

一、 问题重述

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

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长下滑；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。

下面我们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，寻找出影响就业的主要因素或指标。
2. 建立城镇就业人数或城镇登记失业率与上述主要因素或指标之间联系的数学模型。
3. 对上述数学模型从包含主要的经济社会指标、分行业、分地区、分就业人群角度，尝试建立比较精确的数学模型。
4. 利用所建立的关于城镇就业人数或城镇登记失业率的数学模型，根据国家的有关决策和规划对 2009 年及 2010 年上半年的我国就业前景进行仿真。
5. 根据所建立的数学模型和仿真结果，对提高我国城镇就业人口数或减少城镇登记失业率提出咨询建议。

二、 问题的分析

2.1 寻找影响就业的主要因素或指标

我们从失业出发，来寻求影响就业的主要因素或指标。按照西方市场经济国家中较成熟的失业理论，失业一般分为两种类型，一是经济循环所影响的需求不足性失业；另一种是劳动市场供求结合机制不完善的结构性失业和摩擦性失业。解决前者的需求不足性失业，需动用宏观经济政策；解决后者的失业则属于劳动就业政策的范畴。因此我们必须明确失业的类型。

我们先考虑影响失业的宏观经济因素，通过查找相关文献找出与失业尽可能相关的各个宏观经济指标及这些指标的数据。然后先应用相关性分析法，对这些因素进行筛选剔除，找出尽可能多的真正相关的指标。由于各个指标间可能还有相关关系，再利用主成分分析法，对筛选出来的因素进行主成分分析，从而找到我们需要的最重要的相互无关的指标因素组。如果经过分析发现这些宏观经济因素对失业率的影响不是太大，则我们可能考虑到目前我国的失业是第二种类型的失业。然后根据相关文献查找相关失业的其他指标进行分析最后得出结论。

2.2 建立失业率与上述各指标间的数学模型

根据问题一得出的结论，首先查找与失业相关的各个指标的数据，然后对数据进行分析，选择较合理的因素对其建模。

2.3 分行业、分地区、分就业人群建立更精确的模型

由于时间紧迫以及数据的搜集工作比较繁杂，我们只考虑分地区的情况。我们准备选取东西部两个典型地区，对其城镇登记失业率进行分析。

2.4 仿真预测 2009 及 2010 年上半年失业率

根据问题二建立的预测模型，结合近年来的数据，对 2009 年及 2010 年的就业前景进行仿真预测。由于问题二，我们只考虑每年的数据，如果模型也适合于季度或月份数据，我们仍然用这个模型来预测 2010 年失业率；如果不适合，我们将重新拟合模型，使其适合季度或月份数据。

三、 模型假设

- 1、假设我国城镇登记失业率的操作很规范。
- 2、假设 2009 年我国 GDP 增长 8%。
- 3、假设 2010 年之前中国经济呈现复苏现象，国外经济不出现大的波动。
- 4、假设各种统计数据真实。
- 5、假设中国经济运行平稳。

四、模型的建立与解答

问题一：影响失业率（就业）的主要因素

根据相关报告和文献[1]，我们得知影响失业率的因素主要有 GDP、城镇就业人数、固定资产投资总额、社会消费品零售总额、进出口总额、进口总额、出口总额、净出口额、城镇人均收入、CPI、PPI、广义货币供应量 M2、狭义货币 M1、流通现金 M0、汇率等等。于是我们从中国统计局及相关财经网站搜集了以上经济指标从 1990 年至 2008 年的年度的数据。考虑到经济指标的量纲不一致，我们对其进行了归一化处理，进而得到无量纲的统一数据，然后我们利用 Spearman 相关性分析法对各个指标与失业率的相关性进行分析，从中筛选出与失业率相关的主要经济指标。

4.1.1 Spearman 相关性分析的统计模型

模型中的符号意义如下：

- X_i ：第 i 年的失业率。
- Y_{ji} ： $j=1,2,3,\dots$ 第 i 年与失业率可能相关的指标 j ，其中 i 代表年。

- $R(X_i)$: X_i 与其他 X 值相比的秩。
- $R(Y_i)$: 类似 $R(X_i)$ 。
- ρ : 相关性度量

假设: H_0 : X_i 和 Y_{ji} 互相独立 $j=1,2,3\cdots m$ 表示 m 个指标

H_1 : X_i 和 Y_{ji} 不独立, 即有相关关系

X 与 Y 的相关性可用 Spearman 相关系数表示:

$$\rho = \frac{\sum_{i=1}^n R(Y_i)R(X_i) - n\left(\frac{n+1}{2}\right)^2}{\left(\sum_{i=1}^n R(X_i)^2 - n\left(\frac{n+1}{2}\right)^2\right)^{\frac{1}{2}} \left(\sum_{i=1}^n R(Y_i)^2 - n\left(\frac{n+1}{2}\right)^2\right)^{\frac{1}{2}}}$$

此检验用到的统计量为: $\sqrt{n-1}\rho$, 并且: $\sqrt{n-1}\rho \sim N(0,1)$

$$p\text{-值} = 2P(Z \geq |\rho|\sqrt{n-1})$$

如果 ρ 的绝对值大于它的 $1-\alpha/2$ 分位数, 则以水平 α 拒绝 H_0

通过相关性分析, 我们得到与失业率相关的经济指标因素主要有固定投资总额增长率, 社会消费品零售总额增长率, 进出口总额增长率, 净出口增长率, CPI (居民消费价格指数), PPI (工业品出厂价格指数), 城镇人均收入增长率, 广义货币供应量 M2 增长率, 第一产业 GDP 比重。然后我们又对这些指标进行相关性分析发现很多指标间存在相关关系, 所以我们又对这些指标进行主成分分析以找出反映失业的主要指标或综合指标。

4.1.2 对上述经济指标因素进行主成分分析。

1. 主成分分析的基本思想:

主成分分析是一种处理高维数据的方法, 它通过投影的方法, 将高维数据以尽可能少的投影到低维的空间, 使数据将为达到简化数据结构的目的。它也是将多个相关变量以尽可能少的信息损失为原则进行综合化为少数几个不相关变量的方法。

2. 主成分分析的数学模型

设 $X = (x_{ij})$ 为一个 $n \times p$ 的数据阵, 常常列表示变量或指标, 行表示样本或个体, 我们记

$$X = (x_1, \cdots, x_n)^T = (x_{(1)}, \cdots, x_{(p)}),$$

其中 x_i 为 X 的第 i 行, x_j 为 X 的第 j 列, 即 X 既是 n 个 p 维点的集合, 又可视作 p 个 n 维点的集合。考虑 X 是 p 维点的集合时, 我们希望在 R^p 空间的一个低维子空间, 设为 R^q 空间 ($q < p$), 使得这些点到 R^q 子空间的投影点与原始点最接近, 即信息损失最小。

(1) 首先将 $X_{(n \times p)}$ 中心标准化, 即

$$X \rightarrow HXD^{-1},$$

其中 $H = I - \frac{1}{n}11^T$, I 为 n 维单位阵, $1 = (1, \cdots, 1)^T$, $D = \text{diag}\{\|Hx_{(1)}\|, \cdots, \|Hx_{(p)}\|\}$,

变换后的数据仍记为 X 。

(2) 求 $X^T X$ 的特征值 $\lambda_1 \geq \cdots \geq \lambda_p \geq 0$ 和对应的标准正交特征向量 u_1, \cdots, u_p , 这是对 $X^T X$ 的谱分解完成的, 即

$$X^T X = U \Lambda U^T, \text{ 其中 } U = (u_1, \cdots, u_p)。$$

(3) 求主成分对总变差的累计贡献率, 即对 $1 < q \leq p$, 计算

$$\alpha_{(q)} = \sum_{i=1}^q \lambda_i / \sum_{i=1}^p \lambda_i$$

对事先选好的贡献率 c_0 , 确定使 $\alpha_q \geq c_0$ 的最小的 q , 通常取 $c_0 = 85\%$ 。

(4) 计算主成分 $y_{(j)} = Xu_{(j)}, j = 1, \cdots, q$ 即为第 i 个个体 (样本) 在第 j 个主成分上的得分。

(5) 根据实际问题对主成分以及个体特征作出解释, 作图和其他分析。

3. 主成分分析的 SPSS 实现

取 $X = (x_1, \cdots, x_n)^T = (x_{(1)}, \cdots, x_{(p)})$ 为我们筛选出的与失业率相关的经济指标, 其样本数据为各指标从 1990 年到 2008 年的年度数据。对这一样本数据阵 $X_{(19 \times 8)}$, 我们利用 SPSS 软件得到如下结果:

表一 主成分的统计信息

Total Variance Explained						
Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	5.562	69.521	69.521	5.562	69.521	69.521
2	.989	12.369	81.890	.989	12.369	81.890
3	.868	10.850	92.740	.868	10.850	92.740
4	.360	4.496	97.236	.360	4.496	97.236
5	.108	1.345	98.580			
6	.065	.808	99.388			
7	.029	.360	99.748			
8	.020	.252	100.000			

Extraction Method: Principal Component Analysis.

上表显示了各个主成分的贡献率及累积贡献率，由于前三个主成分的累积贡献率就达到了 92.74%，而且第四个主成分的贡献率较小，故我们选取前三个主成分即可。

表二 Component Matrix 信息

Component Matrix^a

	Component			
	1	2	3	4
固定资产投资总额增长率	.742	-.322	-.421	-.113
社会消费品零售总额增长率	.914	-.010	.175	-.326
进出口总额增长率	.578	-.319	.354	.656
净出口增长率	-.609	.260	.623	-.196
CPI	.919	.104	.317	-.101
PPI	.951	-.090	.064	-.122
城镇人均收入增长率	.954	-.100	.188	-.035
广义货币供应量M2增长率	.923	.304	-.112	-.017
一产GDP比重	.502	.788	-.200	.282

Extraction Method: Principal Component Analysis.

a. 4 components extracted.

上表显示，第一主成分包括以下信息：固定资产投资总额增长率，社会消费品零售总额增长率，CPI, PPI, 城镇人均收入增长率，广义货币供应量 M2 增长率，该主成分可以理解为投资和消费对失业的影响；第二主成分主要含有第一产业 GDP

比重的信息，该主成分可以理解为农业或农村劳动力对失业的影响。第三主成分含有净出口总额增长率的信息，该主成分可以理解为净出口对失业的影响。

通过主成分分析可以发现影响就业（失业率）的主要指标按影响程度从大到小排列为 1、投资和消费；2、农业比重或农村剩余劳动力；3、净出口总额（即出口与进口的总额的差额）。然而我们想知道各个主成分对失业率的影响到底有多大，而这可以用主成分回归来实现。

4. 主成分回归分析

由于与失业率相关的主要经济指标之间有很强的相关性，体现出多重共线性，于是经典回归方法作回归系数的最小二乘估计的效果会很差，而采用主成分回归能够克服经典回归带来的不足。

主成分回归的模型如下：

$$Y = \alpha + \beta_1 Z_1^* + \beta_2 Z_2^* + \beta_3 Z_3^* + \varepsilon$$

$$\varepsilon \sim N(0, \sigma^2)$$

其中 $Z_i^*, (i = 1, 2, 3)$ 分别代表第 i 主成分。

通过主成分分析我们可以得到，前三个主成分的累计贡献率达到了 92.74%，于是我们对前三个主成分的预测值在 SPSS 中作回归分析，得到的结果如下：

表三 影响失业率的因素的主成分分析

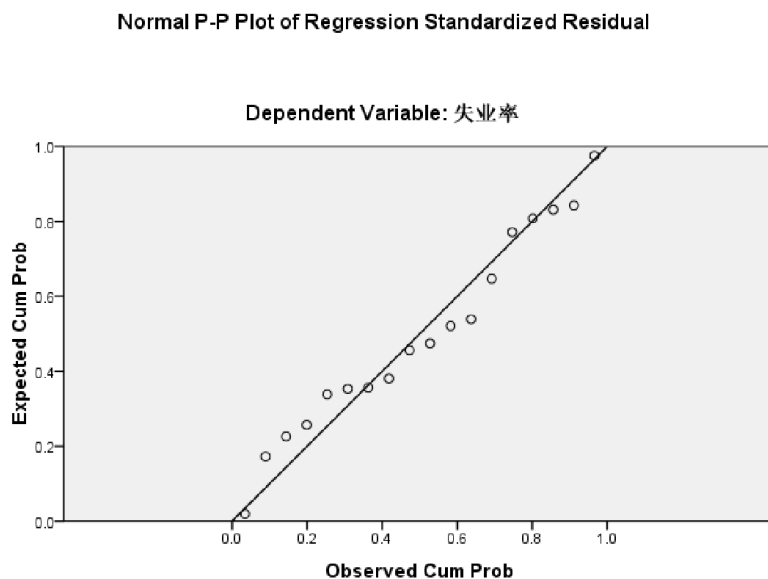
Coefficients ^a					
Model		Unstandardized Coefficients		Standardized Coefficients	
		B	Std. Error	Beta	
1	(Constant)	3.378	.060		55.988
	REGR factor score 1 for analysis 2	-.344	.062	-.495	-5.546
	REGR factor score 2 for analysis 2	-.430	.062	-.618	-6.925
	REGR factor score 3 for analysis 2	.356	.062	.512	5.738

a. Dependent Variable: 失业率

回归系数和回归方程均通过检验，而且效果显著，于是得到回归方程为

$$Y = 3.378 - 0.344Z_1^* - 0.430Z_2^* + 0.356Z_3^* \quad (1)$$

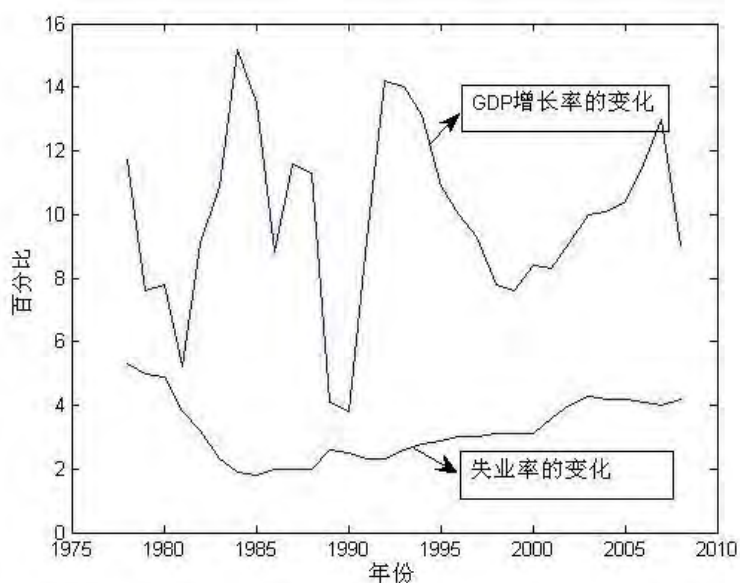
模型检验：模型的拟合优度为 $R^2 = 0.924$



通过标准参差的正态 PP 图大致可以看出，残差是服从正态分布的，于是我可以认定，主成分回归方程的拟合效果是显著的，模型的建立也是可信的。由方程（1）各因素系数的大小我们分析其对失业率的影响，得知净出口总额越大，城镇失业率越大，这与客观实际相违背。

下面我们分析 GDP 增长率与城镇登记失业率的变化关系，见下图：

图一 GDP 增长率与失业率的关系



图一显示，90 年代以后经济增长和城镇失业率的关系不如 80 年代那么密切，GDP 增长的时候，我国城镇失业率反而有上升趋势，这同样有背客观事实，说明单单考虑经济指标与失业率的关系是不对的。也有文献[2]中指出了这样的观点。

结论一：上述宏观经济指标与我国城镇失业率的关系，与实际状况相违背。这说明，我们研究与失业有关的因素时不能简单的只考虑宏观经济因素，而且有文献[2]说明我国的失业还不能达到需求不足性失业，于是我们考虑劳动市场供求结合机制不完善所产生的结构性失业和摩擦性失业的相关因素。

4.1.3 影响失业率的真正因素

根据文献[2]，我们得知（1）我国的经济增长还在高位区，不可能出现需求不足性失业；（2）我国劳动力相对过剩，失业主要由供给过剩造成，与经济波动关系很小。因此，我们考虑影响失业率的上述宏观经济指标以外的因素。

根据相关文献[2]得知影响失业率的因素以下几个方面：

- ◆ 上一年的失业率（ U_{t-1} ），上一年结余的劳动力也是劳动力供给，影响新生劳动力的失业发生和失业流出。
- ◆ 结构变化指标（ STO ），这里区分公有经济和其他经济两组的职工增长差异。表达式如下：

$$STO = \sum_{i=1}^n |\Delta E_i - \Delta E| E_i / E \quad \text{其中，} E_i \text{ 为公有经济（} i=1 \text{）或其他经济}$$

（ $i=2$ ）的就业者人数， ΔE 为就业变化率。

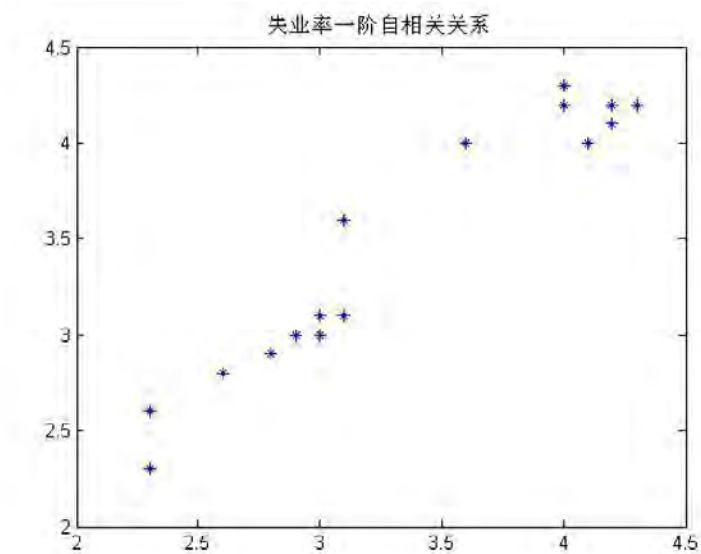
- ◆ 直接接受的农村劳动力占城镇新就业人数的比例（ F ），表示农村劳动力挤占城镇劳动力就业机会的程度，对失业率的影响。由于此数据很难获得，我们考虑与此量相关的指标。考虑到农村劳动力向城镇转移主要受城乡收入差距、物价水平等因素有关，所以我们用城乡收入差距指数来量化此指标：

$$F = \frac{\text{城镇绝对收入}}{\text{城镇CPI}} / \frac{\text{农村绝对收入}}{\text{农村CPI}}$$

- ◆ 第一产业 GDP 所占比例（ X ），一产 GDP 越高说明农村劳动力参与越多，一定程度上能减轻城镇失业率。
- ◆ 城镇新增就业人口，该指标作为影响失业率的因素应该是显然的。

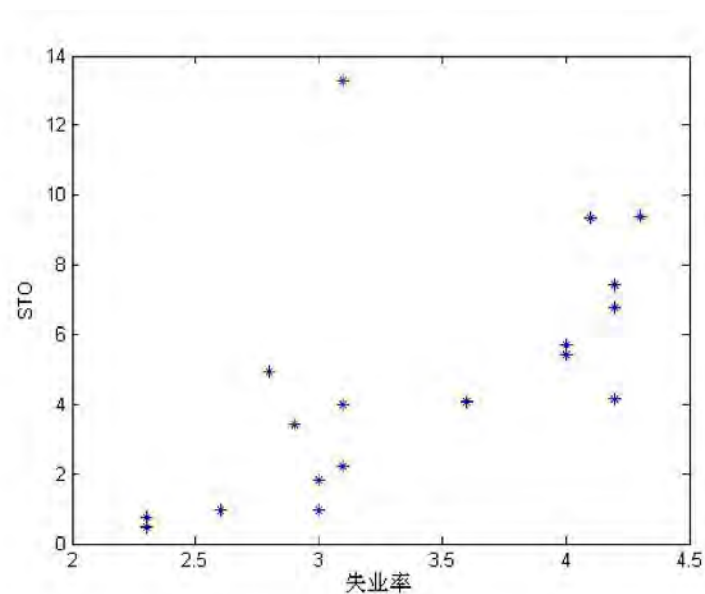
对于上述指标因素与失业率的关系，我们分别做了统计分析，得到的结果如下：

图二 上一年失业率对当年失业率的影响



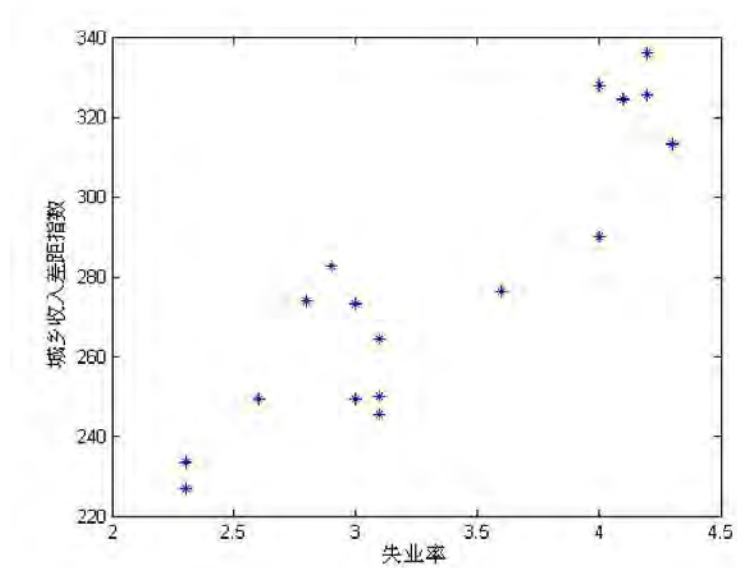
从图二可以看出，上一年的失业率与当年失业率成正相关关系，即上一年失业率越高，第二年失业率也会越高，这是符合实际情况的

图三 结构变化（ STO ）与失业率的关系



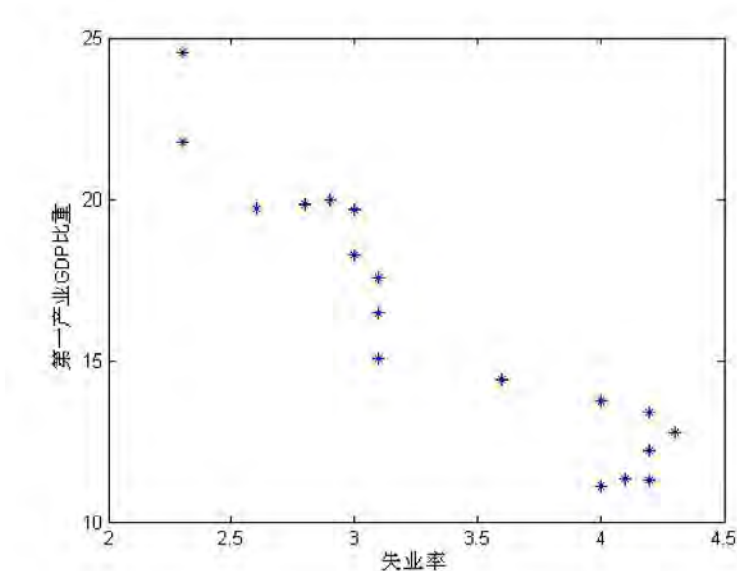
图三表示，劳动力市场供求结构变化与失业率成正相关，这说明，在劳动力市场中，结构变化越厉害，失业率越高。这是因为，结构变化后，劳动力需要重新寻找工作，而在找到新的工作之前需要一段时间，即要失业。

图四 失业率与城乡收入差距指数关系图



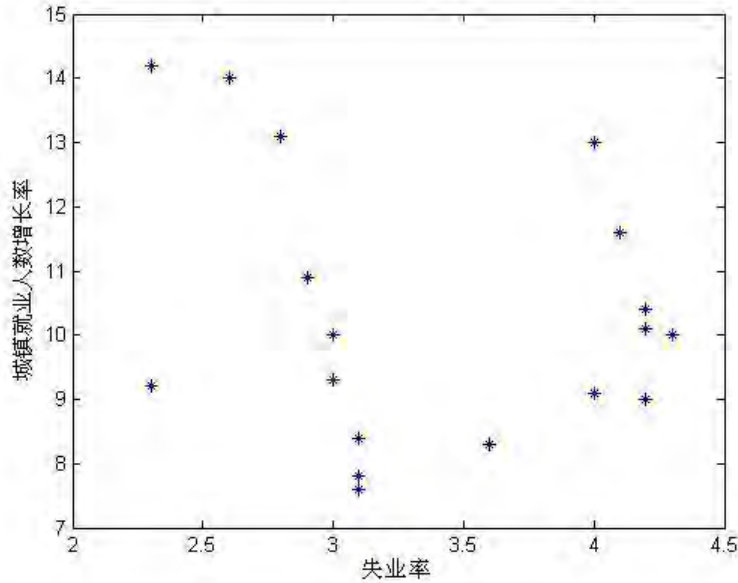
从图四我们看出，城乡收入差距与失业率成正相关，这是因为城乡收入差距越大，农村剩余劳动力就会更加迫切进去城市获取更多经济收入。进而会挤占更多的城镇就业机会，使得城镇失业率变大。

图五 第一产业 GDP 所占比例与失业率的关系



如图五，第一产业 GDP 所占比例增加，失业率降低。这是因为第一产业 GDP 增加，说明劳动力参与到第一产业的比例比较高，因此农村剩余劳动力流向城镇的比例就要减少，进而使城镇失业率降低。

图六 城镇新增就业人口



上图六显示，城镇新增就业人数对失业率的影响时大时小，这一因素对城镇失业率不具有长期影响力，因此我们只考虑上述分析的其他指标。

结论二： 由以上分析结果，我们得到结论，影响城镇失业率的主要因素有：

- 上一年的失业率 U_{t-1}
- 结构变化指标 STO
- 直接接受的农村劳动力占城镇新就业人数的比例 F ，用城乡收入差距指数表示
- 第一产业 GDP 所占比例 X

而且我们发现上述各因素与失业率基本上都成线性关系。

问题二 城镇登记失业率与各主要因素的模型建立

4.2.1 模型的建立

目前，我国的实际就业情况与 GDP 增长的关系很小，但是为了更加全面得考虑影响失业率的因素，我们还是考虑了 GDP 对失业率的影响,这是因为经济学理论认为 GDP 增长是拉动就业岗位的主要因素。因此我们考虑了上一年的失业率 U_{t-1} ，结构变化指标 STO ，直接接受的农村劳动力占城镇新就业人数的比例 F ，用城乡收入差距指数表示，第一产业 GDP 所占比例 X ，以及 GDP 的增长率五个因素。

由问题一的分析，我们知道这五个因素与失业率都成线性关系，因此我们对其建立多元线性回归模型，如下：

$$U_t = \alpha_0 + \alpha_1 U_{t-1} + \alpha_2 GDP + \alpha_3 STO + \alpha_4 F + \alpha_5 X + \varepsilon \quad (2)$$

$$\varepsilon \sim N(0, \sigma^2)$$

我们首先利用统计数据计算了各个指标的值，然后输入 SPSS 软件系统，得到如下结果：

表 4.2.3.1

Coefficients ^a					
Model		Unstandardized Coefficients		Standardized Coefficients	Sig.
		B	Std. Error	Beta	
1	(Constant)	.211	.328		1.527
	上期失业率	.450	.121	.441	3.731
	GDP增长率	-.047	.021	-.138	2.225
	STO	1.448	.573	.070	1.634
	城乡收入差距指数	.823	.290	.431	2.841
	GDP1的比重	-1.604	1.079	-.101	1.572

a. Dependent Variable: 失业率

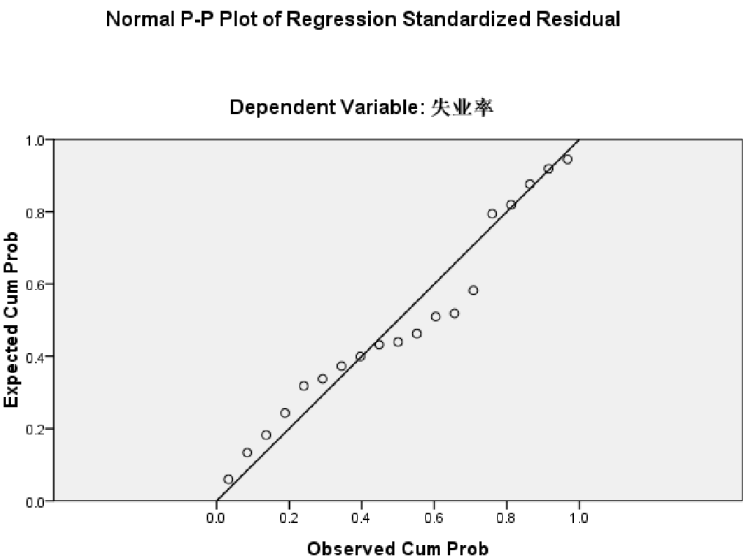
故其回归方程为

$$U_t = 0.211 + 0.450U_{t-1} - 0.047GDP + 1.448STO + 0.823F - 1.604X \quad (3)$$

4.2.2 模型检验：

从表 4.2.3.1 知，每个变量表示系数的 t 值都比较高，具有统计意义，说明所选的变量对我国失业率都有影响，而且正负号也符合我们的合理期待。

拟合优度 $R^2 = 0.976$ ，PP 概率图如下：



通过残差的正态 PP 图大致可以看出，残差是服从正态分布的，于是我们可以认定，上述回归方程的拟合效果是显著的，模型的建立也是可信的。

4.2.3 模型解释：

式 (3) 的回归结果说明，影响我国城镇失业率的五个因素按影响程度由大到

小依次为 一产 GDP 所占比重, 产业结构调整程度 ST0, 城乡收入差距 F, 上一期失业率 U_{t-1} , GDP 增长率。从而我们得知, 影响我国城镇失业率的主要因素是产业结构调整程度和农村剩余劳动力的比例及流动性, 而经济增长的影响是比较小的。

问题三 分地区建立更精确的数学模型

我们只分地区考虑来建立模型, 作为东西部地区的代表, 选取上海和甘肃两地区为例。

结合问题二的模型, 我们建立地区统计模型如下:

$$\begin{aligned} U_t &= \alpha + \beta X + \varepsilon \\ \varepsilon &\sim N(0, \sigma^2) \end{aligned} \quad (4)$$

X 代表问题二中, 我们考虑的全部或部分指标因素组。

4.3.1 上海的模型分析如下:

第一步: 考虑问题二中所有因素, 将上海地区失业率因素的数据进行处理, 输入 SPSS 软件系统, 我们得到如下结果:

表 4.3.1.1

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-3.683	2.066		-1.783	.100
	GDP	.031	.030	.063	1.024	.326
	GDP1比重	.455	.318	.353	1.431	.178
	城乡收入差距指数	2.065	.708	.688	2.916	.013
	上期失业率	.510	.154	.558	3.313	.006
	ST0	5.703	2.487	.178	2.293	.041

a. Dependent Variable: 失业率

从表 4.3.1.1 可以发现, 对于上海的失业率, GDP 的影响系数较小, P 值很大, 因此我们剔除 GDP 的增长率这一变量, 用其他变量做回归。得出如下结果:

表 4.3.1.2

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-2.408	1.652		-1.458	.169
	GDP1比重	.311	.286	.241	1.088	.296
	城乡收入差距指数	1.742	.635	.581	2.742	.017
	上期失业率	.522	.154	.572	3.402	.005
	STO	4.569	2.231	.143	2.048	.061

a. Dependent Variable: 失业率

从表 4.3.1.2 得知，第一产业 GDP 所占比重，其对上海失业率的影响也很小。因此，我们也将其剔除，只考虑剩余变量。

第三步，我们计算剩余因素对上海失业率的影响，得到如下结果：

表 4.3.1.3

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-.687	.477		-1.438	.172
	城乡收入差距指数	1.253	.452	.418	2.771	.015
	上期失业率	.469	.146	.514	3.202	.006
	STO	3.637	2.073	.113	1.754	.101

a. Dependent Variable: 失业率

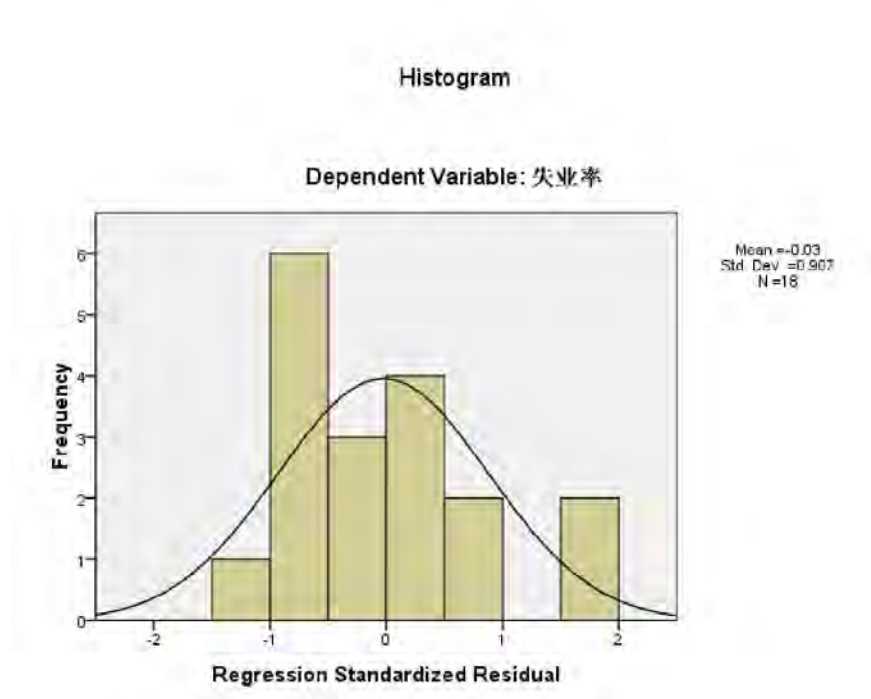
从表 4.3.1.3 可知，影响上海城镇失业率的主要因素为 城乡收入差距指数，上一年失业率，和产业结构变化指数 STO，其中产业结构变化指数 STO 的影响程度最大。

因此，我们针对上海城镇登记失业率的变化建立以下模型：

$$U_t = -0.687 + 1.253F + 0.469U_{t-1} + 3.637STO \quad (5)$$

模型检验：

$$R^2 = 0.961$$



上述结果显示， $R^2 = 0.961$ ，且残差服从正态分布，说明我们的预测模型是合乎统计学要求的。

由以上分析，我们得知影响上海城镇登记失业率的主要因素是城乡收入差距指数，上一年失业率，和产业结构变化指数 ST0，其中产业结构变化指数 ST0 的影响程度最大。

4.3.2 甘肃地区城镇失业率分析

这里与上海失业率分析步骤类似，对于甘肃地区城镇失业率影响因素我们得到如下结果：

表 4.3.2.1
Coefficients^a

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
1 (Constant)	-2.388	1.522		-1.569	.139
城乡收入差距指数	.496	.243	.477	2.043	.060
上期失业率	.694	.092	.958	7.547	.000
STO	3.790	1.929	.256	1.965	.070
GDP1比重	.068	.032	.525	2.112	.053

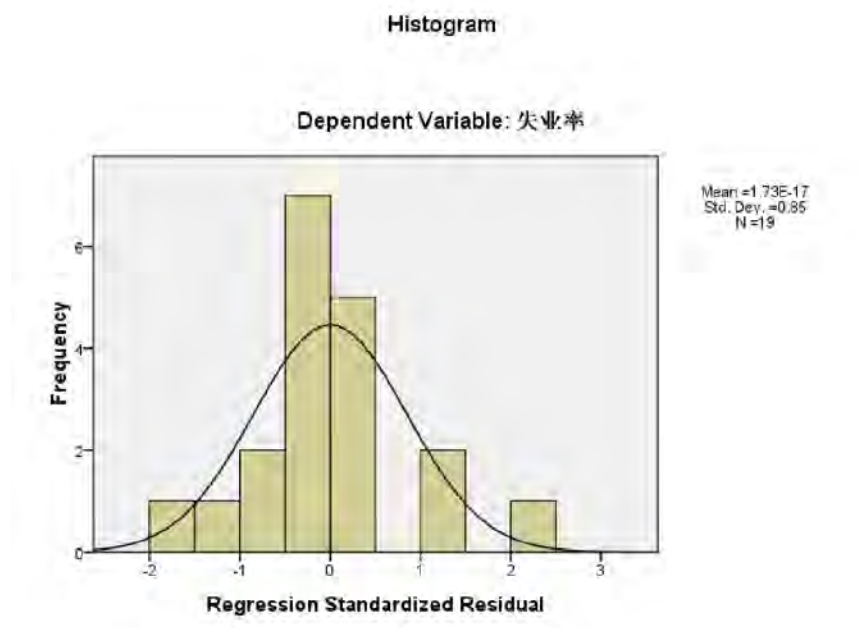
a. Dependent Variable: 失业率

从表 3.2.1 我们得知，影响甘肃城镇登记失业率的主要因素有：城乡收入差距指数，上一年失业率，产业结构变化程度 STO, 一产 GDP 所占比重。

我们得到甘肃城镇就业失业率的模型如下：

$$U_t = -2.388 + 0.496F + 0.694U_{t-1} + 3.970STO + 0.068X \quad (6)$$

模型检验：



通过上表可知变量的系数通过显著性检验，并且拟合优度 $R^2 = 0.848$ ，残差服

从正态分布，拟合较好，检验通过。

同样，我们得知甘肃的城镇登记失业率的影响因素主要有城乡收入差距指数 F ，上一年失业率，产业结构变化程度 $ST0$ ，一产 GDP 所占比重 X 。按照影响因素由大到小是产业结构变化程度 $ST0$ ，上一年失业率，城乡收入差距指数 F 和一产 GDP 所占比重。

结论三：通过对上海和甘肃两个地区的城镇失业率的分析，我们得知

- a、影响两地失业率的共同因素是城乡收入差距指数 F ，上一年失业率，产业结构变化程度 $ST0$ ，这正说明，我们上面认为的观点，即近年来，宏观经济指标对我国失业率的影响是微弱的，然而影响我国失业率的主要因素是上述指标。
- b、两地失业率所受上述因素的影响程度有所不同：一方面，作为经济发达的上海，城乡收入差距 F （即农村劳动力挤占城镇就业机会的比例）对其失业率的影响相对于欠发达的甘肃地区更大；另一方面，由于甘肃地区对第一产业投入比例比较大，加上其城乡差距相对上海要小，因此，其失业率受一产 GDP 比重的影响要比上海严重，事实上，由模型可知，上海失业率几乎不受一产 GDP 比重影响。我们可以理解为，其一产 GDP 比重越大，农村劳动力挤占城镇就业机会就越小，这也符合我们的实际情况。

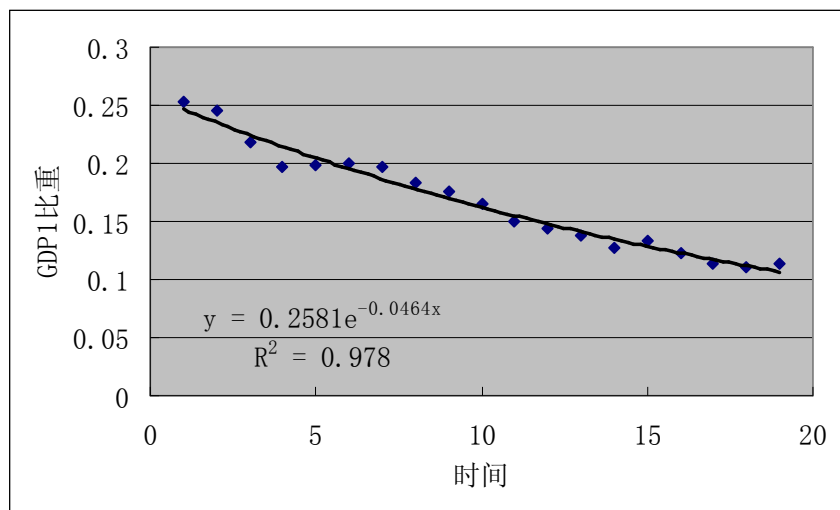
问题四 对 2009 年及 2010 年上半年我国就业前景进行仿真

4.4.1 2009 年失业率的预测

数据的准备：

- GDP 增长率，由于政府的目标是保证 2009 年的 GDP 增长率为 8%，故假设我国 GDP 增长率为 8%。
- 上一年的失业率，由统计数据可以得到 $U_{2008}=4.2\%$
- 结构变化 $ST0$ ，此值较难获得，但是我们可以用上两年 $ST0$ 的平均值来近似表示，此时 $ST0$ 为 0.050072
- 一产 GDP 占总 GDP 的比例，由以下的散点图可以知道它随着时间的变化是有规律的，所以我们首先对其进行拟合，然后利用拟合方程对其进行预测。

图七 一产 GDP 占总 GDP 的比例



经过数据的拟合我们得到拟合方程为

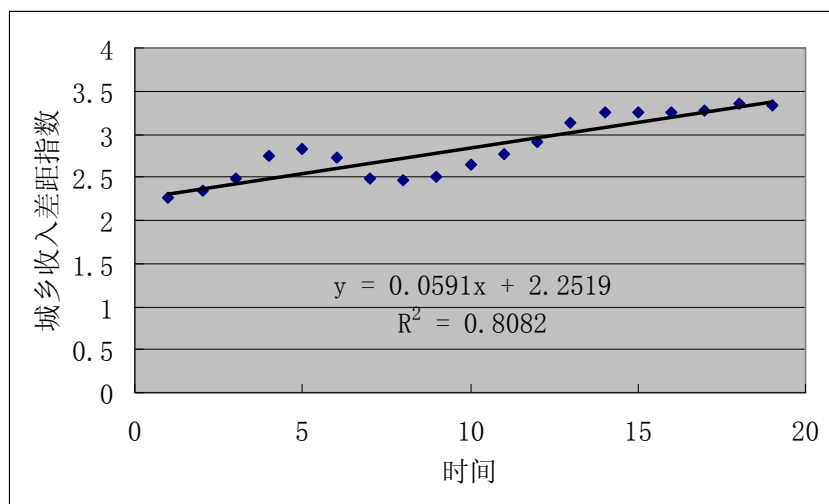
$$y = 0.2581 e^{-0.0464 t}$$

其中拟合优度为 0.987，说明拟合效果很好。

利用此方程我们可以得到 2009 年一产 GDP 所占比例为 10.2%。

- 城乡收入差距指数，我们发现每年的变化是有规律可循的，如图：

图八 城乡收入差距指数预测



同样我们可以对其拟合，拟合方程为 $y = 0.0591x + 2.2519$ ，拟合优度为

0.8082，说明拟合效果不错。利用此方程我们可以得到 2009 年城乡收入差距指数的预测值为 3.4339。

将以上数据带入我们的预测模型

$$U_t = 0.211 + 0.450U_{t-1} - 0.047GDP + 1.448STO + 0.823F - 1.604X \quad (7)$$

得知 $U_{2009} = 4.46$ 。

因此我们得出结论，到 2009 年我国城镇登记失业率将达到 4.46%。

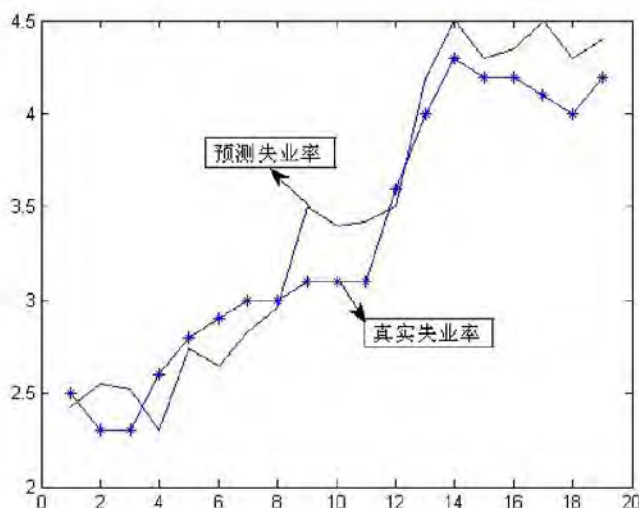
4.4.2 2010 年上半年失业率的预测：

由于问题二我们是对年数据建立的模型，而此问题要求预测上半年的失业率，所以我们考虑到问题二的模型可能不适合于此问题的预测。因此我们想先用半年的数据来检验问题二的模型，如果正确，我们将用问题二的模型进行预测。如果不正确重新建立数学模型。

模型的检验：

从相关网站上我们得到了与问题二的指标有关的数据，经过数据处理最终得到了五个指标的数据，然后将此数据带入问题二的方程得到预测失业率，我们想比较真实失业率与预测失业率的关系，下图为两者的关系图：

图九 真实失业率与预测失业率的关系



从上图我们可以得知用半年的数据预测的失业率与年数据的真实失业率相差不大，所以问题二的模型同样适合半年的数据，因此我们可以用问题二的模型对 2010 年上半年的失业率作出预测。考虑到经济周期的影响我们需要用 2009 年上半年的失业率对 2010 年上半年失业率做预测，另外还要考虑 2010 年上半年的其他四个指标。

数据的准备：

- GDP 增长率，网上相关专家预测 2010 年上半年的 GDP 增长率为 9%，这里我们假设 2010 年上半年我国 GDP 增长率为 9%。
- 2009 上半年的失业率，4.3%（已知）
- 结构变化 ST0 的值，此数据较难获得，同样可以用上两年 ST0 的平均来近似表示。此时 ST0 为 0.05437。
- 一产 GDP 占总 GDP 的比例，与 2009 年失业率的预测相同，经过数据的拟合我们得到拟合方程为

$$y = 0.2641e^{-0.0457t}$$

其中拟合优度为 0.96，说明拟合效果很好。利用此方程我们可以得到 2010 上半年一产 GDP 所占比例为 10.11%。

- 城乡收入差距指数，考虑到由于半年时间城乡收入差距变化不大，所以用 2010 年的年城乡收入差距指数代表上半年的城乡收入差距指数，2010 年的年城乡收入差距指数仍用公式 $y = 0.0591x + 2.2519$ 预测，由此得到值为 3.4909。

$$U_t = 0.211 + 0.450U_{t-1} - 0.047GDP + 1.448STO + 0.823F - 1.604X \quad (8)$$

把上述五个指标代入方程（8），我们得到 2010 年上半年的失业率是 4.446%。

问题五 对减少我国失业率的咨询建议

由我们得到的模型和结论可以看出，到 2009 年，我国城镇登记失业率将会进一步增高，达到 4.46%，为进一步控制失业率的上升，我们提出以下建议：

- ◆ 继续保持 GDP 的持续增长。虽然在我们的模型中 GDP 对失业率的影响比较小，但是 GDP 依然是拉动就业的主要因素。所以要降低失业率应该继续保持我国经济增长，而消费、投资、进出口是拉动经济的三驾马车，因此应该着力扩大国内需求，进一步扩大居民消费，保持投资稳定增长，并且增加进出口的数量。
- ◆ 合理减少劳动力的流动性。有关部门应加强《劳动合同法》及《实施条例》的精神和主要内容，自觉遵守各项劳动法规，努力构建和谐稳定的劳动关系，积极整合人才市场和劳动力市场信息平台，出台国家统一的社会保险接续办法，增强人力资源市场的合理流动性和配置效率。[3]
- ◆ 缩小城乡收入差距，进一步完善小额贷款制度，适当向农民兄弟倾斜。让农民通过贷款实现创业，再通过创业带动就业，这将是缩小城乡差距的一条重要途径。[4]从而，农民挤占城镇就业机会的比例就会降低。
- ◆ 合理优化产业结构，调整一、二、三产的投资比重，以此来控制农村剩余劳动力，避免给城镇就业带来巨大压力。

五、 模型的改进设想

- 1、影响失业率的因素不是很全面，还有待更进一步的探讨和研究。
- 2、我们还没有充分考虑国家宏观调控的因素对失业率的影响。
- 3、我们考虑的是年度数据，考虑更细致的月度或季度数据，模型会更精确。

六、参考文献

- [1] 张兴会, 杜升之, 基于对角 Elman 神经网络的失业预测模型, 南开大学学报, 第 35 卷, 第 2 期, 2002 年 6 月。
- [2] 钱小英, 我国失业率的特征及其影响因素分析, 经济研究, 1998 年第 10 期 第 28-36 页.
- [3] 杨宜勇, 池振合, 2009 年中国就业形势及对策建议, 经济研究参考, 2009 年第一期。
- [4] 肖青, 让小额贷款切实起到促进创业作用, 中国劳动保障报, 2008 年 9 月 9 日。
- [5] N. 格里高利. 曼昆, 宏观经济学(第五版), 北京: 中国人民大学出版社, 2005 年。
- [6] W. J. Conover, 实用非参数统计(第三版), 北京: 人民邮电出版社, 2006 年。
- [7] 姜启源, 谢金星, 叶俊, 数学模型(第三版), 北京: 高等教育出版社, 2003 年。
- [8] 张润楚, 多元统计分析, 北京: 科学出版社, 2006 年
- [9] 宇传华, SPSS 与统计分析, 北京: 电子工业出版社, 2007 年。
- [10] 龚刚敏, 我国矫正口径失业率的预测及相关失业预测模型评价, 财经论丛, 第 6 期(总 120 期), 2005 年 11 月, Page: 61—67。