

全国第六届研究生数学建模竞赛



题目 基于回归分析的城镇登记失业率数学建模

摘 要:

本文主要讨论的是城镇登记失业率与各主要社会经济指标之间定性和定量的关系。在充分理解问题的基础上,利用熵权法、多元混合加权线性回归以及 Logistic 回归方程等方法对各个问题进行分析,建立模型并求解得到所需结果,最后给增加我国城镇就业人口数或降低城镇登记失业率作出一些咨询建议。

首先对国家统计局等网站的相关社会经济运行数据进行统计分析,得到影响就业情况的多种因素和指标。然后以城镇失业率为评价对象,各影响数据为评价指标,采用熵权法对上述影响因素进行重要性加权排名,取权重比较大的作为影响就业的主要因素。

其次,在通过对主要社会经济指标与城镇失业率的关系走势图的分析后,确定上述主要因素与失业率的线性关系。采用线性多元混合加权回归方程来确定上述指标与失业率之间的定量关系并给出相应数学模型。同时,对于一些无法采用一般方法给出其对失业率影响的确定的定量关系的主观因素,如国家政策法规、经济危机等,通过采用 Logistics 线性回归方程等方法给出其对失业率影响的定量分析以及计算方法,从而为精确模型建立打下基础。

再次,在给出对政策和法规等主观因素对城镇失业率的定量分析之后,分别从按行业,按地区,按就业人群等不同角度分别给出主要社会经济指标与城镇登记失业率的精确模型。并对所建立的模型进行仿真实验,从而给出 2009 年至 2010 年上半年的按季度的城镇人口失业率的预测值。

最后,根据所建立的模型和上述对社会经济指标的定性分析,以及模型的仿真结果,我们给出增加我国城镇就业人口数或降低城镇登记失业率的咨询建议。

关键词: 社会经济指标 熵权法 多元混合加权线性回归 Logistic 回归方程

参赛队号 1033704

队员姓名 李浩 孟文超 吴海涛

参赛密码 _____

1 引言

背景分析:

由于经济危机的肆虐，失业问题已经成为当前社会研究的热点。而失业、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[23]。

按照已有研究，就业可以定义为三个月内有稳定的收入或与用人单位有劳动聘用关系。失业的统计方法各国差异较大，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。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。

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，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，因此找出影响失业的各种社会经济因素，进而解决失业问题成为当前人们面临的一个重要课题[1][2][3][4]。

问题重述:

请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就我国就业人数或城镇登记失业率研究如下问题。

1. 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。（由于时间限制，建议适度即可）
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，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（可以根据模型的需要对未来的情况作适当的假设）。
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你们的咨询建议。

2 问题分析

针对问题一，采取走势分析的方法对中国国家统计局官网、国研网[1][2][3][4]等网站发布的近四年国家社会和经济的统计数据进行分析，得到影响就业的因素包括国家经济总量、固定资产投资等社会经济因素的统计数据，然后将就业率作为评价对象，各经济因素作为评价指标，采用熵权法对各因素进行重要性加权排名，确定影响就业的主要因素包括：中国经济总量，海关进出口总量，固定资产投资，国家财政税收，消费品总额[14][15][16][19]。

针对问题二，首先筛选影响就业的五个主要指标的季度统计数据，通过问题一中的走势分析，可以知道就业情况与其主要影响因素之间存在着某种线性联系，再考虑到国家政策、世界经济大环境等外部因素对失业率的影响，建立以城镇登记失业率为因变量，五个主要影响指标为自变量的多元混合加权回归模型[14][16]。

针对问题三，首先按照题目要求按照分区域、分行业、分就业人群的条件，分别选择一个特例，以城镇登记失业率为因变量，主要社会经济指标为自变量分别进行建模。然后将这三个条件综合起来，选择“浙江省 IT 行业高校毕业生失业率与主要社会经济指标之间的联系”作为建模对象，建立比较精确的数学模型。

针对问题四，根据各主要影响因素在 2005 年到 2008 年的统计数据，采用线性回归方程预测出各主要影响因素在 2009 年到 2010 上半年的运行数据，然后将这些数据代入问题二的模型中，得到 2009 年到 2010 上半年城镇登记失业率。

针对问题五，结合中国实际情况，对增加我国城镇就业人口数或降低城镇登记失业率提出一些咨询和建议。

3 模型假设

1. 各经济因素之间相互独立。
2. 2009 年到 2010 年上半年经济发展稳定，没有出现大的波动。
3. 2009 年到 2010 年上半年中国的就业受经济危机影响不大。

4 符号说明

变量	定义
M	影响城镇登记失业率主要指标的个数
N	样本个数
p_i	第 i 个案例的城镇登记失业率
α	常数项
r_{ij}	第 j 个影响因素或指标的增长率
β_i	第 i 个影响因素或指标的回归系数
w	处于经济危机下的稳定权重
p	政策影响就业的概率因子
Δr_i	影响就业的五个主要指标在政策法规颁布或实施之后的变化量
λ_i	五个主要指标对就业情况影响的概率

5 问题一的求解

5.1 影响就业指标的确定

传统上，影响就业情况的因素主要包括国家的经济发展状况以及海关进出口总额等有关社会经济方面的运行状况。通过近年来我国有关的统计数据，得到包括 GDP、海关进出口总额、城镇就业人数、固定资产投资以及 CPI 的季度统计数据。将 06~09 年统计结果列表如下：

表 1 06~09 年中国各经济指标的季度统计数据

	季度	经济总量	海关进出口总量	固定资产投资	财政税收	CPI
2009	第一季度	65745	4287.66	33837.785	13023.58	100
	第二季度	139862	5174.29	168700.983	16506.49	99.057
2008	第一季度	63475	5705.29	26438.23	11527.48	107.667
	第二季度	71251	6632.34	127110.253	10230.14	108.067
	第三季度	73299	7326.24	256950.483	13091.96	107.334
	第四季度	92645	5964.72	388970.458	12370.04	106.29
2007	第一季度	53058	4757.86	21078.62	11027.48	102.433
	第二季度	59400	5232.97	100717.01	12312.14	102.867
	第三季度	61970	5899.15	201603.61	12081.96	103.567
	第四季度	82878	6035.02	306971.84	11081.96	104.6
2006	第一季度	44419.8	3712.45	16902.5	10457.32	101.5
	第二季度	49191.8	4247.84	79817.48	11312.14	101.233
	第三季度	50958	4767.79	159244.79	10881.76	101.233
	第四季度	67353.4	5184.11	242854.98	12011.26	101.367

(GDP、海关进出口总量、固定资产投资、财政税收单位：亿元；CPI 单位：%)

由于上述数据单位不一致，无法进行各数据之间发展走势的比较，因此将上述原始数据去单位化、标准化，得到上述相关数据的走势图，如下图所示：

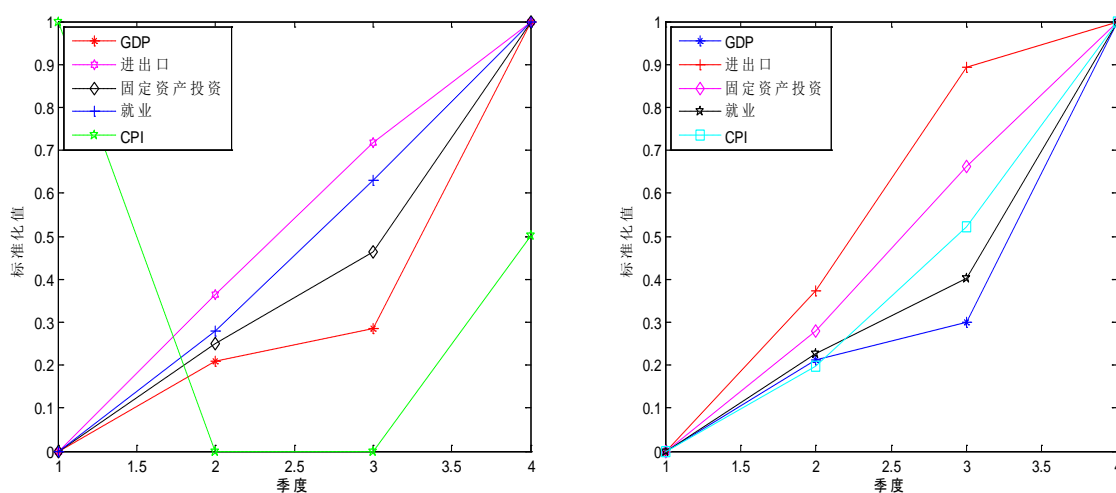


图 1 06、07 年我国主要社会经济因素走势图

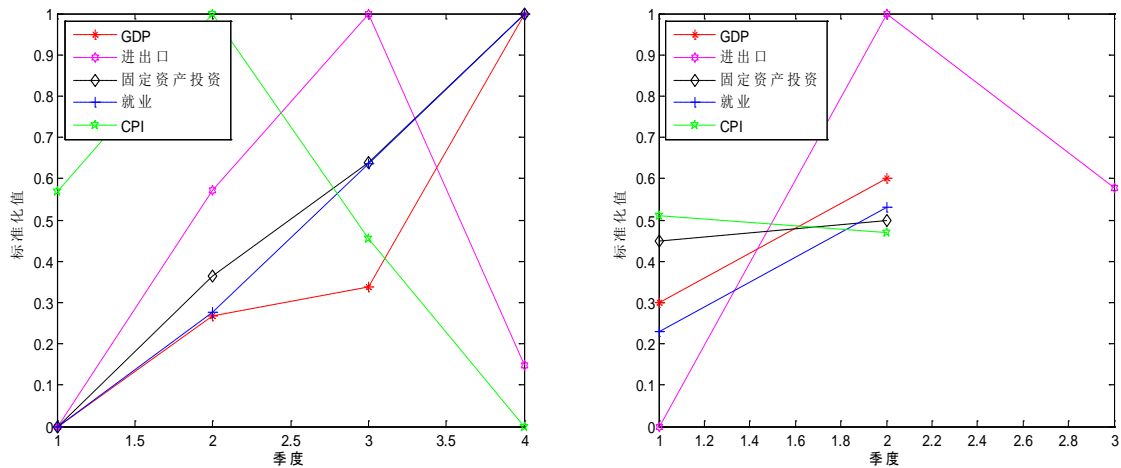


图 2 08、09 年我国主要社会经济因素走势图

从图中不难看出，随着中国经济总量、进出口总额、固定资产投资等经济因素的不断变化，对国内同期的就业人数造成了不同程度的影响，存在着某种线性关系。因此首先对中国国家统计局官网等相关数据统计网站给出的近四年的社会经济统计数据进行分析，得到中国经济总量、海关进出口总量、固定资产投资、国家财政税收、教育与科学投入、人口总量、CPI、消费品总额八个社会经济方面相关的统计数据，然后采用比较客观的熵权法对以上影响因素进行筛选加权排序，取权重值比较大的作为影响就业最主要的因素[17][23]。

熵权法是利用指定指标的信息熵值来判断该指标的有效性和价值，它能够深刻地反映出指标信息熵值的效用价值，其给出的指标权重值比德尔菲法和层次分析法有较高的可信度。

5.1.1 确定评价矩阵 R

把就业率作为评价对象，将近年来中国经济总量、消费者物价指数、海关进出口总量、人口总数、教育与科技、国家财政税收等因素的统计数据作为评价指标。为了便于数据建模，将中国经济总量、CPI、海关进出口总量等八个评价指标进行编号如下：

表 2 编号说明

编号	评价指标
1	中国经济总量 (GDP)
2	海关进出口总量
3	固定资产投资
4	国家财政税收
5	教育和科技
6	人口总数
7	消费者物价指数
8	消费品总额

按照熵权法的定义，评价指标体系有 M 项评价指标，这是有 N 个样本组成，用 M 个指标作综合评价的问题。而因为统计的是最近四年的月度统计数据，因此可建立如下数学模型：

就业率 (评价对象) 集合为：

$$R = \{R_1, R_2, \dots, R_N\}$$

每一个影响因素 (评价目标) 对应的 M 个评价指标的数据集合为：

$$R_i = \{r_{i1}, r_{i2}, \dots, r_{iM}\} \quad (i = 1, 2, \dots, N)$$

其中： $M = 8; N = 48$ 。

得到以下评价矩阵 R ：

$$R = \begin{pmatrix} R_1 \\ R_2 \\ \vdots \\ R_N \end{pmatrix} = \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1M} \\ r_{21} & r_{22} & \cdots & r_{2M} \\ \cdots & \cdots & \cdots & \cdots \\ r_{N1} & r_{N2} & \cdots & r_{NM} \end{pmatrix}$$

- 【符号说明】 (1) 矩阵中 R_i 表示就业率第 i 个月的统计数据；
 (2) r_{ij} 表示第 j 个影响因素在第 i 个月的统计数据；
 (3) $M = 8; N = 48$ 。

5.1.2 矩阵 R 的标准化处理

为使数据之间具有可比性，需要将各评价指标同度量化，做标准化处理。将矩阵 R 标准化，得到标准化后的评价矩阵 $R' = (r'_{ij})_{NM}$ 。按照标准化的流程，每一行的每一个数据 r_{ij} 减去每一行中最小的数据 $\min_i[r_{ij}]$ ，然后除以每一行中最大元素减去最小元素的差值 $\max[r_{ij}] - \min_i[r_{ij}]$ ，得到标准化后的 r'_{ij} ，其中：

$$r'_{ij} = \frac{r_{ij} - \min_{1 \leq i \leq 48} [r_{ij}]}{\max_{1 \leq i \leq 48} [r_{ij}] - \min_{1 \leq i \leq 48} [r_{ij}]} \quad (i = 1, 2, \dots, 48, j = 1, 2, \dots, 8)$$

5.1.3 得到信息熵 e 和信息效用值 d

按照熵的定义，在 M 个评价指标， N 个被评价对象区域系统内，计算第 j 项指标的信息熵值为：

$$e_j = -k \sum_{i=1}^N g_{ij} \ln(g_{ij}) \quad (i = 1, 2, \dots, N; j = 1, 2, \dots, M)$$

其中：

$$g_{ij} = r'_{ij} / \sum_{i=1}^N r'_{ij} \quad (i = 1, 2, \dots, N; j = 1, 2, \dots, M)$$

【符号说明】常数 k 与样本数 N 有关，对于一个信息无序的系统，有序度为零其熵值最大， $e = 1$ ， N 个样本处于完全无序分布时， $g_{ij} = r'_{ij} / \sum_{i=1}^N r'_{ij} \frac{1}{N}$ ，得到

$e_j = -k \sum_{i=1}^N \frac{1}{N} \ln\left(\frac{1}{N}\right) = k \ln(N)$ ，则：

$$k = \frac{1}{\ln(N)}，当 g_{ij} = 0 时，g_{ij} \ln(g_{ij}) = 0。$$

第 j 个准则的评价值数据的信息效用值为信息熵 e_j 与 1 之差：

$$d_j = 1 - e_j \quad (j = 1, 2, \dots, M)$$

【说明】评价指标的信息效用值 d_j 越大，则其对于评价对象的作用越重要。

5.1.4 计算评价指标权重

利用熵权法估算各评价指标的权重，其本质是利用指标信息的价值系数来计算的。其价值系数越高，对评价目标的作用越大[7][8][10][11]。

最后得到第 j 项评价指标的权重为：

$$w_j = d_j / \sum_{j=1}^M d_j \quad (j=1,2,\dots,M)。$$

【说明】其中熵权的大小与熵值成反比，即熵值越小，熵权越大；熵值越大，熵权越小。

将原始数据代入以上数学模型得到各影响因素对就业的影响权重如下表所示：

表 3 各影响因素的权重

编号	评价指标	信息熵	效用值	权重
1	中国经济总量	0.9908	0.0092	0.0924
2	进出口总量	0.9952	0.0048	0.0481
3	固定资产投资	0.9144	0.0732	0.6440
4	国家财政税收	0.9873	0.0072	0.1254
5	教育科技投入	1.0000	0.0000	0.0000
6	人口总数	1.0000	0.0000	0.0000
7	CPI	1.0000	0.0000	0.0000
8	消费品总额	0.9923	0.0052	0.0900

由上表可知，教育和科技、人口总数、CPI 对城镇登记失业率的影响不大，因此舍去这些影响因素，得到影响城镇登记失业率的最主要因素：中国经济总量，进出口总量，固定资产投资，国家财政税收，消费品总额。

6 问题二的求解

6.1 数据的预处理

问题一已经给出影响就业的主要因素和指标，主要包括：中国经济总量，进出口总量，固定资产投资，国家财政税收，消费品总额。为了避免数量级不同造成误差的加大，将上述主要影响因素的统计数据统一转化为增长率。还要考虑国家相关政策以及经济危机对就业的影响，因此还要增加扰动项的影响。为了便于建模，将五个主要影响因素进行编号如下：

表 4 编号说明

编号	影响因素
1	中国经济总量增长率 (GDP)
2	海关进出口总量增长率
3	固定资产投资增长率
4	国家财政税收增长率
5	消费品总额增长率

根据问题一的分析，影响就业的主要是以上因素或指标，并由图（1）和图（2）的走势可以发现，城镇登记失业率的走势与上述主要影响指标走势存在着某种线性关系。因此我们采用建立混合加权多元线性回归分析模型，研究影响中国就业问题的诸多因素与城镇登记失业率之间的关系。

为了尽量保障求解结果的合理性，采用多数据量回归分析的方法，因此，选取所有评价指标从 2005 年-2008 年最近四年内的季度统计数据。共取得 16 条案例用于最终

的建模，如下表所示：

表 5 各影响因素的统计增长率

	季度	GDP 增长率	固定资产投资增长率	进出口增长率	零售商品总额增长率	财政税收增长率
2008	第一季度	9.41	12	15.46	4.68	4.02
	第二季度	12.25	8.75	16.25	9.26	-11.25
	第三季度	2.87	10.21	10.46	5.32	27.97
	第四季度	6.39	9.38	18.83	14	-5.51
2007	第一季度	9.22	13.32	8.22	1.61	-8.19
	第二季度	11.95	6.8	9.99	1.57	11.65
	第三季度	4.33	10.1	12.73	4.45	-1.87
	第四季度	7.74	15.76	2.3	9.53	-8.28
2006	第一季度	8.94	9.3	8.49	11.07	-12.06
	第二季度	10.74	13.25	14.42	22.34	8.17
	第三季度	7.59	8.51	12.24	13.53	-3.8
	第四季度	6.17	9.51	8.73	14.35	10.38
2005	第一季度	6.01	10.4	8.76	10.05	1.17
	第二季度	9.49	9.5	13.45	8.65	3.05
	第三季度	4.54	11.38	14.23	6.95	12.31
	第四季度	9.9	7.9	9.12	13.79	2.67

6.2 模型的建立

6.2.1 多元线性模型

多元线性回归模型是在分析分类因变量时最常使用的统计分析模型之一，多元线性回归属于多变量线性回归，假设存在自变量 x_1, x_2, \dots, x_k ，变量 y 的值由以上自变量经过某种关系变换得到，则多元线性的回归模型为[5][9][14][16][20]：

$$y = \alpha + \sum_{k=1}^k \beta_k \times x_k + \varepsilon_i$$

【符号说明】式中 α 为常数项， β_i 为回归系数， k 为自变量的个数，误差 ε_i 是由于某种主观因素对因变量造成影响而增加的补偿因子。

题目中影响就业的主要包括题目一筛选中出的主要因素以及相关的国家政策，把城镇登记失业率作为因变量，主要影响因素作为自变量，符合多元线性回归的条件。设城镇登记失业率为 $p(0 \leq p \leq 1)$ ，影响因素个数为 M ，就业率样本个数为 N ，建立如下数学模型：

就业率(因变量)的数据集合为：

$$p = \{R_1, R_2, \dots, R_N\}$$

每一个评价目标(因变量)对应的 M 个评价指数(自变量)的数据集合为：

$$p_i = \{r_{i1}, r_{i2}, \dots, r_{iM}\} \quad (i = 1, 2, \dots, N)$$

预建立的多元线性回归模型如下：

$$p_i = \alpha + \beta_1 r_{i1} + \beta_2 r_{i2} + \beta_3 r_{i3} + \beta_4 r_{i4} + \beta_5 r_{i5} + Z_i$$

【符号说明】 r_{i1} = 中国经济总量增长率； r_{i2} = 海关进出口总量增长率； r_{i3} = 固定资

产投资增长率； r_{i4} =国家财政税收增长率； r_{i5} =消费品总额增长率； Z 为政策影响变量。

需要说明的是,从20世纪90年代起,我国开始了建立和完善社会主义市场经济体制的转型时期,许多政策有了较大的变化,宏观经济指标有了不同的走势,这其中包括就业政策和就业指标。在基本经济理论(即定性分析)不变的情况下,这种政策变化的效果往往也要体现在定量分析上。对于这种情况,一般地引入政策影响变量加以分析。为此,本模型统一引入政策影响变量 Z 。

6.2.2 稳定权重的设定

但是由于近几年的经济危机,国家外部经济环境的异常,造成经济危机前后的社会经济运行状况不稳定。因此我们设定一个稳定权重 W_i ,原则如下:

设定一个经济危机爆发的季度月份时间 t ,最后经济危机解除时的季度时间为 T 。计算第 i 季度的就业率时,增加稳定权重 w_i 。当时间 i 在经济危机爆发时间 t 之前或解除时间 T 之后,则其影响为零。当时间 i 处于经济危机期间,其值大小和距离经济危机爆发时间有关,距离经济危机爆发时间越短,则受影响越大,赋予较小权值。反之,则赋予较大的权值。

基于以上考虑,实际应用中经常采用以下面权函数:

$$w_i = \begin{cases} 1 & i < t \text{ 或 } i > T \\ \exp(-\lambda \times (i - t)) & t < i < T \end{cases}$$

【符号说明】 λ 为经济危机对就业影响的经验值, $i, t, T \in [1, 16]$ 且 $t < T$ 。

6.2.3 政策影响变量 Z_i 的确定

通过对政策法规对就业情况影响的相关研究,可以得出国家政策法规对影响是巨大的。在经济学领域,税收制度、救济金制度以及商业和劳动力法规是影响就业的最主要因素,无论东西方国家都纷纷出台了刺激本国就业的政策法规。国家政策对就业情况的影响是不可忽略的,因此在模型之中加入政策影响变量。

就目前来看,还没有将政策影响量化的相关模型,我们提出了基于Logistic回归方程的政策影响量化模型[16][22],定量的分析了国家政策法规对于就业情况的影响。

Logistic回归模型是在分析分类因变量时最常使用的统计分析模型之一,属于概率型非线性回归,假设在自变量 x_1, x_2, \dots, x_k 作用下,某事件的发生概率为 p ,发生概率与不发生概率之比为 $\frac{p}{1-p}$,则Logistic的回归模型为:

$$\ln\left(\frac{p_i}{1-p_i}\right) = \alpha + \sum_{k=1}^k \beta_k \times x_i + \varepsilon_i$$

【符号说明】式中 α 为常数项, β_k 为回归系数,误差 ε_i 是随机变量,且 $E(\varepsilon_i) = 0$, $\text{var}(\varepsilon_i) \geq 0$ 。

本文中因变量为政府政策法规对就业的影响,其结果有两种可能0和1(影响 $Y=1$,不影响 $Y=0$),影响的概率为 p ,主要研究税收制度、救济金制度以及商业和劳动力法规对因变量的影响。

建立的Logistic回归模型如下:

$$\ln\left(\frac{p_i}{1-p_i}\right) = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$$

其中: X_1 = 税收制度, X_2 = 救济金制度, X_3 = 商业和劳动力法规, β_i 为回归系数。

确定影响因子 p 之后,变量 Z 的值可由以下公式确定:

$$Z = p \times (\lambda_1 \times \Delta r_1 + \lambda_2 \times \Delta r_2 + \lambda_3 \times \Delta r_3 + \lambda_4 \times \Delta r_4 + \lambda_5 \times \Delta r_5)$$

【符号说明】

变量	含义
Z	政策影响变量
p	政策影响就业的概率因子
Δr_i	影响就业的五个主要指标在政策法规颁布或实施之后的变化量
λ_i	五个主要指标对就业情况影响的概率（见问题一）

6.2.4 混合加权回归模型的建立

将考虑了经济危机影响因素的稳定权重加入到模型当中，建立混合加权的多元线性回归模型：

$$p_i = \alpha + \beta_1 w_i r_{i1} + \beta_2 w_i r_{i2} + \beta_3 w_i r_{i3} + \beta_4 w_i r_{i4} + \beta_5 w_i r_{i5} + Z$$

【符号说明】

变量	含义
p_i	第 <i>i</i> 个案例的城镇登记失业率
α	常数项
r_{ij}	第 <i>j</i> 个影响因素或指标的增长率
β_i	第 <i>i</i> 个影响因素或指标的回归系数
w_i	处于经济危机下的稳定权重
Z	政策影响变量

将统计数据代入回归模型，得到最终模型[12]：

$$p_i = 5.773 - 0.445r_1 - 0.127r_2 + 0.003r_3 - 0.086r_4 + 0.118r_5 + Z$$

【符号说明】 r_{i1} = 中国经济总量增长率； r_{i2} = 海关进出口总量增长率； r_{i3} = 固定资产投资增长率； r_{i4} = 国家财政税收增长率； r_{i5} = 消费品总额增长率； Z 为政策影响变量。

6.2.5 模型结果的说明及残差分析

当国家经济总量增长时，城镇登记失业率相对就会减少。这与事实基本相符，经济增长就会创造更多的就业机会，因此城镇登记失业率下降。而保留政策影响变量 Z ，也是符合中国国情的。当由于经济发展不顺畅，导致城镇登记失业率上升，国家就会出台相关政策措施，用于影响失业率并使之下降。

为了验证模型是否与实际相符，将原始数据与模型拟合出来的数据放在同一个坐标轴下进行比较，如图所示：

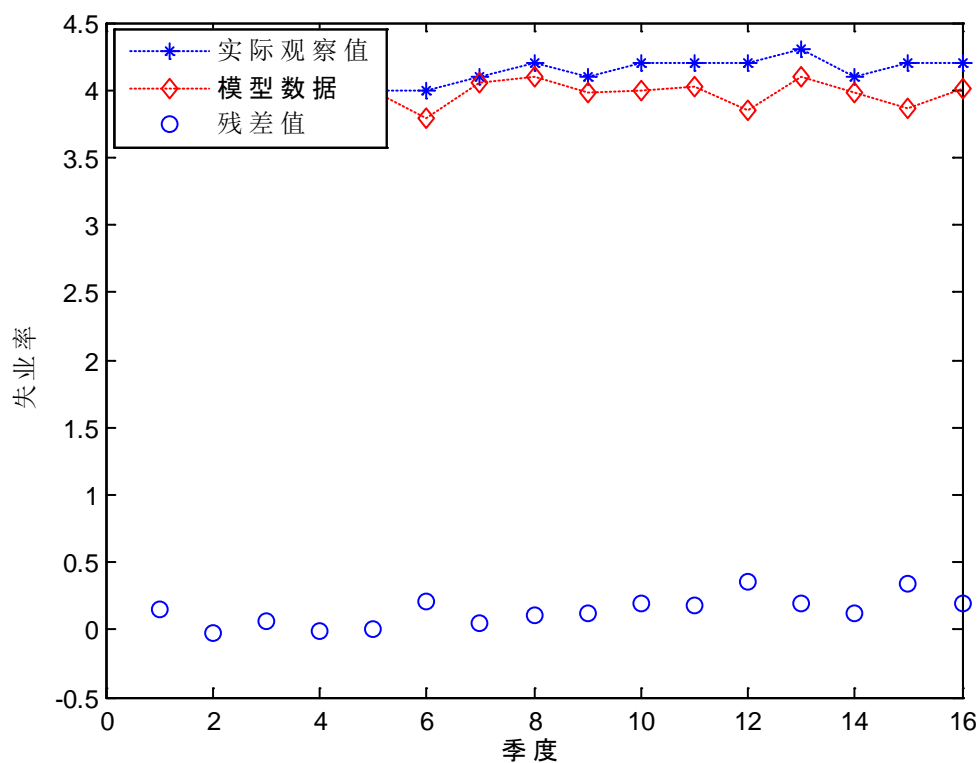


图3 实际数据与模型数据对比图

从图中不难看出，实际观察数据与通过模型拟合出来的数据相差不大，因此模型是比较合理的。

残差分析：残差是指实际观察值与回归估计值的差，即

$$e_i = y_i - \hat{y}_i \quad (i = 1, 2, \dots, n)$$

通常用残差标准差 $\hat{\sigma}$ 来估计 σ 。 $\hat{\sigma}$ 可用下式求得

$$\hat{\sigma}^2 = \frac{1}{n-2} \times S_{\text{残}}$$

由此可得残差置信带：

$$\begin{cases} y - \hat{y} = -2\hat{\sigma} \\ y - \hat{y} = 2\hat{\sigma} \end{cases}$$

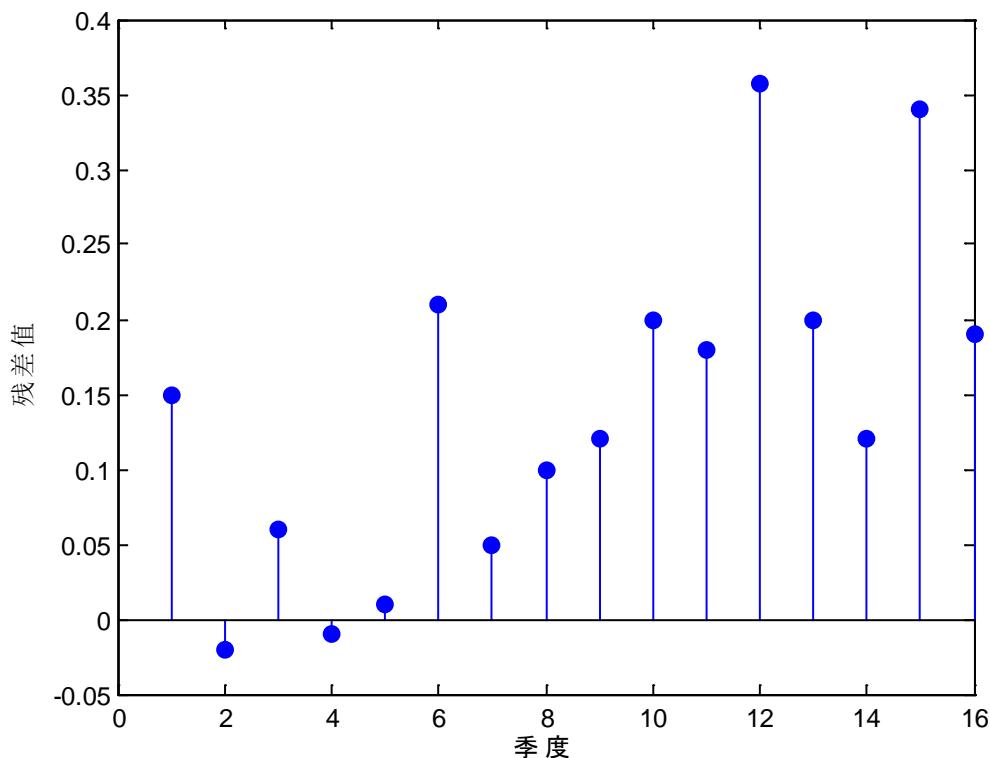


图4 残差分布图

从上图中 05-08 年共 16 个季度的失业率的残差值的离散分布图, 可以看出模型获得残差值近似的服从正态分布曲线, 这也符合统计学上的规律[13][6]。也可从上图中可以看出模型产生的残差, 如果用正态分布检验, 则相应的残差均落在 $\pm \hat{\sigma}$ 的范围内, 也就是以 90%置信度接受该组数据。因此, 从上述不同的角度均证明模型的真实、有效, 可以在一定的假设条件下 (各变量相互独立) 真实的反映主要经济指标也失业率之间的准确的关系, 这也为今后做失业率的预测提供了一个有力的模型工具。

7 问题三的求解

7.1 模型准备

经济社会指标是指反映一定社会经济现象数量方面的名称及其数值。经济现象的名称用经济范畴表述, 经济范畴的数量方面则通过数值反映。例如, 社会总产品和国民收入是经济范畴, 表现这些经济范畴数量方面的名称及其数值“社会总产值××亿元”、“国民收入××亿元”就是经济指标。经济指标在反映经济现象及其发展规律的数量表现时, 是以理论经济学所确定的经济范畴的涵义为依据。在社会主义国家, 根据各种经济事物在经济生活中的作用和存在时间、空间上的位置不同而存在不同联系。与之对应的经济指标可划分为不同的体系。

社会经济指标分为很多体系, 仅选出其中最主要的几个社会经济指标进行统计分析。经济学界学者普遍认为的中国主要社会经济指标包括: 国内生产总值、固定资产投资、国家财政收入、海关进出口总额、消费品销售总额。

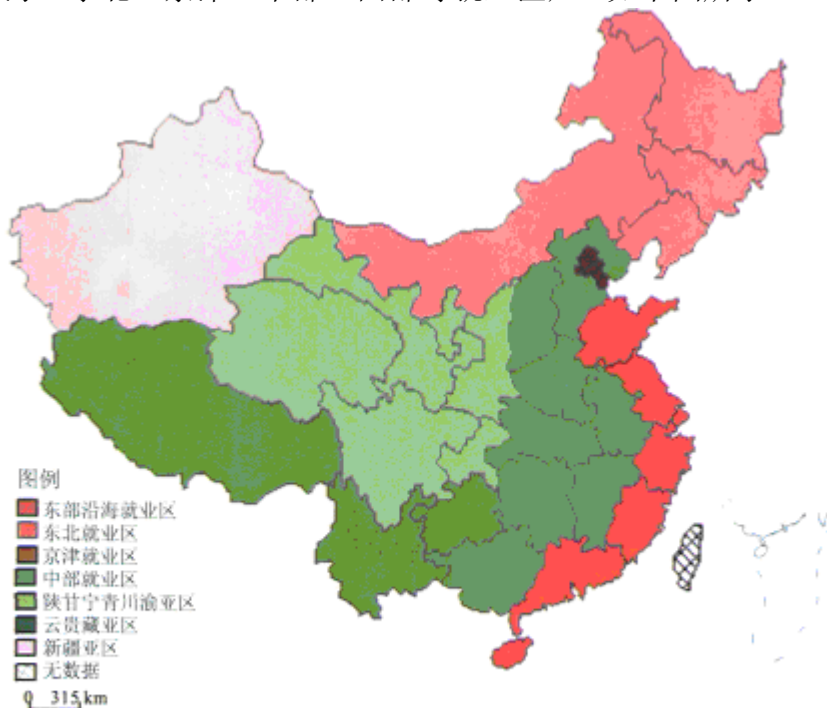
7.2 在划分区域条件下模型的建立

7.2.1 区域的划分

为了从宏观上揭示中国就业的空间分布状况, 服务于国家流动人口的引导及就业政策的制订, 必须将研究结果按省进行统计, 揭示宏观上的分布规律。对中国第一、

第二、第三产业人口主导就业区进行面积统计得出，全国第一产业、第二产业、第三产业人口就业主导区面积分别占总面积的 72.36%、20.89%和 6.75%。以此为标准，并与各省(市、区)对应的比重加以比较，进行全国就业区划分。

考虑到行政区的空间邻接性及就业空间分布的实际情况,全国被划分为 5 个就业区：东部沿海、东北、京津、中部、西部等就业区。如下图所示：



中国就业区划图

The employment regionalization of China

图 5 中国就业区划图

7.2.2 数据准备

不同的就业区，所能提供的就业空间、产业结构都存在着差异，因此各个区域的影响就业率的主要社会经济指标也不相同，因此建立模型也不一样。本文仅选出隶属于东部沿海就业区的浙江省的城镇登记失业率作为建模对象。

依据问题二的模型，统计最近四年（2005 年-2008 年）浙江省城镇登记失业率主要经济指标的统计数据，选择生产总值增长率、固定资产投资增长率、财政收入总额增长率、海关进出口总额增长率、社会消费品零售总额增长率作为城镇登记失业率的主要影响因素，统计数据如下：

表 6 浙江省主要社会经济指标统计数据

	季度	GDP 增加	固定资产投资	进出口	零售总额	财政税收
2008	第一季度	-18.61	-51.71	-0.77	-0.77	44.17
	第二季度	14.88	61.71	21.6	21.6	-1.9
	第三季度	2.39	-1.21	-50.34	-50.34	-16
	第四季度	14.26	29.62	-6.27	-6.27	-17.69
2007	第一季度	-75.08	-53.51	-3.98	-3.98	19.63
	第二季度	14.01	59.2	18.73	18.73	2.17
	第三季度	4.37	0.02	13.62	13.62	-0.89
	第四季度	21.87	48.45	-6.29	-6.29	-13.65
	第一季度	-22.88	-53.38	-27	-27	3.84

2006	第二季度	114.6	54.12	22.01	22.01	2.7
	第三季度	54.38	1.71	12.43	12.43	10.44
	第四季度	42.3	51.91	0.23	0.23	-7.21
2005	第一季度	3.78	11.08	-15	-15	1
	第二季度	21.92	60.19	19.77	19.77	4.26
	第三季度	-12.72	-0.29	7.83	7.83	6.21
	第四季度	47.3	56.71	32.23	32.23	1.52

(单位：%)

为了便于数学建模，将此五个评价指标进行编号如下：

表 7 影响失业率的主要经济指标编号

编号	评价指标
1	生产总值增长率
2	固定资产投资增长率
3	财政收入总额增长率
4	进出口总额增长率
5	社会消费品零售总额增长率

7.2.3 模型建立

根据以上统计数据，建立加入经济危机影响因素后的混合加权的多元线性回归模型：

$$p_i = \alpha + \beta_1 w_i r_{i1} + \beta_2 w_i r_{i2} + \beta_3 w_i r_{i3} + \beta_4 w_i r_{i4} + \beta_5 w_i r_{i5} + Z$$

【符号说明】

变量	含义
p_i	第 i 个案例的城镇登记失业率
α	常数项
r_{ij}	第 j 个影响因素或指标的统计值
β_i	第 i 个影响因素或指标的回归系数
w_i	处于经济危机下的稳定权重
Z	政策影响变量

将统计数据代入回归模型，得到最终模型：

$$p_i = -15.935 + 0.016r_1 + 0.112r_2 - 0.023r_3 + 1.123r_4 + 0.018r_5 + Z$$

【符号说明】 r_{i1} = 生产总值增长率； r_{i2} = 固定资产投资增长率； r_{i3} = 财政收入总额增长率； r_{i4} = 进出口总额增长率； r_{i5} = 社会消费品零售总额增长率； Z 为政策影响变量。

7.3 在划分行业、就业人群条件下模型的建立

由于划分行业以及划分就业人群条件下的模型与划分区域条件下的建模步骤相类似，建模步骤如下：

Step1: 符合条件案例的选定。题目要求分行业和分就业人群，因此这两个模型分别选取一个案例，例如建立符合分行业条件的模型选择 IT 行业作为建模对象，建立符合分就业人群的模型选择高校毕业生作为建模对象。

Step2: 影响因素的确立。找出影响 IT 行业失业率和高校毕业生就业的主要社会经济指标。例如，影响 IT 行业的有国家经济总量、科技投资、进出口总量等；影响大学生就业的有国家教育科技投入、国家经济总量的增长率等。

Step3: 读取相关影响因素的统计数据以及。

Step4:建立混合加权的多元线性回归模型，将各影响因素的统计数据代入求解，得到各影响因素的回归系数，从而得到最终的影响城镇登记失业率的准确模型。本文将不再做数据统计和处理，仅将两者模型与符号说明罗列如下：
模型建立：

$$p_i = \alpha + \beta_1 w_i r_{i1} + \beta_2 w_i r_{i2} + \beta_3 w_i r_{i3} + \beta_4 w_i r_{i4} + \beta_5 w_i r_{i5} + Z$$

【符号说明】

变量	含义
p_i	第 <i>i</i> 个案例的城镇登记失业率
α	常数项
r_{ij}	第 <i>j</i> 个影响因素或指标的统计值
β_i	第 <i>i</i> 个影响因素或指标的回归系数
w_i	处于经济危机下的稳定权重
Z	政策影响变量

7.4 综合案例的分析

最后将包含主要的社会经济指标、分行业、分地区、分就业人群的条件综合到一个案例中进行分析，设置综合案例为“浙江省IT行业高校毕业生失业率与主要社会经济指标的相互关系”。其中浙江省属于五大就业区域中的东部沿海就业区，行业为IT行业，就业人群为高校毕业生。符合问题三中提出的分行业、分地区、分就业人群的综合条件。

7.4.1 数据的准备

影响IT行业主要社会经济指标一般包括经济总量、科技投资、进出口总量、统计浙江省近四年来生产总值、科技投资、财政收入总额、海关进出口总量、社会消费品零售总额的季度统计数据以及IT行业的城镇登记失业率，选择生产总值增长率、科技投资增长率、财政收入总额增长率、海关进出口总额增长率、社会消费品零售总额增长率作为城镇登记失业率的主要影响因素，为了便于数学建模，将此五个评价指标进行编号如下：

表8 影响IT行业失业率的主要经济指标编号

编号	评价指标
1	生产总值增长率
2	科技投资增长率
3	财政收入总额增长率
4	进出口总额增长率
5	社会消费品零售总额增长率

7.4.2 模型建立

根据以上统计数据，建立加入经济危机影响因素以及政策影响变量后的混合加权的多元线性回归模型：

$$p_i = \alpha + \beta_1 w_i r_{i1} + \beta_2 w_i r_{i2} + \beta_3 w_i r_{i3} + \beta_4 w_i r_{i4} + \beta_5 w_i r_{i5} + Z$$

【符号说明】

变量	含义
p_i	第 <i>i</i> 个案例的城镇登记失业率
α	常数项

r_{ij}	第 j 个影响因素或指标的统计值
β_i	第 i 个影响因素或指标的回归系数
w_i	处于经济危机下的稳定权重
Z	政策影响变量

将统计数据代入模型，得到最终模型：

$$p_i = 6.438 - 0.325r_1 - 0.077r_2 + 0.153r_3 - 0.172r_4 + 0.047r_5 + Z$$

【符号说明】 r_{i1} = 生产总值增长率； r_{i2} = 科技投资增长率； r_{i3} = 财政收入总额增长率； r_{i4} = 进出口总额增长率； r_{i5} = 社会消费品零售总额增长率； Z 为政策影响变量。

8 问题四的求解

根据问题的含义，要求对 2009 年及 2010 年上半年的我国就业前景进行仿真，因为是对全国的就业前景进行仿真，问题三中的局部层次的数学模型不适合。因此采用问题二中的数学模型进行仿真。

采用问题二中的模型对 2009 年及 2010 年上半年的我国就业前景进行仿真，需要得到模型中所有影响因素（自变量）在 2009 年及 2010 年上半年的数据，因此，首先要预测各主要影响因素在 2009 年及 2010 年上半年的运行数据。

提出假设：

(1) 2009 年及 2010 年上半年，中国经济稳定发展，没有出现大的波动，按照温家宝总理在工作报告中的发展进度发展。

(2) 2009 年及 2010 年上半年与之前的经济发展趋势相差不大。

首先依据对影响就业情况的主要因素 2005 年-2008 年的运行数据，对 2009 年及 2010 年上半年对应指标进行仿真预测，如下图：

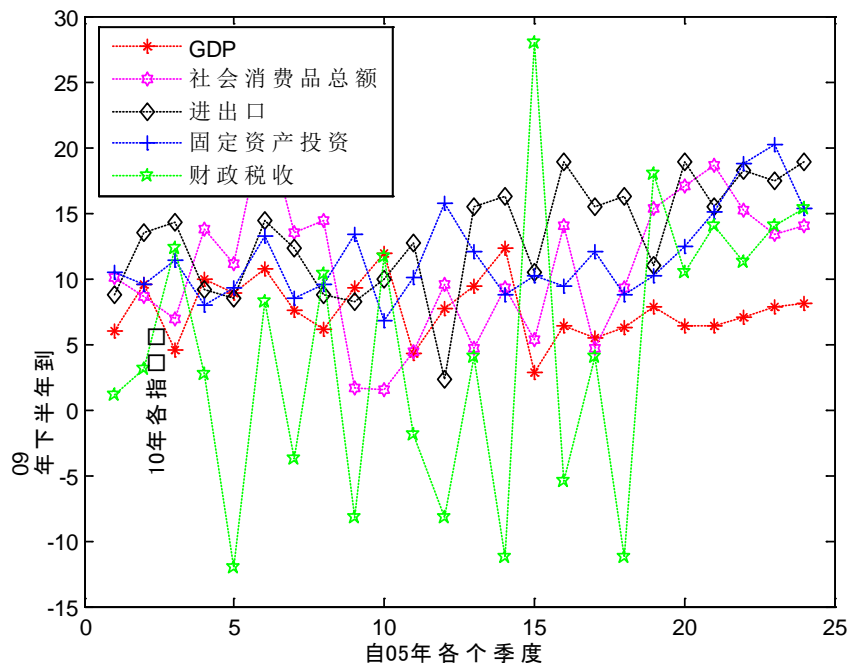


图 6 实际数据与预测数据走势比较图

然后将影响城镇失业率的五个主要指标的预测数据代入问题二中城镇登记失业率与各主要指标的计算模型，得到 2009 年-2010 年上半年的城镇登记失业率的预测数据，其预测数据及走势图表示如下：

表 7 2009 年-2010 年上半年的城镇失业率的预测数据（单位：%）

	09 第一季度	09 第二季度	09 第三季度	09 第四季度	10 第一季度	10 第二季度
失业率	4.1	3.98	3.86	4.01	3.99	4.21

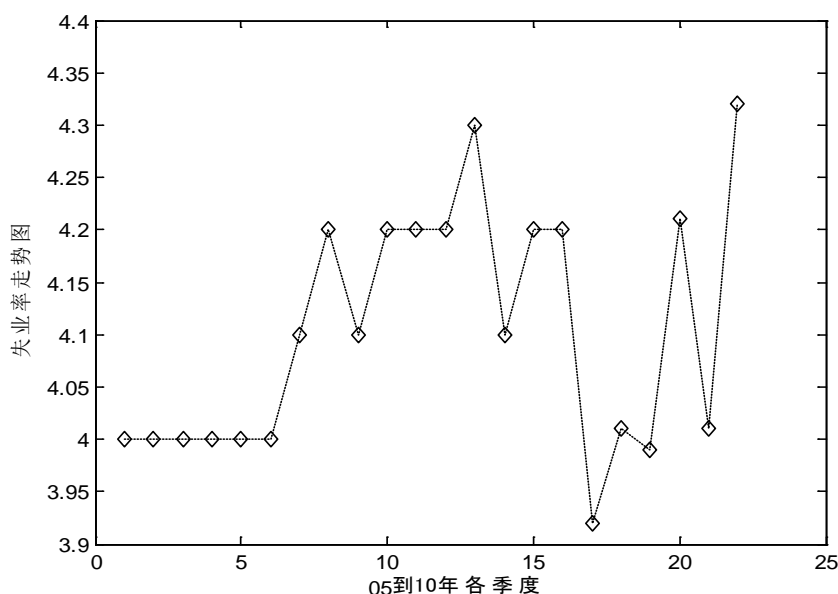


图 7 02 年-10 年城镇失业率的走势图（包含预测数据）

由图表数据不难看出，2009 年-2010 年上半年，在世界各国的共同努力下，经济危机在世界各国的影响不断减弱。经济大环境的好转，加之中国经济的自身复苏，国内外经济发展环境的稳定。国内经济的不断好转，增加了就业岗位，减轻了就业压力，城镇登记失业率也稳定在 4.6% 以下，实现了温总理政府工作报告中提出的 2010 年失业率保持在 4.6% 以下的目标。

9 咨询建议

就业问题不仅是重大的经济社会问题，也是重大的政治问题，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。如果就业问题不能有效解决就会引发一系列社会矛盾和问题。因此我们应从影响就业的主要因素考虑，采取积极有效的就业政策，就以上问题，我们提出一些咨询和建议如下 [18][20][21]：

一、提高社会消费品零售总额、扩大内需

在扩大内需、拉动经济增长以促进就业方面，我国还存在很多问题，一是国际金融危机的前景还有较大不确定性，外部需求下滑的压力依然很大。二是经济发展还缺乏应对危机、克服困难的内在动力和活力。三是因外需减弱使国内一些行业产能过剩的矛盾凸显，对工业和经济增速回升形成较大制约。四是受客观条件限制特别是经济运行规律制约，短期内大幅度扩大内需的难度很大。因此，在扩大内需方面我们要抓住关键、突出重点、统筹兼顾、全面落实。要坚持扩大内需为主和稳定外需相结合，采取强有力的措施扩大国内需求特别是扩大消费需求，以拉动经济增长，降低失业率。

二、增加财政收入，为促进就业提供充足的资金

国家提供一系列措施来缓解就业压力，需要大量资金。因此我们要千方百计增收节支。一方面中央财政应在实施好结构性减税政策的基础上，依法加强税收征管，强化非税收收入管理，努力做到应收尽收，促进财政收入增长。另一方面要牢固树立过紧日子的思想，如压缩公务购车用车、公务接待和出国(境)经费。严格控制党政机关楼

堂馆所建设。全面推行财政科学化、精细化管理，提高财政管理绩效。

三、积极发展进出口贸易

消费品的进口阻碍就业，而资本品的进口有利于就业增加。出口对就业的拉动作用与该国的贸易结构有关：若该国出口劳动密集型商品的数量越大，则对劳动力的需求越大，对就业的拉动作用显著。另一方面，出口间接拉动就业。出口部门往往是一国具有比较优势的部门，即资源使用效率较高的部门，所以出口增加会导致生产资源向效率较高的部门流动，贸易的发展会促进这个部门产量的增加，当企业处于规模经济阶段时，长期平均成本的增加将随着产量的增加而下降，因此企业必然投入更多的资源组织生产，进而促进一国经济增长，增加就业机会。对于我国来说，出口的直接作用表现在我国是劳动丰富的国家，贸易结构以劳动密集型商品为主，因此，出口将有利于拉动劳动就业。

因此我们要实行一系列支持外贸发展的政策措施，如提高出口退税率、加大资金支持力度。改善投资环境，稳定吸收外资的规模，积极发展对外投资和对外工程承包，通过外资、外经工作带动外贸发展。在危机中，广大企业也要苦练内功，增强开拓国内外市场的营销能力，培育自主品牌，增强综合竞争力。商务部门、商会、企业要共同做好应对贸易保护主义的工作，支持企业积极应诉，维护自身正当权益。

四、调整经济结构和企业结构

从经济结构和企业结构的角度来看，为扩大就业需求，我国要大力发展劳动密集型产业和中小企业。当劳动力市场供求矛盾突出、失业压力较大时，各级政府要通过积极的就业投入政策和产业结构调整政策，促进劳动密集型产业的迅速发展，以提高经济增长的就业密度，并大力发展就业容量大、投入少、见效快的第三产业和民营企业。提高积极的就业政策的效率，必须通过政府、雇主和工会组织密切协作，协同促进就业工作。我国目前有再就业工作联系会议制度，但是仍然是政府唱主角。就业不仅有市场机制，还有社会机制，要通过完善政府、雇主和工人（或者工会组织）三方协商机制全面促进就业。

五、培育和发展全国统一有序的劳动力市场

建立全国统一、竞争、有序的劳动力市场，不仅可以缓解失业者求职无门、用人者招人无路，以及一些人没事干和一些事没人干等就业与需求信息不对称的问题，也可以疏通和加快地区之间、城乡之间、行业之间、部门之间的劳动力转移和流动。目前，许多发达国家均已建立了全国统一的劳动力市场和用人信息网络库，随时为失业者提供国家就业计划、政府就业指导、企业招工信息、培训服务信息、专家职业分析、心理咨询、推荐工作岗位、申请失业补助等各种就业信息咨询。这种通过市场机制对劳动力资源及时和有效配置的方式，可以大幅度降低劳动力流动的成本，有效扩大就业。

六、深化教育体制改革，调整人才培养结构

我国目前的就业矛盾既存在总量问题也存在结构问题。结构问题主要反映在教育体制、教育结构和专业设置等方面。如某些专业人才过剩致使大部分学生就业困难，而有些专业如软件人才短缺；在普通教育与职业教育结构上，由于忽视职业技术教育导致技术工人严重不足。因此，应根据社会经济发展对人才的需求结构调整教育结构，缓解目前人才培养结构与市场需求相脱节的矛盾。

从宏观层面上看，积极的就业政策是指以促进就业为取向的宏观政策体系，即不仅要就业作为经济增长的前提和经济运行的结果，而且要将之作为经济发展的基本目标，在产业结构和产业布局的调整以及经济增长方式和增长速度的确定等重要决策中，充分考虑各项措施的就业效应，将能否促进就业增长，作为宏观经济决策的基

本原则。政府的目标是多重的，包括经济增长、市场稳定、社会公平和充分就业等，在经济发展的不同阶段，政府目标的侧重点也有所不同。根据我国当前实际状况，将就业增长作为宏观政策体系的重心，在经济快速增长的同时实现就业增长，有效解决我国因经济结构调整和体制转型带来的严重失业问题。

参考文献

- [1] 国家统计局，中国统计，<http://www.stats.gov.cn/tjsj/ndsj/>，访问时间（2009-9-21）
- [2] 国研网，国务院发展研究中心，<http://www.drcnet.com.cn>，访问时间（2009-9-21）
- [3] 华尔街日报网，华尔街日报，<http://chinese.wsj.com/gb/index.asp>，访问时间（2009-9-21）
- [4] 和讯网，和讯公司，<http://news.hexun.com/>，访问时间（2009-9-21）
- [5] 《组合数学》，卢开澄 卢华明，北京：清华大学出版社，2002、8
- [6] 《Matlab 7.0 使用指南》，苏金明 王永利，西安：电子工业出版社 2004、7
- [7] 《MATLAB 数学实验教程》，胡华，银川：宁夏人民教育出版社，2007、23
- [8] 叶其孝，大学生数学建模竞赛辅导教材，湖南：湖南教育出版社，1998、4
- [9] 吴文虎 王建德，图论的算法与程序设计（Pascal），北京：清华大学出版社，1997、3
- [10] 王沫然，Matlab 与科学计算（第二版），北京：电子工业出版社，2003、9
- [11] 肖铁树，姜启源等，数学试验，北京：高等教育出版社，2000、4
- [12] 谢金星 薛毅，优化建模与LINDO/LINGO软件，北京：清华大学出版社，2005、7
- [13] 李涛等，Matlab 工具箱应用指南——应用数学篇，北京：电子工业出版社，2000、5
- [14] 方芳，朱道元，增长曲线模型中回归系数的局部根方估计，江南大学学报（自然科学版），第6卷第3期：375-378，2007
- [15] 杨雪，欧盟国家就业率变化及相关政策对其的影响，人口学刊，第6期，48-51，2002
- [16] 浦小松，基于logistic 回归模型的加拿大妇女就业影响因素的实证研究及对我国的启示，经济视角，第9期，27-31，2008
- [17] 盛乐，对我国失业问题的微观分析与模型解释，中国经济问题，2002年第1期：28-31，2002
- [18] 张世英 张晖东，我国宏观经济市场非均衡模型，系统工程，1995年第13卷4期，23-30，1995
- [19] 杨东升，经济分析中的预测方法评述，预测，1998年第1期，45-50，1998
- [20] 卢方元，试论经济预测精度问题，经济师，2000 年第11 期，11-12，2000
- [21] 顾煜，优化经济结构促进西部发展，经济师，2000 年第11 期，13-13，2000
- [22] 姜作培 管怀瑾，西方国家的再就业政策及其对我国的启示，晋阳学刊，1998年第3期，47-52，1998
- [23] 陈昌明，金融危机对我国劳动就业的影响分析，中国经贸导刊，2009年第2期，39-40，2000

附录

附录一 全国经济因素统计数据

表 1.1 全国 GDP 统计 (单位: 亿元)

季度	2009 年	2008 年	2007 年	2006 年	2005 年
第一季度	65745	63475	53058	44419.8	38763.3
第二季度	139862	71251	59400	49191.8	42443.2
第三季度		73299	61970	50958	44370.7
第四季度		92645	82878	67353.4	57639.9

表 1.2 社会消费品零售总额统计 (单位: 亿元)

季度	2009 年	2008 年	2007 年	2006 年	2005 年
第一季度	29398	25555.2	21187.8	18440.2	16344
第二季度	29313.2	25487.5	20856	18008	15911.5
第三季度		26843	21783.2	18643.2	16380.8
第四季度		30602	25383	21318.6	18640.3

表 1.3 全国进出口贸易总额统计 (单位: 亿美元)

月份	2009 年	2008 年	2007 年	2006 年	2005 年
1	1417.98	1998.29	1753.6	1204.89	1104.3
2	1249.48	1661.81	1404.36	1058.53	1023.6
3	1620.2	2045.19	1599.9	1449.03	1256.3
4	1707.34	2207.36	1780.17	1434.4	1435.2
5	1641.27	2207.83	1656.53	1332.24	1456.3
6	1825.68	2217.15	1796.27	1481.2	1200.8
7	2002.11	2480.72	1911.31	1460.61	1252.7
8		2410.51	1977.35	1627.36	1408.3
9		2435.01	2010.49	1679.82	1503.5
10		2214.53	1883.97	1524.25	1780.4
11		1898.85	2089.6	1987.88	1245.3
12		1833.34	2061.45	1671.98	1398.12

表 1.4 全国固定资产统计 (单位: 亿元)

月份	2009 年	2008 年	2007 年	2006 年	2005 年
2	10275.79	8121.29	6535.01	5294.1	4221.78

3	23561.9952	18316.94	14543.61	11608.4	9036.68
4	37082.3042	28410.07	22594.41	18005.66	14024.67
5	53520.3249	40264.2	32044.78	25443.47	19719.32
6	78098.354	58435.98	46077.82	36368.35	27967
7	95932.0107	72160.08	56697.83	44771.01	34637.16
8		84919.69	66659	52593.66	41150.9
9		99870.71	78246.78	61880.12	48741.49
10		113189.1	88953.32	70070.52	55792.12
11		127614.1	100604.61	79312.1	63259.85
12		148167.2	117413.91	93472.36	75096.48

附录二 浙江省经济因素统计数据

表 2.1 浙江省财政收入统计（单位：亿元）

季度	2009 年	2008 年	2007 年	2006 年	2005 年
第一季度	1049.58	1071.14	849.77	832.23	712.9
第二季度	1078.32	1050.45	868.23	854.78	743.27
第三季度		882.29	860.45	765.52	789.43
第四季度		726.18	742.95	710.3	801.47

表 2.2 浙江省就业人数统计（单位：千人）

季度	2009 年	2008 年	2007 年	2006 年	2005 年
第一季度	7143.202	6406.3	5682	5031.4	4378.9
第二季度	7150.243	6500.3	5796	5150.3	4464.5
第三季度		6531.6	5840	5213.4	4468.6
第四季度		6498.3	6666.888	6108.4	5311.2

表 2.3 浙江省 GDP 统计（单位：亿元）

季度	2009 年	2008 年	2007 年	2006 年	2005 年
第一季度	4632.13	4601.97	3899.01	3319.38	2746.09
第二季度	5411.41	5286.78	4445.39	7123.39	3347.91
第三季度		5413.17	4639.68	10996.9	2922
第四季度		6185	5654.28	15648.93	4304

表 2.4 浙江省进出口贸易总额统计（单位：亿美元）

季度	2009 年	2008 年	2007 年	2006 年	2005 年
----	--------	--------	--------	--------	--------

第一季度	372.6	462	368.6	280.5	225.09
第二季度	457.7	561.8	437.2	342.25	269.6
第三季度		279	496.77	384.8	290.7
第四季度		465.76	465.6	383.9	384.4

表 2.5 浙江省固定资产投资统计（单位：亿元）

季度	2009 年	2008 年	2007 年	2006 年	2005 年
第一季度	1636.66	1478	1294.68	1169.54	1002.23
第二季度	2750.77	2390	2061.19	1802.52	1605.43
第三季度		2361.17	2061.51	1833.28	1600.7
第四季度		3060.66	3060.39	2784.91	2508.5