

全国第六届研究生数学建模竞赛



题目 城镇登记失业率与主要经济指标的模型

摘 要:

本文从经济学的角度,分析提取了 GDP、通货膨胀率等 12 个与失业率相关的经济指标,通过数据的收集和整理使用因子分析法选出国内生产总值 GDP、社会消费品零售总额、通货膨胀率、进出口贸易总额、国家税收收入等 5 个指标作为分析城镇登记失业率的主因素。

通过对 10 年的相关数据进行分析,从数理统计的角度经过分析后,选择回归分析法建立了多元线性回归模型,并验证了模型的正确性。在此基础上,将各个经济因素按不同地区、不同行业进行细化分析,并结合不同人群的失业系数,得到了一个经过修正后的精确模型。

在划分不同地区的时候,依据各个地区的经济实力和经济增长率,以全国人均 GDP 和名义全国平均 GDP 增长率为标准将各个省(市)分为四类,建立波士顿矩阵。

在预测未来就业前景的时候,本文首先不考虑政府决策和规划的影响,认为经济在一个“无政府”状态下运行,利用 BP 神经网络模型,预测出 2009 年和 2010 年的失业率分别为 7.1160%,和 6.2719%。然后在分析了相关政策对各项经济指标的影响,得出了政策对经济指标的影响系数矩阵,建立了政策影响下的城镇登记失业率模型,预测得到在政策影响下 2009 年和 2010 年的失业率分别为 4.0825%和 3.6256%。

最后,本文结合所建立的模型和得出的结论,对减小我国城镇登记失业率提出了扩大内需、增加出口、增加财政支出的建议,并就当前的相关政策进行了系统分析,提出了具体的建议。

本文所用的数据主要来源于“和讯网”、“中国国家统计局”的官方网站、“华尔街日报”网、“国务院发展研究中心信息网”以及《中国统计年鉴 2008》。

关键字: 因子分析法; 回归模型; 波士顿矩阵; BP 神经网络

参赛队号 9004512

队员姓名 谢小平 尹高扬 李冬

参赛密码 _____

(由组委会填写)

1 问题的重述

失业、经济增长和通货膨胀为宏观经济中特别重要的三个指标，就业（或者失业）是社会、国民经济中极其重要的问题。

1.1 影响就业的因素

从经济学的角度，影响就业（或者失业）的因素很多。从宏观层面上，消费、投资、政府购买和进出口都是重要的因素；而从中观层面，不同地区、不同产业也会表现出不同的特征。当然，中央政府调整宏观经济政策（包括财政政策和货币政策），以及对不同地区和不同产业实行不同的扶持政策都会对就业产生巨大的影响。

1.2 我国的现实情况

2008 年我国经济社会经受了历史罕见的考验，GDP 依然保持 9% 以上平稳较快增长，城镇新增就业 1113 万人，城镇登记失业率为 4.2%。2009 年我国就业面临更大的挑战，一是国际金融危机导致国际市场需求难以在短期内复苏；二是今年我国经济增速下滑；三是国内消费需求乏力；四是一些行业产能过剩与市场预期不确定导致企业投资不足，所以就业形势十分严峻。

1.3 国家的就业政策及目标

中央政府从 08 年 10 月开始实施了 40000 亿元的投资计划，确定了十大产业振兴计划，采取扩大国内消费需求的措施，提高对外开放水平以增加出口。同时，中央财政拟投入 420 亿元资金实施积极的就业政策。09 年我国在就业方面的目标：城镇新增就业 900 万人以上，城镇登记失业率控制在 4.6% 以内。

1.4 需要解决的问题

(1) 参考就业问题的研究成果，利用近年来我国有关的统计数据并结合一年多来我国国民经济的运行数据，对有关统计数据进行分析，寻找影响就业的主要因素或指标。

(2) 建立城镇就业人数或城镇登记失业率与上述主要因素或指标之间联系的数学模型。

(3) 对上述数学模型从包含主要的经济社会指标、分行业、分地区、分就业人群角度，尝试建立比较精确的数学模型。

(4) 利用所建立的关于城镇就业人数或城镇登记失业率的数学模型，对未来的情况作适当的假设，根据国家的有关决策和规划对 2009 年及 2010 年上半年的我国就业前景进行仿真。

(5) 根据所建立的数学模型和仿真结果，对提高我国城镇就业人口数或减少城镇登记失业率提出咨询建议。

2 问题的分析

就业：三个月内有稳定的收入或与用人单位有劳动聘用关系。

城镇登记失业率：城镇登记失业人数同城镇从业人数与城镇登记失业人数之和的比。其中，城镇登记失业人员是指有非农业户口，在一定的劳动年龄内（16 岁以上及男 50 岁以下、女 45 岁以下），有劳动能力，无业而要求就业，并在当地就业服务机构进行求职登记的人员。

2.1 失业的类型和形态

在发展中国家，失业的类型和形态包括以下四类：

(1) 自然失业，是指由于经济中某些难以避免的原因所引起的失业。自然失业包括：摩擦性失业、结构性失业、临时性或季节性失业、技术性失业、求职性失业。

(2) 周期性失业，又称总需求不足的失业，是由于总需求不足而引起的短期失业，它一般出现在经济周期的萧条阶段。这种失业与经济中周期性波动是一致的。

(3) 劳动过剩型失业,是指在资本积累不足、劳动严重过剩的发展中经济体,由劳资配合比例失调引发的、即使工资下降到接近于 0,仍无法消除失业。失业者的主体是农村剩余劳动力,这种失业要到工业化、城市化基本完成后才会消失。

(4) 不充分就业。在“劳动过剩型失业”之前,大量的农村剩余劳动力就经常处于就业和失业之间,就业的时间不足、在非正规部门临时性地打工——就业缺乏起码的稳定性、就业的工资报酬极低——经常处于最低工资标准金下,所有这些,使发展中国家的劳动力市场出现了一个庞大的半就业、半失业群体。

2.2 历史的回顾

在 20 世纪 70 年代末和 90 年代初,我国出现了两次失业高峰。70 年代末的失业人员主要是返乡知青,而 90 年代初则主要是国有企业员工。我国在 70 年代末期,由于需要面对十年动乱所带来的严重的就业问题,重点在于改革统包分配的就业政策和坚持多种经济成分并存,扶植发展集体所有制和个体经济,以提供广泛的就业机会。在政府政策的影响下,在接下来的几年我国城镇失业率逐年下降。

20 世纪 80 年代末期,我国在用工制度上对新就业人员开始了全员劳动合同制试点,而对于双向选择后剩余的劳动力,却未能有很好的解决方法。进入 90 年代后,随着社会主义市场经济体制的逐步建立,经济发展由粗放型转向集约型,劳动就业也由政府主导逐渐转为市场主导,这在一定程度上制约了就业。而国有企业环境的恶化,使大批国有企业员工下岗分流,形成了大批失业人口。

近几年我国在经济一直以 10%左右的速度持续高速增长的同时就业率没有得到相应的增长,相反,就业形势日趋严峻,城镇登记失业率逐年增加。有数据表明:1995~2002 年期间,城镇失业率从 4.0%提高到 6.1%,劳动参与率从 72.9%下降到 66.5%(蔡防等,2004)。因此,国内外的很多专家学者就认为中国正在出现着令人担忧的“无就业增长”格局,例如,罗斯基(Rawski,2001)就把“就业增长几乎为零”,作为质疑中国经济实际增长速度的依据之一^[1]。

2.3 影响我国就业因素分析

我国是处于市场经济条件下的社会主义国家,就业状况是由劳动力市场的供给和需求所决定的,而影响供给和需求的因素很多,既有人口、经济发展速度、技术水平、产业结构等生产力方面的因素,也有社会经济制度、就业及社会保障制度等生产关系方面因素的制约,同时还要受到一些观念意识的影响。

目前我国市场机制尚不健全,劳动力资源市场配置仍存在诸多障碍,全国统一、开放、竞争、有序的劳动力市场体制尚未形成,城乡之间、地区之间分割的格局尚未完全打破,特别是社会保障制度有待完善。此外,工资水平、城镇化发展速度和水平、就业观念等,也是影响就业的重要因素,应给予充分的重视。

2.4 问题的分析

研究城镇登记失业率的主要目的:建立准确的失业率相关因素模型,给政府决策部门提出相关的咨询建议和参考,为构建和谐社会贡献一份力量。

研究城镇登记失业率的主要要求包括:

(1) 收集各方面的资料和数据,归纳总结出所有与登记失业率相关的因素,利用主成分分析法、因子分析法等数学方法,找出主要的相关因素或指标。

(2) 建立主要因素与城镇登记失业率的回归分析模型,得出各个因素与城镇登记失业率的数学模型。

(3) 在我国不同地区经济差异比较大,不同行业的就业前景也各不相同,同时不同人群的就业能力和要求也不同。所以,为了将模型更加的细化,就需要将不同地区、不同行业、不同人群作为考虑因素建立比较详细的数学模型。

(4) 利用所建立的失业率数学模型, 结合国家的有关政策和规划, 考虑使用 BP 神经网络等预测方法, 预测出我国 2009 年及 2010 年的失业率。

(5) 根据所建立的模型和预测结果, 结合国家的失业率控制目标, 提出一些政策措施, 保证国家的失业率控制目标能够实现。

3 模型的假设

- 1) 假设只考虑所列出的经济指标, 忽略其他经济指标对失业率的影响;
- 2) 假设只考虑国家政策对经济指标的正面影响;
- 3) 假设不考虑近年来我国的自然灾害的影响, 社会经济政治稳定;
- 4) 假设所收集到的数据真实可靠。

4 符号的说明

| | |
|------------------|-----------------------------------|
| x_1 | GDP |
| x_2 | 社会消费品零售总额 |
| x_3 | 城镇固定资产投资 |
| x_4 | 广义货币供应量 M2 |
| x_5 | 通货膨胀率 |
| x_6 | 进出口贸易总额 |
| x_7 | CPI |
| x_8 | PPI |
| x_9 | 国家财政预算支出 |
| x_{10} | 国家税收收入 |
| x_{11} | 货运量 |
| x_{12} | 客运量 |
| x_{ij} | 第 i 个评价对象的第 j 个指标的取值 |
| \tilde{x}_{ij} | 标准化标量 |
| \bar{x}_j | 第 j 个指标的样本均值 |
| s_j | 第 j 个指标的样本的标准差 |
| r_{ij} | 第 i 个指标与第 j 个指标的相关系数 |
| \hat{F}_{ij} | 第 i 个样本点对第 j 个因子 F_j 得分的估计值 |
| β_i | 回归系数 |
| G_i | 第 i 省的 GDP 数值 |
| t_i | 各省的通货膨胀率 |
| η_l | 不同人群的失业率系数 |
| b_1 | 国家决策和规划对 GDP 指标的影响系数 |
| b_2 | 拉动内需政策对社会消费品零售总额的影响系数 |
| b_3 | 财政政策和货币政策对通货膨胀率的影响系数 |
| b_4 | 贸易税收政策对贸易进出口总额的影响系数 |
| b_5 | 国家政策和规划对税收收入总额的影响系数 |

5 模型的建立

5.1 问题一的解决

劳动力的充分就业是所有经济发展过程所追求的目标。在我国，就业的增长意味着人民收入的提高和生活水平的改善。就业问题已经是目前我国面临的重要课题，如何在未来一段时间内确保我国就业目标的实现，对于我国保证社会的健康发展有着重要的现实意义^[2]。

5.1.1 相关因素的选择

城镇登记失业率作为宏观经济中特别重要的指标之一，反映着当前宏观经济的状况，被很多因素影响着。本文经过分析采用从以下因素中提取影响就业的主要因素：

(1) GDP: x_1

GDP (Gross Domestic Product), 也就是国内生产总值, 通常定义为: 一定时期内 (一个季度或一年), 一个国家或地区的经济中所生产出的全部最终产品和提供劳务的市场价值的总值。

上榜理由: GDP 是宏观经济中最受关注的经济统计数字, 因为它被认为是衡量国民经济发展情况最重要的一个指标, GDP 反映的是国民经济各部门的增加值的总额。社会

上榜指数: ★★★

(2) 社会消费品零售总额: x_2

社会消费品零售总额 (social retailgoods) 指各种经济类型的批发零售贸易业、餐饮业、制造业和其他行业对城乡居民和社会集团的消费品零售额和农民对非农业居民零售额的总和。它反映一定时期内人民物质文化生活水平的提高情况, 反映社会商品购买力的实现程度, 以及零售, 市场的规模状况。

上榜理由: 社会消费品零售总额由社会商品供给和有支付能力的商品需求的规模所决定, 是研究人民生活水平、社会零售商品购买力、社会生产、货币流通和物价的发展变化趋势的重要资料。

上榜指数: ★★★

(3) 城镇固定资产投资: x_3

固定资产投资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。固定资产再生产过程包括固定资产更新 (局部和全部更新)、改建、扩建、新建等活动。

上榜理由: 城镇固定资产投资额是以货币表现的建造和购置固定资产活动的工作量, 它是反映固定资产投资规模、速度、比例关系和使用方向的综合性指标, 对我国的社会主义现代化建设具有重要意义。

上榜指数: ★★

(4) 广义货币供应量 M2: x_4

M2: M1 的所有项目以及流动性较差、不能直接用作支付工具但可以较方便地变为支付工具的项目。我国将货币供应量划分为三个层次: 一是流通中现金 M0, 即在银行体系外流通的现金; 二是狭义货币供应量 M1, 即 M0 加上企事业单位活期存款; 三是广义货币供应量 M2, 即 M1 加上企事业单位定期存款、居民储蓄存款和其他存款。

上榜理由: M2 流动性偏弱, 但是反应的是社会总需求的变换恶化未来通货膨胀的压力状况。

上榜指数: ★★★

(5) 通货膨胀率: x_5

通货膨胀率是货币超发部分与实际需要的货币量之比, 用以反映通货膨胀、货币贬值的程度。经济学上, 通货膨胀率为: 物价平均水平的上升速度 (以通货膨胀为准)。

上榜理由: 通货膨胀引起收入和财富的再分配, 扭曲商品相对价格, 降低资源配置

效率，促发泡沫经济直接影响就业情况。

上榜指数：★★★★

(6) 进出口贸易总额： x_6

进出口总值是指进出口商品所产生的税金总价值，其中以除去部分免关税的的国内企业，就是国家总的出口商品服务和进口商品服务之总和。

上榜理由：进出口总值用以观察一个国家在对外贸易方面的总规模。

上榜指数：★★

(7) CPI： x_7

CPI 即消费者物价指数(ConsumerPriceIndex)，是对一个固定的消费品篮子价格的衡量，主要反映消费者支付商品和劳务的价格变化情况，也是一种度量通货膨胀水平的工具，以百分比变化为表达形式。

上榜理由：CPI 反映与居民生活有关的产品及劳务价格统计出来的物价变动指标，通常作为观察通货膨胀水平的重要指标。

上榜指数：★★

(8) PPI： x_8

生产者物价指数 (PPI)：生产者物价指数主要的目的在衡量各种商品在不同的生产阶段的价格变化情形。

上榜理由：PPI 是衡量工业企业产品出厂价格变动趋势和变动程度的指数，是反映某一时期生产领域价格变动情况的重要经济指标，也是制定有关经济政策和国民经济核算的重要依据。

上榜指数：★

(9) 国家财政预算支出： x_9

国家财政预算支出是指国家所有部门及各事业单位的行政经费、各项事业经费、社会保障支出、基本建设支出、挖潜改造支出、科技三项费用及其他支出。

上榜理由：国家财政预算支出之间决定了未来经济的走势，是国家宏观调控经济的重要手段之一。

上榜指数：★★

(10) 国家税收收入： x_{10}

税收是指国家为了实现其职能，凭借政治权力，按照法定标准，强制地、无偿地、固定地取得财政收入的一种手段。

上榜理由：税收的作用是税收职能在一定经济条件下的外在表现，税收的多少直接决定政府的支出，影响就业。

上榜指数：★★★★

(11) 货运量： x_{11}

货运量是指在国家区域内进口和出口货物的流通量，只包括进口和出口，不包括转口的数据。

上榜理由：货运量反应了国家经济的繁华性。

上榜指数：★

(12) 客运量： x_{12}

客运量是指在规定时间内运送旅客的数量，计量单位以“人”表示。

上榜理由：客运量反应了国家经济的活跃性。

上榜指数：★

5.1.2 因子分析模型的建立

(1) 对原始数据进行标准化处理

本文中, 进行因子分析的指标变量有 12 个: x_1, x_2, \dots, x_{12} , 共有 1999 年到 2008 年 10 个数据, 即有 10 个评价对象, 第 i 个评价对象的第 j 个指标的取值为 x_{ij} 。将各指标值 x_{ij} 转化为标准化标量 \tilde{x}_{ij}

$$\tilde{x}_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}, (i = 1, 2, \dots, 10; j = 1, 2, \dots, 12)$$

其中, 第 j 个指标的样本均值:

$$\bar{x}_j = \frac{1}{10} \sum_{i=1}^{10} x_{ij}$$

第 j 个指标的样本的标准差:

$$s_j = \frac{1}{10-1} \sum_{i=1}^{10} (x_{ij} - \bar{x}_j)^2, (j = 1, 2, \dots, 12)$$

(2) 计算相关系数矩阵 R

相关系数矩阵 $R = (r_{ij})_{12 \times 12}$

其中

$$r_{ij} = \frac{\sum_{k=1}^{10} \tilde{x}_{ki} \cdot \tilde{x}_{kj}}{10-1}, (i, j = 1, 2, \dots, 12)$$

式中 $r_{ii} = 1$, $r_{ij} = r_{ji}$, r_{ij} 是第 i 个指标与第 j 个指标的相关系数。

(3) 计算初等载荷矩阵

计算相关系数矩阵 R 的特征值 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{12} \geq 0$, 及对应的特征向量 u_1, u_2, \dots, u_{12} , 其中

$$u_j = (u_{1j}, u_{2j}, \dots, u_{10j})^T$$

初等载荷矩阵

$$A = [\sqrt{\lambda_1}u_1 \quad \sqrt{\lambda_2}u_2 \quad \dots \quad \sqrt{\lambda_{12}}u_{12}]$$

(4) 选择 2 个主因子, 进行因子旋转

根据初等载荷矩阵, 计算各个公共因子的贡献率, 并选择 2 个主因子。对提取的因子载荷矩阵进行旋转, 得到矩阵 $B = A_2 T$ (其中 A_2 为 A 的前 2 列, T 为正交矩阵), 构造因子模型

$$\begin{cases} \tilde{x}_1 = b_{1,1}F_1 + b_{1,2}F_2 \\ \dots\dots\dots \\ \tilde{x}_{12} = b_{12,1}F_1 + b_{12,2}F_2 \end{cases}$$

利用 MATLAB 程序计算得到旋转后的因子分析表, 见表 1。

表 1 旋转因子分析表

| 因素 | 主因子 1 | 主因子 2 |
|------------|--------|---------|
| 国内生产总值 GDP | 0.9149 | -0.3969 |
| 社会消费品零售总额 | 0.8980 | -0.4284 |
| 城镇固定资产投资 | 0.9002 | -0.4191 |
| 广义货币供应量 M2 | 0.8956 | -0.4318 |

| | | |
|----------|---------|---------|
| 通货膨胀率 | 0.8557 | -0.1922 |
| 进出口贸易总额 | 0.9127 | -0.3993 |
| CPI | -0.6183 | 0.5970 |
| PPI | 0.5371 | -0.5585 |
| 国家财政预算支出 | 0.2189 | -0.9358 |
| 国家税收收入 | 0.9346 | -0.3375 |
| 货运量 | 0.9133 | -0.3974 |
| 客运量 | 0.9135 | -0.3896 |

(5) 计算因子得分，并进行综合评价

本文用回归方法求单个因子得分函数

$$\begin{cases} \hat{F}_1 = b_{1,1}\tilde{x}_1 + \dots + b_{1,12}\tilde{x}_{12} \\ \hat{F}_2 = b_{2,1}\tilde{x}_1 + \dots + b_{2,12}\tilde{x}_{12} \end{cases}$$

记第 i 个样本点对第 j 个因子 F_j 得分的估计值

$$\hat{F}_{ij} = b_{j1}\tilde{x}_{i1} + b_{j2}\tilde{x}_{i2} + \dots + b_{jp}\tilde{x}_{ip} \quad (i=1,2,\dots,12, \quad j=1,2)$$

则有

$$\begin{bmatrix} b_{11} & b_{21} \\ b_{12} & b_{22} \\ \vdots & \vdots \\ b_{1,12} & b_{2,12} \end{bmatrix} = R^{-1}B$$

且

$$\hat{F} = (\hat{F}_{ij})_{n \times m} = X_0 R^{-1} B$$

其中 X_0 是 10×12 的原始数据矩阵， R 为相关系数矩阵， B 是步骤 (4) 中得到的载荷矩阵。

计算得各个因子得分函数

$$\begin{aligned} F_1 &= -4.5313\tilde{x}_1 + 2.5625\tilde{x}_2 + 1.875\tilde{x}_3 - 0.75\tilde{x}_4 + 0.7656\tilde{x}_5 + 0.1875\tilde{x}_6 \\ &\quad + 0.3728\tilde{x}_7 - 0.1641\tilde{x}_8 + 0.75\tilde{x}_9 + 4.625\tilde{x}_{10} + 8\tilde{x}_{11} - 0.5625\tilde{x}_{12} \\ F_2 &= -0.3438\tilde{x}_1 + 0.7539\tilde{x}_2 - 2.1875\tilde{x}_3 - 0.125\tilde{x}_4 + 0.4297\tilde{x}_5 - 0.75\tilde{x}_6 \\ &\quad + 0.3595\tilde{x}_7 - 0.3125\tilde{x}_8 - 1.0313\tilde{x}_9 + 0.1875\tilde{x}_{10} - 0.625\tilde{x}_{11} + 0.25\tilde{x}_{12} \end{aligned}$$

经过计算得到综合因子得分公式

$$F = 0.7381F_1 + 0.2619F_2$$

根据综合因子的得分公式，得到12个因素与失业率关联的得分情况，见表2。

表2 各因素与失业率关联的得分

| 因素 | 得分 |
|------------|---------|
| 国内生产总值 GDP | 0.3187 |
| 社会消费品零售总额 | 0.0205 |
| 城镇固定资产投资 | -0.1351 |
| 广义货币供应量 M2 | -0.0647 |
| 通货膨胀率 | 1.0169 |
| 进出口贸易总额 | 0.0434 |
| CPI | -0.0079 |

| | |
|----------|---------|
| PPI | 0.0032 |
| 国家财政预算支出 | -0.0784 |
| 国家税收收入 | 0.5025 |
| 货运量 | -0.3755 |
| 客运量 | -0.7049 |

以 0 为分界，选出与失业率关联最大的 5 个因素，得到如图 1 的筛选结果。

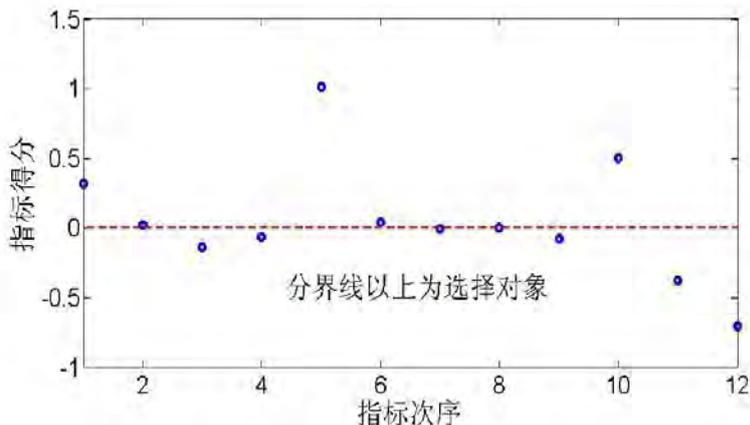


图 1 各因素与失业率关联的得分图

从图 1 中可以明显的看出，国内生产总值 GDP、社会消费品零售总额、通货膨胀率、进出口贸易总额、国家税收收入是与失业率相关的主要因素，这个分析结果与上文中主观标注的上榜指数基本吻合。

5.2 问题二的解决

本小节要解决的问题是得到城镇登记失业率与各相关因素之间联系的数学模型，而本文采用对 10 年的数值进行分析，通常采用曲线拟合的方法来得到各个因素与城镇登记失业率之间的关系。但是从数理统计的角度来看，曲线拟合也是一个估计，有必要对它做假设检验，因此本文选择回归分析来建立一个多元线性回归模型。

5.2.1 模型的建立

根据经验公式得到城镇登记失业率与各个主要因素之间的多元线性回归模型为

$$\begin{cases} y = \beta_0 + \beta_1 x_1 + \cdots + \beta_5 x_{10} + \varepsilon \\ \varepsilon \sim N(0, \sigma^2) \end{cases} \quad (1)$$

式中 $\beta_0, \beta_1, \dots, \beta_5, \sigma^2$ 是与 $x_1, x_2, x_5, x_6, x_{10}$ 无关的未知参数，其中 $\beta_0, \beta_1, \dots, \beta_5$ 称为回归系数。

现有 10 个独立观测数据 $(y_i, x_{i,1}, \dots, x_{i,10})$, $i = 1, \dots, 10$ ，由 (1) 得

$$\begin{cases} y_i = \beta_0 + \beta_1 x_{i,1} + \cdots + \beta_5 x_{i,10} + \varepsilon_i \\ \varepsilon_i \sim N(0, \sigma^2), \quad i = 1, \dots, 10 \end{cases} \quad (2)$$

记

$$X = \begin{bmatrix} 1 & x_{1,1} & \cdots & x_{1,6} & x_{1,10} \\ \vdots & \vdots & & \vdots & \vdots \\ 1 & x_{10,5} & \cdots & x_{10,6} & x_{10,10} \end{bmatrix}, Y = \begin{bmatrix} y_1 \\ \vdots \\ y_{10} \end{bmatrix} \quad (3)$$

$$\varepsilon = [\varepsilon_1 \quad \cdots \quad \varepsilon_{10}]^T, \quad \beta = [\beta_0 \quad \beta_1 \quad \cdots \quad \beta_5]^T$$

因此 (1) 式可以表示为

$$\begin{cases} Y = X\beta + \varepsilon \\ \varepsilon \sim N(0, \sigma^2 E_{10}) \end{cases} \quad (4)$$

其中 E_{10} 为 10 阶单位矩阵。

5.5.2 参数估计

模型 (1) 中的参数 $\beta_0, \beta_1, \dots, \beta_5$ 仍用最小二乘法估计, 即应选取估计值 $\hat{\beta}_j$, 使当

$$\beta_j = \hat{\beta}_j, \quad j = 0, 1, 2, \dots, 5$$

时, 误差平方和 Q 达到最小, 即

$$Q = \min \sum_{i=1}^{10} \varepsilon_i^2 = \min \sum_{i=1}^{10} (y_i - \beta_0 - \beta_1 x_{i,1} - \dots - \beta_5 x_{i,10})^2 \quad (5)$$

令

$$\frac{\partial Q}{\partial \beta_j} = 0, \quad j = 0, 1, 2, \dots, 10$$

得

$$\begin{cases} \frac{\partial Q}{\partial \beta_0} = -2 \sum_{i=1}^{10} (y_i - \beta_0 - \beta_1 x_{i,1} - \dots - \beta_5 x_{i,10}) = 0 \\ \frac{\partial Q}{\partial \beta_j} = -2 \sum_{i=1}^{10} (y_i - \beta_0 - \beta_1 x_{i,1} - \dots - \beta_5 x_{i,10}) x_{ij} = 0, \quad j = 1, 2, \dots, 5 \end{cases} \quad (6)$$

经整理化为以下正规方程组

$$\begin{cases} 20\beta_0 + \beta_1 \sum_{i=1}^{10} x_{i,1} + \beta_2 \sum_{i=1}^{10} x_{i,2} + \dots + \beta_5 \sum_{i=1}^{10} x_{i,10} = \sum_{i=1}^{10} y_i \\ \beta_0 \sum_{i=1}^{20} x_{i,1} + \beta_1 \sum_{i=1}^{20} x_{i,1}^2 + \beta_2 \sum_{i=1}^{20} x_{i,1} x_{i,2} + \dots + \beta_5 \sum_{i=1}^{20} x_{i,1} x_{i,10} = \sum_{i=1}^{20} x_{i,1} y_i \\ \vdots \\ \beta_0 \sum_{i=1}^{20} x_{i,10} + \beta_1 \sum_{i=1}^{20} x_{i,10} x_{i,1} + \beta_2 \sum_{i=1}^{20} x_{i,10} x_{i,2} + \dots + \beta_5 \sum_{i=1}^{20} x_{i,10}^2 = \sum_{i=1}^{20} x_{i,10} y_i \end{cases} \quad (7)$$

正规方程组的矩阵形式为

$$X^T X \beta = X^T Y \quad (8)$$

当矩阵 X 列满秩时, $X^T X$ 为可逆方阵, (8) 式的解为

$$\hat{\beta} = (X^T X)^{-1} X^T Y \quad (9)$$

将 $\hat{\beta}$ 代回原模型得到 y 的估计值

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \dots + \hat{\beta}_5 x_{10} \quad (10)$$

而这组数据的拟合值为 $\hat{Y} = X\hat{\beta}$, 拟合误差 $e = Y - \hat{Y}$ 称为残差, 可作为随机误差 ε 的估计, 而残差平方和

$$Q = \sum_{i=1}^{10} e_i^2 = \sum_{i=1}^{10} (y_i - \hat{y}_i)^2 \quad (11)$$

5.2.3 统计分析

假设上述的模型可以得到以下结果:

(i) $\hat{\beta}$ 是 β 的线性无偏最小方差估计, 即: $\hat{\beta}$ 是 Y 的线性函数; $\hat{\beta}$ 的期望等于 β ; 在 β 的线性无偏估计中, $\hat{\beta}$ 的方差最小。

(ii) $\hat{\beta}$ 服从正态分布, 即

$$\hat{\beta} \sim N(\beta, \sigma^2 (X^T X)^{-1}) \quad (12)$$

记 $(X^T X)^{-1} = (c_{ij})_{10 \times 10}$ 。

(iii) 对残差平方和 Q , 满足

$$EQ = (10 - 5 - 1)\sigma^2$$

且

$$\frac{Q}{\sigma^2} \sim \chi^2(10 - 5 - 1) \quad (13)$$

由此得到 σ^2 的无偏估计

$$s^2 = \frac{Q}{10 - 5 - 1} = \hat{\sigma}^2 \quad (14)$$

s^2 是剩余方差 (残差的方差), s 称为剩余标准差。

(iv) 对总平方和 $SST = \sum_{i=1}^{10} (y_i - \bar{y})^2$ 进行分解, 有

$$SST = Q + U, \quad U = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \quad (15)$$

其中 Q 是由 (5) 定义的残差平方和, 反映随机误差对 y 的影响, U 称为回归平方和, 反映自变量对 y 的影响。

5.2.4 回归模型的假设检验

因变量 y 与自变量 x_1, \dots, x_5 之间是否存在如模型 (1) 所示的线性关系是需要检验的, 显然, 如果所有的 $|\hat{\beta}_j| (j=1, \dots, 5)$ 都很小, y 与 x_1, \dots, x_5 的线性关系就不明显, 所以可令原假设为

$$H_0: \beta_j = 0 (j=1, \dots, 5)$$

当 H_0 成立时由分解式 (15) 定义的 U, Q 满足

$$F = \frac{U/m}{Q/(n-m-1)} \sim F(m, n-m-1) \quad (16)$$

即

$$F = \frac{U/5}{Q/4} \sim F(5, 4)$$

在显著性水平 α 下有上 α 分位数 $F_\alpha(5, 4)$, 若 $F < F_\alpha(5, 4)$, 接受 H_0 ; 否则, 拒绝。

5.2.5 回归系数的假设检验和区间估计

为了进一步验证模型的正确性, 作如下 7 个检验 ($j=0, 1, \dots, 5$):

$$H_0^{(j)}: \beta_j = 0$$

由 (12) 式

$$\hat{\beta}_j \sim N(\beta_j, \sigma^2 c_{jj})$$

其中, c_{jj} 是 $(X^T X)^{-1}$ 中的第 (j, j) 元素, 用 s^2 代替 σ^2 , 由 (12) ~ (14) 式, 当 $H_0^{(j)}$ 成立时

$$t_j = \frac{\hat{\beta}_j / \sqrt{c_{jj}}}{\sqrt{Q/(n-m-1)}} \sim t(n-m-1) \quad (17)$$

即

$$t_j = \frac{\hat{\beta}_j / \sqrt{c_{jj}}}{\sqrt{Q/13}} \sim t(13)$$

对给定的 α ，若 $|t_j| < t_{\frac{\alpha}{2}}(13)$ ，接受 $H_0^{(j)}$ ；否则，拒绝。

5.2.6 模型的计算及验证

经过计算得到回归分析模型中 $\beta_0 \sim \beta_6$ 分别为 -3.4919、0.0002、-0.0002、8.5419、-0.0005、-0.0003，因此回归模型为

$$y = -3.4919 + 0.0002x_1 - 0.0002x_2 - 8.5419x_3 - 0.0005x_4 - 0.0003x_{10}$$

因为 β 的系数都在相应的检验区间内，并且所有置信区间都不包含零点，回归公式满足要求。其中， $R^2 = 0.8668, F = 5.2054, p = 0.0676, s^2 = 0.0618$

取 $\alpha = 0.05$ ，此时 $p < \alpha$ 满足题意，模型成立。

从模型中可以看出，失业率与经济增长成正线性相关，和通货膨胀率成负相关，乍一看来容易使人不能接受。但是仔细分析一下可以发现，中国经济迅速增长的原因是投资拉动型和技术推动型相结合，而大量从农村转移出来的劳动力并不能被一些资本密集型和技术密集型的行业所吸收，所以造成经济增长率较高，失业率高并存现象。

5.3 问题三的解决

上述模型所描述的是全国平均的失业率模型，但是中国有 960 万平方公里，接近 14 亿人口，因此从全国的角度来建立的模型在不同地区、不同人群和不同行业中应用时，会有较大的误差，为此有必要对模型进行更加符合实际的细化，以不同地区、不同行业和不同人群所影响的各项主要的经济指标进行失业率的修正模型。

考虑到数据获得比较困难，并不能得到不同地区、不同行业和不同人群的各项经济指标。为此选择将社会消费品零售总额、进出口总额、税收与不同行业相关联，将 GDP、通货膨胀率与不同地区相关联。

5.3.1 不同地区的划分及权重的确定

经济类型区是根据区内同一性和区外差异性原则确定的，即把全国划分为若干个区域，在每个区域的内部都具有相同或相似的经济特征，而在不同区域之间则有明显的差异。

在我国，经济类型区的划分，除了众所周知的东中西三分法之外，还有：

①以产业结构作为区划标准的“五分法”：把全国分为加工型经济区、加工主导型经济区、资源开发型经济区、资源开发与加工混合型经济区、特殊类型经济区；

②以产业结构作为区划标准的“五分法”：把全国分为重加工业为主的加工区、轻加工业为主的加工区、采掘业为主的资源区、原材料工业为主的资源区、资源与加工并举区；

③在 1960—1970 年代，曾按照战略位置把全国各地划分为一线、二线、三线；

④最近出台的“十一·五”规划纲要草案，把我国国土划分为四大功能区，即：优化开发区域、重点开发区域、限制开发区域、禁止开发区域^[3]。

本文认为按上述划分标准来研究失业率问题不是很适合，因为失业率虽然与产业结构相关，但是相比较经济发展而言，联系更加紧密。因此，本文选择使用经济发展情况，将不同地区进行划分。

表 3 2008 年内地各地区部分经济数据

| 地区 | GDP (亿元) | 人均 GDP (元) | GDP 增幅 (%) | 地区 | GDP (亿元) | 人均 GDP (元) | GDP 增幅 (%) |
|-----|-------------|---------------|---------------|----|-------------|---------------|---------------|
| 广东 | 35696.46 | 37588 | 10.1 | 广西 | 7171.58 | 14786 | 12.8 |
| 山东 | 31072.06 | 33038 | 12.1 | 山西 | 6938.73 | 21103 | 8.3 |
| 江苏 | 30312.61 | 33089 | 12.5 | 陕西 | 6851.32 | 18386 | 15.6 |
| 浙江 | 21486.92 | 42214 | 10.1 | 江西 | 6480.3 | 14727 | 12.6 |
| 河南 | 18407.78 | 18322 | 12.1 | 吉林 | 6424.06 | 18126 | 16 |
| 河北 | 16188.6 | 19588 | 10.1 | 天津 | 6354.38 | 53175 | 16.5 |
| 上海 | 13698.15 | 73124 | 9.7 | 云南 | 5700.1 | 12809 | 11 |
| 辽宁 | 13461.6 | 24945 | 13.1 | 重庆 | 5096.66 | 16014 | 14.3 |
| 四川 | 12506.3 | 11008 | 9.5 | 新疆 | 4203.41 | 17616 | 11 |
| 湖北 | 11330.38 | 18786 | 13.4 | 贵州 | 3333.4 | 7264 | 10.2 |
| 湖南 | 11156.64 | 17521 | 12.8 | 甘肃 | 3176.11 | 11961 | 10.1 |
| 福建 | 10823.11 | 30255 | 13 | 海南 | 1459.23 | 13361 | 9.8 |
| 北京 | 10488 | 63029 | 9 | 宁夏 | 1098.51 | 17540 | 12.2 |
| 安徽 | 8874.2 | 14249 | 12.7 | 青海 | 961.53 | 18940 | 12.7 |
| 黑龙江 | 8310 | 18763 | 11.8 | 西藏 | 395.91 | 13754 | 10.1 |
| 内蒙古 | 7761.8 | 26128 | 17.2 | 全国 | 30.6 万 | 20400 | 9 |

以 2008 年全国各地区的经济实力（人均 GDP）和经济增长率（GDP 增长率）为依据，将国内 31 个省（直辖市、自治区）划分为四类地区，建立波士顿矩阵。从国家统计局公布的数据看到，全国的平均 GDP 增长率要低于各地区的 GDP 增长率，出现了数字“打架”的情况，这是因为各省在计算 GDP 的时候出现了重复计算的情况，导致国家统计局公布数据的“自相矛盾”。鉴于此情况，本文选 10.5% 的名义全国平均 GDP 增长率作为划分标准，即 GDP 增长率高于 10.5% 为一类，低于为一类；而人均 GDP 以全国的人均数据为基准，地区的人均 GDP 超过全国人均 GDP 的为一类，低于的为另一类，由此可以将内地的 31 个省（市）采用波士顿矩阵化为四类。

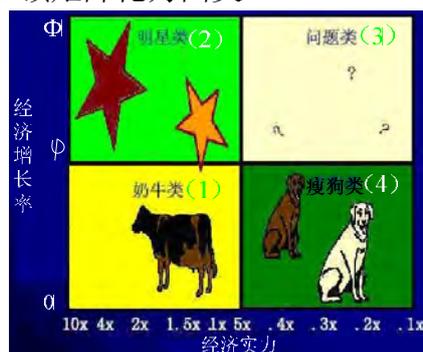


图 2 波士顿矩阵图

波士顿矩阵的说明：

- 1 奶牛类地区：低经济增长率，高经济实力；
- 2 明星类地区：高经济增长率，高经济实力；
- 3 问题类地区：高经济增长率，低经济实力；
- 4 瘦狗类地区：低经济增长率，低经济实力。

根据搜集到的数据并经过整理，按照上面的划分标准，将所有的省分类，得到波士顿逻辑矩阵 $V_{31 \times 4}$ ，矩阵元素是 f_{im} ($f_{im} = 1$ 是指第 i 个地区属于第 m 种地区, 1, 2, 3, 4 分别代表奶牛类地区、明星类地区、问题类地区和瘦狗类地区)。

表 4 波士顿逻辑矩阵

| 省 份 | 1 | 2 | 3 | 4 | 省 份 | 1 | 2 | 3 | 4 | 省 份 | 1 | 2 | 3 | 4 |
|-------|---|---|---|---|-----|---|---|---|---|-----|---|---|---|---|
| 北 京 | 1 | 0 | 0 | 0 | 安 徽 | 0 | 0 | 1 | 0 | 重 庆 | 0 | 0 | 1 | 0 |
| 天 津 | 0 | 1 | 0 | 0 | 福 建 | 0 | 1 | 0 | 0 | 四 川 | 0 | 0 | 0 | 1 |
| 河 北 | 1 | 0 | 0 | 0 | 江 西 | 0 | 0 | 1 | 0 | 贵 州 | 0 | 0 | 0 | 1 |
| 山 西 | 1 | 0 | 0 | 0 | 山 东 | 0 | 1 | 0 | 0 | 云 南 | 0 | 0 | 1 | 0 |
| 内 蒙 古 | 0 | 1 | 0 | 0 | 河 南 | 0 | 0 | 1 | 0 | 西 藏 | 0 | 0 | 0 | 1 |
| 辽 宁 | 0 | 1 | 0 | 0 | 湖 北 | 0 | 0 | 1 | 0 | 陕 西 | 0 | 0 | 1 | 0 |
| 吉 林 | 0 | 0 | 1 | 0 | 湖 南 | 0 | 0 | 1 | 0 | 甘 肃 | 0 | 0 | 0 | 1 |
| 黑 龙 江 | 0 | 0 | 1 | 0 | 广 东 | 1 | 0 | 0 | 0 | 青 海 | 0 | 0 | 1 | 0 |
| 上 海 | 1 | 0 | 0 | 0 | 广 西 | 0 | 0 | 1 | 0 | 宁 夏 | 0 | 0 | 1 | 0 |
| 江 苏 | 0 | 1 | 0 | 0 | 海 南 | 0 | 0 | 0 | 1 | 新 疆 | 0 | 0 | 1 | 0 |
| 浙 江 | 1 | 0 | 0 | 0 | | | | | | | | | | |

为了更加清楚的看出不同地区的经济发展情况，本文将使用不同的颜色将不同地区按上述标准进行着色，如图 3 所示。

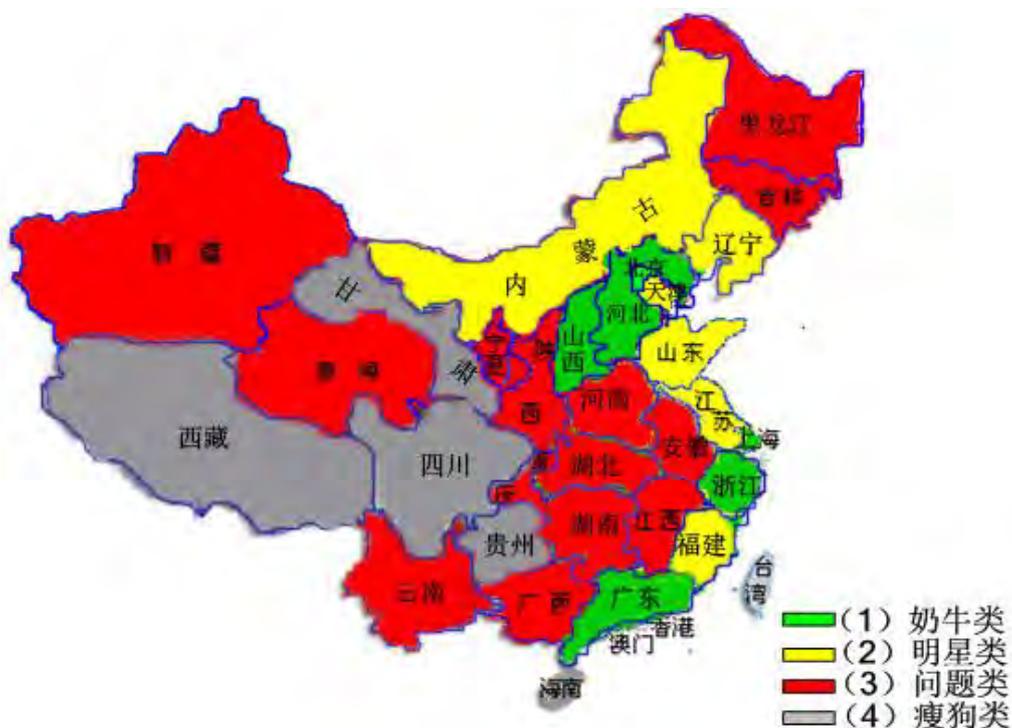


图 3 各个地区的分类地图

从图 3 中可以看出发展快但是经济基础弱的地区（问题类）占主要部分，主要集中在中西部地区，这符合中国目前的西部大开发的实际国情。

5.3.2 不同行业的划分

根据第五次（2000 年）全国人口普查的数据，将上述 19 个行业归纳为 16 个行业，各个行业的就业人群比例见表 5。

表 5 第五次人口普查不同行业就业情况

| | | | |
|---------------|--------|-----------------|-------|
| 农林牧渔业 | 64.38% | 金融保险业 | 0.59% |
| 采掘业 | 1.04% | 房地产业 | 0.23% |
| 制造业 | 12.46% | 社会服务业 | 2.15% |
| 建筑业 | 2.68% | 卫生体育和社会福利业 | 1.06% |
| 电力煤气及水的生产和供应业 | 0.63% | 教育、文化艺术及广播电影电视业 | 2.56% |
| 地质勘查业水利管理业 | 0.13% | 科学研究和综合技术服务业 | 0.22% |
| 交通运输仓储及邮电通信业 | 2.58% | 国家机关政党机关和社会团体 | 2.35% |
| 批发和零售贸易餐饮业 | 6.69% | 其他行业 | 0.25% |

根据 2007 年的三大产业就业人数统计，从事第一产业的人口仍然占绝大多数，同时所占的比例有所减小，而第二产业和第三产业的从业人数都有所提高。从两个数据中可以看出，我国目前还是一个农业大国，但是正走在工业化、社会主义现代化的大路上，具体数据见图 4。

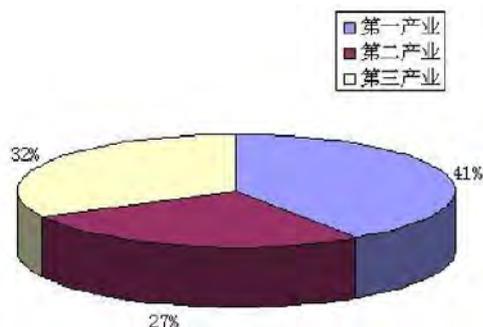


图 4 2007 三大产业从业人数分布

5.3.3 就业人群的划分

政府划分就业人群的时候通常划分为以下几类：下岗失业人员、困难群体、新生劳动力、离岗人员、农民。本文不讨论这样划分的合理性，但是考虑到相关数据的收集问题，因此本文并不能采用这种划分方案。

依据学历划分不失为一种好的划分方法，不论从划分的合理性还是数据的可收集性，对于数理统计的问题来讲，按学历划分劳动力是一个可接受的方案。图 5 表示的是第五次（2000 年）人口普查的，全国劳动力受教育程度。

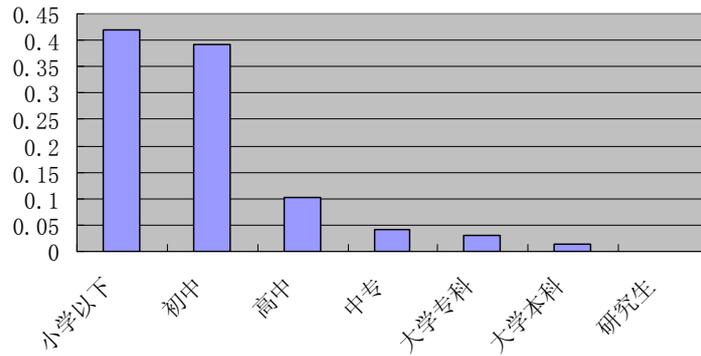


图 5 第五次人口普查时劳动力受教育程度

5.3.4 精确模型的建立

要将第一问中所提取的各项经济指标，分行业、分地区、分就业人群的划分，来建立精确的数学模型，主要有以下几个步骤：

(1) 确定各项经济指标（国内生产总值 GDP、社会消费品零售总额、通货膨胀率、进出口贸易总额、国家税收收入）与行业、地区、就业人群的关联度。

依据数据的可收集性，将 GDP 的数据按照各个地区划分，依据上文将各个地区的划分标准，将同类地区的各省（市）的 GDP 求平均，得到与地区相关的 GDP 数据

$$x_{1j} = \sum_{i=1}^{nj} G_i \quad (j=1,2,3,4)$$

其中， G_i ——第 i 省的 GDP 数值；

j ——1, 2, 3, 4 分别代表奶牛类地区、明星类地区、问题类地区、瘦狗类地区；

nj —— j 类地区包括的省（市）的个数。

将社会消费品零售总额按照行业划分，收集各个行业的社会消费品零售总额，得到各个行业的社会消费品零售总额 x_{2k} ($k=1,2,\dots,16$)，其中， $k=1,2,\dots,16$ 分别代表 16 个行业。

将通货膨胀率按照不同的地区划分，收集各个省市的通货膨胀率的数据，得到各类地区的平均通货膨胀率

$$x_{5j} = \frac{1}{nj} \sum_{i=1}^{nj} t_i \quad (j=1,2,3,4)$$

其中， t_i ——各省的通货膨胀率。

将进出口贸易总额按照行业划分，收集不同行业的进出口贸易总额 x_{6k} ($k=1,2,\dots,16$)，其中， $k=1,2,\dots,16$ 分别代表 16 个行业。

将国家税收收入按照行业划分，收集不同行业的国家税收收入 $x_{10,k}$ ($k=1,2,\dots,16$)。

因为按人群划分的各项数据不能收集到，所以本文暂时只考虑按地区和按行业划分将模型进行初步的精确化。

(2) 建立初步精确模型并求解

因为模型二中验证了失业率与各个经济因素存在线性关系，因此选择最小二乘法建立精确模型，令：

$$f(x) = a_1 x_{1j} + a_2 x_{2j} + a_3 x_{5j} + a_4 x_{6k} + a_5 x_{10k}$$

其中 a_1, a_2, \dots, a_5 是待定系数。

使用最小二乘法求解，记

$$J(a_1, \dots, a_5) = \sum_{i=1}^{10} \delta_i^2 = \sum_{i=1}^{10} [f(x_i) - y_i]^2$$

为求 a_1, a_2, \dots, a_5 使 J 达到最小，只需利用极值的必要条件

$$\frac{\partial J}{\partial a_k} = 0 \quad (k = 1, \dots, 5)$$

得到关于 a_1, a_2, \dots, a_5 的线性方程组

$$\sum_{i=1}^{10} x_i \left[\sum_{k=1}^5 a_k x_{ik} - y_i \right] = 0, \quad (j = 1, \dots, 5)$$

记

$$R = \begin{bmatrix} x_{1,1} & \cdots & x_{5,1} \\ \vdots & \vdots & \vdots \\ x_{1,10} & \cdots & x_{5,10} \end{bmatrix}_{10 \times 5}$$

$$A = [a_1, \dots, a_5]^T, \quad Y = (y_1, \dots, y_{10})^T$$

上述线性方程组可表为

$$R^T R A = R^T Y$$

于是方程组唯一解

$$A = (R^T R)^{-1} R^T Y$$

将所收集到的数据带入计算，得到分行业、分地区的数学模型

$$f(x) = a_1 x_{1j} + a_2 x_{2j} + a_3 x_{5j} + a_4 x_{6k} + a_5 x_{10,k}$$

(3) 模型的修正

因为数据收集的原因，上述初步模型并没有考虑不同人群的失业率问题。本文将就业人群按照学历划分，因为学历的不同直接造成了就业选择的宽度。

按照经验，将上述的初步模型进行修正，得到修正后的就业模型

$$y_2 = \eta_l y_1 \quad (l = 1, 2, \dots, 7)$$

其中， $l = 1, 2, \dots, 7$ 分别代表小学以下、初中、高中、中专、大学专科、大学本科、研究生七类人群；

η_l ——不同人群的失业率系数。

假设不考虑就业者对自己岗位的期望，那么学历越高就业面就越广，因此从这个方面来讲，学历越高失业率就应该越低。但是事实并不完全是这样的，调查显示，用人单位对员工需求素质最看重的前 5 个指标是：专业知识与技艺、敬业精神、学习意愿强、沟通协调能力和基本的解决问题能力。专家表示，工作态度、敬业精神、职业道德、人际关系处理等非认知技能的缺乏是就业者就业的最大障碍。

根据实际的调查情况看来，开始找工作半年内，低学历者就业率要高于高学历者，而中等学历的人群就业率最低。这是因为低学历者对自己的就业期望不高，容易找到工作；而高学历的就要岗位多、同时需求量要大，所以造成高学历的就业者工作好找；对于中等学历的就业人群来说，因为自身就业期望比较高，与用人单位的期望不一致，就业者在选择岗位的时候产生了“高不成低不就”，造成短期内就业的困难^[5]。

根据城镇登记失业率的统计，登记失业人数最多的是低学历人群，高学历人群登记失业的最少。得到不同的就业人群的失业比例

$$\omega_i = \frac{p_i}{P_i} \quad (i = 1, 2, \dots, 7)$$

其中， p_i ——人群*i*的登记失业人数； P_i ——人群*i*就业总人数。

可以依据各个人群的登记失业率的比例来确定不同行业的失业率系数

$$\eta_i = \frac{\omega_i}{\frac{1}{7} \sum_{i=1}^7 \omega_i} = \frac{7\omega_i}{\sum_{i=1}^7 \omega_i}$$

表 6 第五次全国人口普查数据

| 受教育程度区分人群 | 就业人数 | 失业人数 |
|-----------|----------|--------|
| 小学以下 | 27406931 | 114526 |
| 初中 | 27889186 | 588991 |
| 高中 | 6035725 | 324435 |
| 中专 | 2424230 | 76648 |
| 大学专科 | 2198092 | 48250 |
| 大学本科 | 860585 | 9250 |
| 研究生 | 60140 | 319 |

依据第五次全国人口普查（2000 年）的数据，得到不同行业的失业率系数，见表 7。

表 7 各个人群失业率系数 η_i

| 各个人群 | 失业率系数 |
|---------|-------|
| 1. 小学以下 | 0.196 |
| 2. 初中 | 0.979 |
| 3. 高中 | 2.496 |
| 4. 中专 | 1.517 |
| 5. 大学专科 | 1.028 |
| 6. 大学本科 | 0.538 |
| 7. 研究生 | 0.245 |

得到修正后的模型

$$f(x) = \eta_i(a_1x_{1j} + a_2x_{2j} + a_3x_{5j} + a_4x_{6k} + a_5x_{10,k})$$

(4) 模型的例证

以奶牛类地区、农林牧渔业和大学本科生为例，收集相关的经济数据，得到一个精确的失业率模型

$$y = 0.538(0.0001x_1 - 0.0003x_2 - 0.28375x_5 - 0.0028x_6 - 0.0038x_{10})$$

5.4 问题四的解决

要使用上文所建立的失业率模型，根据国家的有关决策和规划预测 2009 年和 2010 年的就业前景，需要考虑不同政策对经济的影响和国家出台的就业政策以及在金融危机背景下的 4 万亿投资的作用。首先不考虑政策的影响，使用能以任意精度逼近的人工神经网络模型，建立 2009 年和 2010 年失业率预测模型，然后考虑不同政策对各个因素影响再建立国家决策和规划下的市盈率模型，并进行预测。

5.4.1 不考虑政策的预测

建立起五种指标数据与失业率的神经网络模型

模型输入数据为：1999~2008 年每年的 GDP、社会消费品零售总额、通货膨胀率、PPI、国家税收收入等五项指标的数据；输出数据为历年的失业率，建立的神经网络结构如图 6 所示：

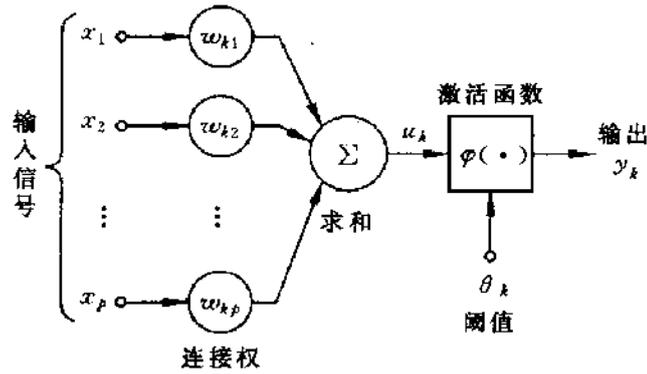


图6 人工神经网络结构

(i) 一组连接（对应于生物神经元的突触），连接强度由各连接上的权值表示，权值为正表示激活，为负表示抑制。

(ii) 一个求和单元，用于求取各输入信号的加权和（线性组合）。

(iii) 一个非线性激活函数，起非线性映射作用并将神经元输出幅度限制在一定范围内（一般限制在(0,1)或(-1,1)之间）。设定一个阈值 θ_k （或偏置 $b_k = -\theta_k$ ）。

以上作用可分别以数学式表达出来：

$$u_k = \sum_{j=1}^p w_{kj} x_j, \quad v_k = u_k - \theta_k, \quad y_k = \varphi(v_k)$$

式中， x_1, x_2, \dots, x_p ——输入信号；

$w_{k1}, w_{k2}, \dots, w_{kp}$ ——神经元 k 之权值；

u_k 为线性组合结果， θ_k 为阈值，激活函数

$$\varphi(v) = \frac{2}{1 + e^{-2v}} - 1$$

y_k 为神经元 k 的输出。

首先按照原始数据进行神经网络结构的搭建，即建立起五种指标与失业率的函数关系，各个权值系数存在建立起的各个节点中。然后利用多项式回归方法建起年份与各个指标的函数关系，考虑到多项式模型精度问题，在此采用三次多项式较适宜，从而对未来年份指标数据进行预测，得到数值。在预测过程中考虑政策的影响，采用折中的方法，对未来年份得到的指标数据进行处理，初步定为增加或减小指标百分比。选定的五种指标为：

- 1) GDP;
- 2) 社会消费品零售总额;
- 3) 通货膨胀率;
- 4) 进出口贸易总额;
- 5) 国家税收收入;

利用BP神经网络建立五个指标因子与失业率的非线性函数关系。考虑到BP网络的一个重要功能就是非线性映射的能力，这一功能非常适合于函数逼近等要求，也就是说，找出两种数据之间的关系，即输入向量P和目标向量T，建立BP网络，找出P和T之间的关系，图7是算法迭代运算进行中的情形。

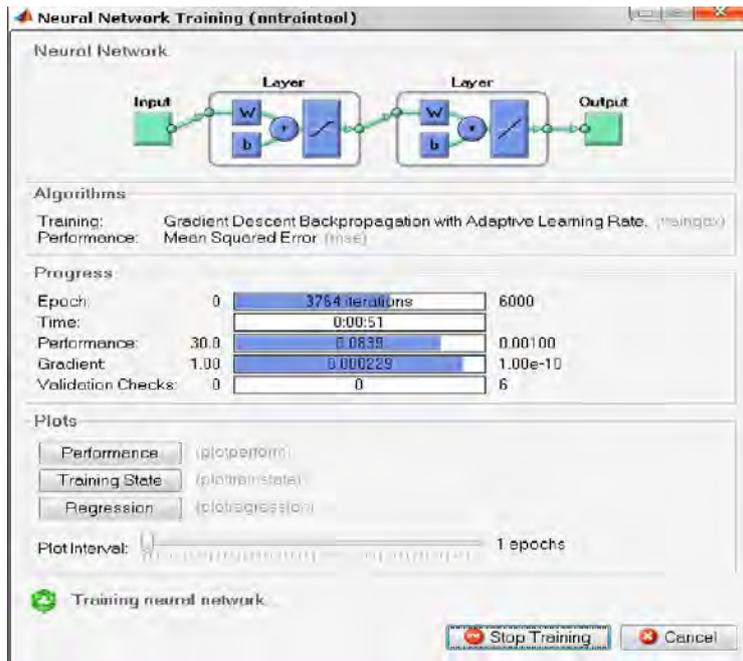


图 7 运算迭代计算中

该网络的输入层和输出层的神经元个数均为 1，根据以上的隐层设计经验公式，以及考虑本例的实际情况，解决该问题的网络的隐层神经元数目可变的 BP 网络，通过误差对比，确定最佳的隐含层神经元个数，并检验隐含层神经元个数对网络性能的影响。网络的隐含层神经元的传递函数为“tansig”，输出层神经元的传递函数为“logsig”。当神经元不同时，拟合误差运行结果见表 8。

表 8 网络误差

| 神经元个数 | 3 | 4 | 5 | 6 | 7 | 8 |
|-------|--------|--------|--------|--------|--------|--------|
| 网络误差 | 0.9967 | 1.1222 | 1.2103 | 1.0152 | 1.1236 | 0.9163 |

上表表明，在经过 6000 次训练后（训练函数采用“traingdx”），隐含层神经元为 8 的 BP 网络对函数的逼近效果最好，因为它的误差最小。所以计算采用的神经元个数为 8 个。

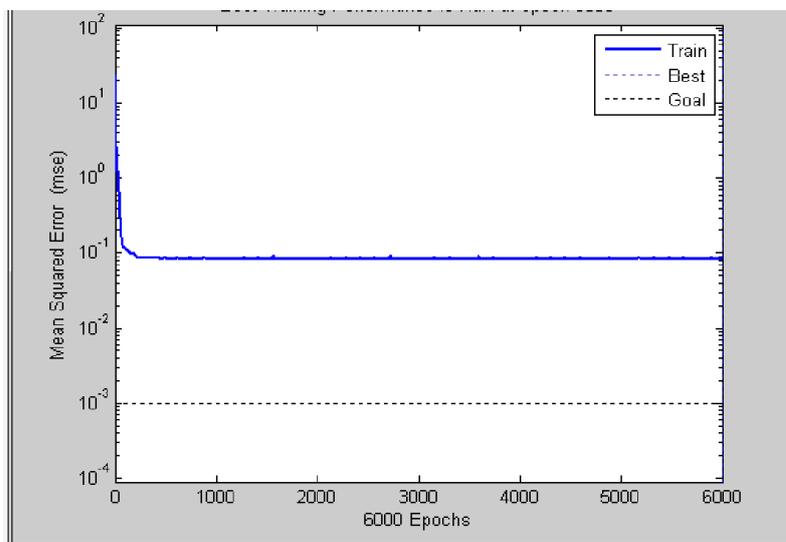


图 8 残差收敛

由图 8 可以看出，计算运行到六千代时，训练曲线残差大约为 10^{-1} ，离训练目标 10^{-3} 还有差距，在不严格情况下，基本达到要求。

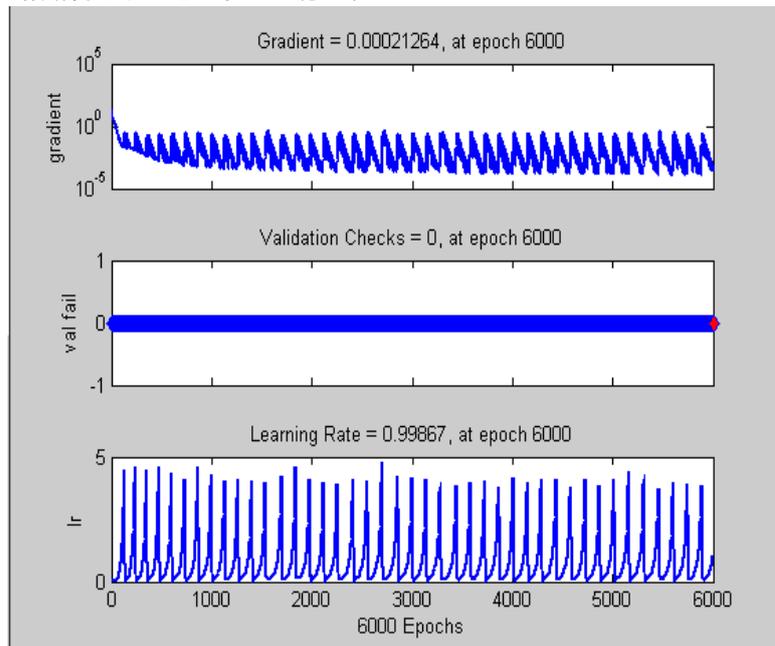


图 9 迭代过程各参数变化情况

图 9 表示采用天然 ‘traingd’ 对网络进行训练，该函数所用的学习算法是普通的梯度下降法，变化比较平稳。

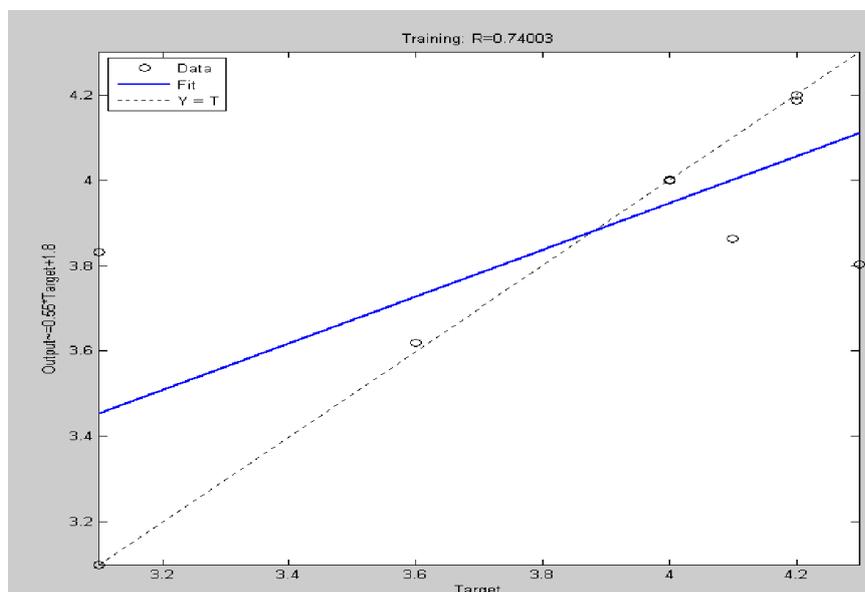


图 10 曲线拟合情况

从图 10 看出，拟合曲线能基本经过目标点，说明拟合的线性度较好，基本上满足拟合要求。利用神经网络模型得到 2009 年的预计失业率为 7.1160%，2010 年的预计失业率为 6.2719%，这个失业率对我国来说是不可以接受的，必须出台相关政策促进就业。

5.4.2 相关政策的影响

因为时间的关系，本文只讨论上述模型研究的各项经济指标的相关政策和关系到就业的决策与规划。在模型一中，提取了国内生产总值 GDP、社会消费品零售总额、通货

膨胀率、进出口贸易总额、国家税收收入五个指标作为分析要点。

(1) GDP 的分析

股市反应了经济的基本面，因此，股市走势必然与 GDP 增速存在一定的相关必然性^[1]。但是中国股市受政策的影响太过于敏感，并不能完全作为中国经济的“晴雨表”。只有当股市的市值在一个合理的范围内，才能满足中国经济“晴雨表”的作用。目前上证指数为 2950 点左右，是一个比较合理的数据，因此目前中国股市基本可以作为预测未来经济走势的一个重要参考依据。

2009 年国家将 GDP 的增长目标定为 8%，鉴于研究与 GDP 相关的政策比较困难，本文转为研究股市的政策。

国家股市政策是针对两市流通市值与银行存款制定的。国家政策降包括：调整印花税、调整存款利息、调整存款准备金、回购股票等力图把握宏观经济的走势^[4]。

(2) 社会消费品零售总额

社会消费品零售总额直接反映了国家的内需情况，在金融危机的背景下，扩大内需对拉动经济促进就业有着至关重要的作用。我国现在的基本状况是因内需不足导致三大产业发展缓慢，进而影响就业。

08 年 10 月份的 4 万亿十项投资计划，就是扩大内需的最直接的决策。十项投资计划分别为：

一、加快建设保障性安居工程。加大对廉租住房建设支持力度，加快棚户区改造，实施游牧民定居工程，扩大农村危房改造试点。

二、加快农村基础设施建设。加大农村沼气、饮水安全工程和农村公路建设力度，完善农村电网，加快南水北调等重大水利工程建设和病险水库除险加固，加强大型灌区节水改造。加大扶贫开发力度。

三、加快铁路、公路和机场等重大基础设施建设。重点建设一批客运专线、煤运通道项目和西部干线铁路，完善高速公路网，安排中西部干线机场和支线机场建设，加快城市电网改造。

四、加快医疗卫生、文化教育事业的发展。加强基层医疗卫生服务体系建设，加快中西部农村初中校舍改造，推进中西部地区特殊教育学校和乡镇综合文化站建设。

五、加强生态环境建设。加快城镇污水、垃圾处理设施建设和重点流域水污染防治，加强重点防护林和天然林资源保护工程建设，支持重点节能减排工程建设。

六、加快自主创新和结构调整。支持高技术产业化建设和产业技术进步，支持服务业发展。

七、加快地震灾区灾后重建各项工作。

八、提高城乡居民收入。提高明年粮食最低收购价格，提高农资综合直补、良种补贴、农机具补贴等标准，增加农民收入。提高低收入群体等社保对象待遇水平，增加城市和农村低保补助，继续提高企业退休人员基本养老金水平和优抚对象生活补助标准。

九、在全国所有地区、所有行业全面实施增值税转型改革，鼓励企业技术改造，减轻企业负担 1200 亿元。

十、加大金融对经济增长的支持力度。

(3) 通货膨胀率

1958 年，经济学家菲利普斯发表了《1861—1957 年英国失业和货币工资变动率之间的关系》，用近 100 年的数据验证了通货膨胀率（原名义工资率）与失业率之间的负相关关系，即低失业的年份往往有高通货膨胀，而高失业的年份往往有低通货膨胀。

要想控制通货膨胀，可通过两种手段，即财政政策手段和货币政策手段。譬如，政府可以通过增税、减少政府购买、增加转移支付等财政政策手段和提高存贷款利率、出

售国债、提高银行的准备金率等货币政策手段来降低通货膨胀水平。

(4) 进出口贸易总额

进出口贸易是我国国民经济的重要组成部分，而目前的出口经济占据国民生产总值的大部分，外贸经济对我国劳动力就业起到了重要作用。

影响进出口贸易的最直接的政策就是贸易税收政策，贸易税收政策主要包括出口退税和关税两部分。

(5) 国家税收收入

税收的作用主要表现在筹集财政资金的作用、调节经济的作用、反映与监督作用。税收是国家财政收入的主要来源，直接影响到国家的财政收入。

一、财政可通过投资性的支出，直接增加就业岗位。政府通过财政拨款兴修水利、进行基础设施等工程、项目的投资，从而创造大量的就业机会，缓解失业压力。

二、政府通过采购各种商品和劳务的购买性支出，增加对社会商品和劳务需求，从而刺激企业的生产，促进企业的投资行为，间接增加就业机会。

三、政府通过各种转移性支出，增加社会成员的收入，这些收入按一定的比例转化成消费和储蓄，从而增加对社会商品和劳务的需求，间接增加就业岗位。

四、政府在再就业培训、职业介绍服务、产业结构转化等方面增加财政支出，如通过财政补贴、税式支出等手段，可以缩短工人寻找工作的时间，增强工人在不同岗位的就业适应能力，减少在流动过程中产生的摩擦性失业以及产业结构转换中造成的结构性失业^[5]。

5.4.3 政策影响经济指标的作用系数

建立国家宏观调控政策对经济指标的影响系数矩阵

$$B = [b_1 b_2 b_3 b_4 b_5]$$

其中， b_1 ——国家决策和规划对 GDP 指标的影响系数；

b_2 ——拉动内需政策对社会消费品零售总额的影响系数；

b_3 ——财政政策和货币政策对通货膨胀率的影响系数；

b_4 ——贸易税收政策对贸易进出口总额的影响系数；

b_5 ——国家政策和规划对税收收入总额的影响系数。

令：

$$C = [c_1 c_2 c_3 c_4 c_5]$$

其中

$$c_1 = a_1 x_1, c_2 = a_2 x_2, c_3 = a_3 x_5, c_4 = a_4 x_6, c_5 = a_5 x_{10}$$

失业率与各项经济指标的数学模型可以表示为

$$y = BC + \beta_0$$

即

$$y = b_1 \beta_1 x_1 + b_2 \beta_2 x_2 + b_3 \beta_3 x_5 + b_4 \beta_4 x_6 + b_5 \beta_5 x_{10} + \beta_0$$

根据相关研究，本文将政策影响的系数矩阵进行赋值，得到赋值后的影响系数矩阵各个元素为

$$c_1 = 1, c_2 = 1.1, c_3 = 1, c_4 = 1.1, c_5 = 1.1$$

得到在相关政策的影响下，2009 年的城镇登记失业率为 4.0825%，2010 年的城镇登记失业率为 3.6256%。

对比政策影响前后的失业率预计值，在没有政策影响下的全国城镇登记失业率偏高，这样不但会影响社会的繁荣发展而且会影响到政局的稳定。但是因为使用宏观调控的手段，制定相应的政策就可以实现将城镇登记失业率控制在一个合理的范围之类，使

经济社会又好又快的发展，早日实现共同富裕的伟大目标。

5.5 对提高就业的几点建议

就业问题关系人民群众的切身利益，关系改革发展稳定的大局，关系全面建设小康社会宏伟目标的实现。群众的利益是具体的，就业是民生之本。只有通过就业，一个人才能获得收入、安居乐业、实现价值，社会才能不断发展和进步。

根据本文所建立的模型可以看出，城镇登记失业率与国内生产总值 GDP、社会消费品零售总额、通货膨胀率、进出口贸易总额、国家税收收入等 5 个经济指标的关联度最大。而在这些经济指标中，与通货膨胀率的关联度又最大，低的通货膨胀率会引起高的失业率，高的通货膨胀率会降低失业率，国家在把握宏观经济的时候需要在通货膨胀和失业率之间进行博弈。

依据本文的模型，要降低城镇登记失业率还可通过提高社会消费品零售总额、提高进出口贸易总额、增加国家税收总额，也就是通过扩大内需、增加出口、增加财政支出的途径提高就业率。

具体建立如下：

一、扩大内需，促进消费

增加社会消费品零售总额有利于降低我国城镇登记失业率。社会消费品零售总额是各种经济类型的批发零售贸易业、餐饮业、制造业和其他行业对城乡居民和社会集团的消费品零售额和农民对非农业居民零售额的总和。它反映一定时期内人民物质文化生活水平的提高情况，反映社会商品购买力的实现程度，以及零售，市场的规模状况。消费的稳步增长对我国经济增长起到了重要的支撑和拉动作用^[6]。

1. 当前促进消费增长的政策

为抵御国际金融危机的影响，促进消费增长，提高消费对经济增长的带动作用，从去年以来中央和各地方先后出台了一系列的政策措施。

这些政策措施大体可以分为 3 类，一是直接提高居民特别是低收入居民收入的政策；二是提高农村居民消费能力和意愿的政策，如家电下乡、汽车下乡政策；三是稳定大宗和热点消费市场的政策，如有关汽车和住房消费的政策。从各方面的情况来看，这些政策措施达到了预期的效果，对今年以来消费的平稳增长起到了不可或缺的重要作用。

当前消费中也还存在一些问题和矛盾，需要逐步加以解决。首先，是消费增长还不均衡，在食品、服装、家具和汽车类商品消费增长较快的同时，一些领域的消费增长比较缓慢。其次，是消费者信心尚需进一步提振。再者，当前的消费增长相当程度是依靠短期政策的刺激和支撑，为保持消费的长期可持续增长，还需要在调整经济结构和完善各项制度方面下大功夫。

2. 进一步促进消费增长的政策建议

(1) 把扩大就业作为稳定居民收入促进消费增长的重要途径。继续增加各级财政对扩大就业的投入，落实好已出台的支持大学生和农民工创业、就业和培训的各项政策措施，创造更多的与改善公共管理、提高公共服务水平和环境保护等相关的公共部门就业岗位。建议对国家投资建设项目进行就业效应评估，在同等条件下对就业带动效应大的项目优先批准实施。

(2) 加快国民收入分配格局调整，培育壮大中等收入阶层，使之成为消费稳定增长的主力。

(3) 加快完善社会保障体系，稳定并改善居民的消费预期。继续通过各种方式加快完善社会保障体系建设，扩大基本养老、基本医疗、失业、工伤等社会保险覆盖面，做实个人账户，不断提高统筹层次，以解除消费者的后顾之忧，稳定消费预期，提高消费者信心。

(4)合理发展消费信贷，促进潜在消费能力释放。消费信贷能够优化居民跨期消费选择，促进潜在消费能力提前释放。鼓励商业银行和其他金融机构在有效控制风险的前提下，结合已经出台的促进消费的政策，开发多样化的消费信贷产品。如可参照房地产消费优惠信贷利率，在一定期限内（暂定1年）对居民购买1.6升及以下排量国产汽车的消费信贷提供优惠利率。探索加快发展居民一般用途个人消费贷款，通过降低准入门槛和实行利率优惠等措施扩大覆盖范围。

(5)改善管理，着力发展文化娱乐旅游等新兴消费。消费的持续稳定增长还需要消费结构的升级转换。文化、娱乐、旅游等服务消费在我国还有很大的增长空间，要改革完善现有管理方式，扩大市场准入和服务供给，加强市场监管，积极引导消费者扩大服务消费，并相应在休假安排、治安等制度方面予以配合。

(6)进一步改善消费环境，让消费者方便放心消费。加强商业网点和电子商务建设，减少消费者受时间和空间的制约。进一步改善农村和落后地区的基础设施建设，为使用家电、通讯等现代消费品创造条件。严厉打击欺诈和假冒伪劣，提高对消费者的财务、健康和人身安全的保障，让消费者能够放心消费。进一步清理限制消费的行政性措施，保障消费者自由选择消费的权利。

(7)密切关注一些市场的价格波动情况，防止价格非正常波动扰乱消费。农产品市场价格易发生较大波动，对消费的稳定有一定的冲击。要建立农产品价格稳定机制，通过生产补贴、储备、进出口调节等方式防止国内农产品价格的大起大落。房地产市场的价格波动影响也较大，对购房家庭的消费影响不可低估。应采取积极措施防止房地产价格泡沫，严格房地产市场秩序管理，严惩虚假交易和哄抬房价行为，使住房消费能够平稳有序增长，并与其他消费增长相协调。

二、推动进出口贸易的发展

进出口贸易是我国国民经济的重要组成部分，对我国就业有很大的影响。进出口贸易是保证国民经济各部门平衡发展和社会再生产顺利进行的重要手段，是节约生产劳动、提高经济效益的可靠途径，是积聚资金、吸收先进技术的有效方式。对推动进出口贸易的政策有以下三点建议：

1. 鼓励双向投资，提高利用外资的质量，积极引进先进技术和管理经验，大胆学习和借鉴人类创造的一切文明成果。
2. 完善涉外经济法律法规，实施保护知识产权的国家战略，努力创造公平的竞争环境，依法保护各类企业和个人的合法利益。
3. 同世界各国开展各领域的交流与合作，开创互利共赢的新局面^[9]。

三、现行促进就业的税收政策的不足与完善建议

符合国情的税收政策可以提高我国城镇就业人口，减少城镇登记失业率。我国现行的促进就业的税收政策有很多，主要是针对各行业，各人群的是税收优惠。但现行税收政策在促进就业方面还有很多不足，需要完善。

1. 现行促进就业的税收政策的不足

(1)并非所有的税收优惠政策对经济的发展都有促进作用，不合理、不科学的税收优惠政策将会抑制经济的发展。

(2)税收优惠政策的优惠内容少。我国促进就业的税收优惠政策主要体现在营业税、所得税里面，而在我国的主体税种增值税里面，相关的优惠政策显得更少。

(3)税收优惠政策的适用对象范围狭窄。一是我国促进就业的税收优惠政策主要针对下岗失业人员，对城镇的富余劳动力、大中专毕业生涉及较少。二是优惠政策中在企业的类型和经营行业门槛上有待放宽。

(4)税收优惠政策的时间范围较短。优惠政策有的规定了截止日期，没有着眼于长

远目标，缺乏稳定性。

(4) 制定的税收优惠政策难以得到切实执行。首先，对于国家规定的有关就业的税收优惠，各地做法不一致。其次，有的地方税务机关不执行税收优惠。

2. 对现行促进就业的税收政策的完善建议

(1) 合理适度增加税收优惠内容。一是优化所得税的减免方法，可以适当提高企业所得税的定额减免额度。对于新办的中小和民营企业，应当给与扶持。二是适当增加优惠税种。增值税的税收优惠对企业有很大的调节作用，适当情况下应给予某些企业一些增值税减免。

(2) 扩大对就业优惠的适用。对于其他类型企业(如制造企业、修理修配企业等)，其他性质企业(如外商投资企业)，只要安置下岗失业人员人数达 30%以上的，均可享受优惠政策，以体现政策的公平和调动各种企业安置事业人员的积极性。

(3) 适当延长优惠期限。未来十几年，我国劳动力的供需矛盾依然十分突出，就业问题将长期作为政府的工作重心之一。促进劳动就业工作是一项持久性工作，不能短期操作。因此，应该允许纳税人适当延长优惠期限。

(4) 统一优惠规定，简化审批手续。对现行就业税收优惠政策进行整理归并，对新、老企业的界定要明确，安置下岗失业人员的优惠政策要统一，使税收促进就业的优惠政策具有系统性和透明度，从而改变目前规定过多、过繁、难于落实的状况；同时，审批手续应电子化、报送资料应尽量简化。

(5) 加强对上述税收政策的管理，提高政策的实施效果。征收管理上的薄弱环节不但会直接削弱税收优惠政策的实施效果，同时还会造成税收收入的流失。

国家的决策和规划对我国未来的经济和社会发展起着重要的指导作用，这需要有有力的政策做支持。如果没有正确的政策引导和约束，国家的经济将陷入混乱，人民无法安居乐业，社会主义现代化建设目标就没法完成。我国正处于社会主义初级阶段，市场经济体系并不完善，因此政策的制定相当的重要。希望政策制定机构和人员在制定相关政策的时候，一定要从长远出发、考虑大局，制定出有利于我国社会经济建设的政策，为构建社会主义和谐社会提供有力的科学支撑。

6 模型的评价

6.1 模型优点

1. 数据来源可靠。所收集的数据来自“和讯网”、“中国国家统计局”的官方网站、“华尔街日报”网、“国务院发展研究中心信息网”以及《中国统计年鉴 2008》，数据的可靠性高。

2. 正确选择了与失业率相关的 12 个经济指标。通过对失业率的分析，从各项经济指标中选择了 12 个关联性大的经济指标，作为模型建立的基础，从最后的模型来看，得到的结果与经济学相符合。

3. 地区分类比较合理。没有按照地域将各个省(市)进行分类，而是使用波士顿矩阵将各个省(市)按照经济实力和经济增速为标准，将各个地区划分为四类。

4. 政策对就业影响分析有力。先利用 BP 神经网络模型预测没有政府政策影响情况下的 2009 年和 2010 年的失业率分别为 7.1160%，和 6.2719%，然后分析了各项经济指标在政策影响下的变化情况，再次预测了 2009 年和 2010 年的失业率 4.0825%和 3.6256%，由此可以看出政策对降低失业率的重要性。

5. 给出的建议针对性强。本文就当前的扩大内需、增加出口、增加财政支出的相关政策进行了系统分析并提出了具体的建议。

6.2 模型的缺点

1. 数据量不够多。十年的数据量做为分析的依据，不够准确的反映各项指标与失业率的关系，应该更多的收集数据。

2. 政策的分析不够系统。各项政策的出台不光对一个经济指标产生影响，各项经济指标之间都存在着联系，一个政策会影响多个经济指标，所以有必要将政策对经济指标的影响做系统的分析。

参考文献

- [1] 万海远，我国自然失业率的测量与解析，当代财经，第 292 期：5-12，2008。
- [2] 陈石，贺晓，中国非加速通货膨胀失业率研究，首都经济贸易学报，2005 年第 6 期，75-79，2005。
- [3] 黑太阳的日志，中国地区划分，<http://blog.163.com/hty-1023/blog/static/174399962007291021150/>，2009 年 9 月 22 日。
- [4] 许业友，我国通货膨胀的收入分配效应分析，价值工程，2009 年第 8 期，15-18。
- [5] 曲顺兰，路春城，税收对就业的影响及其政策选择，税务研究，2004 年底 12 期，2004 年。
- [6] 蔡昉，都阳，高文书，就业弹性、自然失业和宏观经济政策——为什么经济增长没有带来显形就业？，经济研究，2004 年第 9 期，35-40，2004。
- [7] 谢金星，薛毅编著，优化建模与 LINDO/LINGO 软件，北京：清华大学出版社，2005。
- [8] 韩中庚编著，数学建模方法及其应用，北京：高等教育出版社，2005。
- [9] 姜启源，数学模型（第二版），北京：高等教育出版社，1993。
- [10] 中国统计年鉴，2008。

附录

- 附件一：模型一的计算程序及数据
- 附件二：模型二的计算程序
- 附件三：模型三的计算程序及数据
- 附件四：模型四的计算程序