

参赛密码 \_\_\_\_\_  
(由组委会填写)

## 第十届华为杯全国研究生数学建模竞赛

学 校 同济大学

---

参赛队号 10247084

---

1.刘 帅

---

队员姓名 2.顾 帅

---

3.叶尚斌

---

参赛密码 \_\_\_\_\_  
(由组委会填写)



## 第十届华为杯全国研究生数学建模竞赛

题 目 可持续的中国城乡居民养老保险体系的数学模型研究

### 摘 要:

本文针对中国城乡居民养老保险体系的数学建模问题,从养老保险体系发展现状出发,运用真实数据主要完成了以下几方面的工作:

对于问题 1, 本文从国情出发, 建立了三大类模型: 城镇居民养老金收入、支出数学模型; 新农保收入、支出数学模型; 企业年金积累基金、个人储蓄养老保险收入数学模型。具体建模过程基本如下: 首先确定可能影响养老金收入、支出(包括城镇居民、新农保、企业年金、个人储蓄养老保险等)的备选指标并查找相关原始数据; 随后综合运用相关分析和灰色关联分析筛选所有备选指标, 确定若干指标作为自变量(各类模型不尽相同); 接着对自变量进行多重共线性诊断, 建立基于主成分回归的模型; 此外又建立了基于多重线性回归与支持向量机的数学模型。各模型充分考虑现实影响因子, 较好地体现了多个层次在养老保险体系中的地位以及“多缴多得, 长缴多得”的准则。模型拟合度与显著性均较高。

对于问题 2, 首先对养老金缺口定义进行诠释; 然后根据本题涉及到的长期预测背景, 对已建的多重线性回归模型与支持向量机模型进行改进, 改进后模型拟合度与显著性均较好, 并利用改进后模型对从今年至 2035 年我国养老金缺口进行估计, 估计过程中充分考虑其他可能产生影响的变量; 对养老保险收支矛盾最尖锐情况即当年结余由正转负的年份作出预测, 两个模型分别预测为 2017 年与 2019 年; 最后考虑收入倍增计划, 对数学模型需要调整的部分进行了阐释说明。

对于问题 3 与问题 4, 本文出于模型陈述方便与问题连续性考虑, 将这两道题目在同一章内加以解决。首先总结归纳了当今世界各国 5 种主要养老模式, 简要分析其优缺点及值得我国借鉴之处; 之后利用已建立的多重线性回归模型对替代率与缴费率合理区间进行优化选取, 选取 $[0.45, 0.65]$ 为替代率合理区间,

[0.15,0.25]为缴费率合理区间；进而以替代率与缴费率为可调节变量建立优化的多重线性回归数学模型与支持向量机-精算解析模型，并再次利用新建立的支持向量机-精算解析模型探讨替代率与缴费率合理区间，从另一个角度出发，选取[0.3,0.45]为替代率合理区间，[0.15,0.3]为缴费率合理区间；最后针对可能出现的收支矛盾提出了一些政策措施建议，并对政策效果进行了仿真预测。

关键词：养老保险 可持续 学生化残差 相关分析 灰色关联度 多重共线性 主成分回归 多重线性回归 支持向量机 精算解析

# 目录

1 问题重述.....	- 5 -
1.1 问题背景.....	- 5 -
1.2 问题提出.....	- 5 -
2 问题分析.....	- 6 -
3 名词、符号说明及基本假设.....	- 6 -
3.1 名词和符号说明.....	- 6 -
3.2 模型的基本假设.....	- 7 -
4 中国城乡居民养老金收入、支出宏观数学模型.....	- 7 -
4.1 城镇居民养老金收入、支出宏观数学模型.....	- 8 -
4.1.1 数据预处理.....	- 9 -
4.1.2 基于相关分析的指标变量分析.....	- 12 -
4.1.3 基于灰色关联度的指标变量分析.....	- 14 -
4.1.4 多重共线性诊断.....	- 16 -
4.1.5 基于主成分回归的城镇居民养老金收入、支出建模.....	- 16 -
4.1.6 基于多重线性回归的城镇居民养老金收入、支出建模.....	- 19 -
4.1.7 基于支持向量机的城镇居民养老金收入、支出建模.....	- 20 -
4.2 新农保收入、支出宏观数学模型.....	- 23 -
4.2.1 基于多重线性回归的新农保收入、支出建模.....	- 23 -
4.2.2 基于支持向量机的新农保收入、支出建模.....	- 26 -
4.3 多层次养老保险体系数学模型研究.....	- 27 -
4.3.1 企业年金数学模型研究.....	- 27 -
4.3.2 个人储蓄养老保险数学模型研究.....	- 31 -
4.3.3 多层次养老保险体系数学模型探讨.....	- 36 -
4.4 基于“多缴多得，长缴多得”的仿真研究.....	- 36 -
4.4.1 基于多重线性回归模型的仿真研究.....	- 37 -
4.4.2 基于支持向量机模型的仿真研究.....	- 38 -
5 养老金缺口估计与模型调整.....	- 39 -
5.1 养老金缺口估计.....	- 39 -
5.1.1 基于多重线性回归模型的缺口估计.....	- 39 -
5.1.2 基于支持向量机模型的缺口估计.....	- 42 -
5.2 基于收入倍增计划的模型调整.....	- 44 -
6 养老保险体系的可持续性研究.....	- 44 -
6.1 各国养老保险模式分析.....	- 45 -
6.2 我国养老保险体系可持续性研究.....	- 46 -
6.2.1 替代率与缴费率合理区间选取.....	- 46 -
6.2.2 可调节变量的数学模型建立.....	- 48 -
6.2.3 政策措施建议.....	- 53 -
7 结论与建议.....	- 54 -
7.1 模型评价.....	- 54 -
7.2 我国养老保险体系未来发展建议.....	- 55 -
参考文献.....	- 57 -

# 1 问题重述

## 1.1 问题背景

中国共产党第十八次全国代表大会政治报告中提出了“统筹推进城乡社会保障体系建设”的任务：“社会保障是保障人民生活、调节社会分配的一项基本制度。要坚持全覆盖、保基本、多层次、可持续方针，以增强公平性、适应流动性、保证可持续性为重点，全面建成覆盖城乡居民的社会保障体系”。

近年来，我国城镇职工养老保险金制度基本建立，参保覆盖面日益扩大，退休人员待遇逐年提高，养老保险基金规模不断上升，在制度层面上提前完成了城市和农村居民社会保障全覆盖，成为我们社会保障事业发展的重要里程碑。然而也面临巨大的挑战。我国的养老保险体系建设远远滞后于社会发展水平。一方面资金规模相对较少，保障能力有限，地区差距、城乡差距明显；另一方面资金管理分散，运营效率不高，保值增值的压力很大。进一步完善养老保险保障体系仍将是中国政府未来最重要的任务之一。

## 1.2 问题提出

面临人口老龄化以及经济结构性减速、财政收入增长速度下降，社会保障体系可持续问题亟待解决，必须进行前瞻性研究，否则将形成社会问题。我们应该承认历史、立足现实、尊重规律、借鉴国外、留有余地，既量力而行，又尽力而为地完善我国的养老保险体系。本题要求运用数学模型方法研究、解决该问题。为此，请做好以下几项工作。

1，分别建立合乎国情、适应国力的中国城乡居民（含新农保）养老金收入、支出的宏观数学模型，至少包括替代率（基本养老保险人均养老金占城镇单位在岗职工平均工资比率）、缴费率（基本养老保险人均缴费占城镇单位在岗职工平均工资比率）、人口结构、分年龄段死亡率、经济增速、财政补贴、工资水平或物价指数、投资效益等主要因素，要做到模型结合现实，分多个层次（含企业基金等），体现“多缴多得，长缴多得”（不考虑分省、分地区模型）。

2，根据数学模型、对养老金缺口的理解和对未来有关情况的合理估计，估计从今年至 2035 年我国养老金缺口，并说明对养老金缺口分析的合理性。如果全部情况维持不变，按照数学模型预测我国城乡居民养老保险收支矛盾最尖锐的情况发生在什么时间，严重程度如何？考虑到党的十八大提出的收入倍增计划，数学模型需要调整的地方是哪些？

3，养老保险制度也是调节社会分配，请分析各国养老保险的不同模式，取其精华，去其糟粕，根据建立的数学模型和中国的实际情况，利用仿真手段寻找替代率和缴费率的合理区间以保证我国养老保险体系的可持续性（因为人口结构、分年龄段死亡率、经济增速、投资效益等主要因素几乎无法人为较大幅度改动）；在步入良性循环之前，在矛盾最尖锐到来前的过渡期内应该采取哪些政策措施实现平稳过渡并仿真预测相关政策的效果。

4，尝试建立第三问增加可调节变量的数学模型。

## 2 问题分析

养老保险在社会保障体制建设中处于核心地位，因此，养老保险体系的建立与完善关系到社会保障体制建设的全局<sup>[1]</sup>。对养老保险体系的数据进行分析与建模对于深化养老保险改革、完善养老保险保障体系具有重要的参考意义。

问题一要求建立合乎国情、适应国力的中国城乡居民（含新农保）养老金收入、支出的宏观数学模型。本文根据现有文献的研究成果以及检索得到的相关数据，从国情出发，建立了三大类模型：城镇居民养老金收入、支出数学模型；新农保收入、支出数学模型；企业年金积累基金、个人储蓄养老保险收入数学模型。三类模型具有不同的侧重点，且都具有较高的有效性与精确性。具体建模方法主要包括主成分回归、多重线性回归、支持向量机等。

问题二要求估计从今年至 2035 年我国养老金缺口并分析说明，以及预测我国养老保险收支矛盾最尖锐情况，并根据党的十八大提出的收入倍增计划调整数学模型。针对问题要求，本文首先通过查阅文献资料，对养老金缺口定义进行分析与诠释；然后根据本题涉及到的长期预测背景，对已建的多重线性回归模型与支持向量机模型进行改进，改进后模型拟合度与显著性均较好，并利用改进后模型对养老金缺口进行估计；最后考虑收入倍增计划，对数学模型需要调整的部分进行了阐释说明。

问题三要求分析各国养老保险的不同模式，寻找替代率和缴费率的合理区间以保证我国养老保险体系的可持续性，并针对预测矛盾提出相关政策建议。

问题四要求在第三问的基础上建立增加可调节变量的数学模型。

本文出于模型陈述方便与问题连续性考虑，将这两道题目在同一章内加以解决。首先通过查阅大量文献资料，总结归纳了当今世界各国主要养老模式，简要分析其优缺点及值得我国借鉴之处；之后对我国养老保险体系可持续性进行了分析研究，利用已建立的多重线性回归模型对替代率与缴费率合理区间进行优化选取，进而以替代率与缴费率为可调节变量建立优化的多重线性回归数学模型与支持向量机-精算解析模型，并再次利用新建立的支持向量机-精算解析模型探讨替代率与缴费率合理区间选取，最后针对预测情况提出了一些政策措施建议。

## 3 名词、符号说明及基本假设

### 3.1 名词和符号说明

<i>CZYS</i>	城镇养老金收入
<i>CZYZ</i>	城镇养老金支出
<i>TDL</i>	替代率（基本养老保险人均养老金占城镇单位在岗职工平均工资比率）
<i>JFL</i>	缴费率（基本养老保险人均缴费占城镇单位在岗职工平均工资比率）
<i>CBZZ</i>	参保在职人数
<i>CBLT</i>	参保离退休人数

<i>SWL1</i>	0-20 岁人口死亡率
<i>SWL2</i>	21-60 岁人口死亡率
<i>SWL3</i>	61 岁以上人口死亡率
<i>GDPZ</i>	GDP 增长率
<i>GDPRZ</i>	GDP 人均增长率
<i>CZBT</i>	财政补贴
<i>GZSP</i>	工资水平，即城镇单位在岗职工平均工资
<i>XFJZ1</i>	消费价格指数（以上一年为基期）
<i>TZL</i>	通货膨胀率
<i>XFJZ2</i>	消费价格指数（以 2000 年为基期）
<i>SLJZ</i>	商品零售价格指数（以上一年为基期）
<i>TZSYL</i>	投资收益率
<i>TZSYE</i>	投资收益额
<i>QNJJ</i>	企业年金积累基金
<i>SRBS</i>	商业人寿保险费收入
<i>QCZZ</i>	企业参保在职人数
<i>CBQ</i>	参保企业数量

### 3.2 模型的基本假设

对于所研究问题的参数和范围，本文作如下几点假设和限定：

- 1、附件中提供的所有数据和通过网络查询到的相关数据真实有效，并能反映指标所代表的真实含义。
- 2、在预测养老金未来趋势时，假设在预测的时间区间内无政策影响或其他不可预见的事件发生，并忽略提前退休、一次性补缴等不确定因素；
- 3、养老金累计结余的平均投资收益率保持稳定；
- 4、养老金累计结余的投资收益结算时不考虑通胀率的影响。

## 4 中国城乡居民养老金收入、支出宏观数学模型

问题 1 要求建立中国城乡居民养老金收入、支出宏观数学模型。针对该问题，已有大量文献进行了研究。文献[2][3]对城镇居民养老金收支状况进行了分析与预测；文献[4][6]着重分析了“新农保”形势下我国农村养老保险情况；文献[6]-[8]则从多层次乃至多支柱角度出发对我国养老金保障体系改革进行了分析与思考。

本文根据现有文献的研究成果以及检索得到的相关数据，从国情出发，建立了以下三大类模型：城镇居民养老金收入、支出数学模型；新农保收入、支出数学模型；企业年金积累基金、个人储蓄养老保险收入数学模型。三类模型具有不同的侧重点，且都具有较高的有效性与精确性。最后针对“多缴多得，长缴多得”准则进行了仿真研究。

#### 4.1 城镇居民养老金收入、支出宏观数学模型

根据题目建议以及文献查阅可知,对城镇居民养老金收入、支出(CZYS、CZYZ)的影响因素主要有:替代率(TDL)、缴费率(JFL)、参保在职人数(CBZZ)、参保离退休人数(CBLT)、0-20岁人口死亡率(SWL1)、21-60岁人口死亡率(SWL2)、61岁以上人口死亡率(SWL3)、GDP增长率(GDPZ)、GDP人均增长率(GDPRZ)、财政补贴(CZBT)、工资水平(GZSP)、消费价格指数(以上一年为基期)(XFJZ1)、通货膨胀率(TZL)、消费价格指数(以2000年为基期)(XFJZ2)、商品零售价格指数(以上一年为基期)(SLJZ)、投资收益率(TZSYL)、投资收益额(TZSYE)等指标。

本节对上述指标进行考察,首先从题目附件和国家统计局等相关网站查到以上指标的年度数据;然后对数据进行预处理等分析操作;随后采用相关分析、灰色关联度确定各指标之间的相关程度,并从中进一步筛选出最能反映城镇居民养老金收入、支出的指标;接着对筛选得到的变量进行多重共线性诊断以判断多元回归建模的适用性;最后利用选定指标完成城镇居民养老金收入、支出数学模型主成分回归建模、多重线性回归建模与支持向量机建模,并对模型的有效性和精确性进行分析。本节采用技术路线如图1所示。

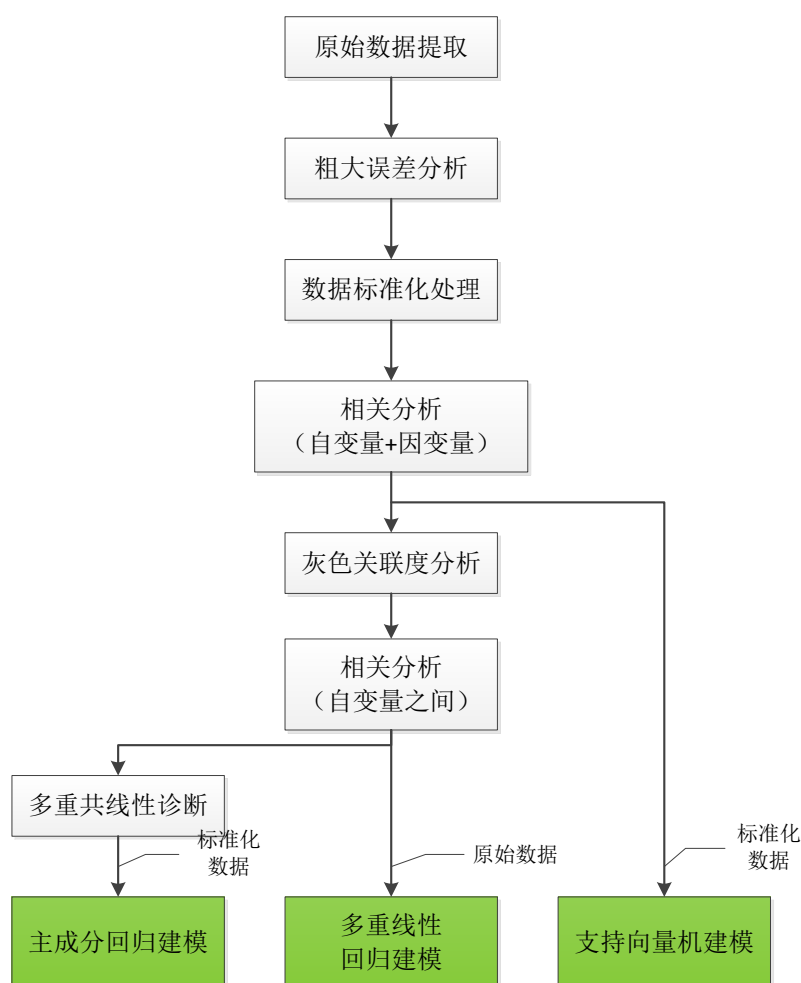


图1 本章技术路线



#### 4.1.1 数据预处理

##### 1) 原始数据提取

由于数据来源于人为测试数据，将不可避免地引入误差，因此首先必须对数据进行预处理。由于某些年份各指标数据不完整，因此经过基本的数据筛选后确定选用 2003 年-2011 年的全国数据作为研究对象。这样虽然数据量有所减少，但却在最大程度上反映了实际测试数据。经过整理后的初选指标数据<sup>[9]-[11]</sup>如表 1 所示。

表 1 各初选指标变量年度数据汇总表

年份		2003	2004	2005	2006	2007	2008	2009	2010	2011
CZYS	(亿元)	3680	4258	5093	6310	7834	9740	11941	13420	16895
CZYZ	(亿元)	3122	3502	4040	4897	5965	7390	8894	10555	12765
TDL	(%)	0.58	0.53	0.5	0.5	0.48	0.48	0.47	0.45	0.44
JFL	(%)	0.19	0.18	0.18	0.18	0.17	0.17	0.16	0.15	0.15
CBZZ	(万人)	11646	12250	13120	14130	15183	16587.5	17743	19402.3	21565
CBLT	(万人)	3860.2	4103	4367.5	4635	4953.7	5303.6	5806.9	6305	6826
SWL1	(ppm)	1124.8	1204.5	1232.47	1148.1	1188.4	1054.6	1035.71	1293.6	983.7
SWL2	(‰)	21.34	20.9	20.73	16.1	15.71	14.26	15.04	16.1	15.29
SWL3	(‰)	411.86	402.65	398.12	353.49	333.94	310.26	325.9	394.49	319.01
GDPZ	(%)	10	10.1	11.3	12.7	14.2	9.6	9.2	10.4	9.3
GDPRZ	(%)	9.3	9.4	10.7	12	13.6	9.1	8.7	9.9	8.8
CZBT	(亿元)	530	614	651	971	1157	1437	1646	1954	2272
GZSP	(元)	14040	16024	18364	21001	24932	29229	32736	37147	42452
XFJZ1	-	101.2	103.9	101.8	101.5	104.8	105.9	99.1	103.3	105.4
TZL	(%)	1.2	3.9	1.8	1.5	4.8	5.9	-0.9	3.3	5.4
XFJZ2	-	101.09	105.04	106.93	108.53	113.74	120.45	119.61	123.55	130.23
SLJZ	(%)	99.9	102.8	100.8	101	103.8	105.9	98.9	103.1	104.9
TZSYL	(%)	3.56	2.61	4.16	29.01	43.19	-6.79	16.12	4.23	0.84
TZSYE	(亿元)	44.71	36.72	71.22	619.79	1453.5	-393.72	850.43	321.22	73.37

##### 2) 粗大误差分析

粗大误差或过失误差是样本中的个别值，其数值明显偏离其它样本值，故亦称作异常值。在数据处理时，如果对混有离群值的测量数据未予剔除，则会造成测量精度偏低的结果；反之如果将正常数据错误地当作离群值予以剔除，就会减少有用的测量信息。因此对测量中离群值的合理判别与处理，是获得正确分析结果的前提条件。

本文采用学生化残差对原始数据中的粗大误差进行分析。学生化残差基本概念<sup>[12]</sup>如下：

设对某量进行  $n$  次独立测量，得到的测量列  $X$  为：

$$X = \{x_1, x_2, \dots, x_i, \dots, x_n\} \quad (i = 1, 2, \dots, n)$$

它的算术平均值  $\bar{x}$  及其残差  $v_i$  分别为：

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad (1)$$

$$v_i = x_i - \bar{x} \quad (2)$$

定义 1: 残差与样本标准差之比为样本学生化残差, 即:

$$y_i = \frac{v_i}{s} \quad (3)$$

式中:  $s$  为样本标准差, 其值为:

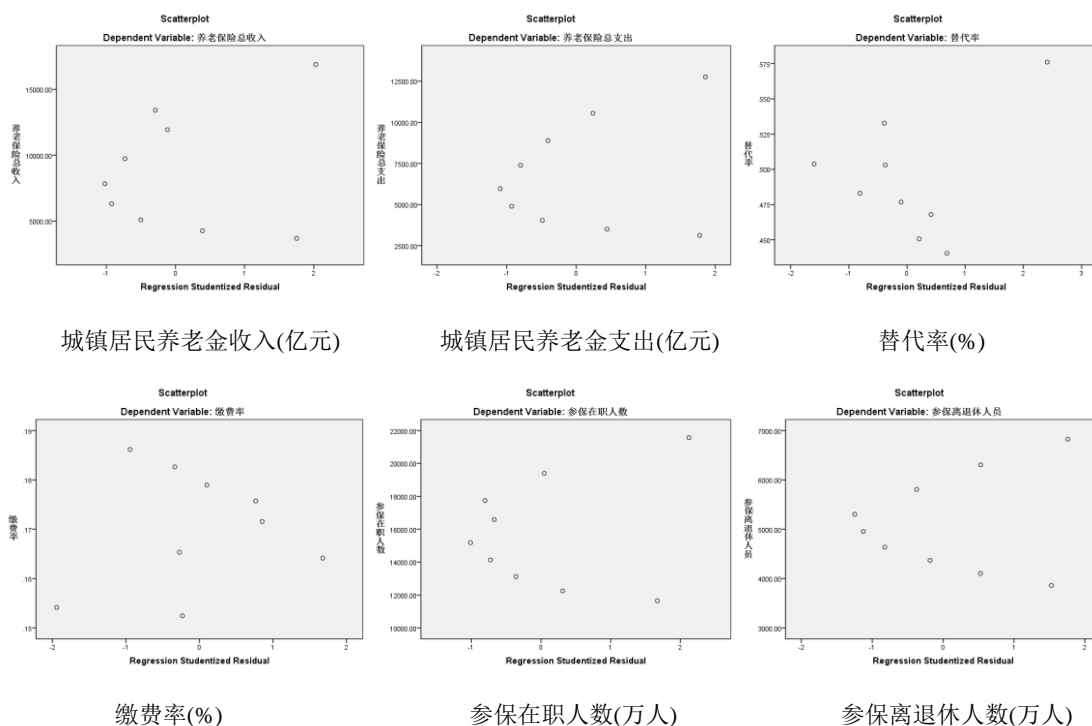
$$s = \sqrt{\frac{\sum_{i=1}^n v_i^2}{n}} \quad (4)$$

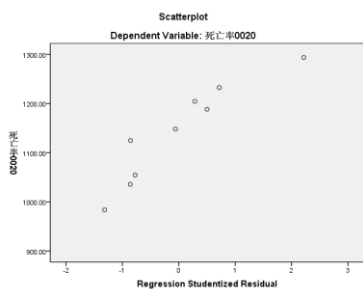
定义 2: 残差的绝对值与样本标准差之比为样本学生化残差的绝对值, 即:

$$|y_i| = \frac{|v_i|}{s} \quad (5)$$

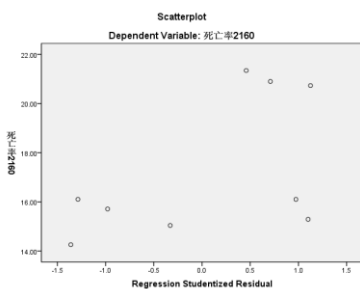
这里所谓的“学生化”主要是针对普通残差不具方差齐性而进行的一种标准化处理, 使之期望为 0, 方差为 1, 从而便于比较。

根据上述学生化残差定义, 对各指标年度数据进行计算, 结果如图 2 所示。当学生化残差的绝对值大于 3 时可认为该数据为变异数据, 即粗大误差, 可予以剔除。但从上图可以看出, 所有指标的学生化残差的绝对值均小于 3, 因此认为以上指标数据中没有粗大误差, 可以直接进行后续的分析处理。

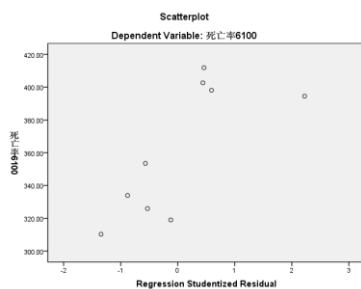




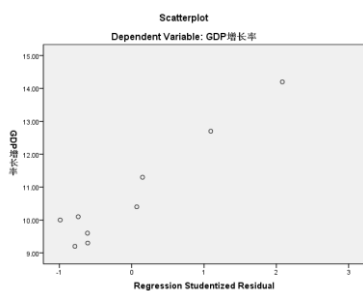
0-20 岁人口死亡率(ppm)



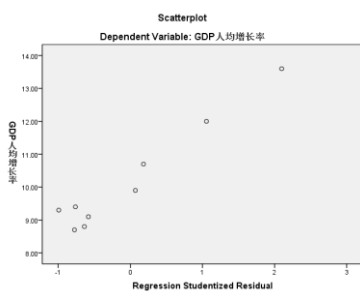
21-60 岁人口死亡率(‰)



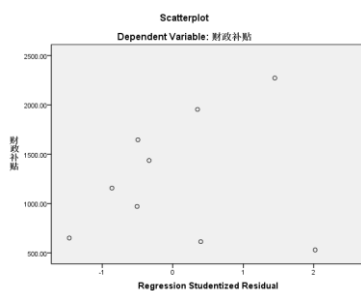
60 岁以上人口死亡率(‰)



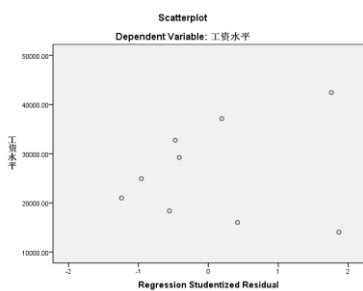
GDP 增长率(%)



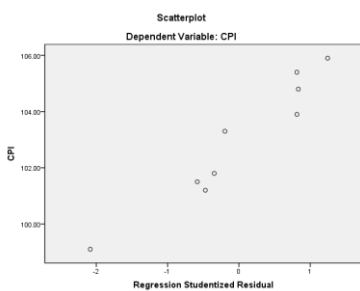
GDP 人均增长率(%)



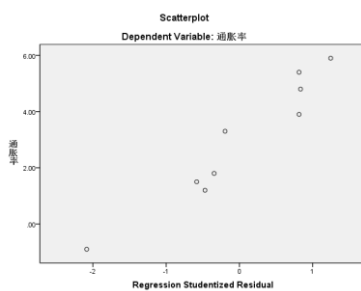
财政补贴(亿元)



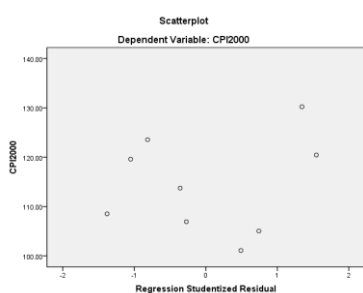
工资水平(元)



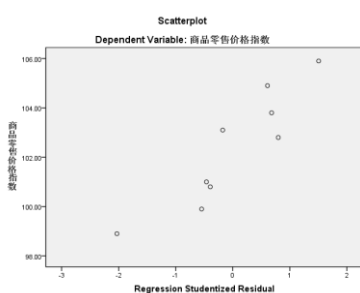
消费价格指数（以上一年为  
基期）



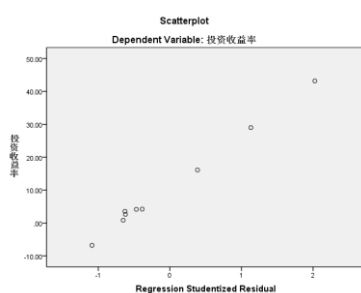
通货膨胀率(%)



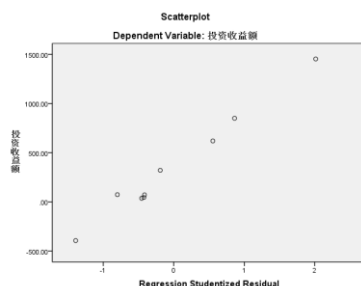
消费价格指数（以 2000 年为基  
期）



商品零售价格指数  
（以上一年为基期）



投资收益率(%)



投资收益额(亿元)

图2 各指标变量数据学生化残差图

### 3) 数据标准化处理

由于以上各种指标的量纲不同，对应的数据的数量级相差很大，为克服其带来的不利影响，对所有指标数据进行标准化处理，其公式如式（6）所示。

$$X'_{ik} = \frac{X_{ik} - \bar{X}_k}{\sigma_k} \quad (6)$$

其中， $X'_{ik}$ 是经标准化处理后的第  $k$  个变量中第  $i$  个数据的数值。 $\bar{X}_k$  是第  $k$  个变量的均值， $\sigma_k$  代表第  $k$  个变量的标准差。

#### 4.1.2 基于相关分析的指标变量分析

相关分析是研究变量之间密切程度的一种统计方法。本小题通过相关分析可以初步了解城镇居民养老金收入、支出与各指标间关系的密切程度，并由此将关系不密切的指标剔除，从而达到减少模型维数的目的。

任意两个变量间的皮尔森相关系数  $\rho$  可由公式（7）计算得到<sup>[13]</sup>。

$$\rho = \frac{\sum_{k=1}^n (X_{ik} - \bar{X}_i)(X_{jk} - \bar{X}_j)}{\sqrt{\sum_{k=1}^n (X_{ik} - \bar{X}_i)^2 \sum_{k=1}^n (X_{jk} - \bar{X}_j)^2}} \quad (7)$$

其中， $\bar{X}$  代表单个基因表达的均值； $X_{ik}$  代表第  $k$  个变量第  $i$  个数据的数值；

$\rho$  为相关系数。 $X_{jk}$  代表第  $k$  个变量第  $j$  个数据的数值。

利用上述经过标准化处理后的数据在 SPSS 软件<sup>[14]</sup>中进行相关分析，可得到各指标之间的相关系数值如表 2、表 3 所示。

表 2 各指标间相关系数矩阵（针对城镇居民养老金收入）

	CZYS	TDL	JFL	CBZZ	CBLT	SWL1	SWL2	SWL3	GDPZ	GDPRZ	CZBT	GZSP	XFJZ1	TZL	XFJZ2	SLJZ	TZSYL	TZSYE
CZYS	1	-.897**	-.979**	.998**	.996**	-.433	-.744*	-.600	-.369	-.329	.995**	.996**	.280	.280	.980**	.408	-.153	.039
TDL	-.897**	1	.938**	-.917**	-.922**	.223	.814**	.655	.090	.044	-.904**	-.922**	-.327	-.327	-.934**	-.464	-.035	-.182
JFL	-.979**	.938**	1	-.989**	-.990**	.293	.783*	.576	.286	.243	-.989**	-.992**	-.333	-.333	-.984**	-.472	.134	-.043

CBZZ	.998**	-.917**	-.989**	1	.999**	-.395	-.762*	-.602	-.328	-.287	.997**	.999**	.305	.305	.986**	.436	-.135	.046
CBLT	.996**	-.922**	-.990**	.999**	1	-.365	-.753*	-.580	-.326	-.285	.995**	.998**	.273	.273	.981**	.405	-.125	.067
SWL1	-.433	.223	.293	-.395	-.365	1	.447	.722*	.465	.456	-.400	-.389	-.066	-.066	-.421	-.129	.170	.144
SWL2	-.744*	.814**	.783*	-.762*	-.753*	.447	1	.879**	-.025	-.065	-.785*	-.776*	-.281	-.281	-.805**	-.436	-.236	-.292
SWL3	-.600	.655	.576	-.602	-.580	.722*	.879**	1	.017	-.016	-.611	-.612	-.325	-.325	-.676*	-.437	-.206	-.223
GDPZ	-.369	.090	.286	-.328	-.326	.465	-.025	.017	1	.999**	-.332	-.330	.083	.083	-.327	.012	.850**	.682*
GDPZ	-.329	.044	.243	-.287	-.285	.456	-.065	-.016	.999**	1	-.291	-.288	.098	.098	-.283	.033	.853**	.694*
CZBT	.995**	-.904**	-.989**	.997**	.995**	-.400	-.785*	-.611	-.332	-.291	1	.997**	.300	.300	.983**	.434	-.122	.060
GZSP	.996**	-.922**	-.992**	.999**	.998**	-.389	-.776*	-.612	-.330	-.288	.997**	1	.300	.300	.989**	.434	-.131	.057
XFJZ1	.280	-.327	-.333	.305	.273	-.066	-.281	-.325	.083	.098	.300	.300	1	1.000**	.401	.980**	-.185	-.262
TZL	.280	-.327	-.333	.305	.273	-.066	-.281	-.325	.083	.098	.300	.300	1.000**	1	.401	.980**	-.185	-.262
XFJZ2	.980**	-.934**	-.984**	.986**	.981**	-.421	-.805**	-.676*	-.327	-.283	.983**	.989**	.401	.401	1	.536	-.164	.005
SLJZ	.408	-.464	-.472	.436	.405	-.129	-.436	-.437	.012	.033	.434	.434	.980**	.980**	.536	1	-.216	-.272
TZSYL	-.153	-.035	.134	-.135	-.125	.170	-.236	-.206	.850**	.853**	-.122	-.131	-.185	-.185	-.164	-.216	1	.939**
TZSYE	.039	-.182	-.043	.046	.067	.144	-.292	-.223	.682*	.694*	.060	.057	-.262	-.262	.005	-.272	.939**	1
**, 在 .01 水平（双侧）上显著相关																		
*, 在 .05 水平（双侧）上显著相关																		

表 3 各指标间相关系数矩阵（针对城镇居民养老金支出）

	CZYZ	TDL	JFL	CBZZ	CBLT	SWL1	SWL2	SWL3	GDPZ	GDPZ	CZBT	GZSP	XFJZ1	TZL	XFJZ2	SLJZ	TZSYL	TZSYE
CZYZ	1	-.890**	-.981**	.997**	.996**	-.401	-.725*	-.565	-.375	-.335	.995**	.995**	.286	.286	.977**	.412	-.170	.023
TDL	-.890**	1	.938**	-.917**	-.922**	.223	.814**	.655	.090	.044	-.904**	-.922**	-.327	-.327	-.934**	-.464	-.035	-.182
JFL	-.981**	.938**	1	-.989**	-.990**	.293	.783*	.576	.286	.243	-.989**	-.992**	-.333	-.333	-.984**	-.472	.134	-.043
CBZZ	.997**	-.917**	-.989**	1	.999**	-.395	-.762*	-.602	-.328	-.287	.997**	.999**	.305	.305	.986**	.436	-.135	.046
CBLT	.996**	-.922**	-.990**	.999**	1	-.365	-.753*	-.580	-.326	-.285	.995**	.998**	.273	.273	.981**	.405	-.125	.067
SWL1	-.401	.223	.293	-.395	-.365	1	.447	.722*	.465	.456	-.400	-.389	-.066	-.066	-.421	-.129	.170	.144
SWL2	-.725*	.814**	.783*	-.762*	-.753*	.447	1	.879**	-.025	-.065	-.785*	-.776*	-.281	-.281	-.805**	-.436	-.236	-.292
SWL3	-.565	.655	.576	-.602	-.580	.722*	.879**	1	.017	-.016	-.611	-.612	-.325	-.325	-.676*	-.437	-.206	-.223
GDPZ	-.375	.090	.286	-.328	-.326	.465	-.025	.017	1	.999**	-.332	-.330	.083	.083	-.327	.012	.850**	.682*
GDPZ	-.335	.044	.243	-.287	-.285	.456	-.065	-.016	.999**	1	-.291	-.288	.098	.098	-.283	.033	.853**	.694*
CZBT	.995**	-.904**	-.989**	.997**	.995**	-.400	-.785*	-.611	-.332	-.291	1	.997**	.300	.300	.983**	.434	-.122	.060
GZSP	.995**	-.922**	-.992**	.999**	.998**	-.389	-.776*	-.612	-.330	-.288	.997**	1	.300	.300	.989**	.434	-.131	.057
XFJZ1	.286	-.327	-.333	.305	.273	-.066	-.281	-.325	.083	.098	.300	.300	1	1.000**	.401	.980**	-.185	-.262
TZL	.286	-.327	-.333	.305	.273	-.066	-.281	-.325	.083	.098	.300	.300	1.000**	1	.401	.980**	-.185	-.262
XFJZ2	.977**	-.934**	-.984**	.986**	.981**	-.421	-.805**	-.676*	-.327	-.283	.983**	.989**	.401	.401	1	.536	-.164	.005
SLJZ	.412	-.464	-.472	.436	.405	-.129	-.436	-.437	.012	.033	.434	.434	.980**	.980**	.536	1	-.216	-.272
TZSYL	-.170	-.035	.134	-.135	-.125	.170	-.236	-.206	.850**	.853**	-.122	-.131	-.185	-.185	-.164	-.216	1	.939**
TZSYE	.023	-.182	-.043	.046	.067	.144	-.292	-.223	.682*	.694*	.060	.057	-.262	-.262	.005	-.272	.939**	1
**, 在 .01 水平（双侧）上显著相关																		
*, 在 .05 水平（双侧）上显著相关																		

一般认为当皮尔森相关系数的绝对值大于 0.8 时，则两个向量之间具有很

强的相关性。根据这一准则，可以看出以上指标中 0-20 岁人口死亡率（ $SWL1$ ）、21-60 岁人口死亡率（ $SWL2$ ）、61 岁以上人口死亡率（ $SWL3$ ）、GDP 增长率（ $GDPZ$ ）、GDP 人均增长率（ $GDPZR$ ）、消费价格指数（以上一年为基期）（ $XFJZI$ ）、通货膨胀率（ $TZL$ ）、商品零售价格指数（ $SLJZ$ ）、投资收益率（ $TZSYL$ ）、投资收益额（ $TZSYE$ ）与城镇居民养老金收入、支出相关关系不强。因此，首先将这 10 个指标剔除，再对其余指标进行下一步分析。

#### 4.1.3 基于灰色关联度的指标变量分析

由表 2、表 3 可以看出，经过相关分析剔除无关变量后仍存在大量指标与城镇居民养老金收入、支出相关，这对于接下来的城镇居民养老金收入、支出建模不利。因此采用灰色关联度对各指标与城镇居民养老金收入、支出的关联程度进行排序，从中选择关系最大的变量进行后续建模使用。

灰色关联度是根据因素之间发展态势的相似或相异程序来衡量因素间关联程度的方法。其解决的主要问题即在某一包含各种因素的系统中，分析主要的、次要的、影响大的或影响小的以及哪些需要发展或需要抑制的。确切说，是发展态势的量化比较和分析。实质上是几种曲线几何形状的分析比较，即认为几何形状越来越接近，则发展态势愈接近，其关联程度就越大。

计算方法：系统中各因素之间的关联程度大小，用关联度量表示，其定义如下：设  $X_0 = \{x_0(k), k=1, 2, 3, \dots, n\}$  为参考序列； $X_i = \{x_i(k), k=1, 2, 3, \dots, n\}$  为比较序列。其中  $i=1, 2, 3, \dots, m$ ，对每个序列作初值化处理，即用每个序列第一个数据去除其它各个数据后得到一个新的序列，即：

$$\gamma_0 = \{y_0(k), k=1, 2, \dots, n\} = \left\{ \frac{x_0(k)}{x_0(1)}, k=1, 2, \dots, n \right\}$$

$$\gamma_i = \{y_i(k), k=1, 2, \dots, n\} = \left\{ \frac{x_i(k)}{x_i(1)}, k=1, 2, \dots, n \right\}$$

其中， $i=1, 2, 3, \dots, m$ ， $\gamma_0$  是新的参考序列， $\gamma_i$  是新的比较序列。

则各时刻比较序列与参考序列的关联系数  $\xi_i(k)$  为：

$$\xi_i(k) = \frac{\min_i \min_k \Delta_i(k) + \rho \max_i \max_k \Delta_i(k)}{\Delta_i(k) + \rho \max_i \max_k \Delta_i(k)} \quad (8)$$

其中， $\Delta_i(k) = |y_0(k) - y_i(k)|$  为新参考序列和第  $i$  个新比较序列的差序列， $\rho$  为分辨系数，取值在 0 与 1 之间。

$$\min_i \min_k |y_0(k) - y_i(k)| \quad \text{称为两极最小差}$$

$$\max_i \max_k |y_0(k) - y_i(k)| \quad \text{称为两极最大差}$$

$$\text{其关联度为: } \psi_i = \frac{1}{n} \sum_{k=1}^n \xi_i(k)$$

$\psi_i$  从小到大排列构成关联度序列，关联度的先后顺序决定影响因素的重要程度。

根据以上定义，对其余指标计算灰色关联度，具体数值如表 4、表 5 所示。

表 4 各指标与城镇居民养老金收入灰色关联度

	<i>TDL</i>	<i>JFL</i>	<i>CBZZ</i>	<i>CBLT</i>	<i>CZBT</i>	<i>GZSP</i>	<i>XFJZ2</i>
灰色关联度	0.627	0.774	0.889	0.864	0.875	0.877	0.747

表 5 各指标与城镇居民养老金支出灰色关联度

	<i>TDL</i>	<i>JFL</i>	<i>CBZZ</i>	<i>CBLT</i>	<i>CZBT</i>	<i>GZSP</i>	<i>XFJZ2</i>
灰色关联度	0.618	0.778	0.852	0.847	0.877	0.844	0.728

根据表 4、表 5 可以对各指标与城镇居民养老金收入、支出关系分别排序如下：

$$CBZZ \succ GZSP \succ CZBT \succ CBLT \succ JFL \succ XFJZ2 \succ TDL \tag{9}$$

$$CZBT \succ CBZZ \succ CBLT \succ GZSP \succ JFL \succ XFJZ2 \succ TDL \tag{10}$$

根据上述结果可以看出，各指标仍然无法直接判断哪些变量可以直接剔除，只能根据自变量之间的相关性和实际常识并结合灰色关联度进行自变量的进一步筛选。

变量具体选取过程如下：

（1）由各指标间的相关系数表 2、表 3 可知，参保在职人数（*CBZZ*）与参保离退休人数（*CBLT*）之间存在很强相关性（相关系数为 0.999），因此剔除参保离退休人数（*CBLT*）；

（2）参保在职人数（*CBZZ*）与工资水平（*GZSP*）之间存在很强的相关性（相关系数为 0.999），因此剔除工资水平（*GZSP*）。

经过上述指标剔除后，得到影响城镇居民养老金收入、支出的指标主要有：替代率（*TDL*）、缴费率（*JFL*）、参保在职人数（*CBZZ*）、财政补贴（*CZBT*）、消费价格指数（以 2000 年为基期）（*XFJZ2*）。对剔除后的指标变量再次计算两两之间的相关系数如表 6 所示。

表 6 剩余变量之间的相关系数矩阵

	<i>TDL</i>	<i>JFL</i>	<i>CBZZ</i>	<i>CZBT</i>	<i>XFJZ2</i>
<i>TDL</i>	1	.938**	-.917**	-.904**	-.934**
<i>JFL</i>	.938**	1	-.989**	-.989**	-.984**
<i>CBZZ</i>	-.917**	-.989**	1	.997**	.986**
<i>CZBT</i>	-.904**	-.989**	.997**	1	.983**
<i>XFJZ2</i>	-.934**	-.984**	.986**	.983**	1
**. 在 .01 水平（双侧）上显著相关					
*.在 .05 水平（双侧）上显著相关					

#### 4.1.4 多重共线性诊断

经上述变量剔除操作后剩余 5 个待选变量，这些变量之间是否存在线性相关项不易从常识判断得出，因此需要在城镇居民养老金收入、支出建模之前，对这些变量进行多重共线性诊断。

共线性可以通过容忍度量，若该值小于 0.1 则可能存在共线性问题。在 SPSS 中对上述 5 个指标变量计算其容忍度如下表所示。

表 7 容忍度计算表

因变量	自变量	共线性统计量		因变量	自变量	共线性统计量		因变量	自变量	共线性统计量	
		容忍度	VIF			容忍度	VIF			容忍度	VIF
TDL	JFL	.018	55.813	JFL	TDL	.109	9.155	CBZZ	TDL	.078	12.892
	CBZZ	.005	184.957		CBZZ	.005	201.328		JFL	.012	85.559
	CZBT	.006	161.777		CZBT	.006	179.594		CZBT	.013	78.605
	XFJZ2	.024	40.834		XFJZ2	.021	46.668		XFJZ2	.022	45.548

因变量	自变量	共线性统计量		因变量	自变量	共线性统计量	
		容忍度	VIF			容忍度	VIF
CZBT	TDL	.103	9.715	XFJZ2	TDL	.082	12.262
	JFL	.015	65.754		JFL	.012	85.436
	CBZZ	.015	67.720		CBZZ	.005	196.215
	XFJZ2	.022	45.424		CZBT	.004	227.136

表 7 的计算结果显示容忍度大都小于 0.1，各变量之间可能存在共线性问题，因此一般不建议直接进行多变量回归建模，需要对变量进行进一步降维处理。

#### 4.1.5 基于主成分回归的城镇居民养老金收入、支出建模

由于上述的指标之间存在较强的共线性，因此采用主成分回归的方法建立城镇居民养老金收入、支出模型。

主成分分析作为一种处理高维数据的方法，通过正交变换方式将高维数据以尽可能少的投影映射到低维空间，使数据降维达到简化数据结构的目的。它也是一种将多个相关变量以尽可能少的信息损失为原则化为少数几个不相关变量的方法。本文的主成分回归建模基于 SPSS 软件实现，具体过程如下：

1) 解释的总方差表如下所示。

表 8 解释的总方差

成份	初始特征值			提取平方和载入		
	合计	方差的 %	累积 %	合计	方差的 %	累积 %
1	4.849	96.982	96.982	4.849	96.982	96.982
2	.122	2.433	99.415			
3	.018	.352	99.767			
4	.009	.184	99.951			
5	.002	.049	100.000			

由上表可知，相关系数矩阵的特征根分别为： $\lambda_1 = 4.849$ ， $\lambda_2 = 0.122$ ，



$\lambda_3 = 0.018$  ,  $\lambda_4 = 0.009$  ,  $\lambda_5 = 0.002$  。只有第一个主成分的特征根大于 1，且其方差占有主成分方差的 96.982%，即包含了原始变量的 96.982%的信息，而一般主成分的累计贡献率达到 80%以上即能令人满意。该结果可以直观地从碎石图看出，如下图所示。基于以上分析，在回归建模时可以取第一个主成分。

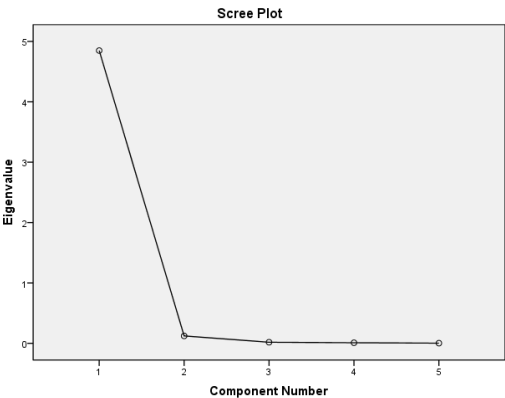


图 3 碎石图

2) 成分矩阵

表9 成分矩阵

指标	成份
	1
<i>TDL</i>	.997
<i>JFL</i>	.968
<i>CBZZ</i>	.986
<i>CZBT</i>	.887
<i>XFJZ2</i>	.957
提取方法 :主成分	

3) 计算特征向量矩阵

根据表9给出的成分矩阵，可以计算出每一特征根对应的单位特征向量，如表10所示。

表10 特征向量矩阵

特征向量
-0.952
-0.995
0.993
0.990
0.993

由上表可进一步得到第一个主成分  $F_1$  的计算公式如下：

$$F_1 = -0.952TDL' - 0.995JFL' + 0.993CBZZ' + 0.990CZBT' + 0.993XFJZ2' \quad (11)$$

其中， $TDL'$ ， $JFL'$ ， $CBZZ'$ ， $CZBT'$ 和 $XFJZ2'$ 分别代表对应变量的标准化得分值。

#### 4) 计算主成分矩阵

根据上述主成分计算公式可以计算出第一个主成分的具体数值如下表所示。

表 11 主成分矩阵

$F_1$	-6.92	-4.93	-3.46	-2.21	-0.27	1.93	2.82	5.42	7.61
-------	-------	-------	-------	-------	-------	------	------	------	------

在计算得到主成分矩阵后，可以对主成分进行线性回归，分别拟合城镇居民养老金收入、支出数据，建立收入、支出模型。结果如下：

表 12 模型参数及检验指标

模型		系数	t	Sig.
养老金收入	(常量)	-1.284E-005	.000	1.000
	$F_1$	.203	15.197	.000
养老金支出	(常量)	-1.282E-005	.000	1.000
	$F_1$	.203	14.353	.000

表13 模型汇总

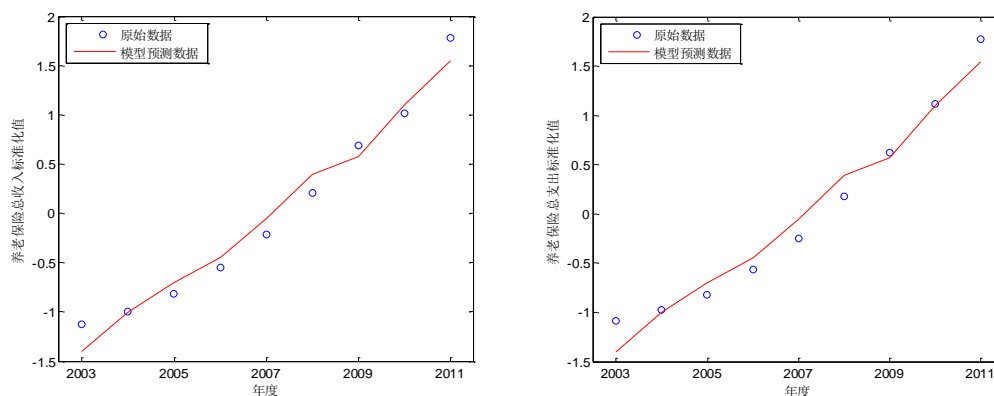
模型	R	R 方	调整 R 方	估计标准误差
养老金收入	.985 <sup>a</sup>	.971	.966	.183355288
养老金支出	.983 <sup>a</sup>	.967	.962	.193796562
a. 预测变量: (常量), $F_1$				

由上述主成分回归分析的结果可知，对于养老金收入与支出而言，调整后的R方分别为0.966与0.962，均大于0.95，说明模型拟合程度很高；变量的显著性水平均为0.000，说明模型显著性很好。

根据以上回归结果，可以得到城镇居民养老金收入、支出的模型为：

$$\begin{cases} CZYS' = -1.284 \times 10^{-5} + 0.203180F_1 \\ CZYZ' = -1.282 \times 10^{-5} + 0.202819F_1 \end{cases} \quad (12)$$

拟合结果如下图所示，从图中也可直观地看出这两个模型拟合程度很高。



(a) 收入

(b) 支出

图 4 主成分回归城镇居民养老金收入、支出模型曲线

#### 4.1.6 基于多重线性回归的城镇居民养老金收入、支出建模

基于主成分回归的城镇居民养老金收入、支出模型具有良好的拟合程度与显著性，但由于模型基于标准化参数，无法对原始数据进行长期仿真。因此本小节采用多重线性回归思想建立城镇居民养老金收入、支出模型。

多重线性回归模型是指含有多个解释变量的线性回归模型，用于表示被解释变量与其他多个解释变量之间的线性关系。其回归方程为：

$$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p \quad (13)$$

本文的多重线性回归建模基于 SPSS 软件实现，具体过程如下：

经过前文灰色关联分析与相关分析筛选，可得到 5 个指标：替代率（*TDL*）、缴费率（*JFL*）、参保在职人数（*CBZZ*）、财政补贴（*CZBT*）、消费价格指数（以 2000 年为基期）（*XFJZ2*）。提取这 5 个指标在表 1 中对应的原始数据，采用逐步回归法建立多重线性回归模型，结果如下：

表 14 模型参数及检验指标

模型		系数	t	Sig.
养老金收入	(常量)	-42909.367	-4.245	.005
	参保在职人数	1.823	11.430	.000
	缴费率	135299.066	3.021	.023
养老金支出	(常量)	-17257.155	-7.799	.000
	参保在职人数	1.140	27.890	.000
	替代率	12412.330	3.804	.009

表15 模型汇总

模型	R	R 方	调整 R 方	估计标准误差
养老金收入	.999 <sup>a</sup>	.998	.998	224.01234
养老金支出	.999 <sup>b</sup>	.998	.998	155.42602
a. 预测变量: (常量), 参保在职人数, 缴费率				
b. 预测变量: (常量), 参保在职人数, 替代率				

上述分析结果说明养老金收入模型以参保在职人数（*CBZZ*）与缴费率（*JFL*）作为自变量，支出模型以参保在职人数（*CBZZ*）与替代率（*TDL*）作为自变量。调整后的R方均为0.998，说明两个模型拟合程度都很高；各个自变量系数均通过了显著性检验（ $p < 0.05$ ），模型显著性水平较高。

根据以上回归结果，可以得到城镇居民养老金收入、支出的模型为：

$$\begin{cases} CZYS = -42909.367 + 1.823 \times CBZZ + 135299.066 \times JFL \\ CZYZ = -17257.155 + 1.140 \times CBZZ + 12412.33 \times TDL \end{cases} \quad (14)$$

拟合结果如下图所示，从图中也可直观地看出这两个模型拟合程度很高。

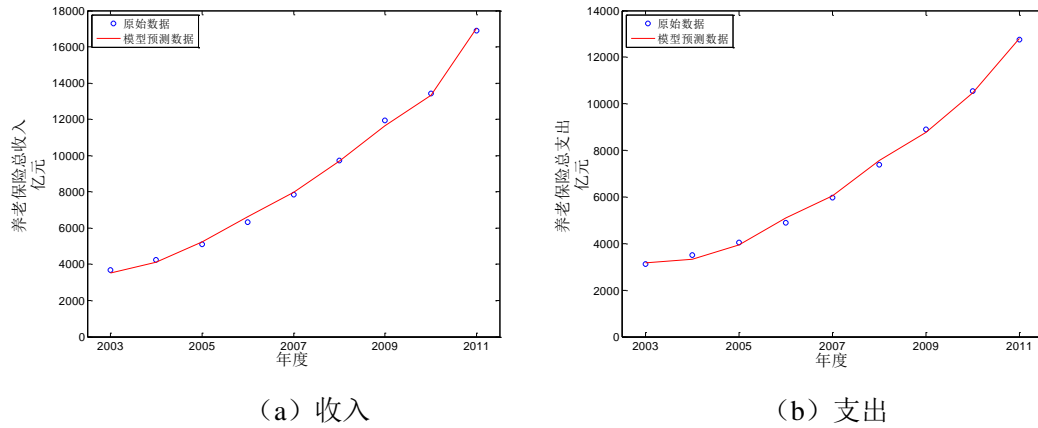


图 5 多重线性回归城镇居民养老金收入、支出模型曲线

#### 4.1.7 基于支持向量机的城镇居民养老金收入、支出建模

##### 1) 支持向量机基本概念

支持向量机 (SVM) 方法是建立在统计学习理论的 VC 维理论和结构风险最小原理基础上的, 根据有限的样本信息在模型的复杂性和学习能力之间寻求最佳折衷, 以期获得最好的泛化能力。它在解决小样本、非线性及高维模式识别中表现出许多特有的优势, 并能够推广应用到函数拟合等其他机器学习问题中<sup>[15]</sup>。

利用 SVM 进行时间序列建模是当前的研究热点, 结合本题的建模要求可知所有数据均为时间序列数据, 因此可采用  $\varepsilon$ -SVR 对城镇居民养老金收入、支出进行建模。

该模型优化函数为:

$$\begin{aligned} \min_{\omega, b, \xi, \xi^*} \quad & \frac{1}{2} \omega^T \omega + C \sum_{i=1}^l \xi_i + C \sum_{i=1}^l \xi_i^* \\ \text{subject to} \quad & \omega^T \phi(x_i) + b - z_i \leq \varepsilon + \xi_i \\ & z_i - \omega^T \phi(x_i) - b \leq \varepsilon + \xi_i^* \\ & \xi_i, \xi_i^* \geq 0, i = 1, \dots, l \end{aligned} \quad (15)$$

最终建立的模型形式 (决策函数) 为:

$$\sum_{i=1}^l (-\alpha_i + \alpha_i^*) K(x_i, x) + b \quad (16)$$

其中, 核函数  $K(x_i, x)$  采用 RBF 核函数, 其函数形式为:

$$K(x_i, x) = \exp\left(-\frac{1}{2\sigma^2} \|x_i - x\|^2\right) \quad (17)$$

$$0 \leq \alpha_i, \alpha_i^* \leq C, i = 1, \dots, l$$

##### 2) 支持向量机城镇居民养老金收入、支出建模

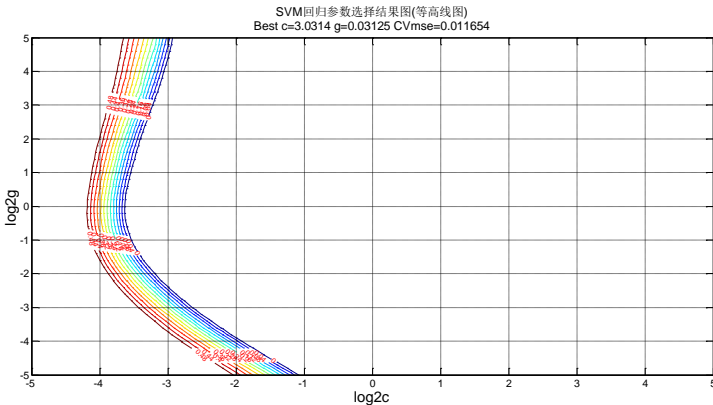
由 4.1.3 小节的剩余指标自变量的相关系数和灰色关联度分析以及 4.1.4 小节

的多重共线性诊断分析可知，自变量之间的相关性比较大，因此自变量之间的冗余性较大，不适合直接进行多变量回归建模。所以本小节建立城镇居民养老金收入、支出与替代率（*TDL*）、缴费率（*JFL*）、参保在职人数（*CBZZ*）、参保离退休人数（*CBLT*）、财政补贴（*CZBT*）、工资水平（*GZSP*）、消费价格指数（以2000年为基期）（*XFJZ2*）7个指标之间的SVM模型。

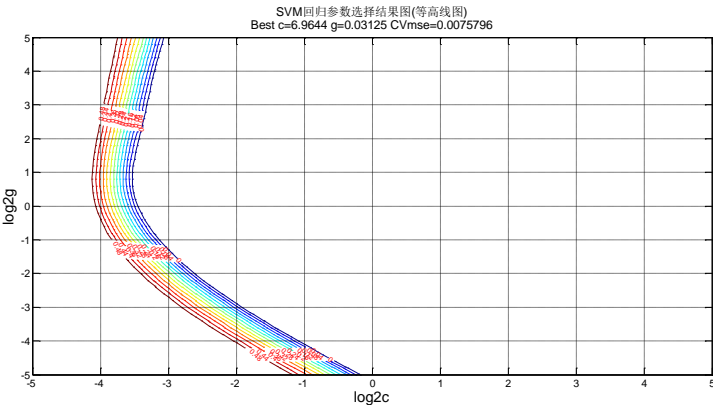
求解过程为：

首先对原始数据进行0-1量化处理，然后将处理后的数据作为训练样本对SVM学习训练，其中惩罚系数*C*、核函数宽度 $\sigma^2$ 的选择是关键。惩罚系数*C*控制经验风险和VC维的平衡，用于实现经验风险和置信范围的折衷，*C*值越大，对函数的拟合程度就越高，误差就越小，过小的*C*会导致训练数据欠学习。核函数的宽度 $\sigma^2$ 控制SVM对输入数据变化的敏感程度，其值越小，SVM对输入数据的反应越不灵敏。此外，不敏感损失函数 $\varepsilon$ 的变化也会对模型的精度产生一定影响。

本文采用libSVM工具箱<sup>[16]</sup>运用网络法<sup>[17]</sup>对*c*、*g*（ $g=\frac{1}{2\sigma^2}$ ）寻优，其优化结果为：对于收入模型，*c*=3.0314，*g*=0.03125；对于支出模型，*c*=6.9644，*g*=0.03125。回归参数选择的等高线和3D图分别如下所示。

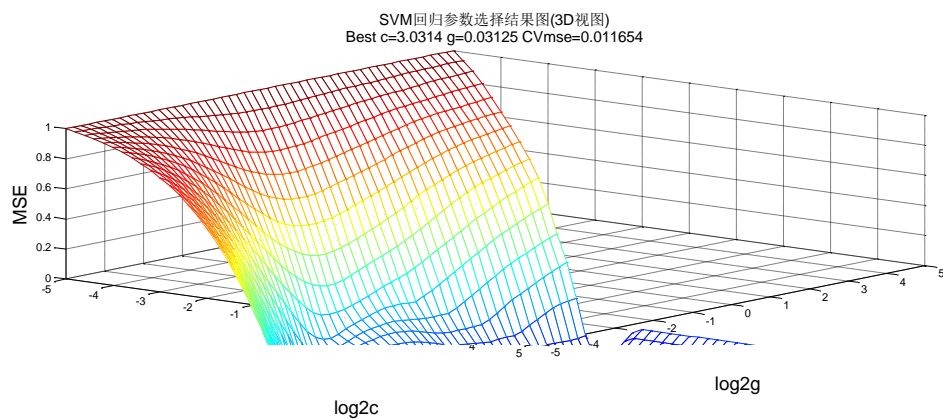


(a) 收入

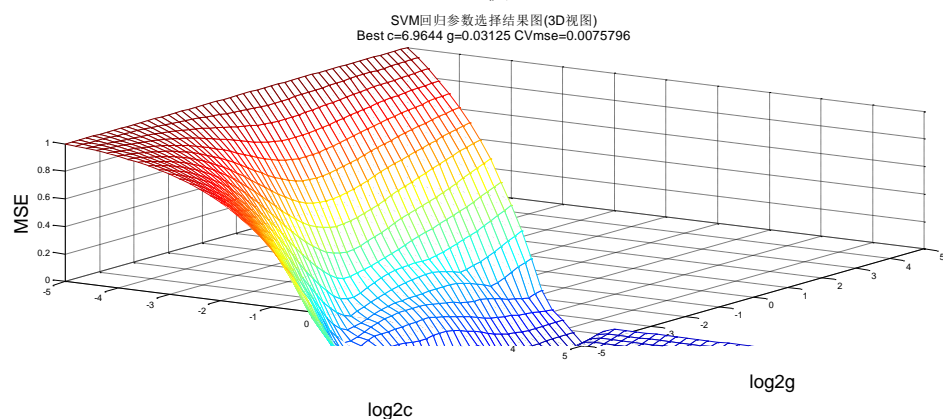


(b) 支出

图6 SVM回归参数选择结果等高线图



(a) 收入

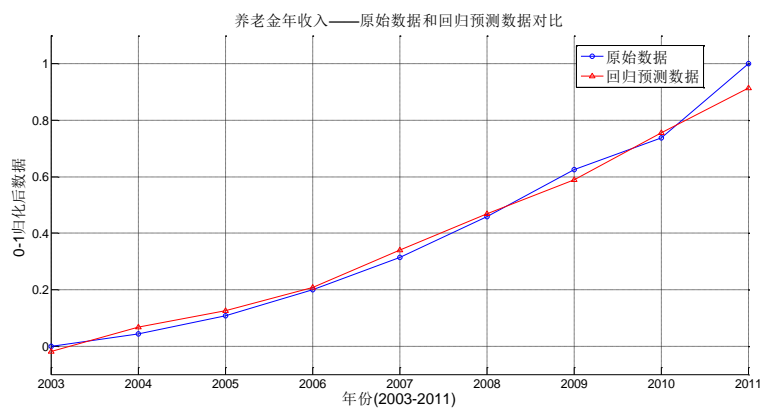


(b) 支出

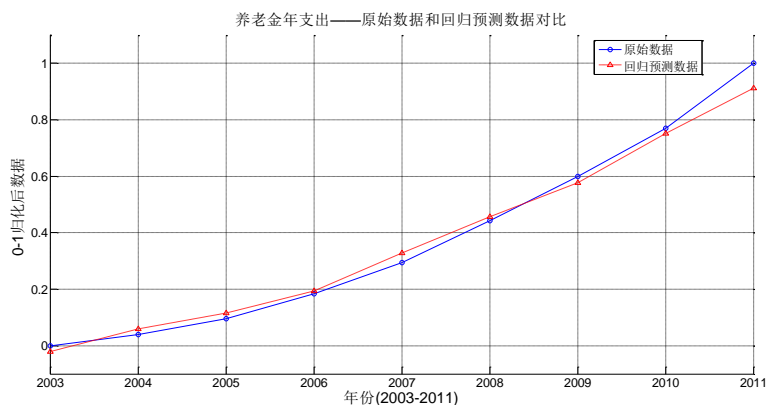
图7 SVM回归参数选择结果3D图

回归模型的预测数据与原始数据的对比如图 8 所示，从图中可知，这两个模型可以较好地反映城镇居民养老金收入、支出的变化趋势。

该模型的指标如下（模型数据均为 0-1 量化数据）：收入模型的均方误差  $MSE = 0.011654$ ，拟合度  $R^2 = 99.11\%$ ；支出模型的均方误差  $MSE = 0.0075796$ ，拟合度  $R^2 = 99.65\%$ 。



(a) 收入



(b) 支出

图 8 城镇居民养老金收入、支出模型

## 4.2 新农保收入、支出宏观数学模型

我国农村社会养老保险制度建设发展曲折，大体经历两个阶段：第一阶段（1986-2004），农村社会养老保险探索阶段；第二阶段（2004至今），新型农村社会养老保险试点创立阶段。构建新型农村社会养老保险是当前及今后我国社会保障制度建设的重要任务<sup>[5]</sup>。对新农保收入、支出进行分析与建模，有助于更深入了解我国社会保障制度，并对社会保障体系的完善具有重要参考意义。

### 4.2.1 基于多重线性回归的新农保收入、支出建模

初选指标的数据<sup>[9]-[11]</sup>如下表所示。

表16 新农保收入、支出初选数据

年份		2010	2011	2012
XNBS	(亿元)	453	1070	1829
XNBCR	(万人)	10277	32643	48370
XNBSL	(万人)	2863	8525	13075
XNBJ	(亿元)	225	415	594
XNBZ	(亿元)	200	588	1150
XNLJ	(亿元)	423	1199	2302
TDL	(%)	0.4506598	0.44049471	0.44157455
JFL	(%)	0.1541474	0.152444703	0.14748538
XFJZ1	-	103.3	105.4	102.7
TZL	(%)	3.3	5.4	2.7
XFJZ2	-	123.55429	130.2262236	133.742332
SLJZ	-	103.1	104.9	102

然后对以上数据进行粗大误差分析，结果如图9所示（部分结果与图2相同，在此不再重复）。从图2与图9可知，以上指标数据中没有粗大误差，可以直接进行后续的分析处理。经标准化处理后，将指标数据进行相关分析，结果如表17所示。

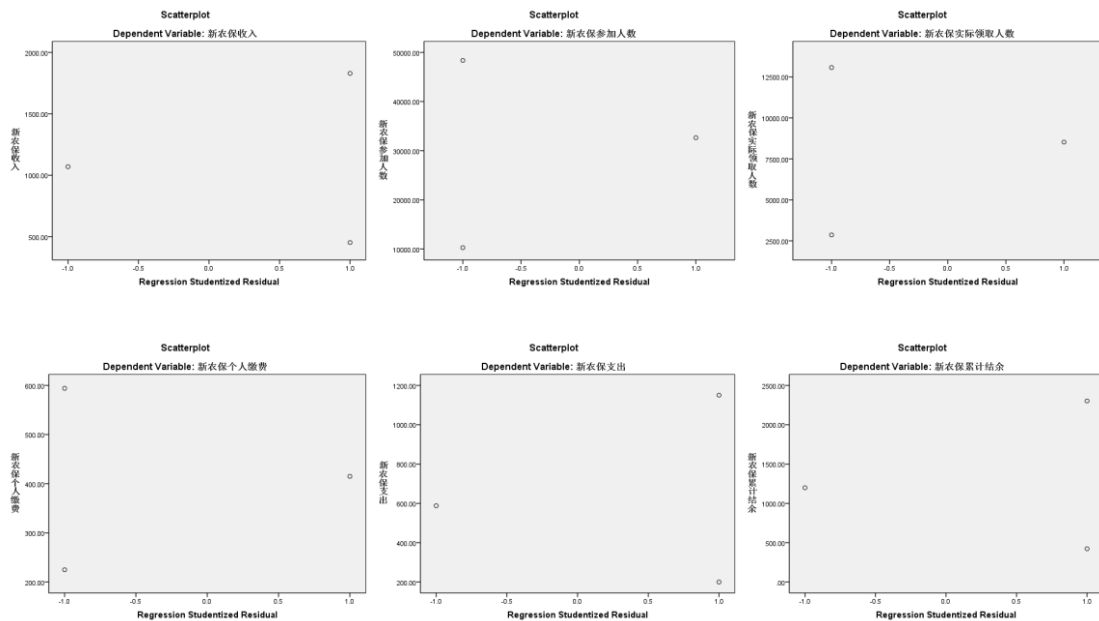


图 9 各指标变量数据学生化残差

表17 各指标间相关系数矩阵（针对新农保）

	<i>XNBS</i>	<i>XNBCR</i>	<i>XNBSL</i>	<i>XNBJ</i>	<i>XNBZ</i>	<i>XNLJ</i>	<i>TDL</i>	<i>JFL</i>	<i>XFJZ1</i>	<i>TZL</i>	<i>XFJZ2</i>	<i>SLJZ</i>
<i>XNBS</i>	1	.999*	.987	.993	.997*	.999*	-.778	-.977	-.269	-.269	.972	-.430
<i>XNBCR</i>	.999*	1	.979	.986	.992	1.000**	-.748	-.986	-.313	-.313	.960	-.471
<i>XNBSL</i>	.987	.979	1	.999*	.997	.980	-.868	-.930	-.113	-.113	.997*	-.281
<i>XNBJ</i>	.993	.986	.999*	1	.999*	.987	-.848	-.943	-.150	-.150	.993	-.317
<i>XNBZ</i>	.997*	.992	.997	.999*	1	.993	-.823	-.958	-.195	-.195	.987	-.360
<i>XNLJ</i>	.999*	1.000**	.980	.987	.993	1	-.751	-.985	-.308	-.308	.962	-.466
<i>TDL</i>	-.778	-.748	-.868	-.848	-.823	-.751	1	.625	-.396	-.396	-.903	-.233
<i>JFL</i>	-.977	-.986	-.930	-.943	-.958	-.985	.625	1	.469	.469	-.900	.613
<i>XFJZ1</i>	-.269	-.313	-.113	-.150	-.195	-.308	-.396	.469	1	1.000**	-.036	.985
<i>TZL</i>	-.269	-.313	-.113	-.150	-.195	-.308	-.396	.469	1.000**	1	-.036	.985
<i>XFJZ2</i>	.972	.960	.997*	.993	.987	.962	-.903	-.900	-.036	-.036	1	-.207
<i>SLJZ</i>	-.430	-.471	-.281	-.317	-.360	-.466	-.233	.613	.985	.985	-.207	1
**. 在 .01 水平（双侧）上显著相关												
*. 在 .05 水平（双侧）上显著相关												

从上表可知，由于指标数据较少，较多自变量与因变量相关性不大。于是结合自变量之间的相关分析，对自变量进行筛选，结果如表 18 所示。

从表 18 中可以看出，*XFJZ1* 与 *XNBCR* 相关性较大，予以剔除；*XNBCR*、*XNBJ* 与 *XNBSL* 相关性较大，予以剔除；*SLJZ*、*XNBS* 与 *XNBZ* 相关性不大，予以剔除。经筛选后剩余 6 个自变量，分别是新农保实际领取人数（*XNBSL*）、新农保累计结余（*XNBLJ*）、替代率（*TDL*）、缴费率（*JFL*）、通货膨胀率（*TZL*）、消费价格指数（以 2000 年为基期）（*XFJZ2*）。

之后进行多重线性回归建模，结果如表 19、表 20 所示。



表18 各指标间相关系数矩阵（针对新农保）

	XNBCR	XNBSL	XNBJ	XNLJ	TDL	JFL	XFJZ1	TZL	XFJZ2	SLJZ
XNBCR	1	.999*	.997	.980	-.868	-.930	-.113	-.113	.997*	-.281
XNBSL	.999*	1	.999*	.987	-.848	-.943	-.150	-.150	.993	-.317
XNBJ	.997	.999*	1	.993	-.823	-.958	-.195	-.195	.987	-.360
XNLJ	.980	.987	.993	1	-.751	-.985	-.308	-.308	.962	-.466
TDL	-.868	-.848	-.823	-.751	1	.625	-.396	-.396	-.903	-.233
JFL	-.930	-.943	-.958	-.985	.625	1	.469	.469	-.900	.613
XFJZ1	-.113	-.150	-.195	-.308	-.396	.469	1	1.000**	-.036	.985
TZL	-.113	-.150	-.195	-.308	-.396	.469	1.000**	1	-.036	.985
XFJZ2	.997*	.993	.987	.962	-.903	-.900	-.036	-.036	1	-.207
SLJZ	-.281	-.317	-.360	-.466	-.233	.613	.985	.985	-.207	1
*. Correlation is significant at the 0.05 level (2-tailed).										
**. Correlation is significant at the 0.01 level (2-tailed).										

表 19 模型参数及检验指标

模型		系数	t	Sig.
新农保收入	(常量)	163.376	3.629	.171
	新农保累计结余	.729	24.600	.026
新农保支出	(常量)	-15.656	-3.909	.159
	新农保累计结余	.506	191.743	.003

表20 模型汇总

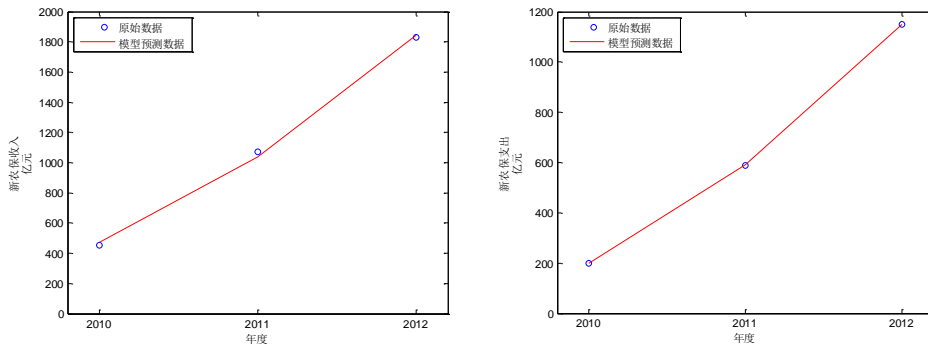
模型	R	R 方	调整 R 方	估计标准误差
新农保收入	.999 <sup>a</sup>	.998	.997	39.58999
新农保支出	1.000 <sup>a</sup>	1.000	1.000	3.52288
a. 预测变量: (常量),新农保累计结余				

上述分析结果说明新农保收入、支出模型均以新农保累计结余（XNBLJ）作为自变量。调整后的R方分别为0.997与1.000，说明两个模型拟合程度都很高；各个自变量系数均通过了显著性检验（ $p<0.05$ ），模型显著性水平较高。

根据以上回归结果，可以得到新农保收入、支出的模型为：

$$\begin{cases} XNBS = 163.375529 + 0.729326 \times XNBLJ \\ XNBZ = -15.655922 + 0.505853 \times XNBLJ \end{cases} \quad (18)$$

拟合结果如下图所示，从图中也可直观地看出这两个模型拟合程度较高。



(a) 收入

(b) 支出

图 10 多重线性回归新农保收入、支出模型曲线

#### 4.2.2 基于支持向量机的新农保收入、支出建模

以下利用支持向量机工具，对新农保收入与支出进行建模研究。对原始数据的处理过程与4.1.7节基本相同，即经过粗大误差分析、数据标准化处理、相关分析后，得到替代率（*TDL*）、缴费率（*JFL*）、参保在职人数（*CBZZ*）、参保离退休人数（*CBLT*）、财政补贴（*CZBT*）、工资水平（*GZSP*）、消费价格指数（以2000年为基期）（*XFJZ2*）7个指标变量，之后建立新农保收入、支出与这7个变量之间的SVM模型。

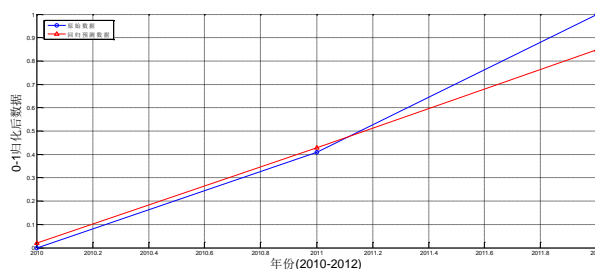
模型优化结果分别为：对收入， $c=2.639$ ， $g=0.0313$ ；对支出， $c=2.639$ ， $g=0.0313$ 。

回归模型的预测数据与原始数据的对比如图11所示，从图中可知，模型可以较好地反映新农保收入、支出的变化趋势。

该模型的指标如下（模型数据均为0-1量化数据）：对收入，均方误差  $MSE = 0.2302$ ，拟合度  $R^2 = 99.4\%$ ；对支出，均方误差  $MSE = 0.2368$ ，拟合度  $R^2 = 99.05\%$ 。



(a) 收入



(b) 支出

图11 新农保收入与支出模型

### 4.3 多层次养老保险体系数学模型研究

早在 1991 年,国务院就提出了要逐步建立起基本养老保险与企业补充养老保险和职工个人储蓄型养老保险相结合的制度<sup>[18]</sup>。十四届三中全会进一步做出建立多层次养老保险体系的重要决策。企业年金和个人储蓄养老保险是多层次养老保险体系的重要组成部分,是对国家法定基本养老保险的必要补充。

本小节针对企业年金与个人储蓄养老保险,分别选取企业年金积累基金(*QNJ*)与商业人寿保险费收入(*SRBS*)两个指标作为衡量标准,选择企业参保在职人数(*QCZZ*)、参保企业数量(*CBQ*)、替代率(*TDL*)、缴费率(*JFL*)、0-20 岁人口死亡率(*SWL1*)、21-60 岁人口死亡率(*SWL2*)、61 岁以上人口死亡率(*SWL3*)、GDP 增长率(*GDPZ*)、GDP 人均增长率(*GDPRZ*)、工资水平(*GZP*)、消费价格指数(以上一年为基期)(*XFJZ1*)、通货膨胀率(*TZL*)、消费价格指数(以 2000 年为基期)(*XFJZ2*)、商品零售价格指数(以上一年为基期)(*SLJZ*)等 14 个指标进行建模,从而对企业年金与个人储蓄养老保险情况进行分析与评估。

#### 4.3.1 企业年金数学模型研究

本小节首先从题目附件和国家统计局等相关网站查到上述指标的年度数据;然后对数据进行预处理等分析操作;随后采用相关分析、灰色关联度确定各指标之间的相关程度,并从中进一步筛选出最能反映企业年金积累基金的指标;接着对筛选得到的变量进行多重共线性诊断以判断多元回归建模的适用性;最后利用选定指标完成企业年金积累基金数学模型主成分回归建模、多重线性回归建模与支持向量机建模,并对模型的有效性和精确性进行分析。

##### 1) 数据预处理

##### 原始数据提取

经过整理后的初选指标的数据<sup>[9]-[11][19]-[21]</sup>如表 14 所示。

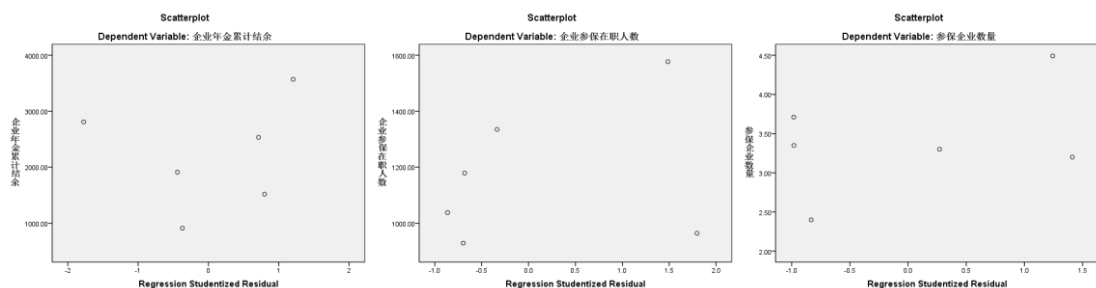
表 21 各初选指标变量年度数据汇总表

年份		2006	2007	2008	2009	2010	2011
<i>QYNJ</i>	(亿元)	910	1519	1911	2533	2809	3570
<i>QCZZ</i>	(万人)	929	1038	1179	1335	1577	1847
<i>CBQ</i>	(万家)	3.2	3.3	3.35	3.71	4.49	5.47
<i>TDL</i>	(%)	0.5	0.48	0.48	0.47	0.45	0.44
<i>JFL</i>	(%)	0.18	0.17	0.17	0.16	0.15	0.15
<i>CBZZ</i>	(万人)	14130	15183	16587.5	17743	19402.3	21565
<i>CBLT</i>	(万人)	4635	4953.7	5303.6	5806.9	6305	6826
<i>SWL1</i>	(ppm)	1148.1	1188.4	1054.6	1035.71	1293.6	983.7
<i>SWL2</i>	(‰)	16.1	15.71	14.26	15.04	16.1	15.29
<i>SWL3</i>	(‰)	353.49	333.94	310.26	325.9	394.49	319.01
<i>GDPZ</i>	(%)	12.7	14.2	9.6	9.2	10.4	9.3
<i>GDPRZ</i>	(%)	12	13.6	9.1	8.7	9.9	8.8
<i>CZBT</i>	(亿元)	971	1157	1437	1646	1954	2272
<i>GZSP</i>	(元)	21001	24932	29229	32736	37147	42452

XFJZ1	-	101.5	104.8	105.9	99.1	103.3	105.4
TZL	(%)	1.5	4.8	5.9	-0.9	3.3	5.4
XFJZ2	-	108.53	113.74	120.45	119.61	123.55	130.23
SLJZ	(%)	101	103.8	105.9	98.9	103.1	104.9
TZSYL	(%)	29.01	43.19	-6.79	16.12	4.23	0.84
TZSYE	(亿元)	619.79	1453.5	-393.72	850.43	321.22	73.37

### 粗大误差分析

根据学生化残差对各指标年度数据进行计算，结果如图 7 所示（其他指标学生化结果如图 1 所示）。



(a) 企业年金积累基金

(b) 企业参保在职人数

(c) 参保企业数量

图12 学生化残差图

从图 1、图 7 可以看出，以上指标数据中没有粗大误差，可以直接进行后续的分析处理。

### 数据标准化处理

根据式（6）对所有指标数据进行标准化处理。

### 2) 基于相关分析的指标变量分析

利用上述经过标准化处理后的数据在 SPSS 软件中进行相关分析，可得到各指标之间的相关系数值如表 15 所示。

表 22 各指标间相关系数矩阵（针对企业年金）

	QNJJ	QCZZ	CBQ	TDL	JFL	SWL1	SWL2	SWL3	GDPZ	GDPRZ	GZSP	XFJZ1	TZL	XFJZ2	SLJZ
QNJJ	1	.947**	.955**	.984**	.961**	-.316	-.186	-.034	-.732	-.721	.995**	.162	.162	.969**	.202
QCZZ	.947**	1	.894*	.924**	.932**	-.289	-.001	.076	-.674	-.670	.960**	.148	.148	.908*	.183
CBQ	.955**	.894*	1	.963**	.919**	-.321	-.221	-.139	-.584	-.571	.952**	.418	.418	.967**	.437
TDL	-.984**	-.924**	-.963**	1	.983**	.172	.124	-.068	.657	.644	-.989**	-.256	-.256	-.967**	-.293
JFL	-.961**	-.932**	-.919**	.983**	1	.115	.103	-.158	.704	.693	-.982**	-.258	-.258	-.959**	-.311
SWL1	-.316	-.289	-.321	.172	.115	1	.696	.877*	.536	.542	-.269	-.033	-.033	-.354	-.060
SWL2	-.186	-.001	-.221	.124	.103	.696	1	.799	.571	.567	-.160	-.268	-.268	-.333	-.340
SWL3	-.034	.076	-.139	-.068	-.158	.877*	.799	1	.230	.230	.021	-.237	-.237	-.127	-.239
GDPZ	-.732	-.674	-.584	.657	.704	.536	.571	.230	1	1.000**	-.734	.078	.078	-.757	-.032
GDPRZ	-.721	-.670	-.571	.644	.693	.542	.567	.230	1.000**	1	-.724	.089	.089	-.746	-.021
GZSP	.995**	.960**	.952**	-.989**	-.982**	-.269	-.160	.021	-.734	-.724	1	.203	.203	.976**	.248
XFJZ1	.162	.148	.418	-.256	-.258	-.033	-.268	-.237	.078	.089	.203	1	1.000**	.360	.991**
TZL	.162	.148	.418	-.256	-.258	-.033	-.268	-.237	.078	.089	.203	1.000**	1	.360	.991**

<i>XFJZ2</i>	.969**	.908*	.967**	-.967**	-.959**	-.354	-.333	-.127	-.757	-.746	.976**	.360	.360	1	.413
<i>SLJZ</i>	.202	.183	.437	-.293	-.311	-.060	-.340	-.239	-.032	-.021	.248	.991**	.991**	.413	1
**. 在 .01 水平（双侧）上显著相关															
*.在 .05 水平（双侧）上显著相关															

由上表可知以上指标中 0-20 岁人口死亡率（*SWL1*）、21-60 岁人口死亡率（*SWL2*）、61 岁以上人口死亡率（*SWL3*）、GDP 增长率（*GDPZ*）、GDP 人均增长率（*GDPZ*）、消费价格指数（以上一年为基期）（*XFJZ1*）、通货膨胀率（*TZL*）、商品零售价格指数（*SLJZ*）与企业年金积累基金相关关系不强。因此，首先将这 8 个指标剔除，再对其余指标进行下一步分析。

### 3) 基于灰色关联度的指标变量分析

对其余 6 个指标计算灰色关联度，具体数值如表 16 所示。

表 23 各指标与企业年金积累基金灰色关联度

	<i>QCZZ</i>	<i>CBQ</i>	<i>TDL</i>	<i>JFL</i>	<i>GZSP</i>	<i>XFJZ2</i>
灰色关联度	0.592	0.778	0.826	0.775	0.885	0.834

根据自变量之间的相关性并结合灰色关联度，并未发现自变量间关联性很显著，故不剔除任何自变量。

### 4) 多重共线性诊断

这一步操作与 4.1.4 节中基本相同，各变量之间可能存在共线性问题，因此不建议直接进行多变量回归建模。

### 5) 基于主成分回归的企业年金积累基金建模

这一步操作与 4.1.5 节中基本相同，第一个主成分  $F_1$  的计算公式如下：

$$F_1 = 0.956 \times QCZZ' + 0.969 \times CBQ' - 0.991 \times TDL' - 0.983 \times JFL' + 0.997 \times GZSP' + 0.983 \times XFJZ' \quad (19)$$

模型结果如下：

表 24 模型参数及检验指标

模型		系数	t	Sig.
企业年金积累基金	(常量)	1.715E-007	.000	1.000
	$F_1$	.172	13.188	.000

表25 模型汇总

模型	R	R 方	调整 R 方	估计标准误差
企业年金积累基金	.989 <sup>a</sup>	.978	.972	.167630269
a. 预测变量: (常量), $F_1$				

由上述主成分回归分析的结果可知，对于企业年金积累基金而言，调整后的 R 方为 0.978，大于 0.95，说明模型拟合程度很高；变量的显著性水平为 0.000，说明模型显著性很好。

根据以上回归结果，可以得到企业年金积累基金模型为：

$$QYNJ' = 1.715 \times 10^{-7} + 0.171647 \times F_1 \quad (20)$$

拟合结果如图 9 所示，从图中也可直观地看出模型拟合程度很高。

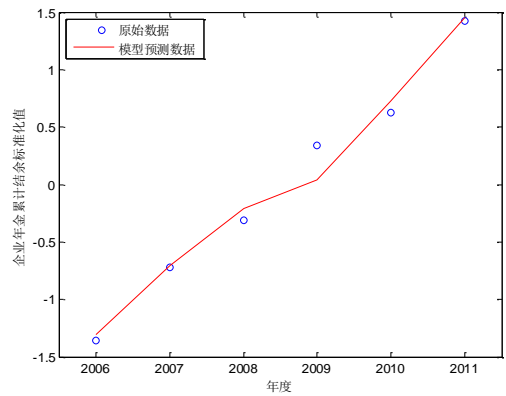


图 13 主成分回归企业年金积累基金模型曲线

#### 6) 基于多重线性回归的企业年金积累基金建模

与 4.1.6 节类似，本小节建立基于多重线性回归的企业年金积累基金模型。具体过程如下：

经过上文灰色关联分析与相关分析筛选，可得到 6 个指标：企业参保在职人数（*QCZZ*）、参保企业数量（*CBQ*）、替代率（*TDL*）、缴费率（*JFL*）、工资水平（*GZSP*）、消费价格指数（以 2000 年为基期）（*XFJZ2*）。提取这 6 个指标在表 1 中对应的原始数据，采用逐步回归法建立多重线性回归模型，结果如下：

表 26 模型参数及检验指标

模型		系数	t	Sig.
企业年金积累基金	(常量)	-1557.261	-8.035	.001
	工资水平	.121	19.941	.000

表27 模型汇总

模型	R	R 方	调整 R 方	估计标准误差
企业年金积累基金	.995 <sup>a</sup>	.990	.988	106.66458
a. 预测变量: (常量), 工资水平				

上述分析结果说明企业年金积累基金模型以工资水平（*GZSP*）作为自变量。调整后的R方为0.988，说明模型拟合程度很高；变量的显著性水平为0.000，说明模型显著性很好。

根据以上回归结果，可以得到企业年金积累基金的模型为：

$$QYNJ = -1557.260985 + 0.120512 \times GZSP \quad (21)$$

拟合结果如下图所示，从图中也可直观地看出模型拟合程度较高。

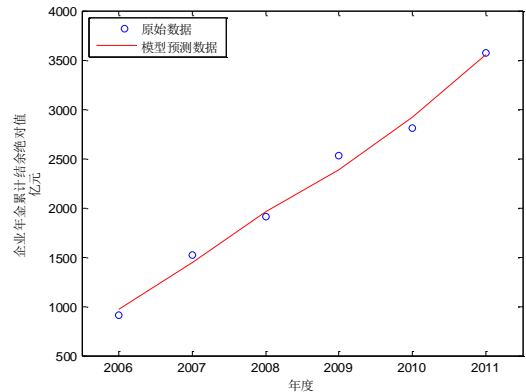


图 14 多重线性回归企业年金积累基金模型曲线

#### 7) 基于支持向量机的企业年金积累基金建模

与4.1.7节类似，本小节建立企业年金积累基金（*QNJJ*）与企业参保在职人数（*QCZZ*）、参保企业数量（*CBQ*）、替代率（*TDL*）、缴费率（*JFL*）、工资水平（*GZSP*）、消费价格指数（以2000年为基期）（*XFJZ2*）6个指标之间的SVM模型。

模型优化结果为： $c=3.0314$ ， $g=0.0313$ 。

回归模型的预测数据与原始数据的对比如图15所示，从图中可知，模型可以较好地反映企业年金积累基金的变化趋势。

该模型的指标如下（模型数据均为0-1量化数据）：均方误差 $MSE = 0.0282$ ，拟合度 $R^2 = 98.71\%$ 。

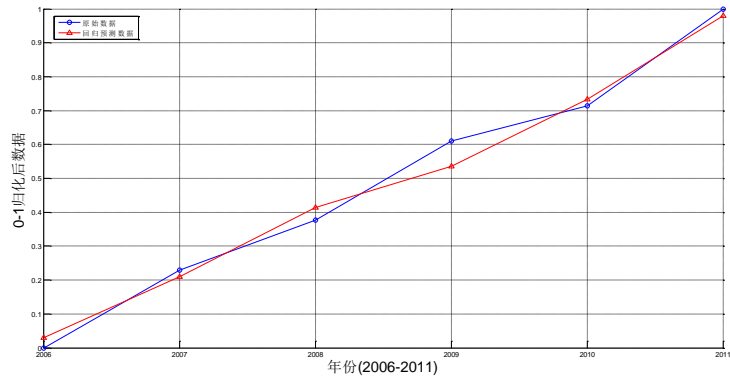


图15 企业年金积累基金模型

### 4.3.2 个人储蓄养老保险数学模型研究

本小节首先从题目附件和国家统计局等相关网站查到各指标年度数据；然后对数据进行预处理等分析操作；随后采用相关分析、灰色关联度确定各指标之间的相关程度，并从中进一步筛选出最能反映商业人寿保险费收入的指标；

接着对筛选得到的变量进行多重共线性诊断以判断多元回归建模的适用性；最后利用选定指标完成商业人寿保险费收入数学模型主成分回归建模、多重线性回归建模和支持向量机建模并对模型的有效性和精确性进行分析。

1) 数据预处理

原始数据提取

经过整理后的初选指标的数据<sup>[9]-[11][19]-[21]</sup>如表 25 所示。

表 28 各初选指标变量年度数据汇总表

年份		2005	2006	2007	2008	2009	2010	2011
SRBS	(亿元)	3247	3593	4463.4	6658.4	7457.4	9679.5	8695.6
TDL	(%)	0.5	0.5	0.48	0.48	0.47	0.45	0.44
JFL	(%)	0.18	0.18	0.17	0.17	0.16	0.15	0.15
CBZZ	(万人)	13120	14130	15183	16587.5	17743	19402.3	21565
CBLT	(万人)	4367.5	4635	4953.7	5303.6	5806.9	6305	6826
SWL1	(ppm)	1232.47	1148.1	1188.4	1054.6	1035.71	1293.6	983.7
SWL2	(‰)	20.73	16.1	15.71	14.26	15.04	16.1	15.29
SWL3	(‰)	398.12	353.49	333.94	310.26	325.9	394.49	319.01
GDPZ	(%)	11.3	12.7	14.2	9.6	9.2	10.4	9.3
GDPRZ	(%)	10.7	12	13.6	9.1	8.7	9.9	8.8
CZBT	(亿元)	651	971	1157	1437	1646	1954	2272
GZSP	(元)	18364	21001	24932	29229	32736	37147	42452
XFJZ1	-	101.8	101.5	104.8	105.9	99.1	103.3	105.4
TZL	(%)	1.8	1.5	4.8	5.9	-0.9	3.3	5.4
XFJZ2	-	106.93	108.53	113.74	120.45	119.61	123.55	130.23
SLJZ	(%)	100.8	101	103.8	105.9	98.9	103.1	104.9
TZSYL	(%)	4.16	29.01	43.19	-6.79	16.12	4.23	0.84
TZSYE	(亿元)	71.22	619.79	1453.5	-393.72	850.43	321.22	73.37

粗大误差分析

根据学生化残差对各指标年度数据进行计算，结果如图 13 所示（除商业人寿保险费收入外其他指标学生化结果如图 1 所示）。

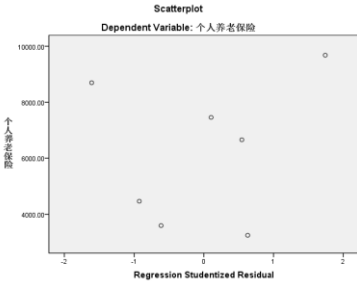


图16 商业人寿保险费收入（亿元）学生化残差图

从图 1、图 16 可以看出，以上指标数据中没有粗大误差，可以直接进行后续的分析处理。

数据标准化处理



根据式（6）对所有指标数据进行标准化处理。

## 2）基于相关分析的指标变量分析

利用上述经过标准化处理后的数据在 SPSS 软件中进行相关分析，可得到各指标之间的相关系数值如表 26 所示。

表 29 各指标间相关系数矩阵（针对商业人寿保险费收入）

	SRBS	TDL	JFL	CBZZ	CBLT	SWL1	SWL2	SWL3	GDPZ	GDPRZ	CZBT	GZSP	XFJZ1	TZL	XFJZ2	SLJZ	TZSYL	TZSYE
SRBS	1	-.952**	-.974**	.937**	.947**	-.233	-.536	-.172	-.695	-.682	.948**	.951**	.207	.207	.928**	.309	-.498	-.252
TDL	-.952**	1	.987**	-.986**	-.988**	.336	.530	.279	.600	.585	-.983**	-.991**	-.335	-.335	-.975**	-.405	.395	.156
JFL	-.974**	.987**	1	-.984**	-.984**	.303	.559	.244	.626	.613	-.987**	-.987**	-.339	-.339	-.972**	-.424	.450	.237
CBZZ	.937**	-.986**	-.984**	1	.998**	-.426	-.552	-.322	-.641	-.630	.995**	.998**	.304	.304	.979**	.380	-.418	-.207
CBLT	.947**	-.988**	-.984**	.998**	1	-.391	-.531	-.281	-.642	-.631	.994**	.997**	.253	.253	.968**	.330	-.406	-.179
SWL1	-.233	.336	.303	-.426	-.391	1	.563	.874*	.523	.526	-.423	-.421	-.120	-.120	-.487	-.176	.224	.202
SWL2	-.536	.530	.559	-.552	-.531	.563	1	.783*	.251	.242	-.618	-.574	-.304	-.304	-.620	-.406	-.043	-.091
SWL3	-.172	.279	.244	-.322	-.281	.874*	.783*	1	.229	.225	-.353	-.334	-.327	-.327	-.431	-.378	-.041	-.033
GDPZ	-.695	.600	.626	-.641	-.642	.523	.251	.229	1	1.000**	-.625	-.650	.056	.056	-.669	-.056	.840*	.659
GDPRZ	-.682	.585	.613	-.630	-.631	.526	.242	.225	1.000**	1	-.613	-.638	.069	.069	-.656	-.043	.841*	.666
CZBT	.948**	-.983**	-.987**	.995**	.994**	-.423	-.618	-.353	-.625	-.613	1	.997**	.297	.297	.979**	.381	-.383	-.174
GZSP	.951**	-.991**	-.987**	.998**	.997**	-.421	-.574	-.334	-.650	-.638	.997**	1	.296	.296	.983**	.376	-.417	-.195
XFJZ1	.207	-.335	-.339	.304	.253	-.120	-.304	-.327	.056	.069	.297	.296	1	1.000**	.422	.987**	-.232	-.327
TZL	.207	-.335	-.339	.304	.253	-.120	-.304	-.327	.056	.069	.297	.296	1.000**	1	.422	.987**	-.232	-.327
XFJZ2	.928**	-.975**	-.972**	.979**	.968**	-.487	-.620	-.431	-.669	-.656	.979**	.983**	.422	.422	1	.505	-.478	-.280
SLJZ	.309	-.405	-.424	.380	.330	-.176	-.406	-.378	-.056	-.043	.381	.376	.987**	.987**	.505	1	-.313	-.394
TZSYL	-.498	.395	.450	-.418	-.406	.224	-.043	-.041	.840*	.841*	-.383	-.417	-.232	-.232	-.478	-.313	1	.934**
TZSYE	-.252	.156	.237	-.207	-.179	.202	-.091	-.033	.659	.666	-.174	-.195	-.327	-.327	-.280	-.394	.934**	1
**, 在 .01 水平（双侧）上显著相关																		
*, 在 .05 水平（双侧）上显著相关																		

由上表可知以上指标中 0-20 岁人口死亡率（SWL1）、21-60 岁人口死亡率（SWL2）、61 岁以上人口死亡率（SWL3）、GDP 增长率（GDPZ）、GDP 人均增长率（GDPRZ）、消费价格指数（以上一年为基期）（XFJZ1）、通货膨胀率（TZL）、商品零售价格指数（SLJZ）、投资收益率（TZSYL）、投资收益额（TZSYE）与商业人寿保险费收入相关关系不强。因此，首先将这 10 个指标剔除，再对其余指标进行下一步分析。

## 3）基于自变量相关分析的指标变量分析

由于剩余的 7 个指标变量与 4.1.3 节中完全相同，因此经相关分析（自变量之间）筛选后得到的影响商业人寿保险费收入的指标同样为替代率（TDL）、缴费率（JFL）、参保在职人数（CBZZ）、财政补贴（CZBT）、消费价格指数（以 2000 年为基期）（XFJZ2）。

## 4）多重共线性诊断

这一步操作与 4.1.4 节中基本相同，各变量之间可能存在共线性问题，因此不建议直接进行多变量回归建模。

## 5）基于主成分回归的商业人寿保险费收入建模

这一步操作与 4.1.5 节中基本相同，第一个主成分  $F_1$  的计算公式如下：

$$F_1 = -0.952 \times TDL' - 0.995 \times JFL' + 0.993 \times CBZZ' + 0.990 \times CZBT' + 0.993 \times XFJZ2' \quad (22)$$

模型结果如下：

表 30 模型参数及检验指标

模型		系数	t	Sig.
商业人寿保险费收入	(常量)	-.404	-2.954	.032
	$F_1$	.239	7.104	.001

表31 模型汇总

模型	R	R 方	调整 R 方	估计标准误差
商业人寿保险费收入	.954 <sup>a</sup>	.910	.892	.328897927
a. 预测变量: (常量), $F_1$				

由上述主成分回归分析的结果可知，对于商业人寿保险费收入而言，调整后的R方为0.892，大于0.85，说明模型拟合程度较高；变量的显著性水平为0.001，说明模型显著性很好。

根据以上回归结果，可以得到商业人寿保险费收入模型为：

$$SRBS' = -0.403707 + 0.238566 \times F_1 \quad (23)$$

拟合结果如图 15 所示，从图中也可直观地看出模型拟合程度较高。

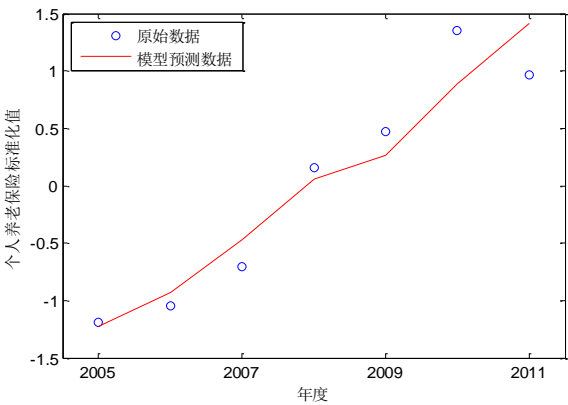


图 17 主成分回归商业人寿保险费收入模型曲线

#### 6) 基于多重线性回归的商业人寿保险费收入建模

与 4.1.6 节类似，本小节建立基于多重线性回归的商业人寿保险费收入模型。具体过程如下：

经过灰色关联分析与相关分析筛选，可得到 5 个指标：替代率（ $TDL$ ）、缴费率（ $JFL$ ）、参保在职人数（ $CBZZ$ ）、财政补贴（ $CZBT$ ）、消费价格指数（以 2000 年为基期）（ $XFJZ2$ ）。提取这 5 个指标在表 1 中对应的原始数据，采用逐步回归法建立多重线性回归模型，结果如下：

表 32 模型参数及检验指标

模型		系数	t	Sig.
商业人寿保险费收入	(常量)	46593.549	11.182	.000
	缴费率	-242929.987	-9.697	.000

表33 模型汇总

模型	R	R 方	调整 R 方	估计标准误差
商业人寿保险费收入	.974 <sup>a</sup>	.950	.939	624.64303
a. 预测变量: (常量), 缴费率				

上述分析结果说明商业人寿保险费收入模型以缴费率（*JFL*）作为自变量。调整后的 R 方为 0.939，说明模型拟合程度较高；变量的显著性水平为 0.000，说明模型显著性很好。

根据以上回归结果，可以得到商业人寿保险费收入的模型为：

$$QYNJ = 46593.549203 - 242929.986528 \times JFL \quad (24)$$

拟合结果如下图所示，从图中也可直观地看出模型拟合程度很高。

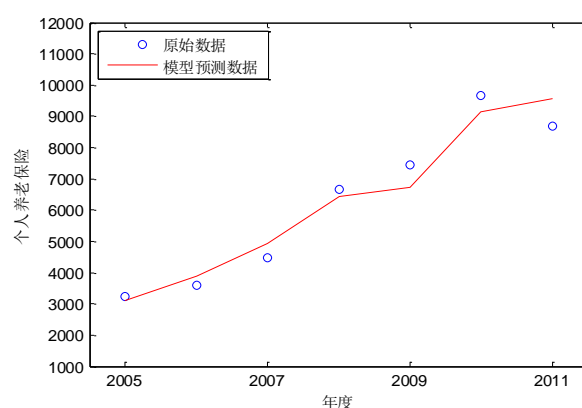


图 18 多重线性回归商业人寿保险费收入模型曲线

### 7) 基于支持向量机的商业人寿保险费收入建模

与4.1.7节类似，本小节建立商业人寿保险费收入（*SRBS*）与替代率（*TDL*）、缴费率（*JFL*）、参保在职人数（*CBZZ*）、参保离退休人数（*CBLT*）、财政补贴（*CZBT*）、工资水平（*GZSP*）、消费价格指数（以2000年为基期）（*XFJZ2*）7个指标之间的SVM模型。

模型优化结果为：c=0.5，g=0.32988。

回归模型的预测数据与原始数据的对比如图18所示，从图中可知，模型可以较好地反映商业人寿保险费收入的变化趋势。

该模型的指标如下（模型数据均为0-1量化数据）：均方误差MSE = 0.02，拟合度R方 = 98.79%。

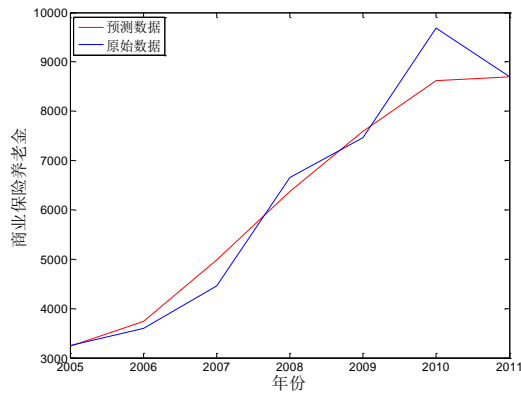


图19 商业人寿保险费收入模型

#### 4.3.3 多层次养老保险体系数学模型探讨

自 20 世纪 90 年代初养老保障体系改革开始以来，多层次养老保障体系历经十多年发展，为保障退休人员基本生活，促进经济发展和社会稳定发挥了重要作用，但随着改革进一步深化，也表现出诸多体制性难题<sup>[7]</sup>。多层次模式各部分功能不清晰，发展不均衡即为其中之一。1995 年国务院《关于深化企业职工养老保险制度改革的通知》指出：“国家在建立基本养老保险、保障离退休人员基本生活的同时，鼓励建立企业补充养老保险和个人储蓄性养老保险，构建保障方式多层次的养老保险体系。”从中可知作为第二层次的企业年金与作为第三层次的个人储蓄养老保险在我国养老保险体系中应当发挥出重要作用。然而实际发展情况却不容乐观。2004 年国家颁布了《企业年金试行办法》和《企业年金基金管理试行办法》，指导企业年金计划发展，并给予了一定税收优惠，但发展十分缓慢，保障水平有限。2009 年《关于企业年金个人所得税征收管理有关问题的通知》明确规定对个人缴费进行征税，企业建立企业年金的积极性空前低落，企业年金发展受到较大负面影响。1995 年《关于深化企业职工养老保险制度改革的通知》指出：“鼓励个人储蓄性养老保险。”但是实际上这些年来并未出台实质性鼓励政策，个人养老性储蓄几乎没有发展。

本文 4.3.1、4.3.2 节对企业年金、个人储蓄养老保险进行了建模研究，旨在探讨与国民经济、人民生活息息相关的诸多影响因子对我国养老保险体系第二、第三层次的影响程度，希望能够从中发掘出一些影响多层次养老保障体系发展的经济学因素。另外，在第 6 章中也探讨了部分与多层次相关的内容。

#### 4.4 基于“多缴多得，长缴多得”的仿真研究

本文对“多缴多得，长缴多得”的理解如下：“多缴”的直观含义是个人年缴费额，“长缴”的直观含义是个人缴费年限，“多得”的直观含义是个人养老金年领取额。进一步分析发现，由于

$$\text{缴费率} = \frac{\text{年总缴费额}}{\text{参保在职人数} \times \text{平均工资}} \quad (25)$$

$$\text{个人年缴费额} = \frac{\text{年总缴费额}}{\text{参保在职人数}} = \text{缴费率} \times \text{平均工资} \quad (26)$$

$$\text{个人缴费年限} = \frac{\text{个人总缴费额}}{\text{个人年缴费额}} = \frac{\text{个人总缴费额}}{\text{年总缴费额}} \times \text{参保在职人数} \quad (27)$$

因此“多缴”实际上可以用调节缴费率或平均工资这两个指标变量来体现，而“长缴”则可用参保在职人数这一指标变量来体现。又由于

$$\text{个人养老金年领取额} = \frac{\text{养老金年支出}}{\text{参保离退休人数}} \quad (28)$$

因此在假设参保离退休人数不变的前提下，“多得”可以用养老金年支出的变化来考量。

根据以上分析，本小节利用上文建立的基于多重线性回归的城镇居民养老金支出模型与基于支持向量机的城镇居民养老金支出模型进行了仿真研究。需要说明的是，由于基于主成分回归的城镇居民养老金收支模型不适合进行原始数据的长期预测，因此在本节中没有选用该模型。

#### 4.4.1 基于多重线性回归模型的仿真研究

由 4.1.6 节可知，养老金支出模型为

$$CZYZ = -17257.155 + 1.140 \times CBZZ + 12412.33 \times TDL \quad (29)$$

亦即养老金支出与参保在职人数（ $CBZZ$ ）和替代率（ $TDL$ ）有关。对于“多缴”而言，该模型与上文分析的缴费率或平均工资并无直接关系，因此还需要作进一步变换。由式（25）可知，

$$\text{缴费率} \times \text{参保在职人数} = \frac{\text{年总缴费额}}{\text{平均工资}} \quad (30)$$

如果缴费率变为原来的  $a$  倍（ $a > 0$ ），则上式变为

$$(a \times \text{缴费率}) \times \text{参保在职人数} = \text{缴费率} \times (a \times \text{参保在职人数}) = a \times \frac{\text{年总缴费额}}{\text{平均工资}} \quad (31)$$

意味着可以将原本描述缴费率变化的系数  $a$  转移到参保在职人数这一变量上来。具体到本模型中，即为式（29）可变形为

$$CZYZ = -17257.155 + 1.140 \times a \times CBZZ + 12412.33 \times TDL \quad (32)$$

式中的  $1.140 \times a$  就是由于缴费率变化而对养老金支出造成的影响。也就是说，在假设  $CBZZ$  和  $TDL$  维持不变的情况下，缴费率的变化会引起养老金支出模型系数的调整，但不会对自变量的选择造成影响。

类似地，对于“长缴”，式（29）也可变形为

$$CZYZ = -17257.155 + 1.140 \times b \times CBZZ + 12412.33 \times TDL \quad (33)$$

其中  $b > 0$  反映缴费年限的变化。

根据以上分析的仿真结果如下图所示（空心圆点为原始数据，实线曲线为原始模型预测曲线，虚线曲线为“多缴”或“少缴”模型预测曲线），其中为简便起见将缴费率的变化倍数和缴费年限的变化倍数取同一个值，即  $a = b = 1.29$ 。从图中可以看出，缴费率提高后养老金年收入得以提高，验证了“多缴多得”准则（假设参保离退休人数不变）。

还需着重说明的一点是，当年缴费率的提高直接影响到的是当年参保在职人员的缴费额度（即“多缴”的直观含义），而当年养老金收入的提高直接影响

到的是当年参保离退休人员的养老金领取额度（即“多得”的直观含义），两者似乎并不对应。然而从下图中可以清楚地看出，“多缴”后的曲线要高于“多缴”前的曲线，且二者之差有不断变大的趋势。因此，当参保在职人员若干年后离退休时，其养老金领取额度也势必会高于“多缴”前所得。也就是说，本模型的仿真结果解释了现下“多缴”与日后“多得”的必然联系，很好地诠释了“多缴多得”对参保人员的长远利益。

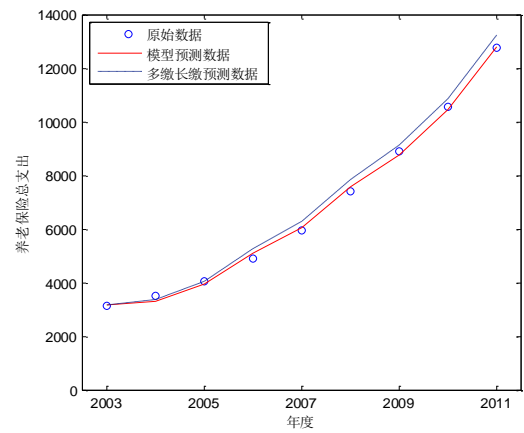


图20 基于多重线性回归模型的仿真研究

#### 4.4.2 基于支持向量机模型的仿真研究

由4.1.7节可知，基于支持向量机的城镇居民养老金支出模型中包含缴费率这一自变量，因此本小节通过直接改变缴费率，观察养老金支出变化情况。

仿真结果如下图所示（红色曲线为提高缴费率前预测曲线，蓝色曲线为提高缴费率后预测曲线，缴费率提高29%）。从图中可以看出，在支持向量机模型下，提高缴费率对养老金支出有所促进，但影响并不明显。分析其原因，在于支持向量机模型具有较高的鲁棒性与自学习能力，尤其在多变量模型中，单独地改变其中一个变量的值难以观测输出量的变化。

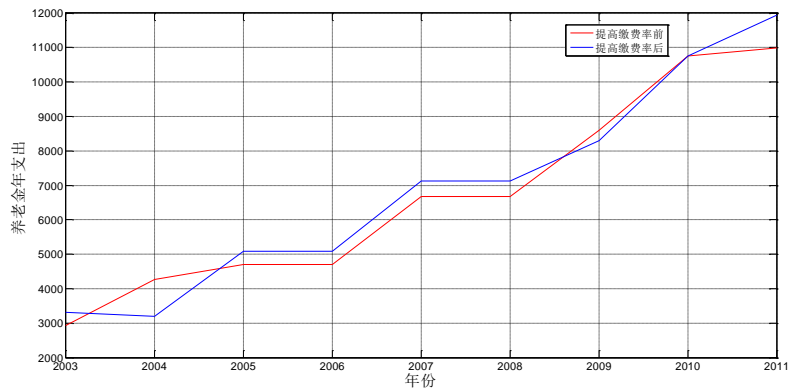


图21 基于多重线性回归模型的仿真研究

## 5 养老金缺口估计与模型调整

问题 2 要求估计从今年至 2035 年我国养老金缺口,预测收支矛盾最尖锐情况发生时间及严重程度,并考虑收入倍增计划对模型做出相应调整。

针对问题要求,本文首先通过查阅文献资料,对养老金缺口定义进行分析与诠释;然后根据本题涉及到的长期预测背景,对已建的多重线性回归模型与支持向量机模型进行改进,改进后模型拟合度与显著性均较好,并利用改进后模型对养老金缺口进行估计;最后考虑收入倍增计划,对数学模型需要调整的部分进行了阐释说明。

### 5.1 养老金缺口估计

本文理解的养老金缺口定义为:养老金缺口=养老金支出-养老金收入-财政补贴。出于数据完整性与模型简便性考虑,本章的养老金缺口估计是在城镇养老金收入与支出模型基础上进行的,由于涉及到长期预测,因此对原有模型做了适当改进。与 4.4 节情况类似,选取了多重线性回归模型与支持向量机模型进行缺口估计。

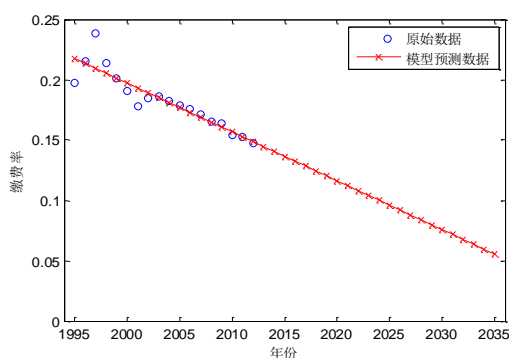
#### 5.1.1 基于多重线性回归模型的缺口估计

由 4.1.6 节可知,养老金收入与支出模型为

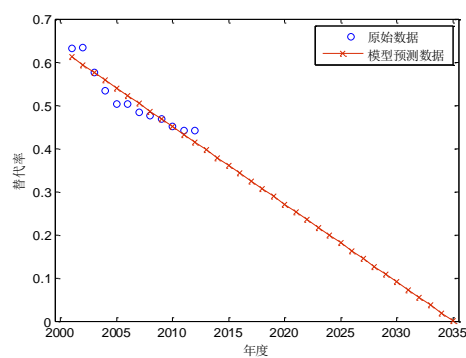
$$\begin{cases} CZYS = -42909.367 + 1.823 \times CBZZ + 135299.066 \times JFL \\ CZYZ = -17257.155 + 1.140 \times CBZZ + 12412.33 \times TDL \end{cases} \quad (34)$$

因此,为进行缺口估计,需要基于时间序列分别建立参保在职人数 ( $CBZZ$ )、缴费率 ( $JFL$ )、替代率 ( $TDL$ )、财政补贴 ( $CZBT$ ) 的多重线性回归模型,预测 2013-2035 年各变量值,进而估计当年养老金缺口。

原始数据可从表 1 中相应选出,具体实现方法与 4.1.6 节基本相同,模型结果如下图所示。



(a) 缴费率



(b) 替代率

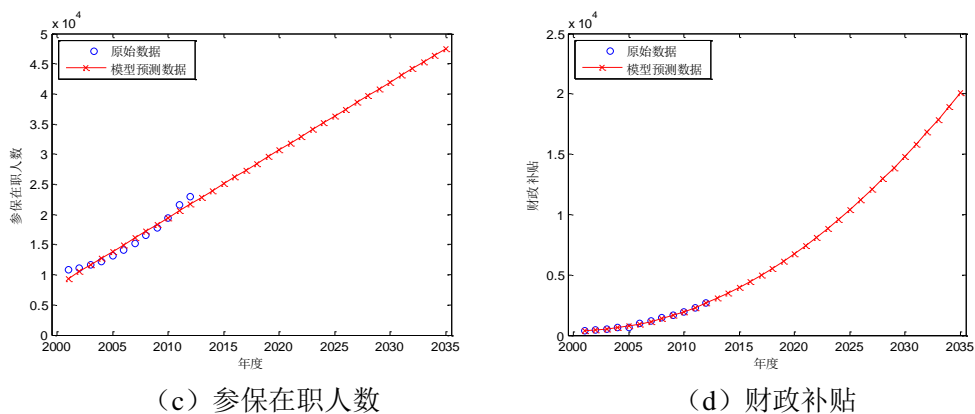


图22 基于多重线性回归模型的自变量预测

分析发现，替代率的变化很不规则，一次线性回归拟合效果不佳。因此又采用指数平滑模型对不规则的时间序列数据加以平滑，从而得到其变化规律与趋势。模型拟合效果如下表所示。

表 34 替代率指数平滑模型拟合效果

拟合参数	平均值
平稳R方	.874
R方	.864

从中可知平稳的R方值为0.874，R方值为0.864，该指数平滑模型的拟合情况较好。

下图为替代率指数平滑模型的拟合图与观测值，替代率序列呈下降趋势，但在近两年几乎稳定在0.44-0.45，据此可对替代率的多重线性回归模型进行调整。

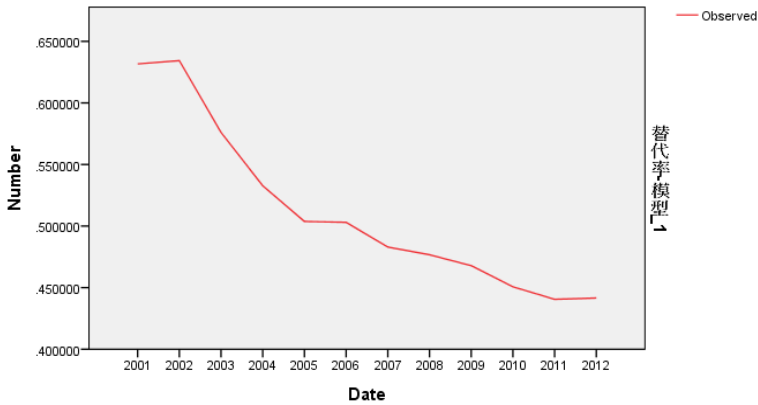


图23 替代率指数平滑模型拟合图与观测值

利用上述 4 个自变量的多重线性回归模型对 2013-2035 年数据进行预测，再根据养老金缺口理解，即可得到 2013-2035 年养老金缺口估计，结果如下表所示。

表 35 2013-2035 年养老金缺口估计

年份	收入-支出	财政补贴	实际结余	年份	收入 - 支	财政补贴	实际结余
----	-------	------	------	----	--------	------	------



	(亿元)	(亿元)	(亿元)		出(亿元)	(亿元)	(亿元)
2013	4050.21	3055.09	995.13	2025	6672.16	10387.76	-3715.61
2014	4268.71	3487.63	781.08	2026	6890.65	11209.79	-4319.14
2015	4487.20	3952.63	534.58	2027	7109.15	12064.28	-4955.13
2016	4705.70	4450.08	255.62	2028	7327.64	12951.22	-5623.58
2017	4924.19	4980.00	-55.80	2029	7546.14	13870.62	-6324.49
2018	5142.69	5542.37	-399.68	2030	7764.63	14822.48	-7057.85
2019	5361.19	6137.20	-776.01	2031	7983.13	15806.80	-7823.67
2020	5579.68	6764.48	-1184.8	2032	8201.62	16823.57	-8621.95
2021	5798.18	7424.22	-1626.05	2033	8420.12	17872.80	-9452.68
2022	6016.67	8116.42	-2099.75	2034	8638.61	18954.49	-10315.9
2023	6235.17	8841.08	-2605.91	2035	8857.11	20068.63	-11211.5
2024	6453.66	9598.19	-3144.53				

从上表可以看出，2017 年将首次出现养老金缺口（即实际结余为负），且之后缺口越来越大，因此将其作为养老金收支矛盾最尖锐的时间；2034 年以后，养老金缺口将达到 1 万亿元。

另外，如果从中长期积累平衡角度考虑，即在表 35 数据中计入 2012 年的养老金积累基金（23941 亿元），则可估计出未来直至 2035 年间的养老金积累基金数目。而这一数目又与投资收益率（*TSL*）息息相关，因此此处对 *TSL* 建立了基于时间序列的指数平滑模型。模型拟合效果如下表所示。

表 36 投资收益率指数平滑模型拟合效果

拟合参数	平均值
平稳R方	.462
R方	.023

从表中可知平稳 R 方值为 0.462，该模型拟合情况一般。

下图为投资收益率指数平滑模型的拟合图与观测值，投资收益率序列呈波动状态。通过指数平滑模型的拟合图可以发现，收益率均值约为 4%。

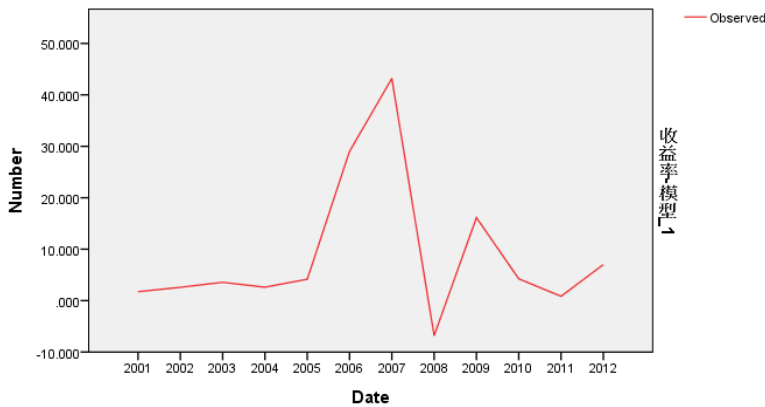


图24 投资收益率指数平滑模型拟合图与观测值

表 37 2013-2035 年养老金积累基金估计

年份	实际结余（亿元）	积累基金（亿元）	年份	实际结余（亿元）	积累基金（亿元）
2013	995.13	25893.77	2025	-3715.61	26719.94
2014	781.08	27710.60	2026	-4319.14	23469.60
2015	534.58	29353.60	2027	-4955.13	19453.25
2016	255.62	30783.36	2028	-5623.58	14607.80
2017	-55.80	31958.89	2029	-6324.49	8867.62
2018	-399.68	32837.57	2030	-7057.85	2164.47
2019	-776.01	33375.06	2031	-7823.67	-5572.62
2020	-1184.80	33525.26	2032	-8621.95	-14194.57
2021	-1626.05	33240.22	2033	-9452.68	-23647.25
2022	-2099.75	32470.08	2034	-10315.9	-33963.13
2023	-2605.91	31162.97	2035	-11211.5	-45174.65
2024	-3144.53	29264.96			

从上表中可知 2031 年养老金积累基金将由正转负，且赤字不断增大，2035 年该负值将超过 4.5 万亿元。

此外，上述分析均未考虑通货膨胀等不利因素，如果考虑这些不利因素，矛盾最尖锐时间可能会更早出现。

#### 5.1.2 基于支持向量机模型的缺口估计

本小节数据来源仍旧是表 1，但与 4.1.7 中模型不同的是采用了循环迭代式 SVM 模型进行参数计算。这样处理主要出于以下两方面考虑：（1）长期预测会导致误差累积，需要在预测过程中不断调整模型，消除误差影响；（2）学习数据量不多，进一步增大了长期预测难度。因此，本小节选用 8 年为一个基本窗口，进行基于前向核函数的  $\varepsilon$ -SVR 自学习，逐年对养老金缺口进行估计。

另外，在学习数据筛选方面也与 4.1.7 中模型有所不同。经过相关分析（具体方法与前文相同）与实际情况分析，最终选择的学习数据如下表所示。需要说明的是，为了提高模型稳定性，本小节将缴费率与平均工资加以结合，作为一个新的指标变量缴费额（JFE），并作为养老金支出的学习数据之一。

表 38 养老金缺口估计的 SVM 学习数据

因变量	自变量
收入	缴费额，参保在职人数，财政补贴，消费价格指数（以2000年为基期）
支出	替代率，参保在职人数，参保退休人数，消费价格指数（以2000年为基期）

最终估计结果如图 25、表 39 所示，其中蓝色曲线为养老金年收入，红色曲线为养老金年支出，绿色曲线为两者之差。

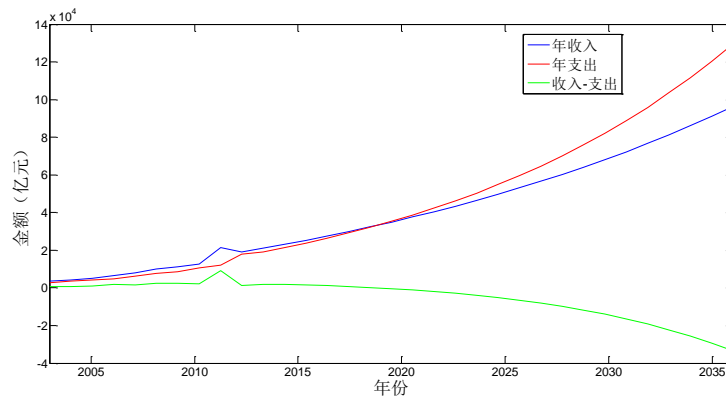


图 25 基于 SVM 模型的养老金缺口估计示意图

表 39 基于 SVM 模型的养老金缺口估计

年份	收入-支出 (亿元)	年份	收入-支出 (亿元)
2013	1896.11	2025	-6503.48
2014	1738.90	2026	-8072.27
2015	1476.45	2027	-9836.36
2016	1142.18	2028	-11813.27
2017	729.54	2029	-14021.94
2018	230.56	2030	-16482.80
2019	-363.38	2031	-19217.90
2020	-1061.66	2032	-22251.04
2021	-1874.40	2033	-25607.89
2022	-2812.59	2034	-29316.12
2023	-3888.09	2035	-33405.59
2024	-5113.75		

表 39 是基于 SVM 模型的养老金缺口估计详细结果，可以看出，2019 年将首次出现养老金缺口（即入不敷出），且之后缺口越来越大，因此将其作为养老金收支矛盾最尖锐的时间；2031 年以后，养老金缺口将突破 2 万亿元大关，是我国经济的巨大负担。

另外，如果从中长期积累平衡角度考虑，即在表 39 数据中计入 2012 年的养老金积累基金（23941 亿元），则可估计出未来直至 2035 年间的养老金积累基金数目。易得 2028 年养老金积累基金将由正转负，2035 年该负值将超过 15 万亿元。

表 40 是对基于两种模型养老金缺口估计的比较，从表中及上文分析可以看出，由于两种模型采用不同的建模思想与数学方法，使得各自的缺口估计不尽相同；而预测结果的差异性，也从另一个角度体现出影响养老金缺口因素的复杂性。

表 40 基于两种模型的养老金缺口估计比较

年份	实际结余-SVM (亿元)	实际结余-多重线 性回归 (亿元)	年份	实际结余-SVM (亿元)	实际结余-多重线 性回归 (亿元)
2013	1896.11	995.13	2025	-6503.48	-3715.61

2014	1738.90	781.08	2026	-8072.27	-4319.14
2015	1476.45	534.58	2027	-9836.36	-4955.13
2016	1142.18	255.62	2028	-11813.27	-5623.58
2017	729.54	-55.80	2029	-14021.94	-6324.49
2018	230.56	-399.68	2030	-16482.80	-7057.85
2019	-363.38	-776.01	2031	-19217.90	-7823.67
2020	-1061.66	-1184.80	2032	-22251.04	-8621.95
2021	-1874.40	-1626.05	2033	-25607.89	-9452.68
2022	-2812.59	-2099.75	2034	-29316.12	-10315.90
2023	-3888.09	-2605.91	2035	-33405.59	-11211.50
2024	-5113.75	-3144.53			

## 5.2 基于收入倍增计划的模型调整

为确保到 2020 年实现全面建成小康社会的目标，十八大报告提出：“实现国内生产总值和城乡居民人均收入比 2010 年翻一番。”<sup>[22]</sup>考虑到党的收入倍增计划，本文建立的数学模型有以下部分需要调整：

(1) 对于主成分回归模型与多重线性回归模型，由于这两个模型的指标变量是根据相关分析与灰色关联分析进行筛选的，因此若考虑收入倍增情况，那么至少工资水平 (*GZSP*) 这一指标变量将发生变化，而且通货膨胀率 (*TZL*) 也会持续按照目前官方数据<sup>[23]</sup>以每年约 3% 速率增加。根据以上分析，各指标变量之间的相关性可能将发生变化。因此模型至少应该以替代率 (*TDL*)、缴费率 (*JFL*)、参保在职人数 (*CBZZ*)、参保离退休人数 (*CBLT*)、财政补贴 (*CZBT*)、工资水平 (*GZSP*)、消费价格指数 (以 2000 年为基期) (*XFJZ2*) 等 7 个指标 (即经灰色关联与自变量相关分析筛选前的待选指标) 作为自变量重新分析建模。

(2) 对于支持向量机模型，由于收入倍增计划最直接影响到的是工资水平这一指标变量，因此可以只改变学习因子中工资水平一项，再进行 SVM 建模。这种调整方式简单明了，但没有考虑工资水平对其他学习因子的影响程度，因此最终结果可能误差较大。更为精确的调整方式则是结合精算模型，计算出收入上升对 SVM 学习数据中因子的影响情况 (主要包括替代率 (*TDL*)、缴费率 (*JFL*)、财政补贴 (*CZBT*)、工资水平 (*GZSP*)、消费价格指数 (以 2000 年为基期) (*XFJZ2*) 等)，再对全面修改后的学习数据重新进行 SVM 建模。

## 6 养老保险体系的可持续性研究

第 3 题要求分析各国养老保险的不同模式，利用仿真手段研究我国养老保险体系的可持续性，并针对潜在矛盾提出相应建议；第 4 题则要求建立第 3 题增加可调节变量的数学模型。

本文出于模型陈述方便与问题连续性考虑，将这两道题目在同一章内加以解决。首先通过查阅大量文献资料，总结归纳了当今世界各国主要养老模式，简要分析其优缺点及值得我国借鉴之处；之后对我国养老保险体系可持续性进行了分析研究，以替代率和缴费率作为可调节变量建立多重线性回归模型与支持向量机-精算解析模型，并利用这两个模型对替代率与缴费率合理区间进行优

化选取，最后针对预测情况提出了一些政策措施建议。

### 6.1 各国养老保险模式分析

自 1889 年德国建立了世界上第一个由国家正式建立的强制性公共养老金计划以来，公共养老金制度在世界上已存在了 100 多年，并由个别国家的尝试逐步变成世界各国普遍实行的一项基本社会经济制度<sup>[24]</sup>。

纵观当今各国养老保险模式，一般可以将其划分为以下几种基本类型<sup>[25]</sup>：

- (1) 普遍保障模式。
- (2) 普遍保障与收入关联模式。
- (3) 收入关联模式。
- (4) 国家管理的公积金模式。
- (5) 国家立法规范，实施商业经营性管理的基金制模式。

代表国家及简要介绍如下表所示<sup>[25]</sup>。

表 41 各国养老保险模式一览

模式	国家	制度体系	财务方式	替代率	开始领取年龄	提早领取年龄	职工缴费比率 (%)	雇主缴费比率 (%)
普遍保障模式	荷兰	基础养老金	现收现付方式	大约 40%~60%	65 岁	无	23.4 (1997 年)	0 (1997 年)
普遍保障与收入关联模式	英国	基础养老金+收入比例养老金	现收现付方式	夫妇大约 60%，单身大约 50%	男 65 岁，女 60 岁	无	2~10 (1996 年)	3~10 (1996 年)
	日本		修正积累方式	夫妇大约 69%，单身大约 50%	65 岁	60 岁	8.675 (1997 年)	8.675 (1997 年)
	加拿大		现收现付方式	夫妇大约 53%，单身大约 40%	65 岁	60 岁	3.2 (1998 年)	3.2 (1998 年)
收入关联模式	德国	收入比例养老金	现收现付方式	大约 70%	男 63 岁，女性与失业人员 60 岁，2018 年 65 岁	62 岁 (2006 年起)	9.3 (1995 年)	9.3 (1995 年)
	美国			单身大约 30%~50%	65 岁，至 2027 年提高到 67 岁	62 岁	6.2 (1998 年)	6.2 (1998 年)
	瑞典		现收现付+积累制方式	大约 70%	65 岁	60 岁	6.95 (1999 年)	11.55 年 (1999 年)
	韩国		积累制方式	40%	65 岁	60 岁	3 (1998 年)	3 (1998 年)

国家管理的公积金模式	新加坡			-	62 岁	一定条件下可以一次性提取	20（1995 年）	10（1995 年）
国家立法规范，实施商业经营性管理的基金制模式	智利	强制储蓄制度	积累制方式	-	男 65 岁，女 60 岁	满足最低额条件可以提早	大约 13（1995 年）	0（1995 年）

这 5 种类型养老保险模式各有利弊，以下作简要分析。

以荷兰为代表的普遍保障模式具有三大优点——收入分配的公平、广泛的覆盖面和完全积累制的补充养老金，在这种模式下，社会保障的目的是保证一种公平的收入替代率，而不仅仅是政府给予的最低福利<sup>[27]</sup>。但缺点则是对国家经济承受力和现行人口膨胀对经济发展的制约要求很高<sup>[25]</sup>。

以英国为代表的普遍保障与收入关联模式优点在于养老保障“三支柱”结构与功能较为平衡，协议退出制度使得养老方式灵活。缺点则为养老保险开支水平过高，面临支付危机，中小企业雇主为员工缴纳的养老保险不足以应付退休后生活等。

收入关联模式旨在通过社会保险机制为在职劳动者建立退休收入保险计划，较为强调劳动者的自我保障意识。但该模式是与工业化、城市化相伴，与劳动就业相关联的养老保障模式，对目前农村人口仍占很大比重的我国而言并不适用<sup>[25]</sup>。

以新加坡为代表的国家管理公积金模式，具有广泛的覆盖范围、高积累率、高住房拥有率等诸多优点，但在应对突发事件、体现社会公平、满足国内需求等方面也存在不足<sup>[28]</sup>。而且根据我国社会保险制度发展现状，如果将现已存在的庞大养老保险制度全面转换到公积金制模式会大幅度增加企业人工成本和国家财政负担，因此该模式可能只具有区域性的意义<sup>[25]</sup>。

以智利为代表的国家立法规范、实施商业经营性管理的基金制模式优势在于采用完全积累方式，减轻了政府责任，同时较好地解决了劳动者代际之间的矛盾，而且投资回报较高。缺点则在于养老金收入差距较大，普及范围有限，而且商业经营性管理更易遭受经济危机打击<sup>[29]</sup>。

综上所述，目前世界各国养老保险模式都存在着需长期维持资金收入与支出的平衡以确保计划的可持续性问题<sup>[31]</sup>。在构造我国养老保险目标模式时，不应简单地照搬他国经验或否定现行的各种养老保险制度，而应根据我国国情，借鉴别国有益经验，分析各种模式的基本内核，选择现实可行的养老保险模式<sup>[25]</sup>。

## 6.2 我国养老保险体系可持续性研究

根据第 5 章及 6.1 节分析，本文选取养老金缺口作为评估我国养老保险体系可持续性的一个观测量。此外，为了更全面地评估养老保险体系的可持续性，与 5.1.1 中类似，本文又选取积累基金作为另一个观测量。

### 6.2.1 替代率与缴费率合理区间选取

1) 基于多重线性回归模型的区间选取

根据 4.1.6 节建立的城镇居民养老金收入、支出多重线性回归模型可知，替代率的调整会影响到养老金支出，替代率过高会加重财政负担，但过低也会影响养老质量。

本文首先设定替代率调整区间为 $[0.4, 0.65]$ ，其中下限根据模型预测未来替代率稳定在 0.45 左右设定，上限则是参照各国替代率水平设定（参见表 41 各国养老保险模式一览）。缴费率调整区间为 $[0.15, 0.25]$ ，其中下限根据预测缴费率在 0.15 左右且逐年下降设定，上限根据 1997 年缴费率为 24% 设定。

具体调整过程如下表所示。

表 42 替代率与缴费率调整过程

	养老金年收入	养老金年支出	当年财政补贴
1	原始缴费率	替代率=0.4	财政补贴
2	缴费率=0.15	原始替代率	财政补贴
3	缴费率=0.20	替代率=0.5	财政补贴
4	缴费率=0.25	替代率=0.65	财政补贴

调整结果如下图所示。

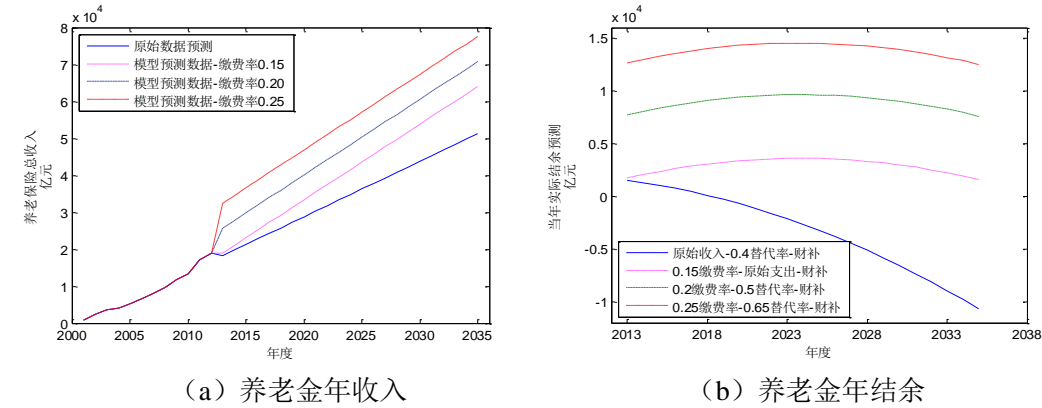


图 26 替代率与缴费率调整效果图

由上图可得以下结论：

- (1) 缴费率不调整（据预测值未来会自动下降），替代率降低。养老金年支出降低，可以延缓养老金缺口最尖锐矛盾到来时间（2 年），但长远来看仍然不足以弥补亏空；
- (2) 维持当前缴费率不变，替代率不调整。可以较好地保持当年结余的可持续性，至少至 2035 年仍可保证当年结余为正；
- (3) 缴费率小幅度提高，替代率小幅度提高。养老金年支出增加，养老质量提高，当年结余可保证在 7500 亿元以上；
- (4) 缴费率大幅度提高，替代率大幅度提高。养老金年支出增加，养老质量大幅度提高，当年结余可保证在 10000 亿元以上。

类似地，根据上文指数平滑模型预测，并计入每年约 4% 的收益率的积累基金预测情况如下图所示。

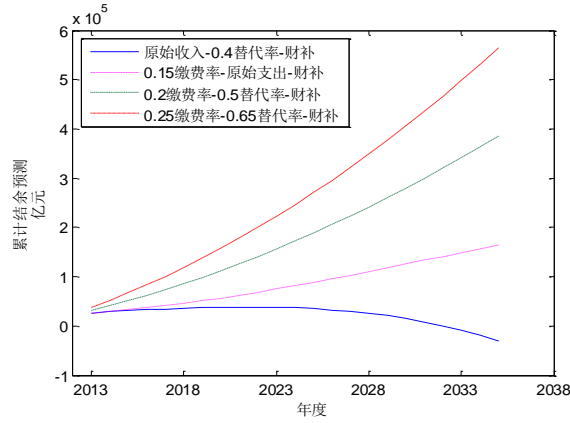


图 27 积累基金调整效果图

与当年结余分析类似，积累基金调整效果再次验证了上述 4 个结论的正确性。

综上所述，当替代率调整区间取 $[0.45, 0.65]$ ，缴费率调整区间取  $[0.15, 0.25]$  时，养老保险体系可以保证可持续发展。

#### 2) 基于支持向量机与精算解析模型的区间选取

为了从更多角度考虑替代率与缴费率调整的合理性，本文又根据支持向量机与精算解析模型对两者的区间选取进行了分析研究。但是由于这一部分内容涉及到下一节可调节变量模型的建立，因此具体过程详见 6.2.2 2) 中所述。

### 6.2.2 可调节变量的数学模型建立

在上一节分析基础上，本小节选取替代率与缴费率作为可调节变量建立数学模型。

#### 1) 基于多重线性回归的可调节变量数学模型建立

由上一节结论可知，替代率合理调整区间为 $[0.45, 0.65]$ ，缴费率合理调整区间为 $[0.15, 0.25]$ 。根据养老金当年结余及积累基金对这两个变量进一步优化，具体过程为：

$$JFL_{n+1} = \begin{cases} (1+0.5\%)JFL_n, & JFL_n < 18\% \\ 18\%, & JFL_n \geq 18\% \end{cases} \quad (35)$$

$$TDL_{n+1} = \begin{cases} (1+1\%)TDL_n, & TDL_n < 65\% \\ 65\%, & TDL_n \geq 65\% \end{cases} \quad (36)$$

优化效果如下图所示，从图中易知可以保证养老保险体系的可持续性，且人民负担增加不大，而养老福利却大幅提升。



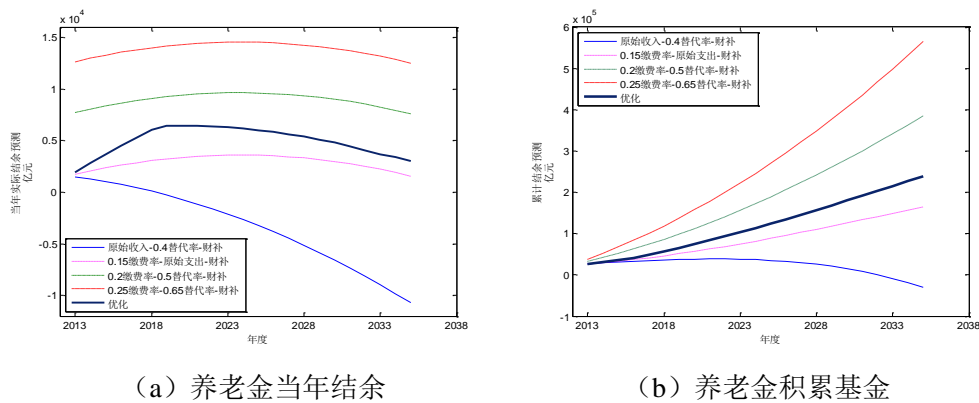


图 28 优化效果图

## 2) 基于支持向量机与精算解析方法的可调节变量数学模型建立

由 5.2 中分析可知, SVM 为数据特征提取与学习型模型, 如果分别建立以替代率和缴费率为可调节变量的 SVM 数学模型, 则由于自变量为学习因子中的一个, 而 SVM 具有很强的鲁棒性, 故自变量变化对模型输出的影响并不显著, 难以进行观测。若采用先进行部分预测值, 然后改变自变量预测值的方法调整模型, 则由于原数据的统计特征与改变后的数据存在较大差异, 输出的可靠性无法保证。要想得到较为精确的模型, 必须先以一定精度计算出自变量变化对 SVM 模型其他学习因子的影响, 然后将考虑上述影响后的数据进行回归学习, 才能直接利用 SVM 算法进行后续预测, 这将大大增加计算量与模型复杂度。

而利用养老金参数定义与经济理论的精算解析法模型, 则可设定模型的任意参数为自变量, 计算其他各参数的变化情况, 更适合本题要求。但精算法需要大量前期数据, 并将其作为基础进行参数预测。每个模型参数的预测需要额外的理论解析, 导致需考虑的变量更加复杂, 处理的数据量更加庞大。

SVM 算法作为一种参数自学习算法, 规避了复杂的参数自变量理论分析过程, 故 SVM 模型强大的拟合与预测功能可以辅助用于题目求解。

综上所述, 本题可以结合 SVM 算法与精算解析方法进行建模。

### 精算解析法理论计算

$$\text{当年年收入} = \text{当年参保在职总人数} \times \text{平均工资} \times \text{缴费率} + \text{当年补贴} \quad (37)$$

$$\text{当年年支出} = \text{参保退休人数} \times \text{等效平均工资} \times \text{替代率} \quad (38)$$

(附: 等效平均工资的解释: 由于规定个人退休养老金水平等于替代率乘以退休前一年的工资水平, 上式中等效平均工资为退休人口整体归一化处理, 从人口普查数据中找出 60 岁以上的人口按年龄的分布, 将其平均化, 得出 68.43 岁。故等效平均工资定义为 8 年前的工资水平。)

按以上的定义, 计算 2003-2011 年的收入与支出水平, 并与实际值进行比较验证, 结果如下表所示。

表 43 2003 年-2011 年收入与支出水平

年份	预测收入 (亿元)	实际收入 (亿元)	预测支出 (亿元)	实际支出 (亿元)
2003	3636.687	3680	1380.931	3122

2004	4147.292	4258	1414.178	3502
2005	4987.842	5093	1645.477	4040
2006	6312.394	6310	1945.81	4897
2007	7592.223	7834	2241.996	5965
2008	9679.213	9740	2748.14	7390
2009	10939.36	11941	3375.094	8894
2010	12765.06	13420	3989.34	10555
2011	16004.16	16895	4818.123	12765
2012	18447.54	20001	6038.017	15562

从上表中可以看出收入的解析预测结果比较好，而支出的预测结果明显小于实际支出。原因主要在于（1）这 9 年的工资水平处于大幅度增长状态，因此用 8 年前的工资水平计算支出太过于保守；（2）这 9 年养老金覆盖率上升明显，且正处于退休高潮期，每年增加的参保退休人数远大于前期（如 90 年代末期到 2002 年），使用线性度较高的模型计算此段时间的模型，准确度较低；（3）此阶段距离中国养老金改革不足 10 年，参保退休人口年龄结构尚不平稳，为模型长期预测增加了难度。

因此考虑采用逐年迭代型模型进行这一特殊时段支出计算，其优点在于迭代计算时使用了近期的数据，可以描述非线性变化情况。

计算公式为

$$\begin{aligned} \text{当年支出} = & \text{当年新增参保人数} \times \text{前一年平均工资} \times \text{当年替代率} \\ & + \text{前一年支出} + \text{前一年新增支出} \end{aligned} \quad (39)$$

结果如下表所示。

表 44 2003 年-2011 年支出水平

年份	预测支出 (亿元)	实际支出 (亿元)
2003	2564.682346	3122
2004	2902.806731	3502
2005	3515.80453	4040
2006	4257.548757	4897
2007	5593.289392	5965
2008	6750.886457	7390
2009	8774.947817	8894
2010	10667.38184	10555
2011	13543.95847	12765
2012	16598.62731	15562

从上表中可以看出采用迭代模型精度提升较大，误差不超过 9%。

需要指出的是，迭代模型的增长趋势呈指数型，这与我国长期人口变化趋势的预测不符，因此迭代模型只用于预测 2018 年以前的支出情况，2018 年以后人口增速放缓、经济增速也将降低，参保退休人数年龄结构也足够稳定，因此可以用前述解析模型预测支出。

*SVM 学习与预测*

根据现有数据，建立 SVM 模型，学习数据根据相关性分析与实际情况分析，应基于 2003-2011 年数据，以及上一章中养老金收入支出的预测结果，逆向学习 2013-2035 前述模型所需的变量数值。

将模型所需变量的 30 年预测数据利用 SVM 预测完毕后，进行了年收入，年支出以及缺口的预测，结果如下图所示。

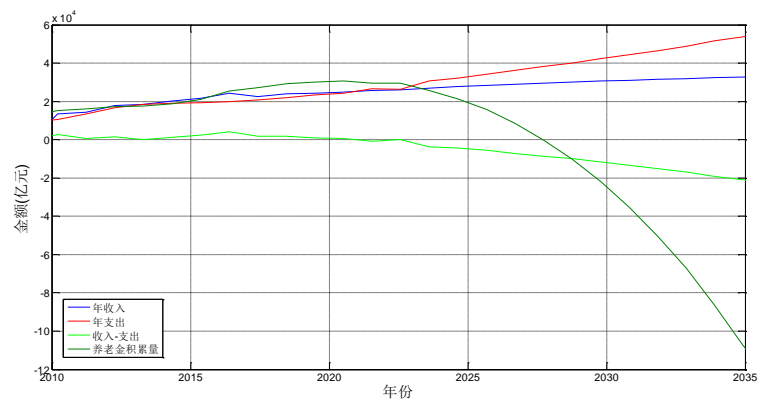


图 29 SVM 预测结果

预测得到的收入减支出如下表所示。

表 45 预测得到的收支差

年份	收支差 (亿元)	养老金积累基金 (亿元)	年份	收支差 (亿元)	养老金积累基金 (亿元)
2013	92.8104354105672	17476.8945706826	2025	-5637.08980388496	15820.1469786569
2014	1219.36093713711	18696.2555078197	2026	-7223.55951686967	8596.58746178726
2015	2412.63402425642	21108.8895320761	2027	-8724.57151990731	-127.984058120051
2016	4237.68185169082	25346.5713837669	2028	-9909.73025138997	-10037.7143095100
2017	1900.47728924131	27247.0486730082	2029	-11733.5858363461	-21771.3001458561
2018	1925.75803520364	29172.8067082119	2030	-13481.4156629679	-35252.7158088240
2019	881.825646424764	30054.6323546366	2031	-15062.4785456729	-50315.1943544969
2020	536.070748632053	30590.7031032687	2032	-16879.6315631980	-67194.8259176949
2021	-948.563157774923	29642.1399454938	2033	-19279.2745703778	-86474.1004880727
2022	-96.1186299505207	29546.0213155433	2034	-21018.6796008907	-107492.780088963
2023	-3842.61460035848	25703.4067151848	2035	-22984.3332238672	-130477.113312831
2024	-4246.16993264288	21457.2367825419			

从上表中可以看出收支差将于 2016 年开始从最大值下降，于 2021 年首次入不敷出，在 2030 年以后将突破 2 万亿元大关，对我国经济是一个巨大的负担。

从中长期积累平衡模型角度出发，即利用上表的数据中，加入 2012 年的养老金积累量 23941 亿元，可算出 2027 年养老金积累量将变为负值，在 2035 年将形成超过 13 万亿元的缺口。

替代率与缴费率合理区间选取分析

a. 保持其他变量不变，只改变 2012 年以后的缴费率（分别为原始预测值，20%，25%，30%），收支缺口图如下。

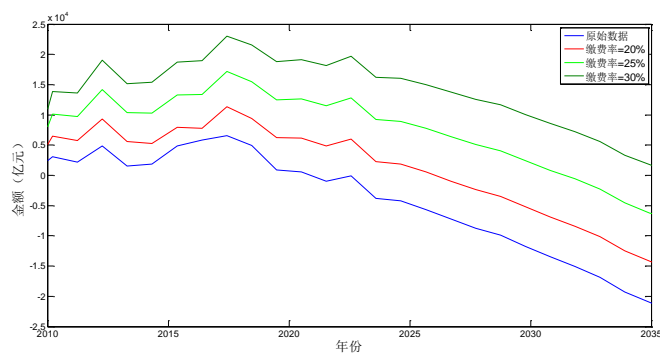


图 30 收支缺口图

积累基金缺口图如下。

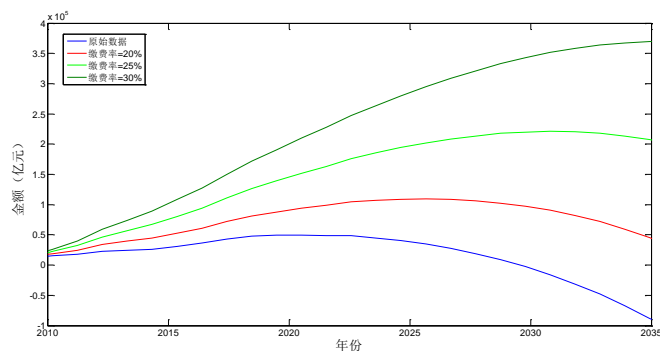


图 31 积累基金缺口图

从中可知：（1）缴费率提高能增大收支差，且基本呈线性关系；（2）单纯提高缴费率并不能扭转收支逆差的趋势，只能暂时减缓缺口增长的速度；（3）只有将缴费率增大到一个极大的值（如大于 30%）才能在未来几十年内积累基金不出现缺口，但可能会大大降低人民生活水平。

b. 保持其他变量不变，只改变 2012 年以后的替代率（分别为原始预测值，40%，35%，30%），收支缺口图如下。

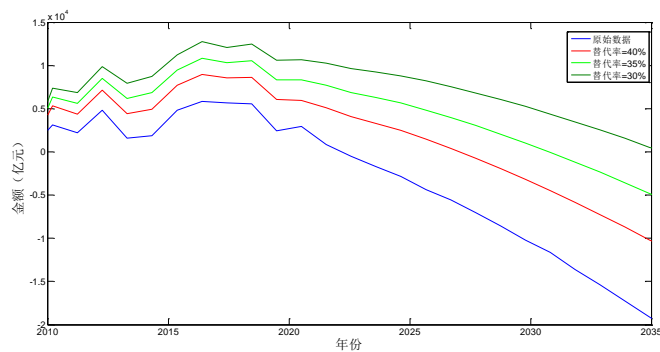


图 32 收支缺口图

积累基金缺口图如下。

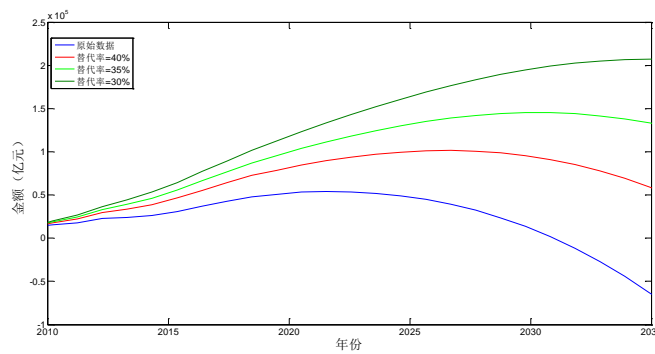


图 33 积累基金缺口图

从中可知：（1）替代率降低能增大收支差，且基本呈线性关系；（2）单纯降低替代率并不能扭转收支逆差的趋势，只能暂时减缓缺口增长的速度；（3）只有将替代率减小到一个极大的值（如小于 30%）才能在未来几十年内基金积累不出现缺口，但可能会大大降低人民生活水平。

### 6.2.3 政策措施建议

根据第 5 章分析，养老金缺口=养老金年支出-养老金年收入-当年财政补贴，因此若想确保可持续的养老金体系，可以从三方面着手：（1）减少养老金年支出；（2）增加养老金年收入；（3）增加当年财政补贴。但三方面中，增加财政补贴会增加国家财政负担，对国家经济承受能力要求较高，所以前两方面是本文考虑的重点。

根据 4.1.6 节建立的城镇居民养老金收入、支出多重线性回归模型可知，减少养老金年支出即减少参保在职人数（ $CBZZ$ ）与替代率（ $TDL$ ），增加养老金年收入即增加参保在职人数（ $CBZZ$ ）与缴费率（ $JFL$ ）。参保在职人数（ $CBZZ$ ）似乎处在自相矛盾的境地。事实上，由式（12）可知，养老金收入模型中参保在职人数（ $CBZZ$ ）的系数大于养老金支出模型中参保在职人数（ $CBZZ$ ）的系数，因此增加参保在职人数（ $CBZZ$ ）虽然会导致养老金年支出增加，但养老金缺口仍将向减小的方向发展。具体仿真预测结果如下图所示。

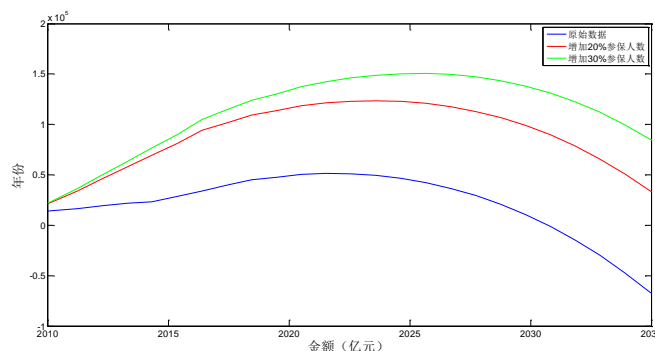


图 34 积累基金图预测

从图中可以看出，参保在职人数的增加有效改善了资金积累的缺口情况，到 2035 年依然存有结余。

另外，通过 6.2.1 分析可知，合理调整替代率与缴费率有助于养老保险体系的可持续性发展。

不过，把希望完全寄托于单纯改变第一层次养老金的替代率与缴费率仍旧很难改变缺口逆差，不能只依靠社会统筹养老金，必须大力发展企业年金与个人储蓄养老保险，配合利好政策的实施，才能切实解决问题。4.3 节的分析也表明作为我国养老保险体系第二、第三层次的企业年金、个人储蓄养老保险在养老保障制度中理应起到重要作用。

综上所述，本文建议采取的政策措施为：

- (1) 提高养老保险体系覆盖率（增加参保在职人数）；
- (2) 制定合理的替代率与缴费率（这一建议政策的效果仿真参见上两节）；
- (3) 出台税收优惠政策，鼓励企业年金发展；
- (4) 出台税收优惠政策，鼓励个人储蓄养老保险发展。

## 7 结论与建议

### 7.1 模型评价

本文针对我国养老金的相关问题进行深入分析，在查阅大量文献和相关数据的基础上，结合指标变量分析统计方法，实现了基于主成分回归、多重线性回归的养老金建模算法，以及研究建立了基于支持向量机的养老金模型。建立的模型涵盖了城镇养老保险、城乡养老保险、新农保等多个实际应用对象，分层次分析仿真社会统筹、企业年金、商业养老保险三大养老支柱，并利用多重线性回归与结合支持向量机的解析算法两种方法预测仿真了养老金缺口问题。最后根据仿真结果与分析，对我国养老金体系今后发展提出了建议。所完成的工作主要具有以下优点：

(1) 针对养老金收入和养老金支出两大模型，综合采用相关分析和灰色关联分析确定建模的自变量，最大限度地去除了冗余信息，有效的降低模型的复杂度。

(2) 采用主成分回归和多重线性回归建立的养老金收入与支出的两个模型的拟合度都非常高，表明本文所建立的模型具有很高精度。此外，模型结构简单，方便相互验证，并通过“多缴多得，长缴多得”的仿真研究，证明模型具有较好实际意义。

(3) 利用支持向量机理论，结合可靠的学习数据以及支持向量机关键参数的选取手段，本文建立具有较高精度且处理手段相对简单的养老金问题的自学习模型，模型可靠度与可延展性较高。

(4) 针对缺口预测问题，设计了基于多重线性回归与结合支持向量机的解析算法。前者节省了运算时间，提高了分析精度，经济意义与实际情况相符，有效地去除了多个次要相关元素，简化了问题难度。后者利用智能算法结合解析方法，对建立的学习集合进行可靠自学习与自验证过程，规避了复杂的理论分析过程的同时也减少了潜在的额外影响因子，其预测结果也通过了验证，具有很好的借鉴意义。

(5) 结合本文所建模型，从模型的数学意义和经济学角度对我国养老金的发展给出较为合理的建议，以减少养老金资金缺口为目的给出了切实有效的解

决办法，并通过仿真验证与理论分析进行了验证。

此外，由于时间等客观原因，本文所完成的工作还存在一些不足之处：

(1) 由于我国养老金发展年份相较于发达国家并不长，统计数据并不充足，同时难以收集到大量的有关新政策的数据，因此本文模型的数据支撑比较薄弱，导致使用本文方法建立的模型不能完全解释少数经济现象或准确预测多年以后的模型发展情况。

(2) 基于多重线性回归的方法难以预测非线性度较大的变量变化情况，模型具有一定的局限性。

(3) 利用支持向量机的自学习功能建立的模型，由于算法的固有缺陷，导致模型的灵活性不足。将其与传统精算法结合可以部分解决该问题，但算法理论解析部分比较简单，故模型难以分析定量分析变量对其他各个变量的影响，即对一些关键因子的综合影响预测难度较大。

## 7.2 我国养老保险体系未来发展建议

在中国城乡一体化改革的背景下，多层次的养老保险是中国社会保障制度改革必然选择。2010年，中央提出建立“广覆盖、保基本、多层次、可持续”的覆盖城乡居民的社会保障体系的十二字方针，这成为我国多层次养老保险体系改革的重要指导。多层次养老保险体系的目标是消除贫困、收入替代和提高生活质量，保障对象是城乡一体化以后包括农村居民在内的全体国民<sup>[6]</sup>。因此结合本文的工作分析结果和相关文献研究内容，提出以下意见和建议：

(1) 全面推进城乡一体化养老金政策，扩大养老金覆盖面。将尚未参加养老保险的适龄在职人口及时纳入全国养老体系，扩大农村居民的养老保障范围，一方面符合平等原则，另一方面有效解决养老金逐年严峻的缺口问题。

(2) 完善多层次养老结构，注重需求决定比例的原则。历史经验与其他国家的实践均表明，单一依靠政府层面的基本养老保障难以从根本上解决养老金负担问题，必须根据需求灵活调整多层次养老的比例，大力发展企业年金养老与个人商业养老事业，辅助建立完整稳固的城乡养老体系，分担养老责任。。

(3) 加大力度完善现有养老金的流通环境。鼓励多种养老金的投资手段，完善投资环境，增大养老金资金积累。根据国家实际经济政治环境，合理调节缴费率与替代率等因子的比例，做到养老金稳固可持续发展。

## 参考文献

- [1] 高连杰. 构建我国新型养老保险体系[D]. 北京交通大学, 2012.
- [2] 王海涛, 刘建政. 我国城镇基本养老金收支状况的预测分析[J]. 中国海洋大学学报, 4:94-96, 2004.
- [3] 沈毅. 基于统筹账户的城镇老年人口养老金收支预测分析[J]. 党政干部学刊, 5: 60-63,15, 2011.
- [4] 周莹. 新型农村社会养老保险中基本养老金仿真学精算模型[J]. 上海经济研究, 7: 17-24, 2009.
- [5] 陶纪坤. “旧农保”与“新农保”方案对比研究[J]. 兰州学刊, 6:90-94, 2010.
- [6] 张慧. 城乡一体化的多层次养老保险体系结构比例研究[D]. 上海工程技术大学, 2012.

- [7] 董克用, 孙博. 从多层次到多支柱: 养老保障体系改革再思考[J]. 公共管理学报, 8 (1): 1-9, 2011.
- [8] 褚福灵. 多支柱养老保险理论与实践[J]. 北京市计划劳动管理干部学院学报, 11 (3): 30-33, 2003.
- [9] 国家统计局. 中国统计年鉴[Z]. <http://www.stats.gov.cn/tjsj/ndsj/>.
- [10] 国家统计局. 中国劳动统计年鉴[Z]. 中国知网: 中国经济与社会发展统计数据库.
- [11] 东方财富网[Z]. <http://data.eastmoney.com/shdata/shdata.aspx?fo=1&fi=17>.
- [12] 王中宇, 张海滨, 刘智敏. 剔出离群值的学生化残差新方法[J]. 仪器仪表学报. 2006, 27(6): 624-628, 637.
- [13] 吴喜之. 统计学: 从概念到数据分析[M]. 北京: 高等教育出版社, 2008.
- [14] 杨维忠, 张甜. SPSS 统计分析与行业应用案例详解[M]. 北京: 清华大学出版社, 2011.
- [15] 王定成. 支持向量机建模预测与控制[M]. 北京: 气象出版社, 2009.
- [16] Chih-Jen Lin's Home Page[Z]. <http://www.csie.ntu.edu.tw/~cjlin/>.
- [17] Matlab 中文论坛. MATLAB 神经网络 30 个案例分析[M]. 北京: 北京航空航天大学出版社, 2010.
- [18] 国务院. 国务院关于企业职工养老保险制度改革的决定[Z]. <http://www.people.com.cn/item/flfgk/gwyfg/1991/112601199121.html>
- [19] 社会保险基金监督司. 2012 年度全国企业年金业务数据摘要 [Z]. [http://www.mohrss.gov.cn/gkml/xxgk/201304/t20130424\\_99278.htm](http://www.mohrss.gov.cn/gkml/xxgk/201304/t20130424_99278.htm).
- [20] 人力资源和社会保障部. 劳动和社会保障事业发展统计公报[Z]. <http://www.mohrss.gov.cn/SYrlzyhshbzb/zwgk/szrs/ndtjsj/tjgb/>.
- [21] 红网[Z]. <http://gov.rednet.cn/c/2006/09/24/990498.htm>.
- [22] 人民网[Z]. <http://cpc.people.com.cn/n/2012/1118/c64094-19612151.html>.
- [23] 中国金融信息网[Z]. <http://news.xinhua08.com/zt/jjsj/201309/#f-0382>.
- [24] 张海波. 中国公共养老金制度的模式选择与完善[D]. 南开大学, 2009.
- [25] 钟仁耀. 养老保险改革国际比较研究[M]. 上海: 上海财经大学出版社, 2004.
- [26] 中国劳动资讯网[Z]. <http://www.51labour.com/labour-law/show-6770.html>.
- [27] 中国金融[Z]. <http://www.cnfinance.cn/magzi/2010-06/21-9777.html>.
- [28] 北斗星社区[Z]. <http://www.bdstar.org/Article/Class31/xgxklm/200508/1750.html>.
- [29] 百度文库[Z]. <http://wenku.baidu.com/view/552e56c008a1284ac9504305.html>.
- [30] 豆丁[Z]. <http://www.docin.com/p-495020664.html>.
- [31] 韩大伟, 厉放, 吴家亨. 建立 21 世纪的养老金体系[M]. 北京: 经济科学出版社, 2007.