

# 全国第六届研究生数学建模竞赛



## 题 目 就业问题的数学建模分析

### 摘 要

就业问题分析的难点在于涉及面广、影响因素多、数据量大,因此本文分宏观、中观、微观三个层次,通过聚类分析、主成分分析和相关分析,从影响就业的因素中归纳出 6 个关键性的就业因素影响指标。依据 6 个指标与城镇登记失业率的高度相关性,建立了城镇登记失业率由 6 个指标线性组合构成的预测模型。考虑到我国城镇登记失业率定义的局限性,按照国际标准对失业率进行重新建模,并以 2008 年的失业率数据对比分析了国际国内两种失业率定义的合理性。

考虑到地区和行业的差异,以 2001~2007 年的数据分别分地区和分行业建立了失业率预测模型和从业人数预测模型,提出了地区内部失业率、预期岗位裕量、行业内部失业率、国家投资回报指数的概念,并以 2008 年的数据验证了模型的准确性。依据分地区精确预测模型和分行业精确预测模型分别对 2009 年和 2010 年上半年的失业率和从业人数进行了预测,分地区预测失业率的结果:2009 年第一季度为 4.38%,第二季度为 4.82%,第三季度为 4.33%,第四季度为 4.91%,2010 年第一季度为 4.57%,第二季度为 4.92%;分行业预测的结果为:2009 年第一季度为 4.32%,第二季度为 4.83%,第三季度为 4.42%,第四季度为 4.09%,2010 年第一季度为 4.85%,第二季度为 4.70%。

依据分地区精确预测模型和分行业精确预测模型得到的预期岗位裕量和国家投资回报指数两个指标,对行业 and 地区进行了划分,总结出从促进就业的角度对政府单位投资反应最灵敏的十大行业和十大地区,并与国家十大产业振兴计划对比分析,验证了国家政策的科学合理性和模型的准确有效性。

关键词:主成分分析 聚类分析 预测模型 预期岗位裕量 投资回报指数

参赛队号 9000221

队员姓名 沈新民 李杰 曹建平

参赛密码 \_\_\_\_\_

(由组委会填写)

## 一、问题重述与分析

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，就业是社会、国民经济中极其重要的问题。我国采用城镇登记失业率，是指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。

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；从微观层面看，劳动者的受教育程度、性别、年龄的差异也会影响就业。

就我国的现实情况，2008 年我国经济社会经受了历史罕见的考验，就业形势十分严峻。据此，中央政府从 08 年 10 月开始实施了 40000 亿元的投资计划，09 年我国在就业方面的目标是城镇新增就业 900 万人以上，城镇登记失业率控制在 4.6% 以内。

就业问题涉及的因素是纷繁复杂的，要在众多的数据中寻找出影响就业的关键因素并用来预测未来的就业前景，对因素进行分类是必需的。本文将影响就业的因素按照宏观、中观、微观三个层面划分，并通过相关分析、聚类分析、主成分分析得到优化重组后的就业影响因子，再用预测的就业影响因子预测就业人数和失业率。这种把就业问题作为一个系统来研究的方法比直接用就业人数和失业率的历史数据预测更加科学合理。同时本文对失业率和就业人数进行分地区和分行业的精确建模，为促进就业提供合理建议。

预测的目的是为了指导国家投资的合理使用，通过对预期岗位裕量和国家投资回报指数分析，得出国家投资对提高就业率最有效的十大行业和十大地区，并给出促进就业的具体建议，将国家的投资用到最合理、最有效的地方，实现就业人数的最大化和失业率的最小化。

## 二、基本假设与符号说明

### 2.1 基本假设

- 假设一：所有的因素都可以用某个指标来表示，因此在下面的叙述中都约定量化的指标来表示各种因素。
- 假设二：每年的数据都可以视为是该年四个季度数据按照某种计算方法得到的，因此可以将按年统计的历史数据插值为以季度为考察单元的数据，并都是以 2001 年第一季度为季度起始计数 1。
- 假设三：由于每个因素都对应很多指标，分析中不可能面面俱到，因此选取最能代表该因素特征的指标来表示，并假设此指标可以完全体现该因素，这是下面很多因素指标选取的依据。
- 假设四：主成分分析的累计贡献率达到 80% 为比较满意，90% 为高度满意，并认为高度满意的情况下主成分分量可以完全的代表原变量。
- 假设五：对于指数、比率等的度量性年度数据，将其视为该年第四季度的季度数据；对于累计性年度数据，同一季度的季度增长率应以年增长率为中心上下波动，且偏离（方差）不大。
- 假设六：未来的数据可以由历史数据准确的预测，因此本文中采用 2001~2007 年的数据进行预测。

## 2.2 符号说明

GPC	国民经济核算指数
JPC	价格指数
CPC	财政税收指数
BPC	金融保险指数
SPC	三大产业指数
MACRO	宏观指标
HPC	分行业指标
DPC	分地区指标
MESO	中观指标
MICRO	微观指标

## 三、问题 1：影响就业的指标的分析

### 3.1 影响就业的指标的分类

从经济学的角度来说，影响就业（或者失业）的指标很多，其分类也是多种多样的，本题中是要在大量数据中提取有用的信息，属于数据挖掘范畴，因此本文首先对影响就业的因素进行了详细的分析。本文采用宏观、中观、微观三层分类法，对影响就业的指标通过聚类分析、主成分分析、相关分析进行优化重组，使用层次分析法逐层计算。

从宏观层面来看，宏观经济学是以国民经济总过程的活动为研究对象，依据这一概念，这里的宏观层面指标本文主要分析国民经济核算、价格指数、财政税收、金融保险、三大产业五大类。

从中观层面来看，中观经济学是以某一部门、某一集团或某一地区的经济活动为研究对象，因此这里的中观层面指标本文主要进行分地区、分行业讨论。

从微观层面来看，微观经济学是以单个经济单位(单个的生产者、单个的消费者、单个市场的经济活动)作为研究对象，因此微观层面指标本文主要分析劳动者的受教育程度、性别、年龄三个方面<sup>[1,2,3]</sup>。

由于历史数据多是以年为单位给出的，近期数据多是以季度或者月份给出的，因此为了实现数据的统一，依据假设二，每年的数据都可以视为是该年四个季度数据按照某种计算方法得到的，本文对按年给出的数据先进行插值处理，并将考察的数据定位在 2001~2007 年，以 2001 年第一季度为季度起始计数 1，通过 7 年 28 个季度的数据寻找规律并对未来进行预测。

### 3.2 宏观层面指标分析

宏观层面指标可以分为国民经济核算、价格指数、财政税收、金融保险、三大产业五大类，每一类又包含了相当多的内容，针对每一类本文依据假设三选取指标，通过主成分分析将其逐步优化得到宏观指标 MACRO。

#### 3.2.1 国民经济核算

国民经济核算又分为国内生产总值、消费、投资、对外经济贸易和国内生产总值指数五个方面，依据假设三，选取国内生产总值增长率、居民消费指数、固定资产投资增长率、进出口贸易总额、人均国内生产总值作为对应的指标，依据国家统计年鉴中 2001~2007 年的统计数据<sup>[1,2]</sup>，按照如下的步骤和流程图 3.2 将其转化成按季度的数据。

为了更好的进行预测，需要季度指标，但收集到的数据大部分是年度数据。在保证一定的准确度的前提下，需要年度数据映射到季度中去。通过对各个指标

的分析，可以将指标分为两类。一类是指数、比率等的度量性数据，比如居民消费指数；另一类是累积性的数据，比如年度生产总值。对于前一类数据，为了处理的方便，依据假设五将年度指标看做是当年第四季度指标。这样，对于时间序列的趋势可以进行拟合分析。

以国内生产总值增长率为例，其原始数据如表 3.1 所示。

表 3.1 国内生产总值增长率原始数据<sup>[2,3]</sup>（单位：%）

时间	2001	2002	2003	2004	2005	2006	2007
国内生产总值增长率	8.3	9.1	10.0	10.1	10.4	11.6	13.0

分别采用线性函数、指数函数、二次函数和 S 函数进行数据拟合，得到的对比结果如表 3.2 所示。

表 3.2 四种不同函数下拟合结果示意图

模型	拟合方程	准确性度量		
		MAPE	MAD	MSD
线性方程	$Y_t = 7.57 + 0.70 \times t$	2.80	0.30	0.13
二次方程	$Y_t = 8.27 + 0.23 \times t + 0.06 \times t^2$	2.55	0.26	0.09
指数方程	$Y_t = 7.85 \times (1.07^t)$	2.60	0.27	0.10
S-曲线方程	$Y_t = (100) / (11.121 - 0.232 \times (1.47^t))$	3.41	0.31	0.18

其中 MAPE、MAD、MSD 分别表示绝对百分数误差均数、绝对差异均数、平方差异均数，其值越小表示拟合得越好。由表 3.2 可知，采用二次方程的拟合结果比较好，拟合的结果如图 3.1 所示。

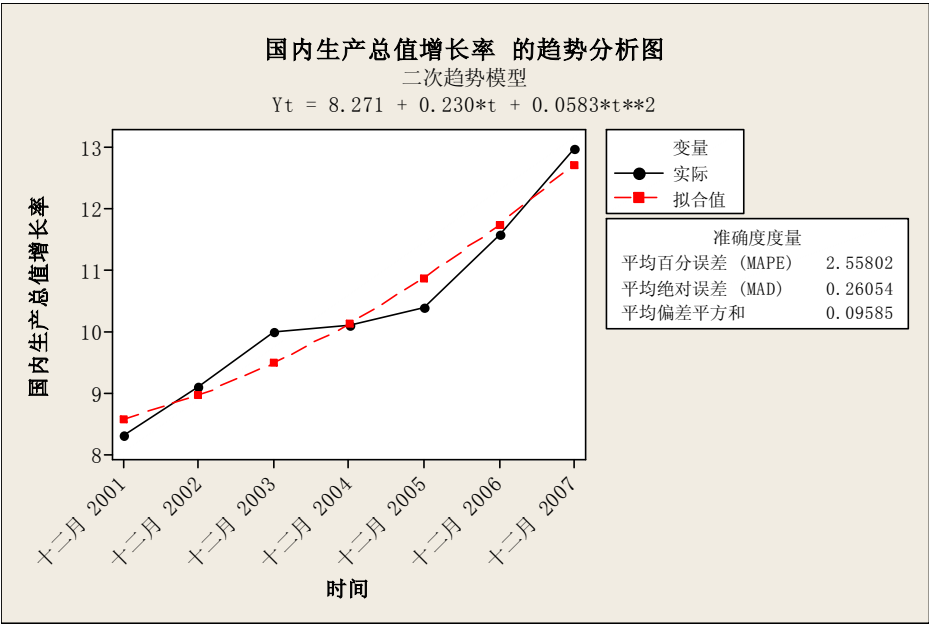


图 3.1 二次方程的拟合结果示意图

仿照上述过程就可以得到其他季度的数据，处理过的数据见附表 1。

对于第二类累计性指标处理起来有一定的复杂度，由已知的年度总数据求四个季度数据相当于由一个方程求四个未知数。依据假设五，利用非线性规划对季度数据进行优化，优化对象为季度同期增长率，优化方程如下所示：

目标函数为：

$$f(R_i) = \text{mean}(R_i) - R y_i + \lambda \text{std}(R_i) \quad (3.1)$$

其中：  $R_i = [r_{i1} \ r_{i2} \ r_{i3} \ r_{i4}]^T$ ，  $r_{ij}$  为第  $i$  年  $j$  季度的同期增长率；

$Ry_i$  为第  $i$  年年增长率；

$\text{mean}(\cdot)$  平均值函数；  $\text{std}(\cdot)$  标准差；  $\lambda$  适应参数，随迭代次数改变。  
约束条件为：

$$\begin{cases} D_{i-1} * R_i = D_i \\ 0 < r_{ij} < 1 \end{cases} \quad (3.2)$$

其中：  $D_i = [d_{i1}, d_{i2}, d_{i3}, d_{i4}]$ ，  $d_{ij}$  为第  $i$  年  $j$  季度的指标。

这样就可以利用最优化求取最优值  $R_i$  使目标函数最小，进而得到各季度数据。

归纳上面的过程，年度数据转化为季度数据的流程如图 3.2 所示。

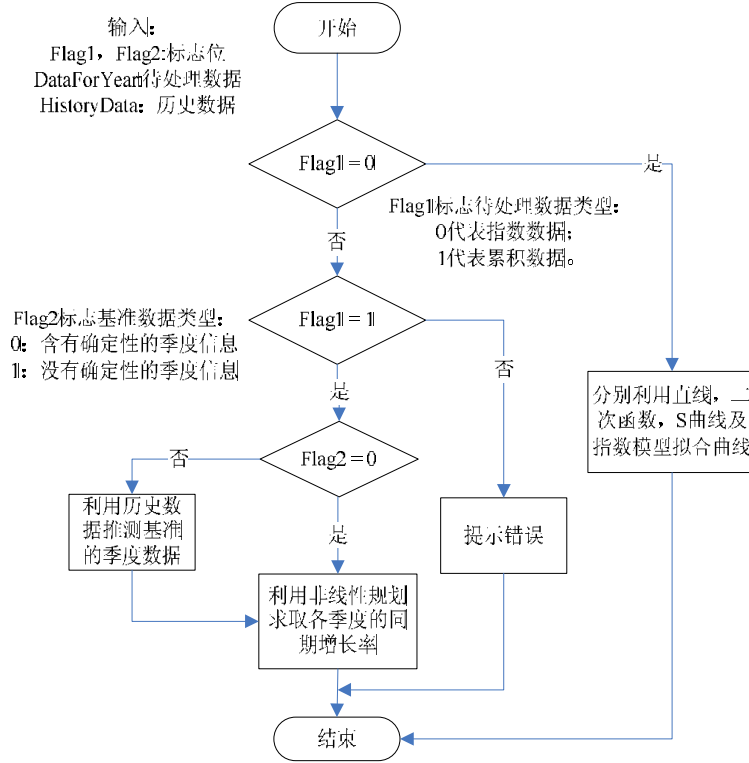


图 3.2 年度数据转化成季度数据的计算流程图

按照如图 3.2 所示的年度数据转化成季度数据的流程图，可以得到 2001～2007 年按季度的国民经济核算统计数据见附表 1。

按照统计学知识进行主成分分析<sup>[4]</sup>，主成分分析的思路如下所述。

主成分分析是把原来多个变量化为少数几个综合指标的一种统计分析方法，从数学角度来看，这是一种降维处理技术。假定有  $n$  个样本，每个样本共有  $p$  个变量描述，这样就构成了一个  $n \times p$  阶的数据矩阵：

$$X = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix} \quad (3.3)$$

如果记原来的变量指标为  $x_1, x_2, \dots, x_p$ ，它们的综合指标即新变量指标为  $z_1, z_2, \dots, z_m (m \leq p)$ 。则

在(3.4)式中, 系数 $l_{ij}$ 由下列原则来决定:

- 这样决定的新变量指标  $z_1, z_2, \dots, z_m$  分别称为原变量指标  $x_1, x_2, \dots, x_p$  的第 1, 2, ..., 第  $m$  主成分。其中,  $z_1$  在总方差中占的比例最大,  $z_2, z_3, \dots, z_m$  的比例依次递减。在实际问题的分析中, 常挑选前几个最大的主成分, 这样既减少变量的数目, 又没有丢失原变量的信息, 简化了变量之间的关系。

主成分分析计算步骤归纳如下:

- $$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1p} \\ r_{21} & r_{22} & \cdots & r_{2p} \\ ..... \\ r_{p1} & r_{p2} & \cdots & r_{pp} \end{bmatrix} \quad (3.5)$$

其中  $r_{ij}$  ( $i, j=1, 2, \dots, p$ ) 为原来变量  $X_i$  与  $X_j$  的相关系数, 其计算公式为

$$r_{ij} = \frac{\sum_{k=1}^n (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j)}{\sqrt{\sum_{k=1}^n (x_{ki} - \bar{x}_i)^2 \sum_{k=1}^n (x_{kj} - \bar{x}_j)^2}} \quad (3.6)$$

因为  $R$  是实对称矩阵(即  $r_{ij} = r_{ji}$ ), 所以只计算其上三角或下三角元素即可。

## (2) 计算特征值与特征向量

首先解特征方程  $|\lambda I - R| = 0$  求出特征值  $\lambda_i$  ( $i=1, 2, \dots, p$ ), 并使其按大小顺序排列, 即  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ ; 然后分别求出对应于特征值  $\lambda_i$  的特征向量  $e_i$  ( $i=1, 2, \dots, p$ )。

## (3) 计算各主成分贡献率及累计贡献率

主成分  $Z_i$  贡献率  $T_i$ :

$$T_i = \lambda_i / \sum_{k=1}^p \lambda_k, (i=1, 2, \dots, p) \quad (3.7)$$

累计贡献率  $S_i$ :

$$S_i = \sum_{k=1}^m \lambda_k / \sum_{k=1}^p \lambda_k \quad (3.8)$$

依据统计学基本原理<sup>[3]</sup>, 取累计贡献率达到 80% 的特征值  $\lambda_1, \lambda_2, \dots, \lambda_m$  所对应的第一, 第二, ……, 第  $m$  ( $m \leq p$ ) 个主成分时是比较满意, 累计贡献率达到 90% 时为高度满意。

## (4) 计算主成分载荷

$$l_{ij} = p(z_i, x_j) = \sqrt{\lambda_i} e_{ij} (i, j=1, 2, \dots, p) \quad (3.9)$$

## (5) 计算主成分综合指标:

$$\begin{bmatrix} z_1 \\ z_2 \\ \dots \\ z_m \end{bmatrix} = \begin{bmatrix} l_{11} & l_{12} & \dots & l_{1p} \\ l_{21} & l_{22} & \dots & l_{2p} \\ \dots & \dots & \dots & \dots \\ l_{m1} & l_{m2} & \dots & l_{mp} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_p \end{bmatrix} \quad (3.10)$$

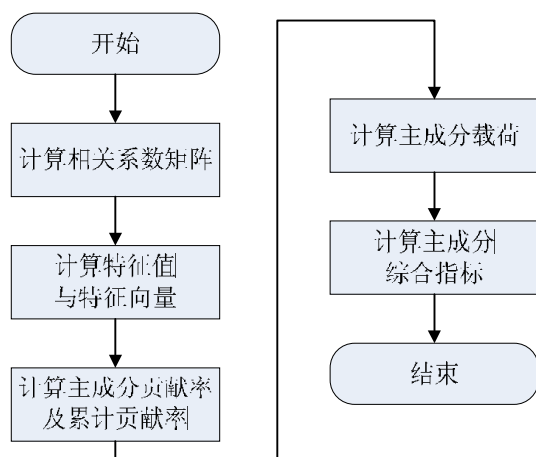


图 3.3 主成分分析的流程图

按照图 3.3 的流程示意图进行主成分分析，得出的主成分分析结果如表 3.3 所示，主成分与原参数之间的关系如表 3.4 所示。

表 3.3 国民经济核算统计数据主成分分析结果示意图

主成分	PC1	PC2	PC3	PC4	PC5
特征值	3.60	0.90	0.47	0.02	0.01
贡献率	0.72	0.18	0.08	0.01	0.01
累积贡献率	0.72	0.90	0.98	0.99	1

表 3.4 国民经济核算统计数据各主成分分量与原参数之间关系示意图

主成分	PC1	PC2	PC3	PC4	PC5
国内生产总值增长率	0.52	-0.11	-0.17	-0.81	0.19
居民消费指数	0.51	-0.11	-0.33	0.23	-0.75
固定资产投资增长率	0.41	-0.12	0.90	0.07	-0.07
进出口贸易总额	0.52	-0.06	-0.23	0.53	0.62
人均国内生产总值	0.19	0.98	0.04	-0.02	-0.03

依据上面主成分分析基本原理，累计贡献率达到 80%才能认为比较满意，本题中第一主成分 PC1 的贡献率为 72%，第一主成分 PC1 和第二主成分 PC2 的累计贡献率达到 90%，因此 PC1 和 PC2 完全可以代表国民经济核算指数，将国民经济核算指数用 GPC1 和 GPC2 表示，由表 3.3 可知

$$\text{GPC1} = \text{PC1} = 0.52 \times \text{国内生产总值增长率} + 0.51 \times \text{居民消费指数} + 0.41 \times \text{固定资产投资增长率} + 0.52 \times \text{进出口贸易总额} + 0.19 \times \text{人均国内生产总值} \quad (3.11)$$

$$\text{GPC2} = \text{PC2} = -0.11 \times \text{国内生产总值增长率} - 0.11 \times \text{居民消费指数} - 0.12 \times \text{固定资产投资增长率} - 0.06 \times \text{进出口贸易总额} - 0.98 \times \text{人均国内生产总值} \quad (3.12)$$

### 3.2.2 价格指数

价格指数又分为消费价格、商品零售价格、工业产品价格、原材料价格、投资价格五个方面，依据假设三，选取居民消费价格指数、商品零售价格指数、工业品出厂价格指数、原材料燃料购进价格指数、固定资产投资价格指数作为代表指标。依据国家统计局年鉴中 2001~2007 年的统计数据<sup>[3]</sup>，按照图 3.1 的年度数据转化成季度数据的流程图，得到 2001~2007 年按季度的价格指数数据见附表 2。

按照图 3.3 的流程示意图主成分分析得到的结果如表 3.5 所示，主成分分量与原参数之间的关系如表 3.6 所示。



表 3.5 价格指数统计数据主成分分析结果示意表

变量	PC1	PC2	PC3	PC4	PC5	PC6
特征值	5.30	0.49	0.15	0.05	0.01	0.00
贡献率	0.88	0.08	0.03	0.00	0.00	0.00
累积贡献率	0.88	0.96	0.99	0.99	1.00	1.00

表 3.6 价格指数各主成分分量与原参数之间关系示意表

变量	PC1	PC2	PC3	PC4	PC5	PC6
城市居民消费价格指数	0.41	0.41	0.26	0.49	-0.29	0.52
农村居民消费价格指数	0.42	0.35	-0.03	0.29	0.55	-0.56
商品零售价格指数	0.41	0.30	0.31	-0.75	-0.24	-0.15
工业品出厂价格指数	0.40	-0.52	0.23	-0.15	0.55	0.43
原材料燃料购进指数	0.39	-0.59	0.11	0.27	-0.49	-0.42
固定资产投资价格指数	0.41	0.00	-0.88	-0.13	-0.10	0.18

依据上面主成分分析基本原理，累计贡献率达到 80%即认为比较满意，本题中第一主成分 PC1 的贡献率就已经达到 88%，因此 PC1 就完全可以代表价格指数，将价格指数用 JPC 表示，由表 3.6 可知

$$\begin{aligned} \text{JPC} = \text{PC1} = & 0.411 \times \text{城市居民消费价格指数} + 0.42 \times \text{农村居民消费价格指数} + 0.415 \times \text{商品} \\ & \text{零售价格指数} + 0.402 \times \text{工业品出厂价格指数} + 0.394 \times \text{原材料燃料购进价格指数} \\ & + 0.408 \times \text{固定资产投资价格指数} \end{aligned} \quad (3.13)$$

### 3.2.3 财政税收

财政税收主要由国内增值税、营业税、国内消费税、关税、农业税、企业所得税表示，依据国家统计年鉴中 2001~2007 年的统计数据<sup>[3]</sup>，按照图 3.1 的年度数据转化成季度数据的流程图，得到 2001~2007 年按季度的财政税收指数数据见附表 3。

按照图 3.3 的流程示意图进行主成分分析，得出的结果如表 3.7 所示。

表 3.7 财政税收指数统计数据主成分分析结果示意表

变量	PC1	PC2	PC3	PC4	PC5	PC6
特征值	5.79	0.10	0.08	0.02	0.00	0.00
贡献率	0.97	0.01	0.01	0.01	0.00	0.00
累积贡献率	0.97	0.98	0.99	1.00	1.00	1.00

表 3.8 财政税收各主成分分量与原参数之间关系示意表

变量	PC1	PC2	PC3	PC4	PC5	PC6
国内增值税	0.41	0.20	-0.06	0.30	-0.67	-0.50
营业税	0.42	0.17	0.00	-0.12	-0.36	0.81
国内消费税	0.41	0.04	-0.06	0.74	0.52	0.11
关税	0.40	-0.74	-0.49	-0.23	0.03	-0.07
农业各税	0.40	-0.26	0.84	-0.20	0.10	-0.13
企业所得税	0.41	0.57	-0.22	-0.51	0.39	-0.23

依据上面主成分分析基本原理，累计贡献率达到 80%即认为比较满意，本题中第一主成分 PC1 的贡献率就已经达到 97%，因此 PC1 就完全可以代表财政税收指数，将财政税收指数用 CPC 表示，由表 3.8 可知

$$\begin{aligned} \text{CPC} = \text{PC1} = & 0.414 \times \text{国内增值税} + 0.415 \times \text{营业税} + 0.413 \times \text{国内消费税} + 0.399 \times \text{关税} + 0.401 \\ & \times \text{农业各税} + 0.407 \times \text{企业所得税} \end{aligned} \quad (3.14)$$

### 3.2.4 金融保险

金融保险主要选择金融机构资金运用合计、金融机构现金收入、金融机构现金支出、上市公司数量、保险公司总资产来表示，依据国家统计局年鉴中 2001~2007 年的统计数据<sup>[3]</sup>，按照图 3.2 的年度数据转化成季度数据的流程图，得到 2001~2007 年按季度的金融保险指数数据见附表 4。

按照图 3.3 的流程示意图进行主成分分析，得出的结果如下表 3.9 所示。

表 3.9 金融保险指数统计数据主成分分析结果示意表

变量	PC1	PC2	PC3	PC4	PC5
特征值	4.90	0.07	0.03	0.00	0.00
贡献率	0.98	0.01	0.01	0.00	0.00
累积贡献率	0.98	0.99	1.00	1.00	1.00

表 3.10 金融保险各主成分分量与原参数之间关系示意表

变量	PC1	PC2	PC3	PC4	PC5
金融机构资金运用合计	0.45	0.27	0.30	0.80	0.01
金融机构现金收入	0.45	-0.26	0.37	-0.31	0.70
金融机构现金支出	0.45	-0.26	0.37	-0.29	-0.71
上市公司总数量	0.45	-0.48	-0.73	0.19	0.00
保险公司总资产	0.44	0.75	-0.32	-0.38	-0.01

依据上面主成分分析基本原理，累计贡献率达到 80%即认为比较满意，本题中第一主成分 PC1 的贡献率就已经达到 98%，因此 PC1 就完全可以代表金融保险指数，将金融保险指数用 BPC 表示，由表 3.10 可知

$$BPC=PC1=0.45 \times \text{金融机构资金运用合计} + 0.45 \times \text{金融机构现金收入} + 0.45 \times \text{金融机构现金支出} + 0.45 \times \text{上市公司总数量} + 0.44 \times \text{保险公司总资产} \quad (3.15)$$

### 3.2.5 三大产业

三大产业是农业、工业和服务业，其代表性的特征量包括农林牧渔业总产值、耕地面积、轻工业总产值、重工业总产值、客运量、货运量，依据国家统计局年鉴中 2001~2007 年的统计数据<sup>[3]</sup>，按照图 3.2 的年度数据转化成季度数据的流程图，得到 2001~2007 年按季度的金融保险指数数据见附表 5。

按照图 3.3 的流程示意图进行主成分分析，得出的结果如下表 3.11 所示。

表 3.11 三大产业指数统计数据主成分分析结果示意表

变量	PC1	PC2	PC3	PC4	PC5
特征值	3.96	1.01	0.02	0.01	0.00
贡献率	0.79	0.20	0.01	0.00	0.00
累积贡献率	0.79	0.99	1.00	1.00	1.00

表 3.12 三大产业指数各主成分分量与原参数之间关系示意表

变量	PC1	PC2	PC3	PC4	PC5
农林牧渔业总产值	0.50	0.03	0.86	-0.11	0.07
农作物总播种面积	0.16	-0.94	-0.02	0.29	0.01
轻工业总产值	0.49	0.19	-0.19	0.35	-0.75
重工业总产值	0.49	0.24	-0.29	0.44	0.66
客运量	0.50	-0.14	-0.38	-0.77	0.03

依据上面主成分分析基本原理，累计贡献率要达到 80%才能认为比较满意，本题中第一主成分 PC1 的贡献率为 79%，第一主成分 PC1 和第二主成分 PC2 的

累计贡献率达到了 99%，因此 PC1 和 PC2 完全可以代表三大产业指数，将三大产业指数用 SPC1 和 SPC2 表示，由表 3.12 可知

$$\text{SPC1}=\text{PC1}=0.50\times\text{农林牧渔业总产值}+0.16\times\text{农作物总播种面积}+0.49\times\text{轻工业总产值}+0.49\times\text{重工业总产值}+0.50\times\text{客运量} \quad (3.16)$$

$$\text{SPC2}=\text{PC2}=0.03\times\text{农林牧渔业总产值}-0.94\times\text{农作物总播种面积}+0.19\times\text{轻工业总产值}+0.24\times\text{重工业总产值}-0.14\times\text{客运量} \quad (3.17)$$

### 3.2.6 宏观层面指标综合分析

由 3.2.1~3.2.5 的分析和公式(3.11)~(3.17)得宏观层面的指标汇总见附表 6。

按照主成分分析的原理再对宏观层面指标进行主成分分析，得到的结果如下表 3.13 所示。

表 3.13 宏观指标统计数据主成分分析结果示意表

变量	PC1	PC2	PC3	PC4	PC5	PC6	PC7
特征值	5.72	0.82	0.46	0.01	0.00	0.00	0.00
贡献率	0.82	0.12	0.07	0.00	0.00	0.00	0.00
累积贡献率	0.82	0.93	1.00	1.00	1.00	1.00	1.00

表 3.14 宏观指标各主成分分量与原参数之间关系示意表

变量	PC1	PC2	PC3	PC4	PC5	PC6	PC7
GPC1	0.41	0.13	0.21	0.01	-0.40	0.06	-0.78
GPC2	-0.41	-0.14	-0.22	-0.09	0.62	0.03	-0.61
JPC	0.30	0.43	-0.85	0.11	0.01	-0.01	0.00
CPC	0.41	0.05	0.22	0.31	0.45	-0.70	0.00
BPC	0.42	0.07	0.08	-0.84	0.29	0.09	0.10
SPC1	0.41	-0.19	0.10	0.41	0.37	0.69	0.08
SPC2	-0.24	0.86	0.36	0.08	0.19	0.18	0.03

依据上面主成分分析基本原理，累计贡献率达到 80%即认为比较满意，本题中第一主成分 PC1 的贡献率为 82%，第一主成分 PC1 和第二主成分 PC2 的累计贡献率达到了 93%，第一主成分 PC1、第二主成分 PC2 和第三主成分 PC3 的累积贡献率达到 100%，因此 PC1、PC2 和 PC3 可以 100%的代表宏观指标，将宏观指标用 MACRO1、MACRO2 和 MACRO3 表示，由表 3.14 可知

$$\text{MACRO1}=\text{PC1}=0.41\times\text{GPC1}-0.41\times\text{GPC2}+0.30\times\text{JPC}+0.41\times\text{CPC}+0.42\times\text{BPC}+0.41\times\text{SPC1}-0.24\times\text{SPC2} \quad (3.18)$$

$$\text{MACRO2}=\text{PC2}=0.13\times\text{GPC1}-0.14\times\text{GPC2}+0.43\times\text{JPC}+0.05\times\text{CPC}+0.07\times\text{BPC}-0.19\times\text{SPC1}+0.86\times\text{SPC2} \quad (3.19)$$

$$\text{MACRO3}=\text{PC3}=0.21\times\text{GPC1}-0.22\times\text{GPC2}-0.85\times\text{JPC}+0.22\times\text{CPC}+0.08\times\text{BPC}+0.10\times\text{SPC1}+0.36\times\text{SPC2} \quad (3.20)$$

### 3.3 中观层面指标分析

中观层面指标主要是进行分地区和分行业的分析，地区和行业都有可以分为很多类的，本文使用聚类分析的方法对行业 and 地区进行划分，再用主成分分析得出中观指标 MESO。

#### 3.3.1 分行业

依据《国民经济行业分类》(GB/T4754—2002)<sup>[5]</sup>，我国的行业划分如表 3.15 所示。

表 3.15 国民经济行业分类表

类别	行业名称	类别	行业名称
A	农林牧渔业	I	金融保险业
B	采掘业	J	房地产业
C	制造业	K	社会服务业
D	电力、煤气及水的供应业	L	卫生、体育和社会福利业
E	建筑业	M	教育、文化和娱乐业
F	地质勘查和水利管理业	N	科学研究和综合技术服务业
G	交通运输仓储和邮电通信业	O	国家党政机关和社会团体
H	批发零售贸易和餐饮业	P	其他

由表 3.15 国民经济行业分类示意表可知,我国行业分为 16 类,依据假设三,以国民经济各个行业中有代表性的指标 2007 年的数据<sup>[2,3]</sup>汇总如表 3.16 所示。

表 3.16 2007 年 16 类国民经济行业代表指标统计数据汇总表

行业	就业人数 (万人)	平均劳动报酬 (元/年)	各行业城镇 投资额(亿元)	各行业城镇投 资增长率(%)
农林牧渔业	31468	8488	1460	31.1
采掘业	534	19381	5256	26.9
制造业	8931	29197	35477	34.8
电力、煤气及水的供应业	307	15845	9089	9.8
建筑业	4375	15555	993	48.5
地质勘查和水利管理业	91	9287	9276	22.3
交通运输仓储和邮电通信业	2459	19932	12997	14.4
批发零售贸易和餐饮业	5571	24192	3781	28.9
金融保险业	549	21650	151	25.4
房地产业	226	22725	28619	32.2
社会服务业	1409	11494	861	28.8
卫生体育和社会福利业	553	18890	1124	13.4
教育、文化和娱乐业	1923	15865	2221	3.9
科学研究和综合技术服务业	257	32094	521	16.7
国家机关党政机关社会团体	1089	32131	2768	7.9
其他	17738	16518	3965	11.6

直接按照图 3.3 的流程示意图进行主成分分析,得到的结果如表 3.17 所示。

表 3.17 16 类国民经济行业代表指标主成分分析结果示意表

变量	PC1	PC2	PC3	PC4
特征值	1.48	1.29	0.69	0.54
贡献率	0.37	0.32	0.17	0.13
累积贡献率	0.37	0.69	0.87	1.00

表 3.18 16 类国民经济行业代表指标主成分分量与变量关系表

变量	PC1	PC2	PC3	PC4
就业人数	-0.61	0.21	0.73	0.25
平均劳动报酬	0.68	0.12	0.31	0.65
各行业城镇投资额	0.32	0.69	0.28	-0.59
各行业城镇投资增长率	-0.25	0.69	-0.55	0.40

由表 3.17 可知，要达到比较满意需要三个主成分分量，达到高度满意需要四个，相当于未进行降维，这样的主成分分析基本没有意义。而且由于行业的种类很多，因此本文使用聚类分析的方法先对其进行归类，聚类分析的基本原理如下所述。

聚类分析是数理统计中研究“物以类聚”的一种方法<sup>[4]</sup>。聚类分析的职能是建立一种分析方法，它是将一批样品或变量，安装它们在性质上的亲疏程度进行分类，而描述其亲疏程度通常有两个途径：一是把每个样品看成是  $m$  维(变量的个数为  $m$  个)空间的一个点，在  $m$  维坐标中，定义点与点之间的距离；另一个是用某种相似系数来描述样品点之间的亲疏程度。本文选用第一种方法。

一般地，设有  $n$  个样本，每个样本有  $p$  个指标描述，其数据矩阵为：

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \quad (3.21)$$

$n$  个样本( $X$  中的  $n$  行)可以看成  $p$  维( $p$  个指标)空间中的  $n$  个点，因而两个样本之间的相近程度可以用  $p$  维空间中两点间的距离来量度。用  $d_{ij}$  表示第  $i$  个样本和第  $j$  个样本间的距离，常用的距离计算方式如表 3.22 所示。

表 3.19 聚类分析常用距离表

名称	公式
绝对值距离	$d_{ij} = \sum_{a=1}^p  x_{ia} - x_{ja} $
欧氏距离	$d_{ij} = (\sum_{a=1}^p  x_{ia} - x_{ja} ^2)^{\frac{1}{2}}$
切氏距离	$d_{ij} = \max_{1 \leq a \leq p}  x_{ia} - x_{ja} $
兰氏距离	$d_{ij} = \frac{1}{p} \sum_{a=1}^p \frac{ x_{ia} - x_{ja} }{x_{ia} + x_{ja}}$

设  $p$  和  $q$  表示两类样本，分别有  $n_1$  和  $n_2$  个样本， $x_i$  和  $x_j$  表示样本点， $x_i \in p$ ， $x_j \in q$ ， $\bar{x}_i, \bar{x}_j$  分别表示  $p$  和  $q$  两类样本的重心， $d(x_i, x_j)$  表示  $x_i$  和  $x_j$  之间的距离， $D_{pq}$  表示两类样本之间的距离。类之间距离的计算通常有以下方式：

(1)最短距离法。两类之间的距离定义为两类中个体(样本)之间距离的最小者

$$D_{pq} = \min d(x_i, x_j) \quad (3.22)$$

(2)最长距离法。两类之间的距离定义为两类中个体(样本)之间距离的最大者

$$D_{pq} = \max d(x_i, x_j) \quad (3.23)$$

(3)重心法。每类中所有样本的均值为该类的重心，两类之间的距离定义为两个重心之间的距离

$$D_{pq} = d(\bar{x}_i, \bar{x}_j) \quad (3.24)$$

(4)类平均法。两类之间的距离定义为这两类所有个体(样本) 两两之间距离的平均值。

$$D_{pq} = \frac{1}{n_1 n_2} \sum_{x_i \in p} \sum_{x_j \in q} d(x_i, x_j) \quad (3.25)$$

聚类分析的流程如图 3.4 所示。

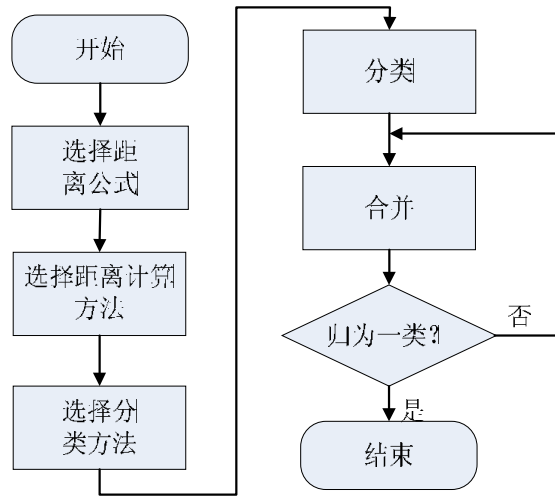


图 3.4 聚类分析流程图

本文选用欧氏距离和最短距离法。

当确定了样品或变量间的距离或相似系数后，就要对样品或变量进行分类，分类的方法很多，一类方法是在样品距离基础上定义类与类之间的距离，其做法是首先将  $n$  个样品自成一类，然后每次将具有最小距离的两类合并，合并后重新计算类与类之间的距离，这个过程一直继续到所有样品归为一类为止，把这个过程作为一张聚类谱系图，这种方法称为系统聚类法(又称为谱系聚类)。另一种分类方法是将  $n$  个样品初步分类，然后根据分类函数尽可能小的原则，对已分类别进行调整，直到分类合理为止，这称为调优法，如动态聚类法就属于这种类型。此外还有模糊聚类、图论聚类、聚类预报等多种方法。本文选用系统聚类法，并将分类的相似度标准定为 90%，即以树状图中相似性为 90 的位置进行分类。

行业聚类分析的依据是表 3.16 中的就业人数和平均劳动报酬，这些指标就代表了该行业的就业容纳能力。聚类分析得到的结果如图 3.5 和表 3.20 所示。

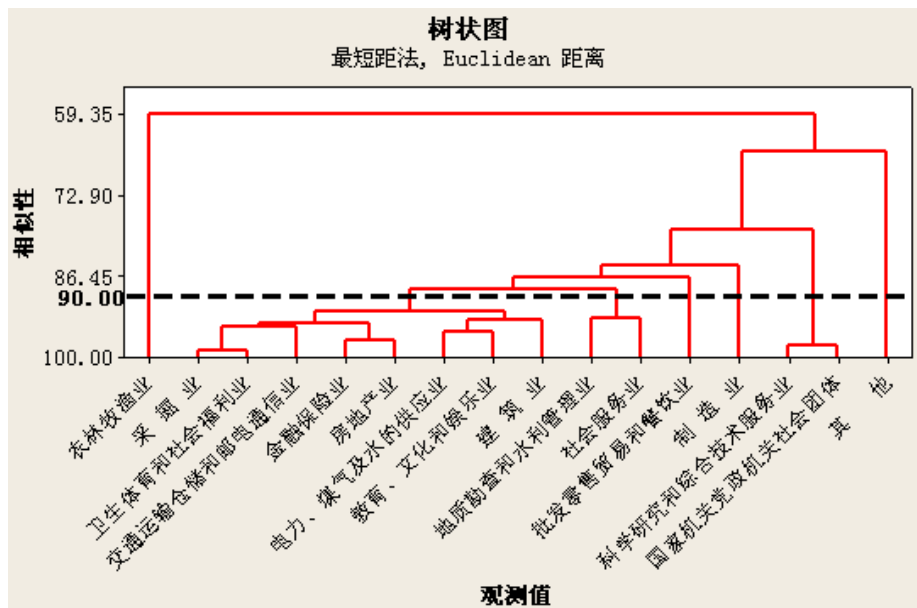


图 3.5 按就业容纳能力进行行业聚类分析的结果示意图

表 3.20 按就业容纳能力聚类分析后行业数据汇总表

行业		就业人数 (万人)	平均劳动报酬 (元/年)	各行业城镇 投资额(亿元)	各行业城镇投 资增长率(%)
第 1 类行业	农林牧渔业	31468	8488	1460	31.1
第 2 类行业	地质勘查和水利管理业	1500	11360	10137	22.8
	社会服务业				
第 3 类行业	批发零售贸易和餐饮业	5571	24192	3781	28.9
第 4 类行业	制造业	8931	29197	35477	34.8
第 5 类行业	科学研究和综合技术服务业	1346	32123	3289	9.3
	国家机关政党机关社会团体				
第 6 类行业	其他	17738	16518	3965	11.6
第 7 类行业	余下的行业	10926	16478	60450	29.7

依据聚类分析的结果再进行主成分分析, 得到的结果如表 3.21 所示。

表 3.21 按就业容纳能力划分指标统计数据主成分分析结果示意表

变量	PC1	PC2	PC3	PC4
特征值	1.86	1.61	0.41	0.12
贡献率	0.47	0.40	0.10	0.03
累积贡献率	0.47	0.87	0.97	1.00

表 3.22 按就业容纳能力划分指标统计数据各主成分分量与原参数之间关系表

变量	PC1	PC2	PC3	PC4
就业人数	-0.09	-0.69	-0.71	-0.05
平均劳动报酬	-0.02	0.71	-0.67	-0.21
各行业城镇投资额	0.71	0.05	-0.18	0.68
各行业城镇投资增长率	0.70	-0.12	0.07	-0.70

由表 3.21 可知，三个主成分分量的累计贡献率达到了 97%，用 HPC 表示分行业指标，则由表 3.22 可知：

$$HPC1=PC1=-0.09 \times \text{就业人数} - 0.02 \times \text{平均劳动报酬} + 0.71 \times \text{各行业城镇投资额} + 0.70 \times \text{各行业城镇投资增长率} \quad (3.26)$$

$$HPC2=PC2=-0.69 \times \text{就业人数} + 0.71 \times \text{平均劳动报酬} + 0.05 \times \text{各行业城镇投资额} - 0.12 \times \text{各行业城镇投资增长率} \quad (3.27)$$

$$HPC3=PC3=-0.71 \times \text{就业人数} - 0.67 \times \text{平均劳动报酬} - 0.18 \times \text{各行业城镇投资额} + 0.07 \times \text{各行业城镇投资增长率} \quad (3.28)$$

### 3.3.2 分地区

我国按照行政区划分为 23 个省、5 个自治区、4 个直辖市、2 个特别行政区，由于获取的数据不完整，这里分析时暂不考虑港澳台和西藏，还有 30 个行政单位，30 个单位有代表性的指标 2008 年的数据<sup>[2,3]</sup>汇总见附表 7。

直接按照图 3.3 的流程示意图进行主成分分析，得到的结果如表 3.23 所示。

表 3.23 30 个行政单位代表指标统计数据主成分分析结果示意表

变量	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
特征值	3.42	2.27	1.13	0.61	0.33	0.16	0.07	0.03
贡献率率	0.43	0.28	0.14	0.08	0.04	0.02	0.01	0.00
累积贡献率	0.43	0.71	0.85	0.93	0.97	0.99	1.00	1.00

表 3.24 30 个行政单位代表指标统计数据各主成分分量与原参数之间关系示意表

变量	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
城镇人口比重	0.40	-0.39	-0.21	-0.17	0.11	-0.08	0.62	0.46
大专以上人数比例	0.51	0.13	-0.06	-0.09	-0.36	-0.32	-0.57	0.40
城镇单位使用的农村劳动力年末人数	0.37	0.19	0.46	-0.20	0.74	-0.10	-0.11	-0.08
城镇登记失业率	-0.31	0.15	-0.59	-0.56	0.34	0.13	-0.24	0.19
城镇登记失业人员数	0.29	0.48	-0.33	-0.19	-0.16	-0.33	0.34	-0.54
职工平均工资	0.35	-0.43	0.04	-0.43	-0.21	0.50	-0.18	-0.43
15 岁及以上文盲人口	-0.34	0.06	0.52	-0.62	-0.33	-0.24	0.19	0.13
总人口	0.17	0.60	0.12	0.02	-0.16	0.67	0.21	0.31

由主成分分析的结果可知，要想比较满意的表征分地区指标的特征，依据假设四，为了完全的表征分地区指标，本文取前四个主成分分量作为分地区指标，累计贡献率达到 93%。据此得到的分地区指标用 dpc 表示，即有：

$$dpc1=PC1=0.40 \times \text{城镇人口比重} + 0.51 \times \text{大专以上人数比例} + 0.37 \times \text{城镇单位使用的农村劳动力年末人数} - 0.31 \times \text{城镇登记失业率} + 0.29 \times \text{城镇登记失业人员数} + 0.35 \times \text{职工平均工资} - 0.34 \times 15 \text{ 岁及以上文盲人口} + 0.17 \times \text{总人口} \quad (3.29)$$

$$dpc2=PC2=-0.39 \times \text{城镇人口比重} + 0.13 \times \text{大专以上人数比例} + 0.19 \times \text{城镇单位使用的农村劳动力年末人数} + 0.15 \times \text{城镇登记失业率} + 0.48 \times \text{城镇登记失业人员数} - 0.43 \times \text{职工平均工资} + 0.06 \times 15 \text{ 岁及以上文盲人口} + 0.60 \times \text{总人口} \quad (3.30)$$

$$dpc3=PC3=-0.21 \times \text{城镇人口比重} - 0.06 \times \text{大专以上人数比例} + 0.46 \times \text{城镇单位使用的农村劳动力年末人数} - 0.59 \times \text{城镇登记失业率} - 0.33 \times \text{城镇登记失业人员数} + 0.04 \times \text{职工平均工资} + 0.52 \times 15 \text{ 岁及以上文盲人口} + 0.12 \times \text{总人口} \quad (3.31)$$

$$dpc4=PC4=-0.17 \times \text{城镇人口比重} - 0.09 \times \text{大专以上人数比例} - 0.20 \times \text{城镇单位使用的农村劳动力年末人数} - 0.56 \times \text{城镇登记失业率} - 0.19 \times \text{城镇登记失业人员数} - 0.43 \times \text{职工平均工资} - 0.62 \times 15 \text{ 岁及以上文盲人口} + 0.02 \times \text{总人口} \quad (3.32)$$



由 30 个单位主成分分析的结果可以看出，各主成分分量对原变量的体现不是很明显，而且地区的数目很多，因此要将其简化成几组来分析。这里本文采用两种简化方式：一种是按照通常理解的按照地域划分为 7 个（港澳台和西藏未考虑在内），如下表 3.25 所示，其对应的统计数据如表 3.26 所示；另一种按地区就业容纳程度聚类分析，聚类分析的依据是附表 7 中的城镇人口比重、大专以上人数比例、职工平均工资和城镇单位使用的农村人数，这四个因素基本体现了地区就业容纳程度，这样聚类分析得到的划分如图 3.6 所示。

表 3.25 30 个单位按照地域划分结果示意表

区域	华北	东北	华东	华中	华南	西南	西北
包含地区	北京、天津、河北、山西、内蒙古	辽宁、吉林、黑龙江	上海、江苏、浙江、安徽、福建、江西、山东	河南、湖北、湖南	广东、广西、海南	重庆、四川、贵州、云南	甘肃、青海、宁夏、新疆

表 3.26 按地域划分代表指标统计数据汇总表

区域	城镇人口比重 (%)	大专以上人数比例 (/万人)	城镇单位使用的农村劳动力人数(万人)	城镇登记失业率 (%)	城镇登记失业人员数(万人)	职工平均工资(元/年)	15 岁及以上文盲人口(万人)	总人口(人)
华北	48.59	1749	208.84	3.26	89.5	23711	25.93	150180352
东北	55.81	2189	67.43	4.22	99.9	21171	12.67	107475239
华东	48.77	2502	835.12	3.53	204.4	25030	65.79	373828445
华中	38.74	2430	185.28	3.99	131.6	20820	21.96	233280553
华南	52.52	2709	321.57	2.88	60.1	26119	18.48	140666406
西南	35.33	1121	224.71	4.13	74.8	21291	51.34	204063389
西北	38.01	1092	114.01	3.83	50.3	21818	64.71	96091010

依据按地域划分的指标数据，进行主成分分析，得到的结果如表 3.27 和表 3.28 所示。

表 3.27 按地域划分代表指标统计数据主成分分析结果示意表

变量	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
特征值	3.67	2.46	1.33	0.39	0.09	0.05	0.00	0.00
贡献率	0.46	0.31	0.17	0.05	0.01	0.01	0.00	0.00
累积贡献率	0.46	0.77	0.93	0.98	0.99	1.00	1.00	1.00

表 3.28 按地域划分代表指标统计数据各主成分分量与原参数之间关系示意表

变量	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
城镇人口比重	0.26	0.44	0.20	0.75	0.10	0.05	-0.13	-0.31
大专以上人数比例	0.39	0.27	0.37	-0.33	-0.72	-0.09	-0.03	-0.10
城镇单位使用的农村劳动力人数	0.47	-0.25	-0.11	0.12	-0.04	0.34	-0.55	0.52
城镇登记失业率	-0.31	-0.32	0.50	0.31	-0.28	0.48	0.34	0.19
城镇登记失业人员数	0.38	-0.30	0.40	0.11	0.28	-0.59	0.29	0.29
职工平均工资	0.41	0.23	-0.44	0.03	-0.02	0.28	0.69	0.19
15 岁及以上文盲人口	0.05	-0.52	-0.43	0.37	-0.48	-0.29	0.03	-0.30
总人口	0.39	-0.39	0.16	-0.26	0.29	0.37	0.04	-0.62

由表 3.27 和表 3.28 可知,为了达到高度满意的结果(累计贡献率超过 90%),可用三个主成分分量 PC1、PC2、PC3 来表征分地区指标,据此得到的分地区指标用 Dpc 表示,则有

$$\text{Dpc1}=\text{PC1}=0.26\times\text{城镇人口比重}+0.39\times\text{大专以上人数比例}+0.47\times\text{城镇单位使用的农村劳动力年末人数}-0.31\times\text{城镇登记失业率}-0.38\times\text{城镇登记失业人员数}+0.41\times\text{职工平均工资}+0.05\times\text{15 岁及以上文盲人口}+0.39\times\text{总人口} \quad (3.33)$$

$$\text{Dpc2}=\text{PC2}=0.44\times\text{城镇人口比重}+0.27\times\text{大专以上人数比例}-0.25\times\text{城镇单位使用的农村劳动力年末人数}-0.32\times\text{城镇登记失业率}-0.30\times\text{城镇登记失业人员数}+0.23\times\text{职工平均工资}-0.52\times\text{15 岁及以上文盲人口}-0.52\times\text{总人口} \quad (3.34)$$

$$\text{Dpc3}=\text{PC3}=0.20\times\text{城镇人口比重}+0.37\times\text{大专以上人数比例}-0.11\times\text{城镇单位使用的农村劳动力年末人数}+0.50\times\text{城镇登记失业率}+0.40\times\text{城镇登记失业人员数}-0.44\times\text{职工平均工资}-0.43\times\text{15 岁及以上文盲人口}+0.16\times\text{总人口} \quad (3.35)$$

下面看先按地区就业容纳程度聚类分析,再进行主成分分析的结果。

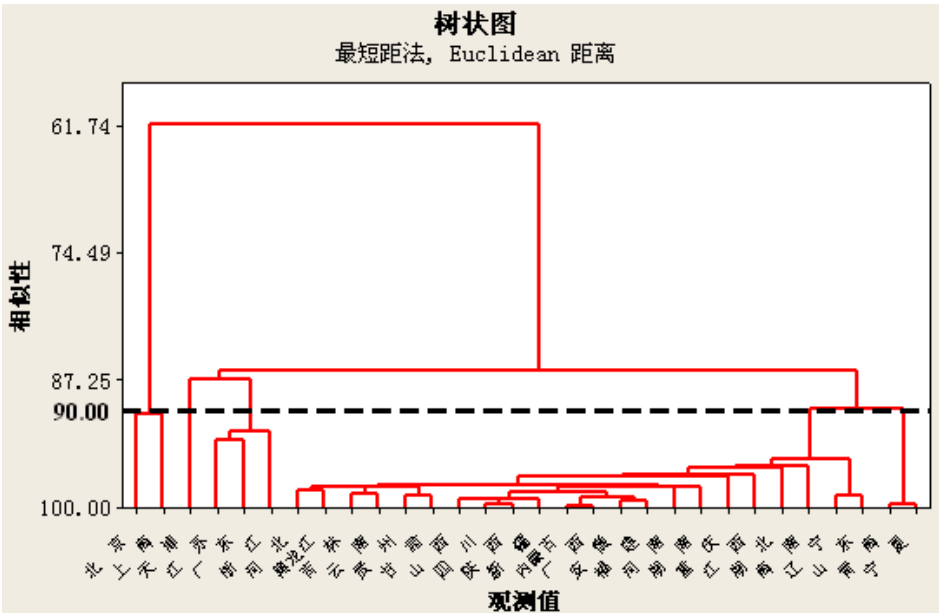


图 3.7 按就业容纳程度聚类分析结果示意图

表 3.29 按就业容纳程度聚类分析数据汇总表

指标		城镇人口比重 (%)	大专以上人数比例 (/万人)	城镇单位使用的农村人数 (万人)	城镇登记失业率 (%)	城镇登记失业人员数 (万人)	职工平均工资 (元/年)	15 岁及以上文盲人口 (万人)	总人口 (人)
区域 1	北京	86.73	3642	160.39	3.05	37.3	47996	7.38	25951084
	上海								
区域 2	天津	76.31	1457	19.14	3.60	15.0	34938	3.85	9641374
区域 3	江苏	58.14	3762	661.04	2.94	104.1	29068	22.43	201694732
	广东								
	浙江								
区域 4	青海	42.23	224	6.30	4.06	8.1	26190	32.21	11347100
	宁夏								
区域 5	其他	41.88	1748	1121.22	3.89	580.9	21625	231.26	1082086820

依据按就业容纳程度划分的指标数据，进行主成分分析，得到的结果如表 3.30 所示。

表 3.30 按就业容纳程度划分指标统计数据主成分分析结果示意表

变量	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
特征值	5.08	2.28	0.57	0.07	0.00	0.00	0.00	0.00
贡献率	0.64	0.31	0.07	0.01	0.00	0.00	0.00	0.00
累积贡献率	0.64	0.94	0.99	1.00	1.00	1.00	1.00	1.00

表 3.31 按就业容纳程度划分指标统计数据各主成分分量与原参数之间关系表

变量	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
城镇人口比重	-0.37	0.24	0.53	-0.69	-0.14	0.02	-0.04	-0.19
大专以上人数比例	-0.16	0.60	-0.24	0.25	-0.54	0.30	-0.33	-0.03
城镇单位使用的农村人数	0.35	0.39	-0.18	-0.11	-0.23	-0.45	0.59	-0.28
城镇登记失业率	0.27	-0.50	0.36	0.12	-0.73	0.03	0.03	-0.06
城镇登记失业人员数	0.40	0.25	0.27	-0.05	0.11	0.64	0.38	0.38
职工平均工资	-0.38	0.20	0.54	0.62	0.08	-0.25	0.27	0.10
15 岁及以上文盲人口	0.42	0.14	0.31	0.21	0.29	0.11	-0.33	-0.68
总人口	0.41	0.24	0.22	-0.09	0.00	-0.48	-0.47	0.52

由表 3.30 和表 3.31 可知，为了达到高度满意的结果（累计贡献率超过 90%），可用三个主成分分量 PC1、PC2 来表征分地区指标，据此得到的分地区指标用 DPc 表示，则有

$$DPc1=PC1=-0.37 \times \text{城镇人口比重} - 0.16 \times \text{大专以上人数比例} + 0.35 \times \text{城镇单位使用的农村劳动力年末人数} + 0.27 \times \text{城镇登记失业率} + 0.40 \times \text{城镇登记失业人员数} - 0.38 \times \text{职工平均工资} + 0.42 \times \text{15 岁及以上文盲人口} + 0.41 \times \text{总人口} \quad (3.36)$$

$$DPc2=PC2=0.24 \times \text{城镇人口比重} + 0.60 \times \text{大专以上人数比例} + 0.39 \times \text{城镇单位使用的农村劳动力年末人数} - 0.50 \times \text{城镇登记失业率} + 0.25 \times \text{城镇登记失业人员数} + 0.20 \times \text{职工平均工资} + 0.14 \times \text{15 岁及以上文盲人口} + 0.24 \times \text{总人口} \quad (3.37)$$

由三组分地区指标对比可以得出如下结论：

(1)按某种方式将初始数据分类以后使用主成分分析的结果满意度更高，而且所需的主成分更少，直接分析需四个主成分分量且满意度只有 93%，按地域划分分析需三个主成分分量且满意度为 93%，按就业容纳程度聚类分析需 2 个主成分分量且满意度达到 94%。

(2)初始数据分类的依据很重要，按就业容纳程度聚类分析是最合理的方式，从主成分分析的结果可以看出比按地域这种笼统的分类方式更加科学合理。

(3)由于按就业容纳程度聚类分析得到的主成分分析结果最合理最能代表初始数据，因此分地区指标用 DPC 表示，且有：

$$DPC1=DPc1=-0.37 \times \text{城镇人口比重} - 0.16 \times \text{大专以上人数比例} + 0.35 \times \text{城镇单位使用的农村劳动力年末人数} + 0.27 \times \text{城镇登记失业率} + 0.40 \times \text{城镇登记失业人员数} - 0.38 \times \text{职工平均工资} + 0.42 \times \text{15 岁及以上文盲人口} + 0.41 \times \text{总人口} \quad (3.38)$$

$$DPC2=DPc2=0.24 \times \text{城镇人口比重} + 0.60 \times \text{大专以上人数比例} + 0.39 \times \text{城镇单位使用的农村劳动力年末人数} - 0.50 \times \text{城镇登记失业率} + 0.25 \times \text{城镇登记失业人员数} + 0.20 \times \text{职工平均工资} + 0.14 \times \text{15 岁及以上文盲人口} + 0.24 \times \text{总人口} \quad (3.39)$$

### 3.3.3 中观层面指标综合分析

由 3.3.1 的分行行业分析和 3.3.2 的分地区分析可以汇总得到中观层面指标的数据汇总如表 3.32 所示。

表 3.32 中观层面指标汇总示意表

按就业容纳程度分行业指标				按就业容纳程度分区域指标		
行业划分	HPC1	HPC2	HPC3	区域划分	DPC1	DPC2
第 1 类行业	-1986.76	-15764.4	-28412.6	区域 1	10543491.03	6214237.6
第 2 类行业	6841.608	7475.56	-10558	区域 2	3910650.979	2312214.8
第 3 类行业	1765.878	13435.87	-20939.4	区域 3	82078472.21	48213446
第 4 类行业	23753.19	16136.57	-32471.7	区域 4	4608390.182	2717372.7
第 5 类行业	1845.64	14174.42	-15773.4	区域 5	440401628.2	258624773
第 6 类行业	894.7312	-430.477	-24484.6			
第 7 类行业	43482.94	6564.592	-29664.4			

这里分地区指标和分行业指标之间没有什么可比性,因此中观指标用 MESO 表示,且定义如下:

MESO1=HPC1; MESO2=HPC2; MESO3=HPC3;

MESO4=DPC1; MESO5=DPC2。

### 3.4 微观层面指标分析

微观层面指标主要分析就业人群的受教育程度、性别、年龄三个方面,通过与城镇登记失业率的相关分析得出三个方面的受影响程度,最后综合得到微观指标 MICRO。

相关分析的基本原理<sup>[4]</sup>如下:

对于两个要素  $x$  与  $y$ , 如果它们的样本值分别为  $x_i$  与  $y_i$  ( $i=1, 2, \dots, n$ ),

则它们之间的相关系数  $r_{xy}$  为:

$$r_{xy} = \frac{\sigma_{xy}^2}{\sqrt{\sigma_x^2} \sqrt{\sigma_y^2}}, r_{xy} \in [-1, 1] \quad (3.40)$$

其中,  $\sigma_{xy}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n} \sum_{i=1}^n x_i y_i - \bar{x} \cdot \bar{y}$ , 为协方差, 用于测度  $x$  与  $y$

的相关性;

$\sigma_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2$ , 为随机变量  $x$  的方差;

$\sigma_y^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2 = \frac{1}{n} \sum_{i=1}^n y_i^2 - \bar{y}^2$ , 为随机变量  $y$  的方差。

依据  $r_{xy}$  的取值不同有以下的判据:

(1)  $r_{xy} > 0$ , 表示正相关, 即同向相关;  $r_{xy} < 0$ , 表示负相关, 即异向相关。

(2)  $r_{xy}$  的绝对值  $|r_{xy}|$  越接近于 1,  $x$  与  $y$  越接近线性关系; 当  $|r_{xy}| = 1$  时,  $y$  完全依赖于  $x$ , 两者为因果关系, 是完全的线性关系。

(3)  $r_{xy}$  的绝对值  $|r_{xy}|$  越接近于 0,  $x$  与  $y$  越没有线性关系; 当  $|r_{xy}|=0$  时, 表示  $x$  与  $y$  之间不存在线性关系(但可能有非线性关系)。

(4) 若  $0 < |r_{xy}| < 1$ , 表示  $x$  与  $y$  有相关关系, 但不是线性关系。  $|r_{xy}| \geq 0.8$ , 视为高度相关;  $0.5 \leq |r_{xy}| < 0.8$ , 视为中度相关;  $|r_{xy}| < 0.3$ , 视为不相关。

### 3.4.1 受教育程度

受教育程度主要考察学历, 2001~2007 年我国国民受教育程度统计数据<sup>[2,3]</sup>见附表 8。对数据进行主成分分析, 分析的结果如表 3.33 所示。

表 3.33 受教育程度统计数据主成分分析结果示意表

变量	PC1	PC2	PC3	PC4	PC5
特征值	3.29	1.63	0.06	0.03	0.00
贡献率	0.66	0.32	0.01	0.01	0.00
累积贡献率	0.66	0.98	0.99	1.00	1.00

表 3.34 受教育程度主成分分量与原变量关系表

变量	PC1	PC2	PC3	PC4	PC5
未上过学	0.38	-0.55	-0.71	-0.04	-0.22
小学	-0.55	-0.06	0.01	-0.16	-0.82
初中	0.05	0.77	-0.54	0.30	-0.15
高中	0.55	-0.03	0.44	0.54	-0.46
大专及以上	0.50	0.31	0.13	-0.77	-0.21

从表中可以看出, 两个主成分就足以表示受教育程度对就业的影响, 用微观层面指标 micro1 和 micro2 来表示, 则有

$$\text{micro1} = \text{PC1} = 0.38 \times \text{未上过小学} - 0.55 \times \text{小学} + 0.05 \times \text{初中} + 0.55 \times \text{高中} + 0.50 \times \text{大专及以上} \quad (3.41)$$

$$\text{micro2} = \text{PC2} = -0.549 \times \text{未上过小学} - 0.55 \times \text{小学} + 0.073 \times \text{初中} - 0.03 \times \text{高中} + 0.312 \times \text{大专及以上} \quad (3.42)$$

### 3.4.2 性别

性别主要是分析男女失业率随总失业率的变化情况, 参考 2000~2007 年我国年度失业率统计情况<sup>[2,3]</sup>, 插值得到以 2001 年第一季度为起始 1 的失业率统计数据见附表 9。

分别计算男女失业率与同期失业率的相关程度, 则有男失业率和同期总失业率的 Pearson 相关系数=0.84, 女失业率和同期总失业率的 Pearson 相关系数=0.74, 即有男失业率和同期总失业率高度相关, 女失业率和同期总失业率中度相关。下面是男女失业率对同期总失业率线性拟合和二次拟合的结果示意图。

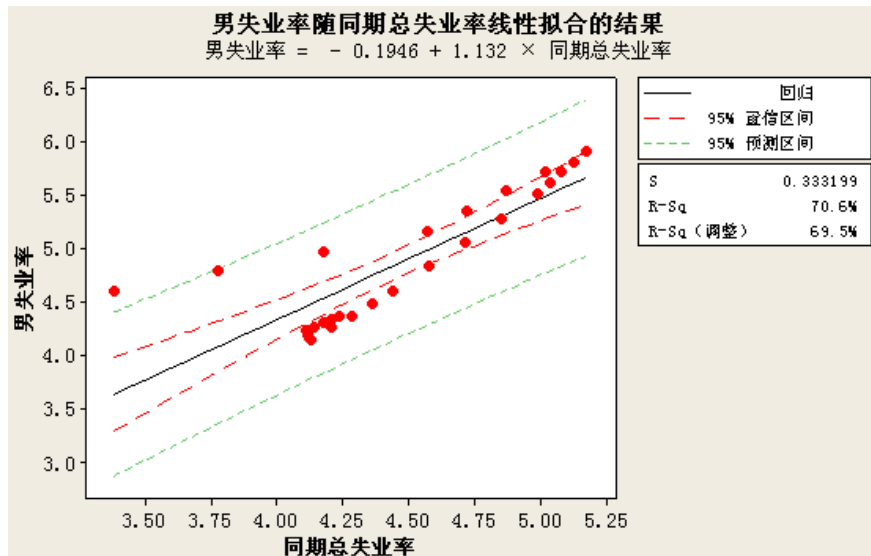


图 3.8 男失业率随同期总失业率线性拟合的结果示意图

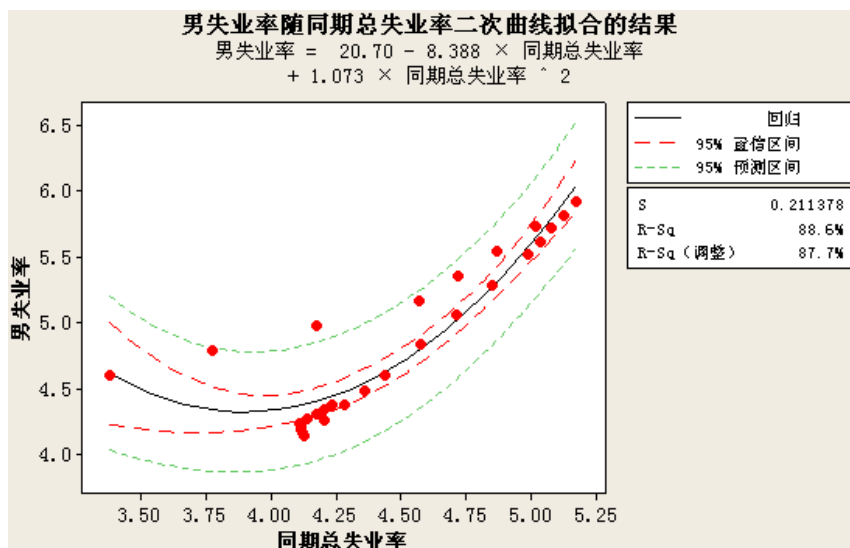


图 3.9 男失业率随同期总失业率二次函数拟合的结果示意图

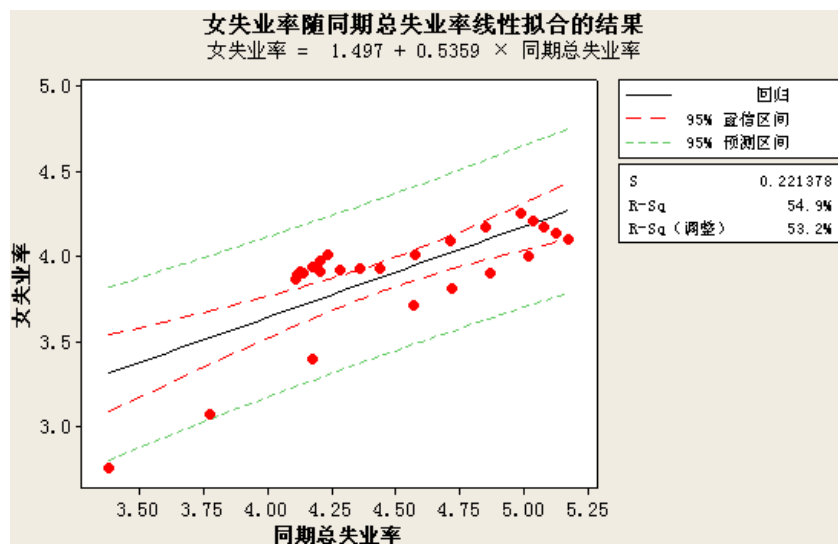


图 3.10 女失业率随同期总失业率线性拟合的结果示意图

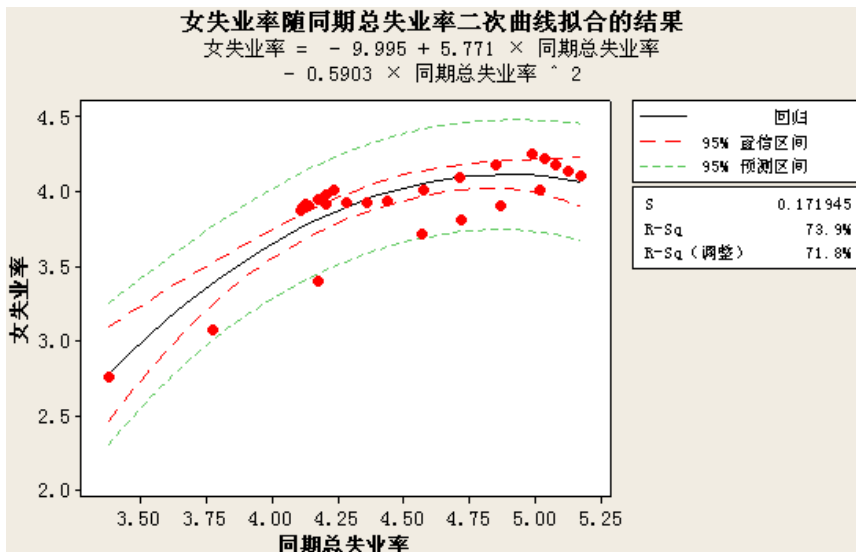


图 3.11 女失业率随同期总失业率二次函数拟合的结果示意图

由拟合结果图 3.8~3.11 可知，面对同样的失业危机时，男性更容易是首先下岗的人，按照线性拟合的结果来看，总失业率每增加一个百分点，男失业率就增加 1.13 个百分点，而女失业率只增加 0.54 个百分点，这说明性别是影响就业率的微观指标，用 micro3 和 micro4 来表示，其中 micro3 表示男失业率，micro4 表示女失业率。

### 3.4.3 年龄

年龄主要分析各年龄段与失业率的关系，依据 2000~2007 年各年龄段失业率转化成的季度数据<sup>[2,3]</sup>见附表 10。

首先对年龄段进行聚类分析，直接以上表的数据为依据，得到的聚类分析结果如图所示。

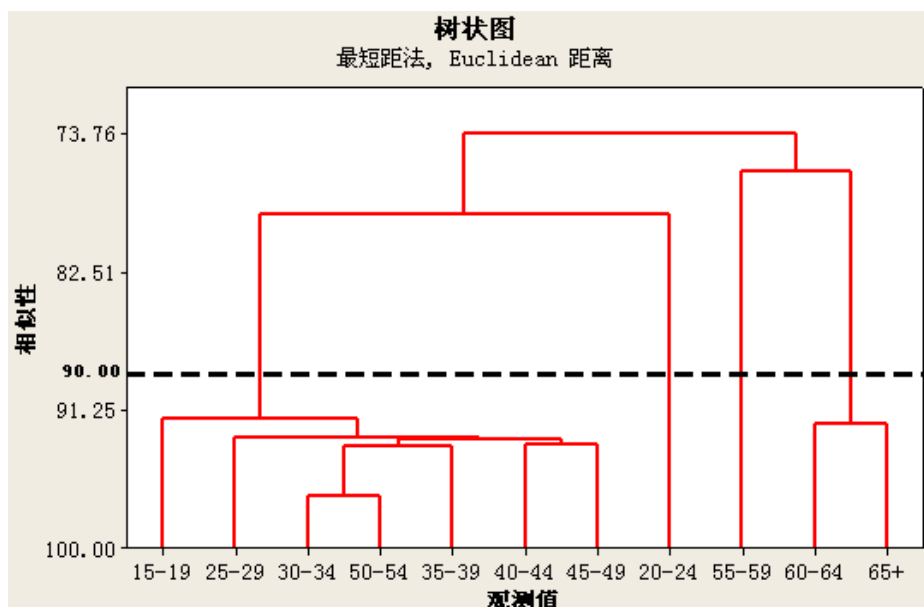


图 3.12 年龄段聚类分析结果示意图

聚类分组后的数据见附表 11，对聚类之后的数据再进行主成分分析，分析的结果如表 3.35 所示。

表 3.35 年龄段分组后主成分分析结果示意表

变量	PC1	PC2	PC3	PC4
特征值	2.33	1.39	0.28	0.01
贡献率	0.58	0.35	0.07	0.00
累积贡献率	0.58	0.93	1.00	1.00

表 3.36 年龄段分组后主成分分量与原变量关系示意表

变量	PC1	PC2	PC3	PC4
年龄段（15-49 除去 20-24）	-0.59	-0.31	-0.40	-0.63
年龄段（20-24）	0.51	-0.48	0.46	-0.54
年龄段（55-59）	0.24	0.79	-0.10	-0.56
年龄段（60 以上）	-0.57	0.23	0.79	-0.07

有表可知，取第一和第二主分量即可表征年龄段特征，用 micro5 和 micro6 来表示，其中：

$$\text{micro5} = \text{PC1} = -0.59 \times \text{年龄段（15-49 除去 20-24）} + 0.51 \times \text{年龄段（20-24）} + 0.24 \times \text{年龄段（55-59）} - 0.57 \times \text{年龄段（60 以上）} \quad (3.43)$$

$$\text{micro6} = \text{PC2} = -0.31 \times \text{年龄段（15-49 除去 20-24）} - 0.48 \times \text{年龄段（20-24）} + 0.79 \times \text{年龄段（55-59）} + 0.23 \times \text{年龄段（60 以上）} \quad (3.44)$$

#### 3.4.4 微观层面指标综合分析

微观层面指标一共有六个，将其汇总后的数据见附表 12。对 micro1~micro6 进行主成分分析，结果如下表所示。

表 3.37 微观层面指标主成分分析结果示意表

变量	PC1	PC2	PC3	PC4	PC5	PC6
特征值	2.84	2.39	0.51	0.22	0.03	0.00
贡献率	0.47	0.40	0.09	0.04	0.01	0.00
累积贡献率	0.47	0.87	0.96	1.00	1.00	1.00

表 3.38 微观层面指标主成分分量与原变量关系示意表

变量	PC1	PC2	PC3	PC4	PC5	PC6
micro1	0.49	0.34	-0.20	0.08	-0.36	-0.68
micro2	-0.26	0.52	-0.28	-0.76	0.10	0.02
micro3	0.14	-0.58	-0.47	-0.34	-0.53	0.19
micro4	0.55	0.03	-0.52	0.12	0.59	0.27
micro5	-0.42	0.35	-0.51	0.54	-0.30	0.25
micro6	0.44	0.39	0.36	-0.06	-0.38	0.61

由表 3.37 和 3.38 可知，可以用三个主分量表示微观层面指标，用微观指标 MICRO1、MICRO2、MICRO3 来表示，则有

$$\text{MICRO1} = \text{PC1} = 0.49 \times \text{micro1} - 0.26 \times \text{micro2} + 0.14 \times \text{micro3} + 0.55 \times \text{micro4} - 0.42 \times \text{micro5} + 0.44 \times \text{micro6} \quad (3.45)$$

$$\text{MICRO2} = \text{PC2} = 0.34 \times \text{micro1} + 0.52 \times \text{micro2} - 0.58 \times \text{micro3} + 0.03 \times \text{micro4} + 0.35 \times \text{micro5} + 0.39 \times \text{micro6} \quad (3.46)$$

$$\text{MICRO3} = \text{PC3} = -0.20 \times \text{micro1} - 0.28 \times \text{micro2} - 0.47 \times \text{micro3} - 0.52 \times \text{micro4} - 0.54 \times \text{micro5} + 0.36 \times \text{micro6} \quad (3.47)$$

### 3.5 综合评价

由 3.1~3.4 的分析可知，就业的影响因素可以归结为 3 个宏观指标、5 个中



观指标和 3 个微观指标，如表 3.39 所示，其中宏观指标和微观指标主要用于失业率和就业人数的模型的建立与就业前景预测，中观指标主要用于国家宏观政策的评价和分析。

表 3.39 就业影响因素指标汇总表

就业影响因素指标	宏观指标	MACRO1
		MACRO2
		MACRO3
	中观指标	MESO1
		MESO2
		MESO3
		MESO4
		MESO5
	微观指标	MICRO1
		MICRO2
		MICRO3

## 四、问题 2：建立城镇登记失业率模型

### 4.1 按国家标准进行建模分析

依据上面分析得到的因素指标 MACRO1~MACRO3、MICRO1~MICRO3 和查得的数据城镇就业人数 N、城镇登记失业率 u，汇总数据得到如表 4.1 所示。

表 4.1 按国家标准得到的数据汇总表

季度 n	城镇就业 人数 N (人)	城镇登记 失业率 u (%)	MACRO1	MACRO2	MACRO3	MICRO1	MICRO2	MICRO3
1	1805242	4.12	136445.1	-98926.8	-1963.43	-11.457	5.50952	-7.8036
2	1811350	4.27	148172.8	-107570	-1725.71	-10.4437	5.72927	-8.25447
3	1817458	4.42	142384.5	-103233	-1621.28	-9.43403	5.9562	-8.69977
4	1823567	4.57	165789.6	-120530	-1562.2	-8.41747	6.17768	-9.15493
5	1829675	4.72	146426.9	-101196	256.3649	-7.51238	6.39232	-9.4874
6	1835783	4.87	159005.2	-110040	1328.427	-6.61849	6.61174	-9.83144
7	1841892	5.02	152791.8	-105602	1865.168	-5.72447	6.83534	-10.1591
8	1848000	5.17	177897.9	-123301	2376.969	-4.82803	7.04956	-10.5003
9	1831116	5.13	156358	-97318	3221.599	-4.69102	7.35824	-10.6171
10	1814233	5.08	169874.1	-105812	3480.507	-4.54209	7.66907	-10.7294
11	1797349	5.04	163342.5	-101521	3401.605	-4.39883	7.97213	-10.8351
12	1780465	4.99	190223.6	-118538	3885.704	-4.26182	8.28081	-10.9519
13	1806936	4.85	180310.4	-103563	6954.071	-4.34819	8.64834	-10.9291
14	1833406	4.72	195873.6	-112605	7532.385	-4.4376	9.00616	-10.9147
15	1859877	4.58	188346	-108033	7304.557	-4.5284	9.36976	-10.8957
16	1886348	4.44	219303	-126152	8420.28	-4.61477	9.73729	-10.873
17	1893190	4.36	194787.8	-105258	10245.73	-4.62463	9.54691	-10.6797
18	1900033	4.29	211567.3	-114458	11102.81	-4.64999	9.36156	-10.4853
19	1906876	4.21	203458.3	-109804	10746.68	-4.65985	9.17118	-10.292

20	1913719	4.13	236830.1	-128240	12412.53	-4.67162	8.98967	-10.0909
21	1930578	4.13	217367.9	-110086	14728.33	-4.76315	9.19435	-10.217
22	1947436	4.12	236101.7	-119708	15970.85	-4.84668	9.39836	-10.3476
23	1964295	4.12	227107.5	-114825	15442.42	-4.93399	9.59953	-10.4687
24	1981154	4.11	264326.7	-134116	17865.42	-5.01636	9.80412	-10.5918
25	1983865	4.14	247814.6	-115963	20545.98	-4.98089	10.05816	-10.7948
26	1986577	4.18	269347.2	-126070	22322.21	-4.95111	10.31985	-10.98
27	1989288	4.21	259344.1	-120874	21608.02	-4.90699	10.57431	-11.1743
28	1992000	4.24	301900.6	-141182	25044.36	-4.87721	10.836	-11.3595

首先对九个变量进行相关性分析，分析得到的矩阵如表 4.2 所示。

表 4.2 相关性分析结果示意表

	城镇就 业人数 N	城镇登 记失业 率 u	MACRO1	MACRO2	MACRO3	MICRO1	MICRO2	MICRO3
城镇就业 人数	1	-0.61	0.907	-0.958	0.937	0.875	0.896	-0.928
城镇登记 失业率	-0.61	1	-0.886	0.897	-0.897	0.975	-0.946	-0.87
MACRO1	0.907	-0.886	1	-0.12	0.066	0.036	0.101	-0.118
MACRO2	-0.958	0.897	-0.12	1	-0.054	-0.118	-0.085	0.1
MACRO3	0.937	-0.897	0.066	-0.054	1	0.059	0.122	-0.135
MICRO1	0.875	0.975	0.036	-0.118	0.059	1	0.053	-0.148
MICRO2	0.896	-0.946	0.101	-0.085	0.122	0.053	1	-0.104
MICRO3	-0.928	-0.87	-0.118	0.1	-0.135	-0.148	-0.104	1

从相关分析的结果可以看出，城镇就业人数 N 和城镇登记失业率 u 与因素指标 MACRO1~MACRO3、MICRO1~MICRO3 之间高度相关，而因素指标之间的相关性都很弱，因此可以得出下面两个结论：

- (1) 因素指标 MACRO1~MACRO3、MICRO1~MICRO3 之间是相互独立的；
- (2) 城镇就业人数 N 和城镇登记失业率 u 由因素指标 MACRO1~MACRO3、MICRO1~MICRO3 加权求和得到。

基于上面的两个结论，则有城镇就业人数 N 和城镇登记失业率 u 关于因素指标 MACRO1~MACRO3、MICRO1~MICRO3 的函数关系：

$$N = w_{11} \times MACRO1 + w_{12} \times MACRO2 + w_{13} \times MACRO3 + w_{14} \times MICRO1 + w_{15} \times MICRO2 + w_{16} \times MICRO3 + c_1 \quad (4.1)$$

$$u = w_{21} \times MACRO1 + w_{22} \times MACRO2 + w_{23} \times MACRO3 + w_{24} \times MICRO1 + w_{25} \times MICRO2 + w_{26} \times MICRO3 + c_2 \quad (4.2)$$

其中：  $w_{ij}$  为权重系数，  $i=1,2$ ；  $j=1\sim6$ 。

$c_k$  为常数，  $k=1,2$ 。

由表 4.1 中的数据，按照最小二乘的方法进行参数辨识，得出的结果如表 4.3 所示。

表 4.3 最小二乘参数辨识的结果

参数	$w_{11}$	$w_{12}$	$w_{13}$	$w_{14}$	$w_{15}$	$w_{16}$	$c_1$
辨识结果	0.3881	-0.015	0.00001	50	0.1994	-0.3998	-17.5837
参数	$w_{21}$	$w_{22}$	$w_{23}$	$w_{24}$	$w_{25}$	$w_{26}$	$c_2$
辨识结果	0.0245	0.0001	0.0118	169.6729	0.00001	0.0001	-168.5368

则得到了 N 和 u 为：

$$N = 0.3881 \times \text{MACRO1} - 0.015 \times \text{MACRO2} + 0.00001 \times \text{MACRO3} + 50 \times \text{MICRO1} + 0.1994 \times \text{MICRO2} - 0.3998 \times \text{MICRO3} - 17.5837 \quad (4.3)$$

$$u = 0.0245 \times \text{MACRO1} + 0.0001 \times \text{MACRO2} + 0.0118 \times \text{MACRO3} + 169.6729 \times \text{MICRO1} + 0.00001 \times \text{MICRO2} + 0.0001 \times \text{MICRO3} - 168.5368 \quad (4.4)$$

因素指标 MACRO1~MACRO3、MICRO1~MICRO3 与季度 n 的函数关系按照三次多项式拟合得到，下面是 MACRO1 对季度 n 三次多项式拟合得到的结果示意图。

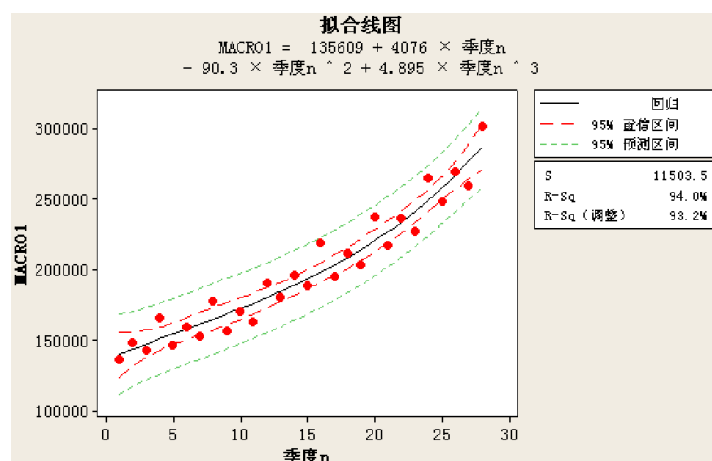


图 4.1 MACRO1 对季度 n 三次多项式拟合的结果示意图

则类似可以得到 MACRO1~MACRO3、MICRO1~MICRO3 关于季度 n 的函数关系式为：

$$\begin{cases} \text{MACRO1} = 135609 + 4076 \times n - 90.3 \times n^2 + 4.895 \times n^3 \\ \text{MACRO2} = -102663 - 1436 \times n + 109.1 \times n^2 - 3.447 \times n^3 \\ \text{MACRO3} = -3069 + 644.3 \times n - 4.08 \times n^2 + 0.5885 \times n^3 \\ \text{MICRO1} = -13.23 + 1.607 \times n - 0.09118 \times n^2 + 0.001598 \times n^3 \\ \text{MICRO2} = 4.816 + 0.402 \times n - 0.01159 \times n^2 + 0.000161 \times n^3 \\ \text{MICRO3} = -6.776 - 0.8182 \times n - 0.05288 \times n^2 - 0.001056 \times n^3 \end{cases} \quad (4.5)$$

将公式(4.5)带入(4.4)中，得到城镇登记失业率 u 关于季度 n 的函数关系式为：

$$u = 3.66 + 0.36 \times n - 0.03 \times n^2 + 0.0006 \times n^3 \quad (4.6)$$

表 4.4 是 2008 年四个季度和 2009 年前两季度的国家发布的城镇登记失业率统计数据<sup>[8]</sup>和模型预测的失业率数据。从表中可以看出，模型预测的数据与真实数据之间的差别不大，验证了模型的准确性；但是略高于真实数据，也说明了模型存在一定的局限性。

表 4.4 2008 年四个季度和 2009 年前两个季度的城镇登记失业率统计数据

	2008 年 第一季度	2008 年 第二季度	2008 年 第三季度	2008 年 第四季度	2009 年 第一季度	2009 年 第二季度
国家发布的城镇 登记失业率(%)	4.0	4.0	4.0	4.2	4.3	4.3
模型预测的失业 率数据(%)	4.23	4.31	4.17	4.44	4.52	4.40

#### 4.2 按国际标准进行建模分析

按照国际劳工组织（ILO）对失业者的界定为：“失业者是指在一定年龄以上，在一定时间内（如四周或者 3 个月），那些没有工作，目前可以工作而且正在寻找工作的人。”我国失业统计口径中的失业人口通常指城镇登记失业人口，其界定为：“有非农业户口，在一定劳动年龄内（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”。我国失业率的计算方法为：

$$\text{失业率} = \frac{\text{城镇登记失业人数}}{(\text{城镇从业人数} + \text{城镇登记失业人数})} \times 100\% \quad (4.7)$$

上述失业统计口径存在很多缺陷，首要的就是只反映城镇失业情况，未反映农村失业情况，而对于我国这样的农业人口占绝大多数的国家来说，对农村失业情况的关注更加有实际意义<sup>[6]</sup>。就城镇失业而言我国的失业统计口径也存在三个缺陷：没有包括未登记失业人员；没有反映我国严重的隐性失业问题；最重要的缺陷是没有包括下岗职工中尚未重新就业的人员。因此考虑到隐性失业人口和未登记失业人口的统计难度大，我国的城镇失业人口和城镇失业率应该定义为：

$$\text{城镇失业人口} = \text{城镇登记失业人口} + (\text{下岗职工总数} - \text{重新就业人数}) \quad (4.8)$$

$$\text{城镇失业率} = \frac{\text{城镇失业人口}}{(\text{城镇从业人口} + \text{城镇失业人口})} \quad (4.9)$$

查国家统计数据可以得到 2000～2007 年城镇登记失业人口、下岗职工总数、重新就业人数和城镇从业人口，得到更新的数据如表 4.4 所示。

表 4.5 更新数据表

季度	下岗职工人数 (人)	重新就业 人数(人)	登记失业 人数(人)	登记从业 人数(人)	登记失业率 u (100%)	国际标准失 业率 U (%)
1	780964	375227	744342	17322208	4.12	6.23
2	848874	407856	809067	18138644	4.27	6.45
3	814919	391541	776705	16795797	4.42	6.67
4	950739	456798	906155	18922189	4.57	6.89
5	820712	387705	943715	19050239	4.72	6.74
6	892078	421418	1025777	20037401	4.87	6.95
7	856395	404561	984746	18631702	5.02	7.16
8	999127	471988	1148870	21072986	5.17	7.37
9	889713	447409	1093684	20225684	5.13	7.06
10	967080	486315	1188786	22212523	5.08	6.99
11	928396	466862	1141235	21502317	5.04	6.94
12	1083129	544672	1331441	25350741	4.99	6.87

13	999612	505895	1224031	24013722	4.85	6.68
14	1086535	549886	1330468	26857423	4.72	6.50
15	1043073	527890	1277250	26610300	4.58	6.31
16	1216919	615872	1490125	32071243	4.44	6.12
17	1038593	542552	1267120	27795258	4.36	5.97
18	1128906	589730	1377304	30727684	4.29	5.87
19	1083750	566141	1322212	30084245	4.21	5.76
20	1264374	660498	1542580	35808036	4.13	5.66
21	799902	459646	1588063	36863822	4.13	4.97
22	869459	499616	1726155	40170817	4.12	4.96
23	834680	479631	1657109	38563984	4.12	4.96
24	973794	559569	1933294	45105487	4.11	4.95
25	820754	458279	1641367	38005188	4.14	5.01
26	892123	498130	1784095	40897604	4.18	5.06
27	856439	478204	1712731	38969720	4.21	5.09
28	999178	557905	1998186	45128851	4.24	5.13

按照第 4.1 部分的步骤进行分析，最终得到国际标准失业率  $U$  为

$$U = 5.71 + 0.43 \times n - 0.038 \times n^2 + 0.0008 \times n^3 \quad (4.10)$$

#### 4.3 对比分析合理性

将国际标准失业率  $U$  和登记失业率  $u$  绘制在同一幅图上，如图 4.2 所示。

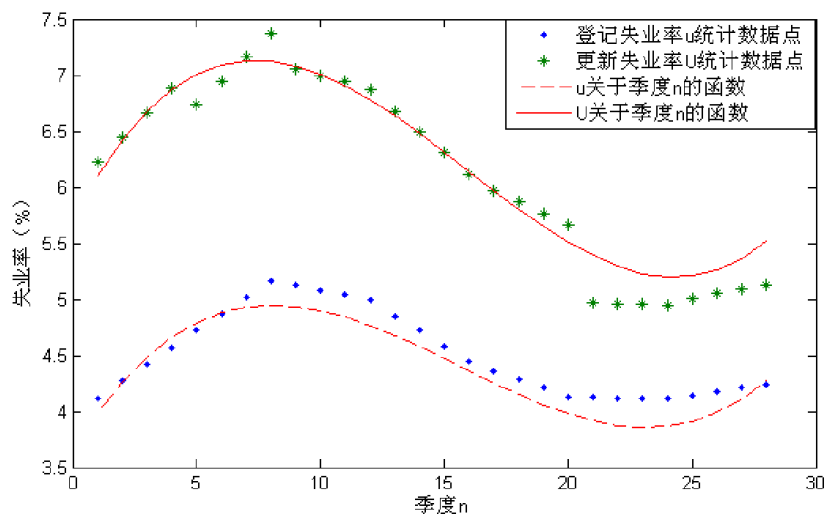


图 4.2 国际标准失业率  $U$  和登记失业率  $u$  对比示意图

由上面的分析和图 4.2 可以得出如下的结论：

- (1) 国际标准失业率明显高于登记失业率，考虑到我国失业率统计偏低的事实，有理由认为按照国际标准得到的失业率更加接近我国的真实情况；
- (2) 国际标准失业率和登记失业率的变化趋势是基本一致的，这也证明了决定它们的影响指标是一致的，即为  $\text{MACRO1} \sim \text{MACRO3}$  和  $\text{MICRO1} \sim \text{MICRO3}$ 。
- (3) 失业率的拟合公式和统计点的分布基本是一致的，这也说明了影响指标对失业率的影响是决定性的。

## 五、问题 3：建立精确数学模型

### 5.1 通过问题二中模型细化得到精确数学模型

问题二中模型使用的影响指标 MACRO1~MACRO3、MICRO1~MICRO3 是按照下面的流程图得到的。

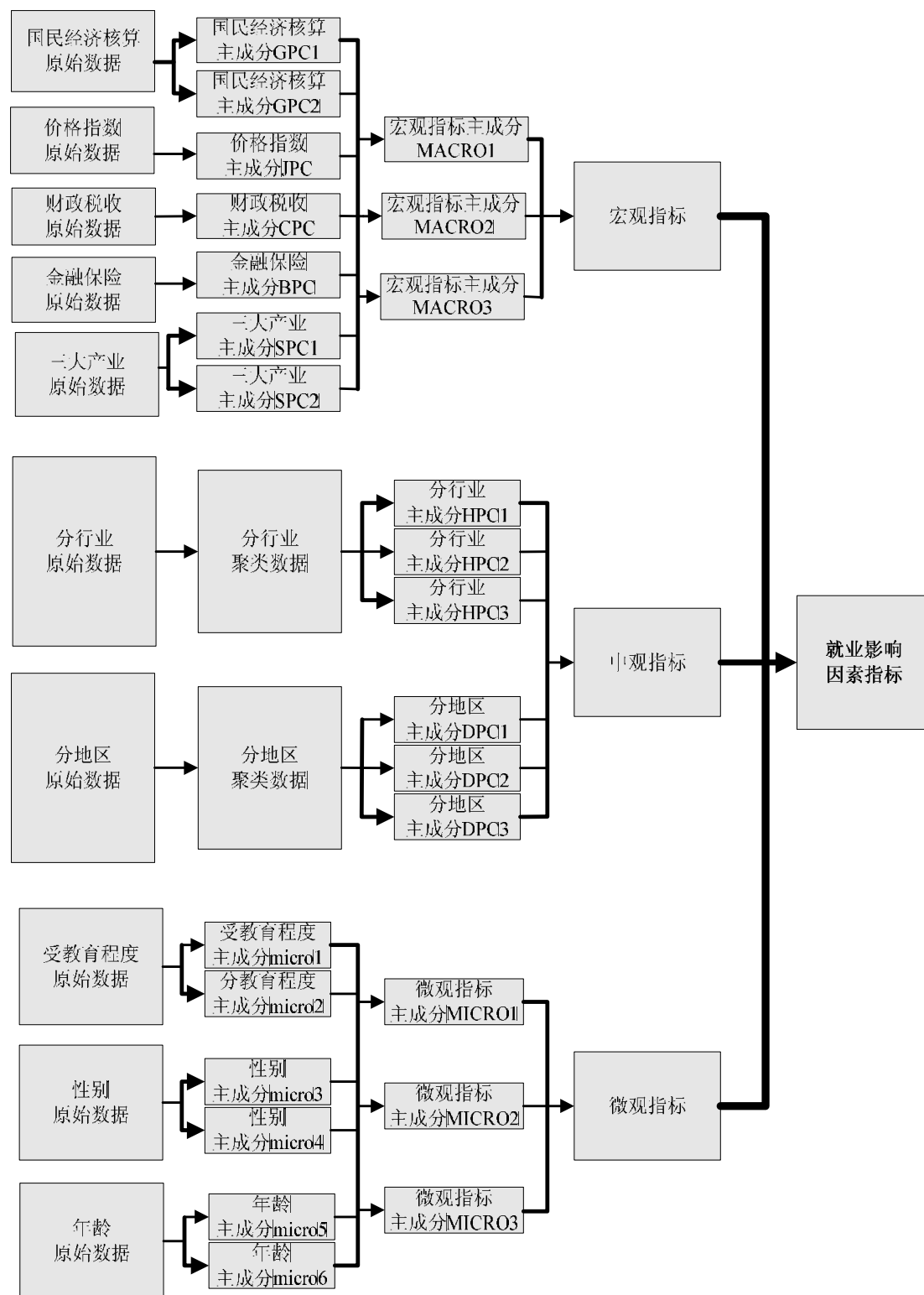


图 5.1 就业影响因素指标获取路线图

本文认为城镇登记失业率不是一个孤立的因素，他是众多因素共同作用的结果。从图中可以看出，从原始数据到就业影响因素指标可以分为五级，问题二中使用的最后一级即直接用的是就业影响因素指标来建模的，而数据每经过一次主成分分析晋级就会丢失一些信息，因此最精确的模型是基于原始数据的。

但是原始数据类型多数据量大，而且很多指标之间是有相关性的，因此对通过问题二中模型细化得到精确数学模型这一途径本文以定性的分析为主，不进行定量的分析，将主要的建模精力放在分行业、分地区精确建模上。

## 5.2 分行业建立精确数学模型（以农林牧渔业为例）

对某一行业的数据分析遵循如下的原则：

- (1)每个行业作为一个相对独立的系统来研究；
- (2)人力成本比例定义为总产值中用于给劳动者发报酬的部分，总产值定义为某一季度该行业的总收入；
- (3)预期提供岗位数=预期总产值×预期人力成本比例/预期平均劳动报酬；
- (4)预期求职人数、预期总产值、预期人力成本比例、预期劳动报酬可以由现在的就业人数、总产值、人力成本比例、平均劳动报酬预测估计出来；
- (5)岗位裕量=预期提供岗位数-预期求职人员；
- (6)国家投资算在总产值里面；
- (7)国家投资回报指数以每单位亿元投资新增的岗位数来度量，相当于总产值为 1，数学转化后就等于人力成本比例/平均劳动报酬×10000
- (8)行业内部失业率=（去年就业人数-今年就业人数）/去年就业人数，当行业内部失业率为正时表示今年就业人数减少，为负时表示今年就业人数增加。
- (9)最终某季度该行业的岗位增加包括行业自然增长的岗位裕量和国家投资产生的岗位数两部分，这些指标都是可正可负的。

下面以农林牧渔业为例来说明，农林牧渔业的数据整理<sup>[2,3]</sup>如表 5.1 所示。

表 5.1 农林牧渔业数据汇总表

季度 n	总产值 S(亿元)	就业人数 P(万人)	平均劳动报酬 M(元/季度)	人力成本比例 r
1	6044.31	7584.02	1403.94	0.1762
2	6569.90	8243.50	1420.63	0.1783
3	6307.10	7913.76	1437.31	0.1803
4	7358.29	9232.72	1454.00	0.1824
5	6299.88	7472.01	1470.69	0.1744
6	6847.70	8121.75	1487.38	0.1764
7	6573.79	7796.88	1504.06	0.1784
8	7669.42	9096.36	1520.75	0.1804
9	6829.11	7374.26	1541.63	0.1665
10	7422.95	8015.50	1562.50	0.1687
11	7126.03	7694.88	1583.38	0.1710
12	8313.70	8977.36	1604.25	0.1732
13	8334.99	7331.02	1633.44	0.1437
14	9059.78	7968.50	1662.63	0.1462
15	8697.38	7649.76	1691.81	0.1488
16	10146.95	8924.72	1721.00	0.1514
17	9073.71	7299.97	1759.31	0.1415
18	9862.73	7934.75	1797.63	0.1446

19	9468.22	7617.36	1835.94	0.1477
20	11046.25	8886.92	1874.25	0.1508
21	9386.48	7208.20	1906.13	0.1464
22	10202.70	7835.00	1938.00	0.1488
23	9794.59	7521.60	1969.88	0.1513
24	11427.02	8775.20	2001.75	0.1537
25	11245.41	7237.64	2031.81	0.1308
26	12223.28	7867.00	2061.88	0.1327
27	11734.34	7552.32	2091.94	0.1346
28	13690.07	8811.04	2122.00	0.1366

由上表中的数据通过三次多项式拟合可以得到预期总产值、预期求职人数、预期劳动报酬、预期人力成本比例关于季度的函数关系：

$$\begin{cases} S = 5929 + 169.6 \times n - 0.92 \times n^2 + 0.1275 \times n^3 \\ P = 7826 + 112.0 \times n - 10.9 \times n^2 + 0.2636 \times n^3 \\ M = 5656 + 11.45 \times n + 5.639 \times n^2 - 0.08589 \times n^3 \\ r = 0.1819 - 0.0004 \times n - 0.000149 \times n^2 + 0.000004 \times n^3 \end{cases} \quad (5.1)$$

$$\text{预期提供岗位数（万人）} Q = \frac{S \times r}{M} \times 10000 \quad (5.2)$$

$$\text{预期岗位裕量（万人）} R = \text{预期提供岗位数 } Q - \text{预期求职人数 } P \quad (5.3)$$

$$\text{国家投资回报指数（万人）} L = \frac{r}{M} \times 10000 \quad (5.4)$$

$$\text{行业内部失业率(\%)} C = \frac{P(n-1) - P(n)}{P(n-1)} \times 100\% \quad (5.5)$$

要考察的指标就是预期岗位裕量、国家投资回报指数和行业内部失业率，预期岗位裕量表征了该行业可以提供的额外工作岗位，国家投资回报指数表征了国家每单位投资增加了多少的岗位数，行业内部失业率表征了该行业的发展前景。这三个指标与季度  $n$  的函数关系如下所示：

$$\begin{aligned} \text{预期岗位裕量（万人）} R = & \frac{(5929 + 169.6 \times n - 0.92 \times n^2 + 0.1275 \times n^3) \times (0.1819 - 0.0004 \times n - 0.000149 \times n^2 + 0.000004 \times n^3)}{5656 + 11.45 \times n + 5.639 \times n^2 - 0.08589 \times n^3} \\ & \times 10000 - (7826 + 112.0 \times n - 10.9 \times n^2 + 0.2636 \times n^3) \end{aligned} \quad (5.6)$$

国家投资回报指数（万人）  $L =$

$$\frac{0.1819 - 0.0004 \times n - 0.000149 \times n^2 + 0.000004 \times n^3}{5656 + 11.45 \times n + 5.639 \times n^2 - 0.08589 \times n^3} \times 10000 \quad (5.7)$$

行业内部失业率(\%)  $C =$

$$\frac{-113.1636 + 22.5908 \times n - 0.7908 \times n^2}{7826 + 112.0 \times (n-1) - 10.9 \times (n-1)^2 + 0.2636 \times (n-1)^3} \times 100\% \quad (5.8)$$

下面是据此得到三个指标随季度  $n$  的变化示意图。



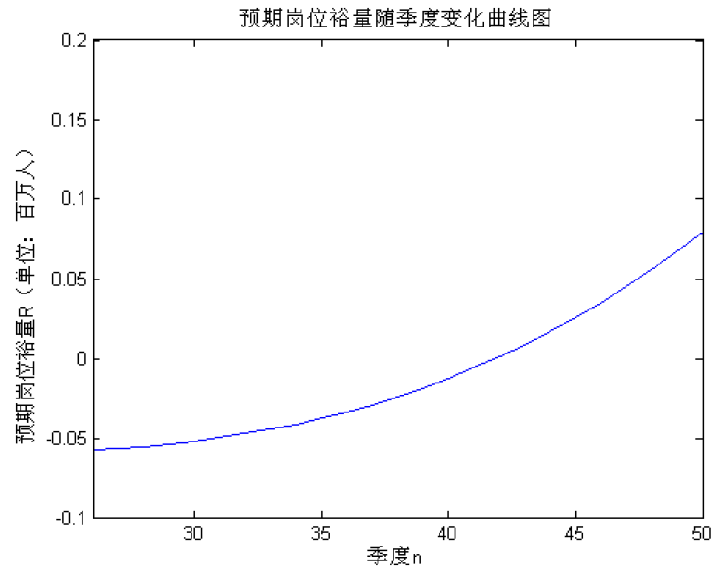


图 5.2 预期岗位裕度 R 随季度变化的曲线图

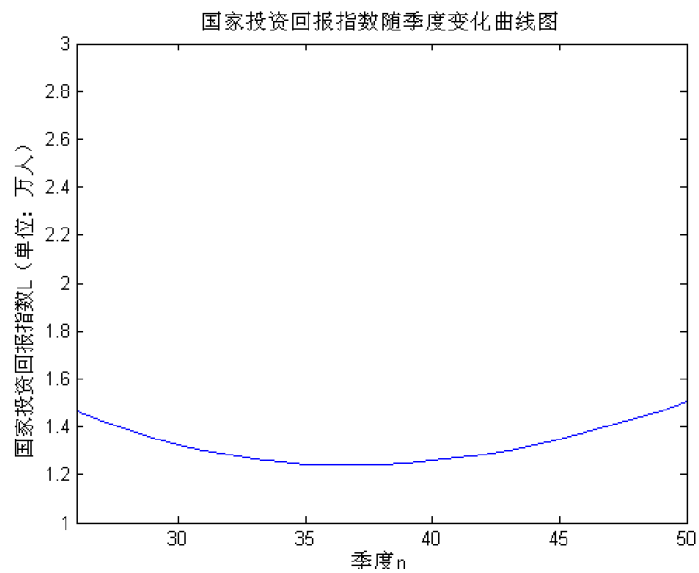


图 5.3 国家投资回报指数 L 随季度变化的曲线图

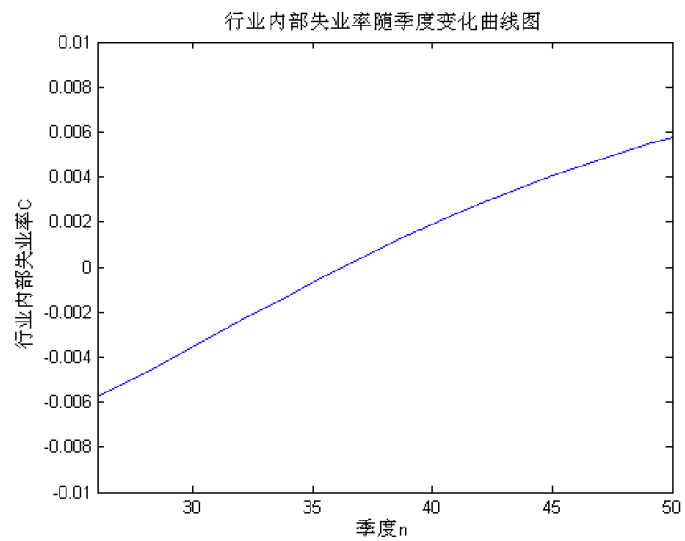


图 5.4 行业内部失业率 C 随季度变化的曲线图

按照同样的方法分析其他的行业，得到预期岗位裕量、国家投资回报指数和行业内部失业率三个指标。每个行业的预期岗位裕量求和可以得到未来全国的岗位裕量，行业国家投资回报指数指标有利于国家宏观调控，通过每个行业的内部失业率和从业人数就可以求得全国的失业率，这将在问题四和问题五的分析中加以详细的分析和讨论。

### 5.3 分地区建立精确数学模型（以北京为例）

分地区建模遵循下面的原则：

- 每个地区作为一个孤立的系统来分析；
- 人力成本比例定义为总产值中用于给劳动者发报酬的部分，总产值定义为某一季度该地区的总收入；
- 预期提供岗位数、预期求职人数、预期总产值、预期人力成本比例、预期劳动报酬、岗位裕量、国家投资、国家投资回报指数的定义与分行业建模一致。
- 这里的预期劳动报酬即为预期的职工平均工资；
- 地区内部失业率=（去年就业人数-今年就业人数）/去年就业人数，当地区内部失业率为正时表示今年就业人数减少，为负时表示增加。
- 最终某季度该地区的岗位增加包括地区自然增长的岗位裕量和国家投资产生的岗位数两部分，这些指标都是可正可负的。

北京地区的数据<sup>[2,3]</sup>汇总如表 5.2 所示。

表 5.2 北京地区数据汇总表

季度 n	登记就业人数 P（万人）	生产总值 S（万元）	职工平均工资 M(元/季度)	人力成本比例 r
1	489.945	5650614.7	18227.75	0.1582
2	480.32	6141972.5	18745.5	0.1705
3	470.695	5896293.6	19263.25	0.1626
4	461.07	6879009.2	19781	0.1886
5	451.445	7045130	20298.75	0.1922
6	441.82	7657750	20816.5	0.2082
7	432.195	7351440	21334.25	0.1993
8	422.57	8576680	21852	0.2322
9	440.1775	8425130	22717	0.2106
10	457.785	9157750	23582	0.2121
11	475.3925	8791440	24447	0.1891
12	493	10256680	25312	0.2055
13	493.125	9851613	26402.5	0.1892
14	493.25	10708275	27493	0.1974
15	493.375	10279944	28583.5	0.1822
16	493.5	11993268	29674	0.2047
17	492.6025	15838513.23	30803.25	0.2610
18	491.705	17215775.25	31932.5	0.2741
19	490.8075	16527144.24	33061.75	0.2546
20	489.91	19281668.28	34191	0.2878
21	494.8325	18101652.05	35672.5	0.2564

22	499.755	19675708.75	37154	0.2649
23	504.6775	18888680.4	38635.5	0.2422
24	509.6	22036793.8	40117	0.2695
25	523.9375	21512636	41714.75	0.2461
26	538.275	23383300	43312.5	0.2507
27	552.6125	22447968	44910.25	0.2261
28	566.95	26189296	46508	0.2483

由上表中的数据通过三次多项式拟合可以得到预期总产值 S、预期求职人数 P、预期职工平均工资 M、预期人力成本比例 r 关于季度的函数关系：

$$\begin{cases} S = 6487768 - 237015 \times n + 55294 \times n^2 - 826.3 \times n^3 \\ P = 481.8 - 7.081 \times n + 0.5707 \times n^2 + 0.008327 \times n^3 \\ M = 4512 + 71.87 \times n + 6.963 \times n^2 - 0.01527 \times n^3 \\ r = 0.1736 - 0.000228 \times n - 0.000461 \times n^2 + 0.000013 \times n^3 \end{cases} \quad (5.9)$$

按照预期岗位裕量、国家投资回报指数和地区内部失业率的定义，则有  
预期岗位裕量（万人）Q=

$$\frac{(6487768 - 237015 \times n + 55294 \times n^2 - 826.3 \times n^3) \times (0.1736 - 0.000228 \times n - 0.000461 \times n^2 + 0.000013 \times n^3)}{4512 + 71.87 \times n + 6.963 \times n^2 - 0.01527 \times n^3} \times 10000 - (481.8 - 7.081 \times n + 0.5707 \times n^2 + 0.008327 \times n^3) \quad (5.10)$$

国家投资回报指数（万人）L=

$$\frac{0.1736 - 0.000228 \times n - 0.000461 \times n^2 + 0.000013 \times n^3}{4512 + 71.87 \times n + 6.963 \times n^2 - 0.01527 \times n^3} \times 10000 \quad (5.11)$$

地区内部失业率(%)C=

$$\frac{7.643373 - 1.116419 \times n + 0.024981 \times n^2}{475.281373 - 8.197419 \times n + 0.545719 \times n^2 + 0.008327 \times n^3} \times 100\% \quad (5.12)$$

下面是据此得到三个指标随季度 n 的变化示意图。

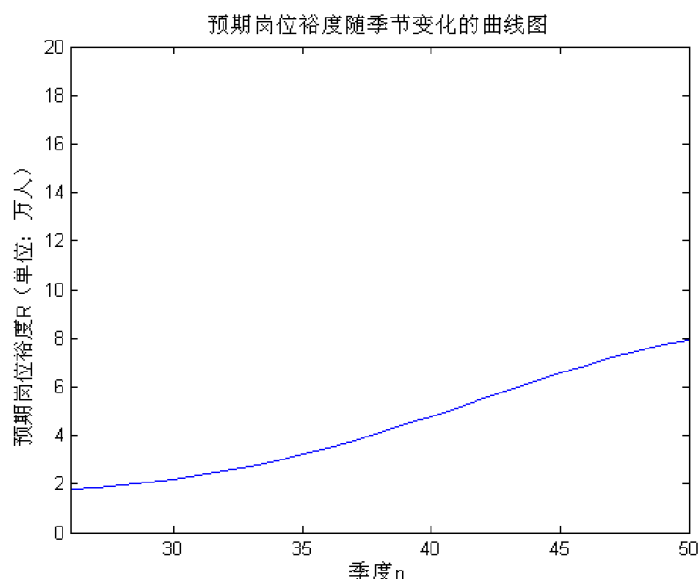


图 5.5 分地区讨论预期岗位裕度 R 随季度变化的曲线图

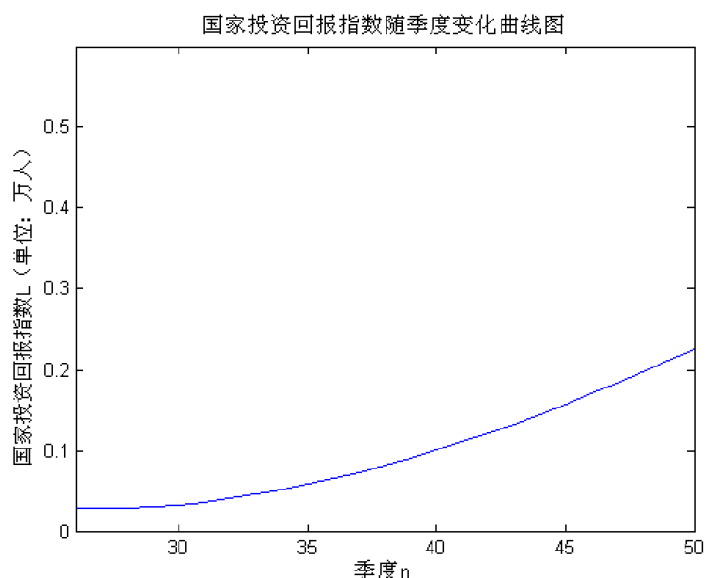


图 5.6 分地区讨论国家投资回报指数 L 随季度变化的曲线图

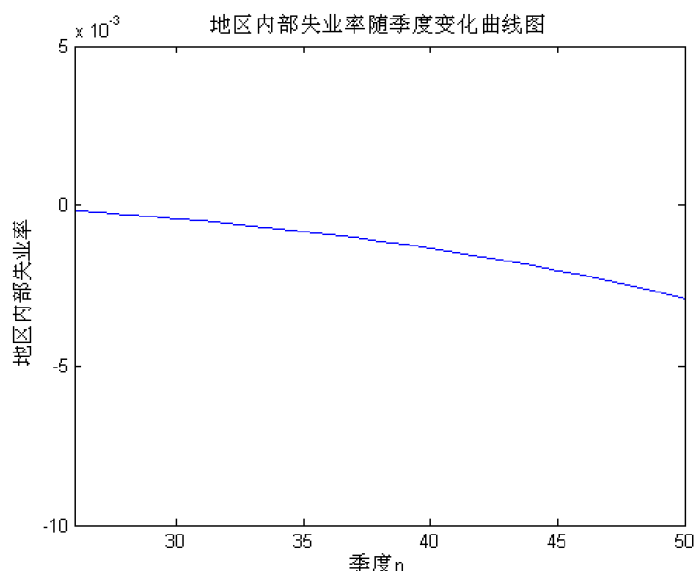


图 5.7 分地区讨论地区内部失业率 C 随季度变化的曲线图

按照同样的方法分析其他的地区，得到预期岗位裕量、国家投资回报指数和地区内部失业率三个指标，就可以的进一步预测全国的就业前景。

## 六、问题 4：就业前景预测

### 6.1 用问题二中国家标准建立的模型进行预测

由公式(4.5)可知，按照国家标准计算失业率的公式为

$$u = 3.66 + 0.36 \times n - 0.03 \times n^2 + 0.0006 \times n^3 \quad (6.1)$$

由于季度计数是以 2001 年第一季度为 1，所以 2009 年和 2010 年上半年对应为  $n=33 \sim 38$ ，带入公式(6.1)中，计算结果如表 6.1 所示。

表 6.1 按照国家标准建立的模型进行失业率预测数据表

季度	失业率 (%)
2009 年第一季度	4.52
2009 年第二季度	4.40
2009 年第三季度	4.24
2009 年第四季度	4.73
2010 年第一季度	4.30
2010 年第二季度	4.94

## 6.2 用问题二中国际标准建立的模型进行预测

由公式(4.7)可知，按照国际标准计算失业率的公式为

$$U = 5.71 + 0.43 \times n - 0.038 \times n^2 + 0.0008 \times n^3 \quad (6.2)$$

则计算得到的 2009 年和 2010 年上半年的预测失业率如下表 6.2 所示。

表 6.2 按照国际标准建立的模型进行失业率预测数据表

季度	失业率 (%)
2009 年第一季度	5.27
2009 年第二季度	5.85
2009 年第三季度	5.51
2009 年第四季度	6.27
2010 年第一季度	6.12
2010 年第二季度	6.08

## 6.3 用分行业建立的精确数学模型来预测

仿照第 5.2 部分的失业率模型预测我国 16 个行业的失业率情况，得到各行业 2009 年和 2010 年上半年的预测失业率和预测就业人数如表 6.3 和 6.4 所示。

表 6.3 各个行业 2009 年和 2010 年上半年的预测失业率（单位：%）

行业 \ 时间	2009 年 第一季度	2009 年 第二季度	2009 年 第三季度	2009 年 第四季度	2010 年 第一季度	2010 年 第二季度
A	1.02	0.58	0.64	0.89	1.32	1.59
B	1.34	2.03	0.45	1.26	1.56	1.83
C	4.63	4.89	5.32	5.97	6.81	6.34
D	1.42	1.56	1.74	1.93	2.03	2.36
E	5.04	5.72	6.37	6.71	6.84	6.87
F	0.36	0.63	1.03	1.30	1.24	1.41
G	3.52	3.63	3.79	4.29	4.51	4.75
H	4.91	4.28	4.62	5.94	5.48	6.32
I	5.36	5.61	5.77	5.92	5.98	6.11
J	3.87	3.25	4.07	4.26	4.71	4.89
K	4.57	4.61	4.97	5.02	5.25	5.64
L	0.32	0.65	1.39	1.06	2.08	2.43
M	2.98	3.42	2.45	3.18	3.59	4.31
N	1.54	0.87	1.74	2.01	1.38	2.14
O	0.21	0.11	1.34	1.36	1.31	2.05
P	4.30	3.96	4.48	4.63	5.29	5.71

表 6.4 各个行业 2009 年和 2010 年上半年的预测就业人数（单位：万人）

行业\时间	2009 年 第一季度	2009 年 第二季度	2009 年 第三季度	2009 年 第四季度	2010 年 第一季度	2010 年 第二季度
A	7832.15	7379.26	7538.03	7629.65	7237.64	7108.25
B	133.59	141.85	138.36	135.63	145.17	140.86
C	2246.94	2365.39	2417.43	2382.96	2479.84	2595.06
D	156.49	163.08	162.56	177.84	179.27	185.83
E	1102.93	1193.04	1160.39	1232.61	1284.70	1307.35
F	22.78	24.06	24.39	23.18	25.49	28.21
G	619.25	657.20	651.93	689.41	700.39	694.39
H	1394.25	1437.02	1441.86	1473.49	1603.42	1643.98
I	139.48	141.26	149.35	146.57	158.73	169.97
J	48.07	47.36	51.73	55.67	60.51	63.29
K	357.82	376.25	374.68	398.06	413.87	4407.51
L	139.25	135.07	149.23	155.74	162.87	175.08
M	491.03	500.48	513.29	563.44	593.27	605.49
N	64.89	63.27	66.49	67.85	70.32	72.47
O	277.35	294.57	312.14	341.12	353.98	357.34
P	4523.51	4702.43	5286.32	5309.35	5515.38	5803.57

用  $m(i)$  表示某行业内部失业率 ( $i=1\sim 16$ , 代表十六个行业),  $n(i)$  表示该行业预测就业人数, 则有

$$\text{该行业总人数 } p(i) = \frac{n(i)}{1-m(i)} \quad (6.3)$$

$$\text{该行业失业人数 } q(i) = m(i) \times p(i) \quad (6.4)$$

$$\text{总的失业率为 } U = \sum_{i=1}^{16} q(i) / \sum_{i=1}^{16} p(i) \times 100\% \quad (6.5)$$

$$\text{总的就业人数为 } N = \sum_{i=1}^{16} n(i) \quad (6.6)$$

据此的得到的全国（全行业）失业率数据和就业人数数据如表 6.5 所示。

表 6.5 分行业建模分析得到的全国（全行业）失业率数据和就业人数数据

	2009 年 第一季度	2009 年 第二季度	2009 年 第三季度	2009 年 第四季度	2010 年 第一季度	2010 年 第二季度
失业率 (%)	4.32	4.83	4.42	4.09	4.85	4.70
就业人数 (万人)	19549.78	19621.59	20438.18	20384.51	20984.85	21358.65

#### 6.4 用分地区建模建立的精确数学模型来预测

首先仿照第 5.3 部分北京的失业率模型预测要考察的我国其他 29 个地区的失业率情况, 得到各个地区 2009 年和 2010 年上半年的预测失业率如表 6.5 所示, 预测就业人数如表 6.6 所示。

表 6.6 各个地区 2009 年和 2010 年上半年的预测失业率（单位：%）

季度 地区	2009 年 第一季度	2009 年 第二季度	2009 年 第三季度	2009 年 第四季度	2010 年 第一季度	2010 年 第二季度
北 京	2.21	2.41	2.62	2.87	3.15	3.47
天 津	3.83	4.2	4.64	5.13	5.7	6.34
河 北	4.30	4.66	5.09	5.56	6.12	6.74
山 西	3.99	4.32	4.72	5.16	5.68	6.25
内蒙古	4.41	4.78	5.21	5.70	6.27	6.91
辽 宁	4.61	5.00	5.46	5.97	6.56	7.23
吉 林	3.90	4.23	4.62	5.05	5.55	6.12
黑龙江	4.54	4.91	5.36	5.87	6.45	4.54
上 海	4.28	4.64	5.07	5.54	6.09	6.71
江 苏	4.11	4.46	4.87	5.32	5.85	6.44
浙 江	3.48	3.77	4.12	4.50	4.95	5.45
安 徽	4.32	4.69	5.12	5.59	6.15	6.77
福 建	4.20	4.55	4.97	5.43	5.97	6.58
江 西	3.59	3.89	4.24	4.64	5.10	5.62
山 东	3.38	3.66	3.99	4.37	4.80	5.29
河 南	3.67	3.98	4.34	4.75	5.22	5.75
湖 北	4.41	4.78	5.21	5.70	6.27	6.91
湖 南	4.54	4.91	5.36	5.87	6.45	7.11
广 东	2.64	2.86	3.12	3.41	3.75	4.13
广 西	4.01	4.34	4.74	5.18	5.70	6.28
海 南	4.76	5.15	5.63	6.15	6.77	7.45
重 庆	4.22	4.57	4.99	5.46	6.00	6.61
四 川	4.43	4.80	5.24	5.73	6.30	6.94
贵 州	4.22	4.57	4.99	5.46	6.00	6.61
云 南	4.42	4.79	5.23	5.72	6.29	6.92
陕 西	4.22	4.57	4.99	5.46	6.00	6.61
甘 肃	3.48	3.77	4.12	4.50	4.95	5.45
青 海	4.01	4.34	4.74	5.18	5.70	6.28
宁 夏	4.54	4.91	5.36	5.87	6.45	7.11
新 疆	4.11	4.46	4.87	5.32	5.85	6.44

表 6.7 各个地区 2009 年和 2010 年上半年的预测就业人数（单位：万人）

季度 地区	2009 年 第一季度	2009 年 第二季度	2009 年 第三季度	2009 年 第四季度	2010 年 第一季度	2010 年 第二季度
北 京	469.04	492.83	462.59	500.88	545.19	530.85
天 津	376.64	388.67	379.31	358.72	346.89	313.09
河 北	651.94	614.66	587.18	629.69	621.80	670.21
山 西	387.31	395.22	416.63	429.68	427.20	458.47
内蒙古	401.20	409.62	394.37	425.23	395.40	375.97
辽 宁	919.83	949.18	975.41	973.10	902.23	914.77
吉 林	588.17	550.92	604.53	664.10	668.41	622.86

黑龙江	663.02	681.13	681.52	664.26	694.25	707.36
上海	596.80	557.45	607.36	611.18	598.79	578.57
江苏	916.07	923.33	983.90	921.19	895.11	923.50
浙江	793.07	812.64	880.51	880.85	843.02	904.33
安徽	601.77	624.14	575.84	566.88	608.68	616.91
福建	340.03	352.10	374.08	386.08	395.84	433.88
江西	695.65	748.08	809.16	837.26	793.92	840.26
山东	1141.60	1030.38	959.57	1047.34	1147.48	1067.75
河南	717.85	690.63	638.44	599.09	615.92	656.95
湖北	1138.72	1202.28	1265.44	1167.04	1103.98	1035.95
湖南	1639.39	1576.21	1646.13	1667.55	1728.02	1776.06
广东	866.97	940.96	969.49	1060.46	1095.63	1132.67
广西	370.40	383.64	403.13	364.73	338.09	356.49
海南	122.59	112.15	115.81	124.39	112.51	109.81
重庆	304.18	278.07	299.39	271.06	258.17	255.15
四川	783.22	706.77	674.56	677.20	625.26	623.14
贵州	261.69	247.41	243.42	228.44	208.76	213.28
云南	317.83	318.86	300.56	313.52	335.65	313.90
陕西	582.32	577.46	523.82	497.70	465.88	419.48
甘肃	227.52	236.77	216.93	235.76	213.71	226.12
青海	77.88	79.16	84.71	78.57	82.24	82.47
宁夏	92.61	92.78	89.82	90.21	90.87	85.65
新疆	272.72	249.50	247.82	267.40	255.42	235.17

用  $M(i)$  表示某地区内部失业率 ( $i=1\sim 30$ , 代表 30 个地区),  $N(i)$  表示该地区预测就业人数, 则有

$$\text{该行业总人数 } P(i) = \frac{N(i)}{1 - M(i)} \quad (6.7)$$

$$\text{该行业失业人数 } Q(i) = M(i) \times P(i) \quad (6.8)$$

$$\text{总的失业率 } U = \sum_{i=1}^{30} Q(i) / \sum_{i=1}^{30} P(i) \times 100\% \quad (6.9)$$

$$\text{总的就业人数为 } A = \sum_{i=1}^{30} N(i) \quad (6.10)$$

据此的得到的全国失业率数据和就业人数数据如表 6.8 所示。

表 6.8 分地区建模得到的全国失业率数据和就业人数数据

	2009 年 第一季度	2009 年 第二季度	2009 年 第三季度	2009 年 第四季度	2010 年 第一季度	2010 年 第二季度
失业率 (%)	4.38	4.82	4.33	4.91	4.57	4.92
就业人数 (万人)	18422.3	18924.81	18962.04	19199.59	19419.86	20422.3



## 6.5 预测结果的对比分析

对比表 6.5、表 6.8 和表 4.4 中 2009 年第一季度和第二季度的失业率数据可以看出，模型预测的结果和统计数据还是比较一致的。

对比表 6.1、表 6.2、表 6.5、表 6.8 的结果可以看出，我国的就业人数和失业率都是在不断缓慢增加的，这就意味着我国的可用劳动力增长超过了市场消化接纳新劳动力的能力，这就要求为增加就业提供更多的投资。

## 七、问题 5：建议

### 7.1 预期岗位裕量和国家投资回报指数分析

依据 5.2 和 5.3 中分行业、分地区建模分析，考察预期岗位裕量（万人）和国家投资回报指数（万人）这两个指标，16 个行业、30 个地区在 2009 年到 2010 年上半年预期岗位裕量（万人）和国家投资回报指数（万人）如表 7.1~7.4 所示。

表 7.1 16 个行业 2009 年到 2010 年六个季度的预期岗位裕量（单位：万人）

行业代号	09 年一季度	09 年二季度	09 年三季度	09 年四季度	10 年一季度	10 年二季度
A	3025.25	2972.41	3159.29	3302.73	3387.27	3416.63
B	37.05	39.52	46.48	42.57	51.56	59.36
C	936.84	1146.83	1286.34	1145.39	1265.93	1305.28
D	24.59	31.04	19.34	20.59	15.34	12.93
E	524.65	513.95	492.73	425.98	395.03	450.29
F	6.46	8.10	7.48	8.94	9.76	11.39
G	194.83	215.37	239.09	208.74	284.39	310.27
H	493.58	519.47	532.96	509.37	593.06	632.38
I	39.4	52.94	41.43	49.37	42.40	38.36
J	21.49	22.46	19.48	20.58	14.96	17.38
K	153.68	148.74	168.93	159.37	174.39	164.18
L	38.59	41.39	50.85	47.29	55.37	61.60
M	179.05	199.47	191.62	225.49	241.06	249.75
N	25.78	20.59	28.51	33.59	34.09	38.27
O	94.67	85.08	104.78	119.52	122.94	143.68
P	1685.38	1904.38	1764.93	1975.75	1840.61	1759.07

表 7.2 16 个行业 2009 年到 2010 年六个季度的国家投资回报指数（单位：万人）

行业代号	09 年一季度	09 年二季度	09 年三季度	09 年四季度	10 年一季度	10 年二季度
A	0.35	0.31	0.36	0.42	0.48	0.41
B	0.76	0.82	0.73	0.94	0.99	0.87
C	1.56	1.67	1.83	1.78	1.89	1.95
D	0.34	0.23	0.28	0.22	0.23	0.25
E	0.19	0.17	0.20	0.18	0.15	0.21
F	0.45	0.49	0.51	0.47	0.54	0.54
G	1.14	1.22	1.19	1.26	1.32	1.31
H	1.35	1.48	1.41	1.52	1.58	1.55
I	0.05	0.06	0.05	0.08	0.06	0.06
J	0.32	0.28	0.25	0.34	0.31	0.33
K	1.52	1.56	1.49	1.46	1.61	1.66

L	0.87	0.73	0.79	0.89	0.88	0.83
M	1.09	1.20	1.17	1.24	1.29	1.28
N	0.56	0.64	0.60	0.65	0.72	0.73
O	0.21	0.18	0.22	0.23	0.19	0.25
P	0.46	0.44	0.46	0.49	0.51	0.55

表 7.3 30 个地区 2009 年到 2010 年六个季度的预期岗位裕量（单位：万人）

	09 一季度	09 二季度	09 三季度	09 四季度	10 一季度	10 二季度
北 京	90.05	92.89	98.76	100.72	96.71	94.70
天 津	36.41	40.05	39.73	33.14	35.45	28.51
河 北	60.16	64.82	63.02	58.80	59.69	63.39
山 西	110.57	111.41	115.75	121.07	105.76	107.38
内 蒙 古	40.61	36.87	39.47	43.01	36.47	36.18
辽 宁	114.12	115.71	120.06	125.03	127.46	130.94
吉 林	60.72	49.66	63.99	72.42	61.46	65.04
黑 龙 江	67.85	67.45	67.45	62.86	68.00	77.16
上 海	61.89	52.35	61.03	60.86	57.88	53.59
江 苏	91.79	97.64	91.93	92.61	97.46	84.38
浙 江	81.70	83.19	91.55	93.24	94.01	93.65
安 徽	60.36	56.37	54.35	53.21	57.93	57.57
福 建	34.72	37.96	34.46	41.77	35.97	46.96
江 西	66.06	68.72	69.11	60.80	61.55	62.91
山 东	51.51	51.43	55.13	54.54	56.46	59.27
河 南	69.17	72.63	67.52	63.11	64.60	62.66
湖 北	115.87	117.36	121.00	127.15	128.17	129.81
湖 南	71.81	92.30	95.55	100.20	107.56	105.17
广 东	82.44	94.17	98.88	115.04	104.11	114.38
广 西	39.29	38.86	36.34	34.28	31.40	38.80
海 南	12.67	11.81	11.79	12.87	11.59	10.61
重 庆	120.46	121.36	130.19	121.42	125.12	123.56
四 川	100.29	105.45	109.53	115.60	116.48	117.84
贵 州	25.93	26.24	23.43	23.28	22.67	21.68
云 南	34.36	31.70	28.44	32.08	32.29	33.78
陕 西	75.70	84.31	91.50	94.83	104.43	94.04
甘 肃	20.77	24.05	20.82	25.85	21.53	22.07
青 海	7.75	8.18	8.76	8.48	7.67	8.62
宁 夏	10.16	9.61	9.77	9.37	8.56	7.99
新 疆	29.58	27.16	24.00	26.42	24.10	25.66

表 7.4 30 个地区 2009 年到 2010 年六个季度的国家投资回报指数（单位：万人）

	09 一季度	09 二季度	09 三季度	09 四季度	10 一季度	10 二季度
北 京	5.60	6.10	6.65	6.97	7.14	7.32
天 津	2.26	2.60	2.11	2.07	2.16	1.51
河 北	4.22	3.51	3.61	3.56	3.68	3.45
山 西	7.12	7.19	7.80	7.33	7.61	7.79

内蒙古	2.17	2.01	2.50	2.61	2.32	1.93
辽 宁	8.05	8.42	9.00	9.55	9.93	8.89
吉 林	3.91	2.71	3.50	4.63	3.30	3.70
黑龙江	4.35	3.75	3.82	3.69	4.17	4.90
上 海	3.44	3.02	3.40	3.29	3.71	3.03
江 苏	5.90	5.32	4.88	5.64	5.34	4.71
浙 江	5.03	5.64	5.70	6.10	6.20	6.85
安 徽	3.23	3.60	3.30	2.96	3.13	3.62
福 建	2.08	2.18	1.86	2.57	2.10	2.59
江 西	4.90	4.64	4.66	4.88	4.52	4.59
山 东	4.74	4.02	4.13	4.08	4.70	4.37
河 南	3.81	3.86	4.14	3.49	3.61	3.56
湖 北	7.24	7.43	7.52	7.97	7.69	7.61
湖 南	4.55	9.58	8.67	10.71	10.95	11.73
广 东	5.13	5.96	5.79	6.09	6.30	7.08
广 西	2.09	2.15	2.04	1.95	2.01	2.45
海 南	0.81	0.76	0.72	0.73	0.73	0.61
重 庆	7.74	7.80	7.95	7.57	7.64	8.65
四 川	5.4	4.45	4.20	3.55	4.01	3.70
贵 州	1.67	1.53	1.39	1.35	1.21	1.17
云 南	1.88	1.86	1.79	2.01	1.92	1.94
陕 西	5.01	5.05	6.21	6.84	6.59	7.02
甘 肃	1.17	1.29	1.26	1.37	1.36	1.30
青 海	0.42	0.46	0.54	0.52	0.48	0.56
宁 夏	0.58	0.53	0.61	0.59	0.50	0.46
新 疆	1.58	1.73	1.54	1.62	1.50	1.44

预期岗位裕量（万人）代表某个行业或地区自然产生的岗位数，国家投资回报指数（万人）表示国家投资 1 亿元能够产生的岗位数，这两个指标对增进我国就业人数、降低失业率有重大意义。本文选择了国家应该优先投资的十个行业和十个地区，如表 7.5 所示，行业和地区选择的依据是 2009 年到 2010 年六个季度的这两个指标绝对值大且处于上升过程中。

表 7.5 国家应该优先投资的十个行业和十个地区

国家优先投资的十个行业	国家优先投资的十个地区
制造业	辽宁
交通运输仓储和邮电通信业	重庆
农林牧渔业	湖北
采掘业	山西
地质勘查和水利管理业	四川
社会服务业	北京
科学研究和综合技术服务业	广东
教育、文化和娱乐业	浙江
卫生、体育和社会福利业	陕西
批发零售贸易和餐饮业	湖南

## 7.2 国家经济振兴计划探讨

国家经济振兴计划又称为十大产业振兴计划<sup>[7]</sup>，它是我国积极应对国际金融危机、进一步扩大内需的重要举措。国务院确定的十大产业如表 7.6 所示。

表 7.6 国务院确定振兴的十大产业

编号	产业名称	隶属的行业
1	汽车工业	制造业
2	钢铁工业	制造业、采掘业
3	纺织工业	制造业
4	装备制造业	制造业
5	船舶工业	制造业、交通运输业
6	信息产业	邮电通信业、科学研究和综合技术服务业
7	轻工业	制造业、农林牧渔业
8	石化产业	地质勘查业、交通运输业
9	有色金属产业	采掘业、地质勘查业
10	物流业	社会服务业、批发零售贸易业、社会福利业

对比表 7.5 和表 7.6，从行业的角度看，国务院确定的十大振兴产业大部分属于制造业，而且其相关行业也全部在表 7.5 确定的国家应该优先投资的十个行业中；从地区来看，辽宁、重庆、湖北、四川、陕西都是我国的重工业基地，山西是我国的能源基地，北京是我国信息产业的中心，广东、浙江是现代制造业和轻工业发展最快的城市，湖南是社会服务业发展最快的中部城市。因此不管是从行业的角度还是从地区的角度看，国家的振兴计划与模型预测分析得到的结果是比较吻合的，既说明了国家振兴计划制定的合理性，也验证了模型的正确性。

## 7.3 对我国促进就业政策的建议

经济增长是扩大就业的最根本途径，但是经济增长并不必然带来就业机会的增加，其原因就在于就业弹性系数的差别。就业弹性系数定义为 GDP 每增长 1%，就业需求量相应变动的百分比<sup>[6]</sup>，计算公式为：

$$\text{就业弹性系数} = \text{就业增长率} / \text{经济增长率（即 GDP 增长率）} \quad (7.1)$$

经济增长与就业增长的不一致性主要表现在经济的快速增长伴随就业增长率的逐步下降。表 7.7 是我国 1991～2002 年间的就业弹性系数变化。

表 7.7 我国 1991～2002 年间的就业弹性系数变化

年份	就业增长率（%）	GDP 增长率（%）	就业弹性系数
1991	1.39	9.2	0.15
1992	1.17	14.2	0.08
1993	1.25	13.5	0.09
1994	1.24	12.6	0.10
1995	1.13	10.5	0.11
1996	1.33	9.6	0.14
1997	1.09	8.8	0.12
1998	0.86	7.8	0.11
1999	0.89	7.1	0.12
2000	0.80	8.0	0.10
2001	1.28	7.5	0.17
2002	0.97	8.0	0.12

从表 7.7 中可以看出,经济增长并不完全带来就业增长,因此在我国经济快速发展的大环境下更加要解决就业问题,降低失业率。

我国的十大产业振兴计划主要是针对关乎国家经济命脉的支柱产业,这对提高我国的综合实力和国际竞争力有重大意义,但是关乎国计民生的行业诸如教育、文化和娱乐业、卫生、体育和社会福利业、批发零售贸易和餐饮业等也需要增加投入,以提高人民的生活水平。综合上面的分析,对我国促进就业政策的建议归纳为以下几点:

(1) 从宏观层面上来看,影响就业的主要是国民生产总值、居民价格指数、财政税收、金融保险和三大产业的发展等五个方面。在国民生产总值方面,要不断发展经济,打牢增加就业的经济基础;在居民价格指数方面,通过宏观调控不断提高人民的生活水平和生活素质;在财政税收方面,建立更加合理的税收体系,促进经济社会健康发展;在金融保险方面,建立完善的金融体系和社会保险体系;在三大产业方面,实现第一产业、第二产业、第三产业的协调发展。

(2) 从中观层面上看,影响就业的主要是地区和行业。按照第 7.1 部分分析的结果,在兼顾协调发展的前提下应该重点增加对提供工作岗位最有力的地区和行业增加投资。

(3) 从微观层面上来看,受教育程度是应该关注而且可以提高的因素。目前我国的劳动力市场是结构性失衡,一方面需要大量的高素质技术工人,另一方面是大量的大学毕业生找不到工作,这就需要国家加大政策引导力度,完善劳动者技能培训与进修体系,将知识转化成技能,将人力转化成生产力。

我国的积极的就业政策就是三位一体的“劳动者自主择业,市场调节就业,政府促进就业”。只要宏观上政府合理引导,中观上市场健康运行,微观上劳动者积极的提高自身素质,就业问题就会随着我国经济社会的发展逐步解决。

## 参考文献

- [1] 百度百科, 经济学, <http://baike.baidu.com/view/31551.htm>, 2009 年 9 月 18 日~2009 年 9 月 22 日
- [2] 和讯财经新闻, 宏观数据, <http://news.hexun.com/>, 2009 年 9 月 18 日~2009 年 9 月 22 日
- [3] 中国国家统计局, 统计年鉴, <http://www.stats.gov.cn/tjsj/ndsj/>, 2009 年 9 月 18 日~2009 年 9 月 22 日
- [4] 陈建成, 统计数据分析理论与方法, 北京: 中国林业出版社, 2006.10
- [5] 中国国家统计局, 国民经济行业分类, <http://www.stats.gov.cn/tjbz/>, 2009 年 9 月 18 日~2009 年 9 月 22 日
- [6] 夏杰长, 增长 就业与公共政策, 北京: 社会科学文献出版社, 2005.7
- [7] CCTV 财经频道, 聚焦国务院行业振兴计划, <http://finance.cctv.com/special/10dazhenxingjihua/01/>, 2009 年 9 月 22 日
- [8] 杭州市改革和发展委员会, 经济形势分析, <http://www.hzdpc.gov.cn/hzdpc/0902/12022.htm>, 2009 年 9 月 22 日

## 附录

附表 1 2001~2007 年按季度的国民经济核算统计数据表

季度	国内生产总值增长率 (%)	居民消费指数 (%)	固定资产投资增长率 (%)	进出口贸易总额 (亿元)	人均国内生产总值 (元)
1	8.38	24.73	10.98	1172.20	1978
2	8.35	26.88	11.65	1274.13	2150
3	8.33	25.80	12.33	1223.16	2064
4	8.30	30.11	13.00	1427.02	2408
5	8.50	24.92	13.98	1427.77	1994
6	8.70	27.09	14.95	1551.93	2167
7	8.90	26.00	15.93	1489.85	2080
8	9.10	30.34	16.90	1738.16	2427
9	9.33	25.15	19.60	1957.27	2012
10	9.55	27.34	22.30	2127.47	2187
11	9.78	26.24	25.00	2042.37	2099
12	10.00	30.62	27.70	2382.77	2449
13	10.03	25.17	27.43	2655.47	2014
14	10.05	27.36	27.15	2886.38	2189
15	10.08	26.26	26.88	2770.92	2101
16	10.10	30.64	26.60	3232.74	2451
17	10.18	25.25	26.45	3270.39	2020
18	10.25	27.45	26.30	3554.78	2196
19	10.33	26.35	26.15	3412.58	2108
20	10.40	30.74	26.00	3981.35	2459
21	10.70	25.53	25.48	4048.91	2042
22	11.00	27.75	24.95	4400.99	2220
23	11.30	26.64	24.43	4224.95	2131
24	11.60	31.08	23.90	4929.11	2486
25	11.95	25.62	24.13	4999.81	2050
26	12.30	27.85	24.35	5434.58	2228
27	12.65	26.74	24.58	5217.19	2139
28	13.00	31.19	24.80	6086.72	2495

附表 2 2001~2007 年按季度的价格指数统计数据表

季度 (以 2001 年第一季度为 1)	城市居民消费价格指数 (上年=100)	农村居民消费价格指数 (上年=100)	商品零售价格指数 (上年=100)	工业品出厂价格指数 (上年=100)	原材料、燃料购进价格指数 (上年=100)	固定资产投资价格指数 (上年=100)
1	101.83	101.70	99.58	99.38	101.83	101.70
2	101.45	101.40	99.45	99.15	101.45	101.40
3	101.08	101.10	99.33	98.93	101.08	101.10

4	100.70	100.80	99.20	98.70	100.70	100.80
5	100.33	100.50	99.08	98.48	100.33	100.50
6	99.95	100.20	98.95	98.25	99.95	100.20
7	99.58	99.90	98.83	98.03	99.58	99.90
8	99.20	99.60	98.70	97.80	99.20	99.60
9	99.70	100.10	99.00	98.93	99.70	100.10
10	100.20	100.60	99.30	100.05	100.20	100.60
11	100.70	101.10	99.60	101.18	100.70	101.10
12	101.20	101.60	99.90	102.30	101.20	101.60
13	101.88	102.40	100.63	103.25	101.88	102.40
14	102.55	103.20	101.35	104.20	102.55	103.20
15	103.23	104.00	102.08	105.15	103.23	104.00
16	103.90	104.80	102.80	106.10	103.90	104.80
17	103.38	104.15	102.30	105.80	103.38	104.15
18	102.85	103.50	101.80	105.50	102.85	103.50
19	102.33	102.85	101.30	105.20	102.33	102.85
20	101.80	102.20	100.80	104.90	101.80	102.20
21	101.73	102.03	100.85	104.43	101.73	102.03
22	101.65	101.85	100.90	103.95	101.65	101.85
23	101.58	101.68	100.95	103.48	101.58	101.68
24	101.50	101.50	101.00	103.00	101.50	101.50
25	102.33	102.48	101.70	103.03	102.33	102.48
26	103.15	103.45	102.40	103.05	103.15	103.45
27	103.98	104.43	103.10	103.08	103.98	104.43
28	104.80	105.40	103.80	103.10	104.80	105.40

附表 3 2001~2007 年按季度的财政税收指数统计数据表

季 度	国内增值 税（亿元）	营业税 （亿元）	国内消费 税（亿元）	关 税 （亿元）	农业各税 （亿元）	企业所得 税（亿元）
1	1232.14	474.74	213.90	193.32	1232.14	474.74
2	1339.28	516.02	232.50	210.13	1339.28	516.02
3	1285.71	495.38	223.20	201.72	1285.71	495.38
4	1500.00	577.95	260.40	235.35	1500.00	577.95
5	1421.03	563.58	240.65	161.98	1421.03	563.58
6	1544.60	612.58	261.58	176.07	1544.60	612.58
7	1482.81	588.08	251.12	169.02	1482.81	588.08
8	1729.95	686.09	292.97	197.20	1729.95	686.09
9	1664.40	654.22	271.92	212.32	1664.40	654.22
10	1809.14	711.11	295.57	230.78	1809.14	711.11
11	1736.77	682.67	283.74	221.55	1736.77	682.67
12	2026.23	796.45	331.03	258.48	2026.23	796.45
13	2074.13	823.85	345.44	240.07	2074.13	823.85
14	2254.49	895.49	375.48	260.94	2254.49	895.49

15	2164.31	859.67	360.46	250.50	2164.31	859.67
16	2525.02	1002.95	420.53	292.26	2525.02	1002.95
17	2482.19	973.47	375.78	245.22	2482.19	973.47
18	2698.03	1058.12	408.45	266.54	2698.03	1058.12
19	2590.11	1015.79	392.11	255.88	2590.11	1015.79
20	3021.79	1185.09	457.47	298.53	3021.79	1185.09
21	2940.51	1179.60	433.71	262.61	2940.51	1179.60
22	3196.20	1282.18	471.42	285.45	3196.20	1282.18
23	3068.35	1230.89	452.57	274.03	3068.35	1230.89
24	3579.75	1436.04	527.99	319.70	3579.75	1436.04
25	3558.15	1513.90	507.57	329.49	3558.15	1513.90
26	3867.56	1645.54	551.71	358.14	3867.56	1645.54
27	3712.86	1579.72	529.64	343.82	3712.86	1579.72
28	4331.66	1843.01	617.91	401.12	4331.66	1843.01

附表 4 2001~2007 年按季度的金融保险指数统计数据表

季度	金融机构资 金运用合计 (亿元)	金融机构现 金收入(亿元)	金融机构现 金支出(亿元)	上市公司总 数量(个)	保险公司总 资产(亿元)
1	35621.50	73917.45	74155.75	1112	5086.25
2	38719.03	80345.05	80604.08	1128	5262.50
3	37170.26	77131.25	77379.91	1144	5438.75
4	43365.31	89986.46	90276.56	1160	5615.00
5	42325.64	84086.41	84451.91	1176	5791.25
6	46006.13	91398.28	91795.55	1192	5967.50
7	44165.88	87742.34	88123.73	1208	6143.75
8	51526.86	102366.07	102811.02	1224	6320.00
9	51822.06	104771.42	105339.06	1240	7012.00
10	56328.33	113881.98	114498.98	1256	7704.00
11	54075.19	109326.70	109919.02	1271	8396.00
12	63087.72	127547.81	128238.85	1287	9088.00
13	60430.22	130612.22	131008.35	1310	9804.50
14	65685.03	141969.80	142400.38	1332	10521.00
15	63057.62	136291.01	136704.36	1355	11237.50
16	73567.23	159006.18	159488.42	1377	11954.00
17	69469.84	144205.06	144794.61	1378	12787.00
18	75510.70	156744.63	157385.45	1379	13620.00
19	72490.27	150474.84	151090.03	1380	14453.00
20	84571.98	175553.98	176271.70	1381	15286.00
21	84002.92	162318.61	163018.04	1394	16390.50
22	91307.53	176433.28	177193.53	1408	17495.00
23	87655.22	169375.94	170105.78	1421	18599.50
24	102264.43	197605.27	198456.75	1434	19704.00



25	104481.59	189712.30	190471.90	1463	22109.75
26	113566.95	206209.03	207034.68	1492	24515.50
27	109024.27	197960.66	198753.29	1521	26921.25
28	127194.98	230954.11	231878.84	1550	29327.00

附表 5 2001~2007 年按季度的三大产业指数统计数据表

季度	农林牧渔业总产值（亿元）	农作物总播种面积（千公顷）	轻工业总产值（亿元）	重工业总产值（亿元）	客运量（万人）
1	6021.31	35812.84	8656.51	13296.76	352848.06
2	6544.90	38927.00	9409.25	14453.00	383530.50
3	6283.10	37369.92	9032.88	13874.88	368189.28
4	7330.29	43598.24	10538.36	16187.36	429554.16
5	6299.88	35566.28	9971.88	15506.83	369874.50
6	6847.70	38659.00	10839.00	16855.25	402037.50
7	6573.79	37112.64	10405.44	16181.04	385956.00
8	7669.42	43298.08	12139.68	18877.88	450282.00
9	6829.11	35055.45	11614.54	21108.02	365124.31
10	7422.95	38103.75	12624.50	22943.50	396874.25
11	7126.03	36579.60	12119.52	22025.76	380999.28
12	8313.70	42676.20	14139.44	25696.72	444499.16
13	8334.97	35317.19	14410.42	28650.41	406514.19
14	9059.75	38388.25	15663.50	31141.75	441863.25
15	8697.36	36852.72	15036.96	29896.08	424188.72
16	10146.92	42994.84	17543.12	34878.76	494886.84
17	9073.70	35762.18	18094.79	38805.83	424814.14
18	9862.72	38871.93	19668.25	42180.25	461754.50
19	9468.21	37317.06	18881.52	40493.04	443284.32
20	11046.25	43536.56	22028.44	47241.88	517165.04
21	9386.49	34994.27	21814.57	51000.89	465556.26
22	10202.71	38037.25	23711.49	55435.75	506039.41
23	9794.60	36515.76	22763.03	53218.32	485797.83
24	11427.03	42601.72	26556.87	62088.04	566764.14
25	11245.38	35296.72	27517.29	65673.45	512385.08
26	12223.24	38366.00	29910.10	71384.19	556940.30
27	11734.31	36831.36	28713.69	68528.82	534662.69
28	13690.03	42969.92	33499.31	79950.29	623773.14

附表 6 宏观层面指标汇总示意表

季 度	国民经济核算		价格指数	财政税收	金融保险	三大产业	
	GPC1	GPC2	JPC	CPC	BPC	SPC1	SPC2
1	2502.189	-439.389	246.8026	1163.301	85375.04	194239.1	-79559.7
2	2585.786	-450	246.1529	1264.457	92645.2	211129.4	-86477.9
3	2589.71	-454.14	245.5032	1213.879	89137.92	202684.3	-83018.8

4	2726.421	-469.065	244.8535	1416.192	103593	236465	-96855.3
5	2756.524	-480.55	244.2038	1340.998	97935.3	204490.5	-80997.7
6	2851.853	-492.604	243.5541	1457.607	106268	222272.3	-88040.9
7	2850.269	-496.112	242.9044	1399.303	102229.5	213381.4	-84519.3
8	3010.206	-513.864	242.2547	1632.52	118809.7	248944.9	-98605.9
9	3162.593	-541.099	244.1503	1511.283	121475.8	205857	-78177
10	3291.12	-558.022	246.0459	1642.699	132034.4	223757.7	-84975
11	3287.023	-562.077	247.9415	1576.991	127225.5	214807.4	-81576
12	3503.967	-587.58	249.8371	1839.823	148029.1	250608.6	-95172
13	3709.33	-623.56	252.1302	1892.694	149788.5	232189.1	-82018.1
14	3894.571	-649.426	254.4233	2057.276	162712.7	252379.5	-89150.1
15	3900.103	-656.688	256.7164	1974.985	156741.8	242284.3	-85584.1
16	4205.149	-694.957	259.0095	2304.15	182262.6	282665	-99848.1
17	4292.142	-716.723	257.4793	2271.26	167519.7	248450.1	-81944.9
18	4508.166	-746.601	255.9491	2468.76	181910.2	270054.5	-89070.5
19	4503.012	-752.832	254.4188	2370.01	175269.1	259252.3	-85507.7
20	4866.488	-798.474	252.8886	2765.012	203680.6	302461	-99759
21	4986.409	-825.528	252.3773	2749.766	192000.2	276420.9	-83484.8
22	5254.867	-863.224	251.8661	2988.876	208505.8	300457.5	-90744.3
23	5249.641	-870.912	251.3548	2869.321	200995.8	288439.2	-87114.6
24	5700.555	-928.614	250.8435	3347.541	233511.6	336512.4	-101634
25	5856.554	-963.235	251.9798	3396.99	228440.8	310533.2	-85896.5
26	6202.266	-1013.26	253.1162	3692.381	248475.5	337536.1	-93365.7
27	6210.168	-1025.48	254.2525	3544.685	240076.1	324034.6	-89631.1
28	6781.087	-1100.71	255.3888	4135.466	279067	378040.4	-104570

附表 7 2008 年 30 个行政单位代表指标统计数据汇总表

地区	城镇人口 比重 (%)	大专以上 人数比例 (/万人)	城镇单位使 用的农村人 数 (万人)	城镇登记 失业率 (%)	城镇登记 失业人员 数 (万人)	职工平均 工资 (元/年)	15 岁及以 上文盲人 口 (万人)	总人口 (人)
北 京	84.50	4039	97.90	1.8	10.6	46507	3.34	12162468
天 津	76.31	1457	19.14	3.6	15.0	34938	3.85	9641374
河 北	40.25	1766	43.90	3.8	29.3	19911	6.25	70315592
山 西	44.03	1480	35.08	3.2	16.1	21525	4.26	33923444
内蒙古	50.15	1038	12.82	4.0	18.5	21884	8.23	24137474
辽 宁	59.20	3151	46.96	4.3	44.5	23202	3.76	42317050
吉 林	53.16	1548	8.19	3.9	23.9	20513	4.55	26960574
黑龙江	53.90	1577	12.28	4.3	31.5	19386	4.36	38197615
上 海	88.70	3292	62.49	4.2	26.7	49310	4.04	13788616
江 苏	53.20	4359	132.88	3.2	39.3	27374	8.19	73540810
浙 江	57.20	2364	242.78	3.3	28.6	31086	10.12	46593398
安 徽	38.70	1075	29.95	4.1	27.2	22180	16.83	66757129
福 建	48.70	1013	184.77	3.9	14.9	22283	11.20	34406569

江 西	39.80	1235	35.59	3.4	24.3	18400	7.20	45286615
山 东	46.75	3174	146.66	3.2	43.5	22844	8.21	93455308
河 南	34.34	2130	75.02	3.4	33.1	20935	7.91	103630009
湖 北	44.30	2977	56.17	4.2	54.1	19818	8.70	60849108
湖 南	40.45	2397	54.09	4.3	44.4	21534	5.35	68801436
广 东	63.14	4035	285.38	2.5	36.2	29443	4.12	81560524
广 西	36.24	981	31.68	3.8	18.5	21898	5.82	50613316
海 南	47.20	277	4.51	3.5	5.4	19357	8.54	8492566
重 庆	48.34	516	40.14	4.0	14.1	23098	8.00	32353229
四 川	35.60	1618	99.56	4.2	34.5	21312	10.62	88151705
贵 州	28.24	502	27.35	4.0	12.1	20668	16.59	39850386
云 南	31.60	1129	57.65	4.2	14.0	20481	16.13	43708069
陕 西	40.62	1575	21.05	4.0	21.0	21296	8.89	37827142
甘 肃	31.59	661	11.48	3.3	9.5	20987	19.33	26492398
青 海	40.07	183	2.11	3.8	3.7	26166	18.40	5218135
宁 夏	44.02	258	4.19	4.3	4.4	26210	13.80	6128965
新 疆	39.15	1241	17.53	3.9	11.7	21434	4.29	20424370

附表 8 2001~2007 年我国国民受教育程度

季度	未上过学 (%)	小 学 (%)	初 中 (%)	高 中 (%)	大专及以 上 (%)	同期失业 率 (%)
1	6.63	44.25	39.16	6.47	3.49	4.12
2	7.15	42.92	38.95	7.32	3.67	4.27
3	7.66	41.59	38.73	8.18	3.84	4.42
4	8.17	40.27	38.51	9.03	4.01	4.57
5	8.69	38.94	38.30	9.89	4.19	4.72
6	9.20	37.61	38.08	10.74	4.36	4.87
7	9.71	36.29	37.86	11.60	4.54	5.02
8	10.23	34.96	37.65	12.45	4.71	5.17
9	10.09	34.58	37.74	12.68	4.91	5.13
10	9.96	34.19	37.84	12.91	5.10	5.08
11	9.82	33.81	37.94	13.14	5.29	5.04
12	9.68	33.42	38.04	13.37	5.49	4.99
13	9.55	33.16	38.35	13.38	5.56	4.85
14	9.42	32.90	38.66	13.38	5.63	4.72
15	9.29	32.64	38.98	13.39	5.70	4.58
16	9.16	32.38	39.29	13.40	5.77	4.44
17	9.46	32.61	39.06	13.16	5.72	4.36
18	9.77	32.83	38.82	12.92	5.67	4.29
19	10.07	33.06	38.58	12.68	5.61	4.21
20	10.37	33.28	38.35	12.44	5.56	4.13
21	9.97	33.23	38.51	12.56	5.73	4.13
22	9.58	33.18	38.67	12.68	5.89	4.12

23	9.18	33.12	38.83	12.81	6.05	4.12
24	8.79	33.07	38.99	12.93	6.22	4.11
25	8.60	32.75	39.30	13.05	6.30	4.14
26	8.40	32.43	39.61	13.17	6.39	4.18
27	8.21	32.12	39.92	13.29	6.47	4.21
28	8.01	31.80	40.22	13.41	6.56	4.24

附表 9 2001~2007 年分性别的失业率数据

季度	男失业率 (%)	女失业率 (%)	同期总失业率 (%)
1	4.598	2.758	3.385
2	4.785	3.075	3.780
3	4.973	3.393	4.175
4	5.160	3.710	4.570
5	5.348	3.808	4.720
6	5.535	3.905	4.870
7	5.723	4.003	5.020
8	5.910	4.100	5.170
9	5.810	4.137	5.125
10	5.710	4.175	5.080
11	5.610	4.213	5.035
12	5.510	4.250	4.990
13	5.283	4.170	4.853
14	5.055	4.090	4.715
15	4.828	4.010	4.578
16	4.600	3.930	4.440
17	4.485	3.925	4.363
18	4.370	3.920	4.285
19	4.255	3.915	4.208
20	4.140	3.910	4.130
21	4.163	3.900	4.125
22	4.185	3.890	4.120
23	4.208	3.880	4.115
24	4.230	3.870	4.110
25	4.265	3.905	4.143
26	4.300	3.940	4.175
27	4.335	3.975	4.208
28	4.370	4.010	4.240

附表 10 各年龄段失业率季度变化数据表

季度	15-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65+
1	0.101	0.184	0.123	0.101	0.103	0.121	0.114	0.096	0.051	0.017	0.003
2	0.099	0.178	0.122	0.101	0.105	0.125	0.116	0.095	0.049	0.016	0.003
3	0.097	0.172	0.121	0.101	0.107	0.128	0.117	0.094	0.047	0.016	0.003
4	0.096	0.167	0.120	0.101	0.109	0.132	0.119	0.093	0.045	0.015	0.003
5	0.094	0.161	0.117	0.102	0.111	0.133	0.120	0.095	0.048	0.016	0.004

6	0.092	0.156	0.115	0.102	0.113	0.133	0.121	0.097	0.050	0.016	0.005
7	0.091	0.150	0.112	0.103	0.115	0.134	0.122	0.099	0.053	0.017	0.005
8	0.089	0.145	0.109	0.103	0.117	0.134	0.123	0.101	0.056	0.017	0.006
9	0.086	0.143	0.109	0.104	0.116	0.135	0.126	0.102	0.058	0.017	0.005
10	0.083	0.140	0.108	0.104	0.116	0.135	0.129	0.104	0.060	0.016	0.005
11	0.079	0.138	0.108	0.104	0.115	0.136	0.131	0.106	0.062	0.016	0.004
12	0.076	0.136	0.108	0.105	0.115	0.136	0.134	0.108	0.064	0.015	0.003
13	0.075	0.134	0.104	0.102	0.113	0.136	0.137	0.111	0.067	0.017	0.004
14	0.075	0.132	0.101	0.099	0.111	0.135	0.141	0.114	0.070	0.019	0.004
15	0.074	0.129	0.097	0.096	0.109	0.135	0.144	0.117	0.074	0.020	0.005
16	0.074	0.127	0.094	0.093	0.107	0.135	0.147	0.119	0.077	0.022	0.005
17	0.073	0.131	0.095	0.092	0.104	0.134	0.148	0.120	0.078	0.021	0.005
18	0.072	0.136	0.095	0.091	0.101	0.133	0.148	0.121	0.078	0.021	0.004
19	0.071	0.140	0.096	0.090	0.099	0.132	0.148	0.121	0.079	0.020	0.004
20	0.069	0.144	0.097	0.089	0.096	0.131	0.149	0.122	0.080	0.019	0.003
21	0.071	0.148	0.098	0.089	0.097	0.126	0.148	0.121	0.081	0.018	0.002
22	0.073	0.152	0.098	0.090	0.098	0.121	0.148	0.119	0.081	0.018	0.002
23	0.075	0.156	0.098	0.091	0.099	0.116	0.148	0.118	0.082	0.017	0.001
24	0.076	0.160	0.099	0.091	0.100	0.111	0.148	0.117	0.082	0.016	0.000
25	0.078	0.164	0.097	0.092	0.102	0.109	0.149	0.113	0.082	0.015	0.000
26	0.081	0.167	0.095	0.093	0.104	0.106	0.149	0.109	0.082	0.015	0.000
27	0.083	0.171	0.092	0.093	0.106	0.104	0.150	0.105	0.082	0.014	0.000
28	0.085	0.174	0.090	0.094	0.108	0.101	0.151	0.100	0.082	0.014	0.000

附表 11 聚类后四个年龄段的数据汇总表

	第一年龄段	第二年龄段	第三年龄段	第四年龄段
	(15-49 除去 20-24)	(20-24)	(55-59)	(60-65+)
1	0.108	0.184	0.051	0.010
2	0.109	0.178	0.049	0.010
3	0.109	0.172	0.047	0.009
4	0.110	0.167	0.045	0.009
5	0.110	0.161	0.048	0.010
6	0.110	0.156	0.050	0.010
7	0.111	0.150	0.053	0.011
8	0.111	0.145	0.056	0.012
9	0.111	0.143	0.058	0.011
10	0.111	0.140	0.060	0.010
11	0.111	0.138	0.062	0.010
12	0.112	0.136	0.064	0.009
13	0.111	0.134	0.067	0.010
14	0.111	0.132	0.070	0.011
15	0.110	0.129	0.074	0.013
16	0.110	0.127	0.077	0.014
17	0.109	0.131	0.078	0.013

18	0.109	0.136	0.078	0.012
19	0.108	0.140	0.079	0.012
20	0.108	0.144	0.080	0.011
21	0.107	0.148	0.081	0.010
22	0.107	0.152	0.081	0.010
23	0.106	0.156	0.082	0.009
24	0.106	0.160	0.082	0.008
25	0.106	0.164	0.082	0.008
26	0.105	0.167	0.082	0.007
27	0.105	0.171	0.082	0.007
28	0.104	0.174	0.082	0.007

附表 12 微观层面指标

季度	micro1	micro2	micro3	micro4	micro5	micro6
1	-14.48	25.09	4.60	2.76	0.04	-0.08
2	-13.02	24.75	4.79	3.08	0.03	-0.08
3	-11.55	24.40	4.97	3.39	0.03	-0.08
4	-10.08	24.05	5.16	3.71	0.03	-0.08
5	-8.61	23.70	5.35	3.81	0.02	-0.07
6	-7.14	23.36	5.54	3.91	0.02	-0.07
7	-5.67	23.01	5.72	4.00	0.02	-0.06
8	-4.20	22.66	5.91	4.10	0.02	-0.06
9	-3.82	22.89	5.81	4.14	0.02	-0.06
10	-3.43	23.11	5.71	4.18	0.02	-0.05
11	-3.04	23.33	5.61	4.21	0.01	-0.05
12	-2.66	23.56	5.51	4.25	0.01	-0.05
13	-2.51	23.91	5.28	4.17	0.01	-0.04
14	-2.36	24.26	5.06	4.09	0.01	-0.04
15	-2.21	24.61	4.83	4.01	0.01	-0.04
16	-2.06	24.96	4.60	3.93	0.01	-0.03
17	-2.24	24.59	4.49	3.93	0.01	-0.03
18	-2.42	24.22	4.37	3.92	0.02	-0.04
19	-2.60	23.85	4.26	3.92	0.02	-0.04
20	-2.77	23.48	4.14	3.91	0.02	-0.04
21	-2.74	23.87	4.16	3.90	0.03	-0.04
22	-2.70	24.27	4.19	3.89	0.03	-0.04
23	-2.67	24.66	4.21	3.88	0.03	-0.04
24	-2.63	25.05	4.23	3.87	0.03	-0.04
25	-2.40	25.43	4.27	3.91	0.04	-0.05
26	-2.18	25.82	4.30	3.94	0.04	-0.05
27	-1.95	26.20	4.34	3.98	0.04	-0.05
28	-1.73	26.59	4.37	4.01	0.04	-0.05