL带来负面响应，第7期后响应几近为0。

(3)方差分解

方差分解是分析预测残差的标准差由不同新息的冲击影响的比例，是进一步评价各内生变量对预测方差的贡献度。下面我们利用方差分解方法分析序列 X_iL 对YL变化的贡献度，其结果如表4-7、图4-9。

表 4-7 YL的方差分解数值表

Effect of Cholesky (d.f. adjusted) One S.D. YL Innovation					
Period	X1L	X2L	X3L	X4L	X5L
1	0.006555 (0.00816)	0.013036 (0.00973)	0.106550 (0.01811)	0.006944 (0.00368)	0.006273 (0.02307)
2	-0.019570 (0.01681)	-0.023814 (0.01535)	-0.002568 (0.03269)	0.007987 (0.00552)	-0.043691 (0.04386)
3	-0.033335 (0.02209)	-0.019688 (0.01617)	0.019533 (0.03170)	0.009676 (0.00593)	-0.087741 (0.05485)
4	-0.027381 (0.02272)	-0.005739 (0.01353)	0.018842 (0.02866)	0.009183 (0.00615)	-0.080624 (0.05391)
5	-0.019559 (0.02030)	-0.004286 (0.01041)	0.024457 (0.02601)	0.009120 (0.00613)	-0.048003 (0.04579)
6	-0.005054 (0.01872)	0.000792 (0.00904)	0.031259 (0.02497)	0.006646 (0.00581)	-0.018111 (0.03943)
7	0.001717 (0.01852)	-0.000914 (0.00822)	0.021327 (0.02269)	0.004054 (0.00534)	-0.003409 (0.03734)
8	0.004268 (0.01840)	-0.000701 (0.00738)	0.013663 (0.02031)	0.002910 (0.00516)	-0.002528 (0.03552)
9	0.003823 (0.01758)	-0.000883 (0.00627)	0.008567 (0.01810)	0.002959 (0.00518)	-0.003584 (0.03171)
10	0.003874 (0.01624)	-0.001103 (0.00478)	0.009640 (0.01785)	0.003332 (0.00519)	-0.001674 (0.02616)

Cholesky Ordering: YL X1L X2L X3L X4L X5L
Standard Errors: Analytic

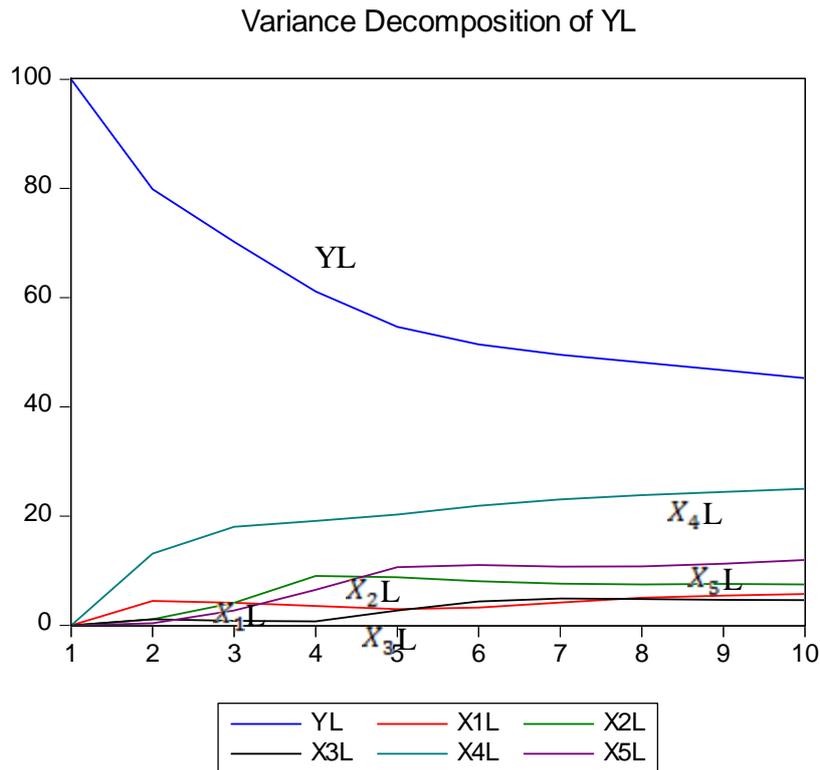


图 4-9 YL 的方差分解趋势图

由图4-9可以直观地看出，序列 X_iL 对 YL 的贡献度在前期均存在一个上升趋势，在后期处于平稳状，这与脉冲响应函数分析的结果一致，而且 X_4L 对 YL 的贡献率最大，短期内 X_2L 对 YL 的贡献率也比较大。

3.结论与建议

3.1.结论

本文利用 1985–2013 年的数据，对广州财政收入与社会从业人员、地区生产总值、税收收入等五个因素间建立了协整关系分析关键影响因素；接着，利用 VAR 模型预测 2015 年的广州财政总收入，并将财政总收入分为个人所得税、企业所得税、营业税、增值税、其他税收收入、非税收收入和政府性基金这七类收入，利用时间序列预测法和神经网络模型进行细化预测；最后，在 VEC 模型、脉冲响应函数和方差分析的基础上，进一步探析广州财政收入与社会从业人员、地区生产总值、税收收入等因素间的关系，得到如下结论：

(1)通过JJ协整检验，财政收入与地区生产总值、第二第三产业产值比、税收收入、社会从业人员、固定资产投资额之间存在长期稳定的协同关系，而且社会从业人员、地区

生产总值、税收收入为关键影响因素；

(2)运用GM(1, 1)白化预测模型，预测出2015年五个影响因素的预测值，再代入VAR模型，求得2015年广州财政总收入预测值为23978544.28万元；利用时间序列预测法和神经网络模型，个人所得税、企业所得税、营业税、增值税、其他税收收入、非税收收入和政府性基金的其预测值分别为346249.60、982475.60、1708531.43、2353663.07、4209813.22、2929400.73、15986526.18万元。

(3)VEC模型表明地区生产总值、第二第三产业产值比、税收收入、社会从业人员、固定资产投资额对财政收入存在反向修正机制；

(4)根据脉冲响应函数分析，财政收入对社会从业人员、地区生产总值、税收收入等因素一个新息冲击的响应在整个分析期内呈现出波动性的影响，后期均趋于平缓；

(5)根据方差分解，社会从业人员、地区生产总值、税收收入等因素对财政收入的贡献度在前期均存在一个上升趋势，在后期处于平稳状，这与脉冲响应函数分析的结果一致，而且社会从业人员对财政收入的贡献度最大，短期内税收收入的贡献度也比较大。

3.2.建议

针对问题三，结合社会经济发展和广州市近几年的财政收入及支出情况，在上述结论的基础上，向广州市财政局提出如下的政策建议：

(1)实现充分就业，促进经济快速增长。根据JJ检验结果，社会从业人员是影响广州市财政收入的关键因素，而促进就业一直以来是中央和地方经济工作会议的重要议题，广州市政府相关部门应优化就业环境，开展各种类型的职业教育培训，建立公平竞争的劳动力市场。此外，也要加大预算支出，鼓励创业尤其是大学生自主创业，以创业带动就业，这与新形势下“三严三实”讲话中的“创业要实”是相符合的。

(2)提高经济发展水平，以经济增长带动财政收入增长。根据JJ检验结果，地区生产总值也是关键影响因素之一，广州市政府应该大力发展生产力，毕竟发展才是硬道理，通过提高生产总值，促进财政收入增长。

(3)提高税收征管执行力度。根据方差分解，短期内税收收入的贡献度比较大，税务机关要加强税源监控，依法治税征税，保持税收增长。

(4)优化产业结构，实现产业结构化升级。根据JJ检验结果，第二第三产业产值比虽不是影响广州财政收入的关键因素，但从方差分解结果可知，它对财政收入的贡献度较大，所以可以通过优化产业结构，加快第三产业发展，尤其是金融保险和文化旅游等现代服

务业，进一步促进财政收入增长。

4.参考文献

- [1] 李子奈 潘文卿.计量经济学（第三版）[M].北京：高等教育出版社，2010 年
- [2] 庞浩.计量经济学（第二版）[M].北京：科学出版社，2010 年
- [3] 熊平.数据挖掘算法与 Clementine 实践[M].北京：清华大学出版社，2011.04
- [4] 蒋盛益 李霞 郑琪.数据挖掘原理与实践[M].北京：电子工业出版社，2011.08
- [5] 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.(4):78-83
- [6] 何邓娇.广州市财政收入影响因素的实证研究——基于 VAR 模型分析[J].经济研究导刊.2014.(15):104-106,121