

全国第六届研究生数学建模竞赛



题 目

城镇登记失业率影响因素的探讨

摘 要：

本文就城镇登记失业率的相关问题，建立模型解决了寻找影响因素以及影响因素和城镇登记失业率关联的问题，并利用所建模型对未来我国的就业前景作出仿真预测，得出了相应结论。

首先，根据计量经济学的滞后变量模型和格兰杰因果关系检验理论，利用 Eviews 软件对相关统计数据进行分析，确定了影响城镇登记失业率的八个主要因素。

针对第二个问题，本文利用主成分分析法，建立了各影响因素与城镇登记失业率之间相互关联的模型。然后，通过研究相关理论，对城镇登记失业率和各影响因素进行协整分析并建立了基于序列相关性的误差修正模型 ECM：

$$\Delta y_t = \beta_0 + \beta_1 \Delta x_t + (\beta_2 - 1) \left(y - \frac{\beta_1 + \beta_3}{1 - \beta_2} x \right)_{t-1} + \varepsilon_t$$

针对第三个问题，本文根据不同地区、不同产业表现出的不同特征，以河南省为例，引入各产业的劳动生产率 L_i 和结构偏离度 P_i ，就当地的产业结构和就业结构的关系问题建立向量误差修正模型 VEC，进一步揭示了两者之间在一定时期内的协整关系。

针对第四个问题，作为对比，本文首先利用了灰色预测模型对未来两年我国就业前景进行了预测。然后，利用所建立的 ECM 模型对就业前景作了进一步仿真并得出相应结论。

最后，结合模型所得到的结论和实际情况，就减少城镇登记失业率提出了我们的咨询建议。

关键词：格兰杰因果关系检验 主成分分析法 误差修正模型 ECM 灰色预测

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。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。但由于统计口径不同，存在一定的差异，有些历史数据也较难获得。

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长速下滑；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。

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

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。
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你们的咨询建议。

二、模型假设

1. 各个因素的数据均真实可靠;
2. 短时间内各经济指标保持稳定;
3. 此处失业一般指正常情况下的失业, 不考虑特殊情况下的失业;
4. 假设所有参与统计的样本都为正常的经济人;
5. 假设所有参与登记失业的人群均为有劳动能力且愿意继续就业的劳动力;
6. 不考虑不同统计口径所存在的差异;
7. 假设 2009 年及 2010 年上半年间不出现重大异常情况。

三、符号说明

- s_t 第 t 年登记失业率;
- b_t 第 t 年高校毕业生规模;
- e_t 第 t 年货物进出口总额;
- g_t 第 t 年国内生产总值 GDP;
- i_t 第 t 年全社会固定资产投资总额;
- j_t 第 t 年居民消费水平;
- r_t 第 t 年人民币汇率;
- z_t 第 t 年职工平均工资;
- c_t 第 t 年城镇人口;
- Y_{n-m} 被解释变量 Y 的第 m 期滞后;
- X_{n-i} 解释变量的第 i 期滞后;
- α_0 短期或即期乘数;
- α_i 动态乘数或延迟系数;
- RSS_R 受约束的残差平方和;
- RSS_{UR} 无约束的残差平方和;
- Z 标准化阵;
- R 相关系数矩阵;
- U_p 称为第 p 主成分;
- d_u 临界上限;
- d_l 临界下限;
- L_i 第 i 产业劳动生产率;
- P_i 第 i 产业结构偏离度;

四、问题分析

本题是一个开放性的探讨预测问题。

根据题目要求，我们需要在搜集处理出相关数据的基础之上，建立模型，解决寻找影响因素以及影响因素和城镇登记失业率关联的问题，并利用所建模型对未来我国的就业前景作出仿真预测，得出相应结论。

根据相关资料，失业理论有很多，包括劳动力市场失灵理论、凯恩斯失业理论、内部人——外部人理论、效率工资理论以及劳动力市场缺失的隐性失业理论等^[2]。综合以上理论，影响失业的主要因素不少于以下几种：(1)经济增长：奥肯定律表明经济增长与失业率是负相关的，经济的增长能够创造更多的就业机会，从而降低失业率；(2)通货膨胀：菲利普斯曲线说明失业率和通货膨胀率表现为一种替代关系。通货膨胀率越高，失业率越低；(3)产业结构：第二三产业能够吸收从第一产业中转移出来的劳动力，从而对失业率产生重要影响。(4)人力资本结构：人力资本结构如果与经济发展模式、职位需求不一致，会导致结构性失业，从而增加失业水平。

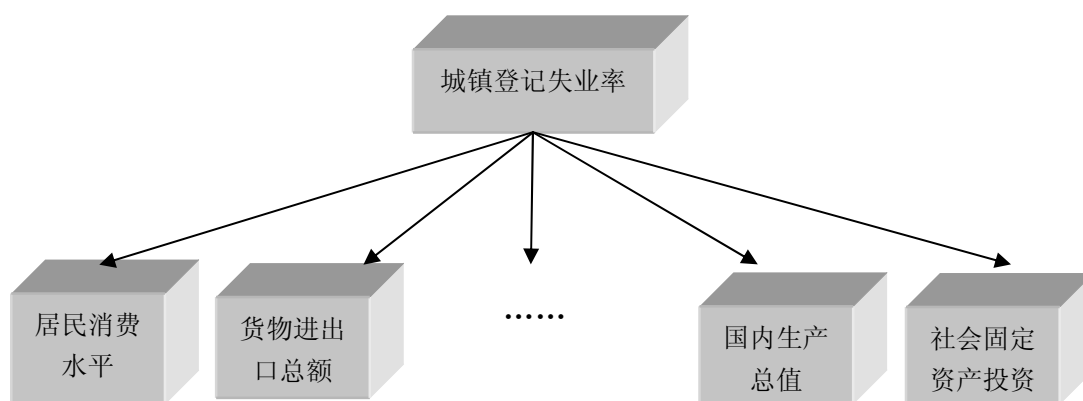


图 4-1 可能影响城镇登记失业率的因素

由于我国正处于体制转型期，同时经济二元结构的现状使我国存在大量的隐性失业，从而使得失业率很难有一个良好的度量。本文从就业总量出发来研究各种因素对就业的影响。经济增长与就业增长之间具有一致性的变动关系，索洛模型和奥肯定律从不同的角度对这种一致性做出了阐述。因此，在参照以上理论的基础上，通过查询年鉴，暂列出以下几个具体指标：

(1) 经济增长对就业的影响：由于国内生产总值是衡量经济发展水平的重要指标之一，因此，选取其作为影响就业总量的首选指标。同时根据投入、产出原理，还需加入全社会固定资产投资额、财政支出这两个指标来反映内需；加入进出口总额，反映外需。另外，由于经济活动与人的消费、需求是十分不开的，因此，还需加入居民消费水平，反映消费与积累的关系。

(2) 通货膨胀、货币政策对就业的影响：鉴于目前的研究主要是通过工资、消费来反映其对就业的影响，除选取居民消费水平外，还需加入反映职工工资水平的平均工资作为影响就业总量的指标。

(3) 产业结构对就业的影响：借鉴美国经济学家西蒙库兹涅茨的研究方法，选取三次产业的总产值，即第一产业产值、第二产业产值、第三产业产值的总和作为研究产业结构的变化对就业总量的影响指标。

(4) 人口对就业的影响：鉴于我国人口对就业的影响主要是由于人口的基数太大

而造成的就业压力，因此，选取年末城镇人口数，即从人口总量出发来研究人口对就业总量的影响。

经初步分析，建立模型应从以下几个方面考虑：

1. 收集相关数据并确定相应指标

怎样选取以及选取什么样的影响指标至关重要。首先，所选取的指标应尽可能的反映失业率的情况；其次，根据数据处理的要求，样本点必须大于指标数，且各个变量间要统一量纲；第三，指标过多时，会影响模型预测的精度；最后，对于各影响指标要求尽可能的独立，避免多重相关性。经分析，模型中可能用到居民消费水平、全社会固定资产投资、职工平均工资、货物进出口总额、国内生产总值、城镇人口、人民币汇率和当年高校毕业生数等数据。各种数据为后续讨论提供依据，所以必须保证所用原始数据的权威性、合理性和计算的科学性。

2. 建立一个影响失业率的数学模型

首先把问题中的各个因素条理化、归属化，再充分考虑实际可能因素的影响，修改建立较为完善的影响失业率的数学模型。

3. 从模型应用结果中得出有用结论

这部分是文章建立数学模型的出发点和落脚点。通过所建模型应当得出对解决实际问题的指导意义。能否得到有用结论，关键在于考虑因素是否全面、所建模型是否合理，必要时可兼用定性分析法。

五、影响因素的因果关系确定

本文暂将影响失业率的因素初步定为居民消费水平、全社会固定资产投资、职工平均工资、货物进出口总额、国内生产总值、城镇人口、人民币汇率和当年高校毕业生数等方面。根据《中国统计年鉴 2008》中的相关统计资料^[3]，我们得出从 1998 至 2007 年十年间这些影响因素及失业率的值，如表 5-1 所示。

表 5-1 1998 至 2007 失业率及其影响因素

指标 年份	影响失业率的因素								失业率 s (%)
	居民消 费水平 j (元)	全社会固 定资产投 资 i (亿元)	职工 平均 工资 z (元)	货物进 出口总 额 e (亿 元)	国内生产 总值 GDP g (亿元)	城镇 人口 c (万 人)	人民币汇 率 r (人民 币元/100 美元)	当年高 校毕业 生数 b (万人)	
1998	3159	28406.2	7479	26849.7	84402.3	41608	827.91	83.0	3.1
1999	3346	29854.7	8346	29896.2	89677.1	43748	827.83	84.8	3.1
2000	3632	32917.7	9371	39273.2	99214.6	45906	827.84	95.0	3.1
2001	3869	37213.5	10870	42183.6	109655.2	48064	827.70	103.6	3.6
2002	4106	43499.9	12422	51378.2	120332.7	50212	827.70	133.7	4.0
2003	4411	55566.6	14040	70483.5	135822.8	52376	827.70	187.7	4.3
2004	4925	70477.4	16024	95539.1	159878.3	54283	827.68	239.1	4.2
2005	5463	88773.6	18364	116921	183217.4	56212	819.17	306.8	4.2
2006	6138	109998.2	21001	140971	211923.5	57706	797.18	377.5	4.1
2007	7081	137323.9	24932	166740	249529.9	59379	760.40	447.8	4.0

通过计量经济学的相关理论及模型，以从上表中收集整理的数据为例，来计算寻找影响就业的主要因素，即探寻以上指标和我国就业人数或城镇登记失业率之间的关系。首先来了解一下计量经济学的相关理论及模型。

5.1 滞后变量模型

时间滞后效应，即动态性，广泛存在于经济活动中。某些经济变量不仅受到同期各种因素的影响，而且也受到过去某些时期的各种因素甚至自身的过去值的影响。通常把这种过去时期的具有滞后作用的变量叫做滞后变量，含有滞后变量的模型称为滞后变量模型。

滞后变量模型考虑了时间因素的作用，使静态分析的问题有可能成为动态分析。含有滞后被解释变量的模型，又称动态模型。

5.1.1 滞后效应

在很多情况下,被解释变量与解释变量的因果关系不一定就在同时发生,可能存在的时间的滞后,或者说解释变量的变化可能需要经过一段时间才能完全对被解释变量产生影响。同样地,被解释变量当前的变化也可能受其自身过去水平的影响,这种被解释变量受自身或另一解释变量的前几期值影响的现象称为滞后效应,表示前几期值的变量称为滞后变量。例如,在研究消费函数时,通常认为,本期的消费除了受本期的收入水平影响之外,还受前一期收入以及前一期消费水平的影响:

$$Q_n = \lambda_0 + \lambda_1 P_n + \lambda_2 P_{n-1} + \lambda_3 Q_{n-1} + \mu_n \quad (5-1)$$

这就是含有滞后变量的模型,其中 P_{n-1} , Q_{n-1} 为滞后变量。

现实经济生活中,产生滞后变量的原因众多,主要有心理原因,技术原因,制度原因等很多方面,在此不作过多解释。

5.1.2 滞后变量模型

以滞后变量作为解释变量,就得到滞后变量模型。它的一般形式为

$$Y_n = \alpha_0 + \alpha_1 Y_{n-1} + \alpha_2 Y_{n-2} + \cdots + \alpha_m Y_{n-m} + \beta_0 X_n + \beta_1 X_{n-1} + \cdots + \beta_i X_{n-i} + \mu_n \quad (5-2)$$

其中, m , i 为滞后时间间隔, Y_{n-m} 为被解释变量 Y 的第 m 期滞后, X_{n-i} 为解释变量的第 i 期滞后。由于模型既含有 Y 对自身滞后变量的回归,还包括着解释变量 X 分布在不同时期的滞后变量,因此一般称为自回归分布滞后模型。若滞后期长度有限,称模型为有限自回归滞后模型;若滞后期无限,称模型为无限自回归分布滞后模型。

1. 分布滞后模型

如果滞后变量模型中没有滞后被解释变量,仅有解释变量的当前值及其若干期的滞后值,称为分布滞后模型。分布滞后模型的一般形式为:

$$Y_n = \alpha + \sum_{i=0}^m \alpha_i X_{n-i} + \mu_n \quad (5-3)$$

分布滞后模型各系数体现了解释变量的当前值和各期滞后值对被解释变量的不同影响程度,因此也称为乘数。 α_0 称为短期或即期乘数,表示本期 X 变化一个单位对 Y 平均值的影响程度; α_i ($i=1,2,\cdots,m$) 称为动态乘数或延迟系数,表示各滞后期 X 的变化对平均值影响的大小; $\sum_{i=0}^m \alpha_i$ 则称为长期或均衡乘数,表示 X 变动一个单位,由于滞后效应而形成的对 Y 平均值总影响的大小。

由(5-3)式知,如果 X 各期的值保持不变,则 X 与 Y 间的长期或均衡关系为:

$$E(Y) = \alpha + \left(\sum_{i=0}^m \alpha_i \right) X \quad (5-4)$$

2. 自回归模型

如果滞后变量模型中的解释变量仅包含 X 的当前值与被解释变量 Y 的一个或多个滞后值,则称为自回归模型。自回归模型的一般形式为

$$Y_n = \alpha_0 + \alpha_1 X_n + \sum_{i=1}^m \beta_i Y_{n-i} + \mu_n \quad (5-5)$$

其中，滞后期长度 m 也称为自回归模型的阶数。而

$$Y_n = \alpha_0 + \alpha_1 X_n + \alpha_2 Y_{n-1} + \mu_n \quad (5-6)$$

称为一阶自回归模型。

5.2 格兰杰因果关系检验

计量经济模型的建立过程，本质上是用回归分析工具处理一个经济变量对其他经济变量的依存性问题，但这并不是暗示这个经济变量与其他经济变量之间必然存在着因果关系。关于因果性的研究，最早是由美国著名计量经济学家 Granger(格兰杰)在 1969 年提出的。他提出了一个简单的检验程序，即著名的格兰杰因果关系检验，其具有很强的操作性，目前已被国际上广泛应用于分析经济变量之间的因果关系。

5.2.1 因果关系分类

因果关系其实就是指变量之间的依赖性，被解释变量是由解释变量所决定的，解释变量的变化引起被解释变量的变化。在回归方程中，解释变量是被解释变量的因，但是，这一因果关系实际上是先验设定的，或者是在回归之前就已确定的。比如，之所以在回归方程中以降雨量为解释变量，以农作物产量为被解释变量而不是相反，并不是出于统计上的原因，而是普遍常识提示我们不能把关系颠倒过来，因为用农作物产量来控制降雨量是不可能的。

但是，在许多情况下，变量之间的因果关系并不总像农作物产量和降雨量之间那么一目了然，或是因为没有充分的知识使我们认清变量之间的因果关系，而很多时候，弄清变量之间的因果关系往往是我们所关心的。即使某一经济理论宣称了一种因果关系，也许仍然需要给以经验上的支持。格兰杰从预测的角度给出了因果关系的一种定义，并将这种定义下的因果关系称为格兰杰因果关系。

格兰杰因果性检验假定有关 Y 和 X 每一变量的预测的信息全部包含在这些变量的时间序列之中，检验要求估计以下回归

$$Y_n = \sum_{i=1}^m \alpha_i X_{n-i} + \sum_{j=1}^m \beta_j Y_{n-j} + \mu_{1n} \quad (5-7)$$

$$X_n = \sum_{i=1}^k a_i X_{n-i} + \sum_{j=1}^k b_j Y_{n-j} + \mu_{2n} \quad (5-8)$$

其中，随机干扰项 μ_{1n} 和 μ_{2n} 假定为不相关。

式(5-7)假定当前 Y 与 X 自身以及 X 的过去值有关，而式(5-8)对 X 也假定了类似的行为。

对式(5-7)而言，其零假设 $H_0: \alpha_1 = \alpha_2 = \cdots = \alpha_m = 0$

对式(5-8)而言，其零假设 $H_0: b_1 = b_2 = \cdots = b_k = 0$

现分四种情形讨论：

(I) X 是引起 Y 变化的原因，即存在由 X 到 Y 的单项因果性。若式(5-7)中滞后的

X 的系数估计值在统计上整体的显著不为零，同时式(5-8)中滞后的 Y 的系数估计值在统计上整体的显著为零，则称 X 是引起 Y 变化的原因。

(2) Y 是引起 X 变化的原因，即存在由 Y 到 X 的单项因果性。若式(5-8)中滞后的 Y 的系数估计值在统计上整体的显著不为零，同时式(5-7)中滞后的 X 的系数估计值在统计上整体的显著为零，则称 Y 是引起 X 变化的原因。

(3) X 和 Y 互为因果关系，即存在：到 Y 的单向因果性，同时也存在 Y 到 X 的单向因果性。若式(5-7)中滞后的 X 的系数估计值在统计上整体的显著不为零，同时，式(5-8)中滞后的 Y 的系数估计值在统计上整体的显著不为零，则称 X 和 Y 间存在双向因果关系。

(4) X 和 Y 是独立的，或 X 与 Y 间不存在因果性。若式(5-7)中滞后的 X 的系数估计值在统计上整体的显著为零，同时，式(5-8)中滞后的 Y 的系数估计值在统计上整体的显著为零，则称 X 与 Y 间不存在因果关系。

一般地，如果变量 X 是变量 Y 的格兰杰原因，则 X 的变化应先于 Y 的变化。因此，在做 Y 对其他变量(包括自身的过去值)的回归时，如果把 X 的滞后值包括进来能够显著的改进对 Y 的预测，则可以说 X 是 Y 的格兰杰原因。类似的可定义 Y 是 X 的格兰杰原因。

5.2.2 格兰杰因果关系检验

为了检验 X 是引起 Y 的原因，格兰杰因果关系检验步骤如下：

(1) 将当前的 Y 对所有的滞后项 Y 以及别的什么变量(如果有的话)做回归，即 Y 对 Y 的滞后项 $Y_{n-1}, Y_{n-2}, \dots, Y_{n-m}$ 。及其他变量的回归，但在这一回归中没有把 X 滞后项包括进来，这是一个受约束的回归。然后从此回归得到受约束的残差平方和 RSS_R 。

(2) 做含有 X 滞后项的回归，即在前面的回归之中加进 X 滞后项，这是一个无约束的回归，由此回归得到无约束的残差平方和 RSS_{UR} 。

(3) 零假设是 $H_0: \alpha_1 = \alpha_2 = \dots = \alpha_m = 0$ ，即 X 滞后项不显著。

(4) 为了检验此假设，用 F 检验，即：

$$F = \frac{(RSS_R - RSS_{UR})/m}{RSS_{UR}/(n-k)} \quad (5-9)$$

它遵循自由度为 m 和 $(n-k)$ 的 F 分布。在这里， n 是样本容量， m 等于 X 的滞后项的个数，即有约束回归方程中待估参数的个数， k 是无约束回归中待估参数的个数。

(5) 如果在选定的显著性水平 α 上计算的 F 值超过临界值 F_α ，则拒绝零假设，这样 X 的滞后项就属于此回归，表明 X 是 Y 的原因。

(6) 同样，为了检验 Y 是否是 X 的原因，可将变量 Y 与 X 相互替换，重复上述步骤(1)~(5)。

需要指出的是，格兰杰因果关系检验对于滞后期长度的选择有时很敏感。不同的滞后期可能会得到完全不同的检验结果。因此，一般而言，常进行不同滞后期长度的检验，以检验模型中随机干扰项不存在序列相关的滞后期长度来选取滞后期。

由于假设检验的零假设是不存在因果关系，在该假设下统计量服从 F 分布，因此严格的说，该检验应称为格兰杰非因果关系检验。^[4]

5.3 失业率影响因素的确定

依据上述经济学的相关理论及模型,我们给出本文关于各影响指标与城镇登记失业率之间因果关系检验的具体步骤:

(I) 对表 5-1 中各影响指标及城镇登记失业率进行 ADF 单位根检验,也即进行序列的平稳性检验。利用 Eviews 软件进行 ADF 单位根检验,得出结果如表 5-2 所示。

表 5-2 居民消费水平(JMXF) 指标 ADF 单位根检验结果

Null Hypothesis: D(JMXF)

Exogenous: Constant

Lag Length: 0 (Automatic based on SIC, MAXLAG=1)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	1.502213	0.9964
Test critical values:		
1% level	-6.582648	
5% level	-4.320969	
10% level	-3.801384	

*MacKinnon (1996) one-sided p-values.

Warning: Probabilities and critical values calculated for 20

observations and may not be accurate for a sample size of 8

Augmented Dickey-Fuller Test Equation

Dependent Variable: D(JMXF,2)

Method: Least Squares

Date: 09/20/09 Time: 11:48

Sample (adjusted): 2000 2007

Included observations: 8 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(JMXF(-1))	0.314229	0.209178	1.502213	0.1837
C	-22.51111	85.33421	-0.263799	0.8008
R-squared	0.273313	Mean dependent var		94.50000
Adjusted R-squared	0.152198	S.D. dependent var		107.0607
S.E. of regression	98.57742	Akaike info criterion		12.23188
Sum squared resid	58305.04	Schwarz criterion		12.25174
Log likelihood	-46.92752	F-statistic		2.256644
Durbin-Watson stat	2.347259	Prob(F-statistic)		0.183726

已知当显著性水平为 10%时, t 统计量的临界值为-3.15^[5]。由表 5-2 可知, 在显

著性水平为 10%时，居民消费水平采用 ADF 检验所得 t 统计量值小于-3.15，较为平稳，可以进行序列的因果关系及相关性检验。

经验证，在显著性水平为 10%时，各因素采用 ADF 检验所得的 t 统计量值都小于临界值-3.15，如表 5-3 所示。所以本文所选取的数据较为平稳，可以进行序列的因果关系及相关性检验。

表 5-3 全部因素 ADF 单位根检验结果

变量	居民消费水平 j	全社会固定资产投资 i	职工平均工资 z	货物进出口总额 e
差分次数	1	1	1	1
滞后变量次数	0	0	0	0
ADF Test Statistics	1.502213	4.957826	1.884538	3.971818
1% level	-6.582648	-4.886101	-5.582648	-5.420595
5% level	-4.32096	-3.995865	-4.320969	-4.259808
10% level	-3.801384	-3.599088	-3.801384	-3.771129
变量	国内生产总值 g	城镇人口 c	人民币汇率 r	当年高校毕业生数 b
差分次数	1	1	1	0
滞后变量次数	0	1	2	1
ADF Test Statistics	0.534383	1.824597	-0.865187	-0.167154
1% level	-5.582648	-5.803492	-7.119808	-6.582648
5% level	-4.320969	-4.403313	-5.519595	-5.320969
10% level	-3.801384	-3.841819	-3.898418	-4.801384

(2) 针对各影响指标，为了检验其是否是影响城镇登记失业率的原因，分别建立“无限制条件回归模型”

$$X_n = \sum_{i=1}^m \alpha_i X_{n-i} + \sum_{i=1}^k \beta_i Y_{n-i} + \mu_{1n} \quad (5-10)$$

同时对各模型进行格兰杰因果关系检验，得出结果如表 5-4 至表 5-11 所示。

表 5-4 当年高校毕业生数与失业率的格兰杰因果关系检验结果

Pairwise Granger Causality Tests			
Date: 09/20/09 Time: 14:46			
Sample: 1998 2007			
Lags: 2			
Null Hypothesis:	Obs	F-Statistic	Probability
BYS does not Granger Cause SYL	8	7.09971	0.07285
SYL does not Granger Cause BYS		0.00485	0.99517

表 5-4 所得结果表明拒绝原假设，得出当年高校毕业生数是引起城镇登记失业率变化的原因，即存在由当年高校毕业生数到城镇登记失业率的单项因果性。

表 5-5 居民消费水平与失业率的格兰杰因果关系检验结果

Pairwise Granger Causality Tests

Date: 09/20/09 Time: 15:04

Sample: 1998 2007

Lags: 2

Null Hypothesis:	Obs	F-Statistic	Probability
JMXF does not Granger Cause SYL	8	1.35621	0.38058
SYL does not Granger Cause JMXF		1.74453	0.31435

表 5-5 所得结果表明拒绝原假设，而同时也拒绝了城镇登记失业率不是居民消费水平变动原因的假设。得出城镇登记失业率和居民消费水平互为因果关系，即存在居民消费水平到城镇登记失业率的单向因果性，同时也存在城镇登记失业率达到居民消费水平的单向因果性。

表 5-6 全社会固定资产投资（I）与失业率的格兰杰因果关系检验结果

Pairwise Granger Causality Tests

Date: 09/20/09 Time: 14:51

Sample: 1998 2007

Lags: 2

Null Hypothesis:	Obs	F-Statistic	Probability
I does not Granger Cause SYL	8	7.42689	0.06888
SYL does not Granger Cause I		0.00491	0.99511

表 5-6 所得结果表明拒绝原假设，得出全社会固定资产投资是引起城镇登记失业率变化的原因，即存在由全社会固定资产投资到城镇登记失业率的单项因果性。

表 5-7 城镇人口（CZRK）与失业率的格兰杰因果关系检验结果

Pairwise Granger Causality Tests

Date: 09/20/09 Time: 14:48

Sample: 1998 2007

Lags: 2

Null Hypothesis:	Obs	F-Statistic	Probability
CZRK does not Granger Cause SYL	8	1.11376	0.43475
SYL does not Granger Cause CZRK		0.40363	0.69946

表 5-7 所得结果表明拒绝原假设，得出城镇人口是引起城镇登记失业率变化的原因，即存在由城镇人口到城镇登记失业率的单项因果性。但同时也可以看出其因果关系并不是很显著。

表 5-8 货物进出口总额（EI）与失业率的格兰杰因果关系检验结果

Pairwise Granger Causality Tests

Date: 09/20/09 Time: 14:55

Sample: 1998 2007

Lags: 2

Null Hypothesis:	Obs	F-Statistic	Probability
EI does not Granger Cause SYL	8	7.33467	0.06996
SYL does not Granger Cause EI		0.69395	0.56532

表 5-8 所得结果表明拒绝原假设，得出货物进出口总额是引起城镇登记失业率变化的原因，即存在由货物进出口总额到城镇登记失业率的单项因果性。

表 5-9 国内生产总值与失业率的格兰杰因果关系检验结果

Pairwise Granger Causality Tests

Date: 09/20/09 Time: 14:55

Sample: 1998 2007

Lags: 2

Null Hypothesis:	Obs	F-Statistic	Probability
GDP does not Granger Cause SYL	8	0.94357	0.48095
SYL does not Granger Cause GDP		0.02223	0.97817

表 5-9 所得结果表明拒绝或接受原假设都可以，但同时接受城镇登记失业率不是国内生产总值变化的原因的结论。

表 5-10 人民币汇率与失业率的格兰杰因果关系检验结果

Pairwise Granger Causality Tests

Date: 09/20/09 Time: 14:57

Sample: 1998 2007

Lags: 2

Null Hypothesis:	Obs	F-Statistic	Probability
RMBHL does not Granger Cause SYL	8	4.24736	0.13333
SYL does not Granger Cause RMBHL		0.03196	0.96887

表 5-10 所得结果表明拒绝原假设，得出人民币汇率是引起城镇登记失业率变化的原因，即存在由人民币汇率到城镇登记失业率的单项因果性。

表 5-11 职工平均工资与失业率的格兰杰因果关系检验结果

Pairwise Granger Causality Tests

Date: 09/20/09 Time: 14:52

Sample: 1998 2007

Lags: 2

Null Hypothesis:	Obs	F-Statistic	Probability
SYL does not Granger Cause ZGGZ	8	3.19071	0.18083

表 5-11 所得结果表明拒绝原假设，得出职工平均工资是引起城镇登记失业率变化的原因，即存在由职工平均工资到城镇登记失业率的单项因果性。

通过以上分析发现，虽然有些因素如城镇人口和货物进出口总额等对失业率的影响并不是很显著，并且居民消费水平和城镇登记失业率互为因果关系，但这些并不能说明这些因素不对城镇登记失业率造成影响。本文所选取的居民消费水平、全社会固定资产投资、职工平均工资、货物进出口总额、国内生产总值、城镇人口、人民币汇率等因素都是城镇登记失业率变化的原因。因此，本文将这些因素全部保留作为影响城镇登记失业率的重要指标，以便后期能够作出较为精确的失业率影响模型。

六、失业率影响模型的探讨

由第五部分分析可知，影响城镇登记失业率的主要因素有居民消费水平、全社会固定资产投资、职工平均工资、货物进出口总额、国内生产总值、城镇人口、人民币汇率等。结合表 5-1 中的数据，我们建立以下城镇登记失业率与上述影响因素之间关联的数学模型。

6.1 基于主成分分析法的失业率影响模型

首先，由以上分析可知，城镇登记失业率的影响因素较多而且在问题研究中，为了更全面、系统地分析问题，我们必须考虑这些众多的影响因素。因为每个因素都在不同程度上反映了所研究问题的某些信息，但是这些指标之间彼此有一定的相关性，因而所得的统计数据反映的信息在一定程度上有重叠。在用统计方法研究多变量问题时，变量太多会增加计算量和增加分析问题的复杂性，我们希望在定量分析的过程中，涉及的变量较少，得到的信息量较多。因此首先本文应用主成分分析法来建立初步的模型。主成分分析法可以用较少的指标来代替原来较多的指标，并使这些较少的指标尽可能地反映原来指标的信息，从根本上解决了指标间的信息重叠问题，又大大简化了原指标体系的指标结构。

用主成分分析法分析影响城镇登记失业率因素的优势主要体现在：

(1) 全面性。能够消除评价指标的相互影响，在条件满足的情况下，不限制指标的个数，可以综合评价我国的经济发展状况，主成分分析的降维处理技术能较好地解决多指标评价的要求，在选择了 m 个主成分后，仍能保留原是数据信息的 85% 以上，因此这一方法综合评价经济发展水平比较全面，可以克服片面追求个别经济指标而忽略全面经济发展指标的倾向；

(2) 可加性。在综合评价经济发展水平对城镇登记失业率影响时，所建立的评价指标量纲往往不同，变差不能直接综合，主成分分析法避免了此现象的发生，因为在计算过程中，主成分分析法把各个指标进行了标准化处理，这就使得各个经济指标之间具有可比性即可加性；

(3) 客观性。在层次分析法计算过程中，通过专家打分来确定权重，也就是说在确定权重的问题上具有了人为因素，而主成分分析法在确定综合因子的权重时，克服了某些评价方法中人为确定权重的缺陷，使得综合评价结果唯一；

(4) 可行性。随着电子计算机技术的发展，SPSS、SAS 等计算机软件的推进与使用，使得主成分分析法在综合评价实践中的广泛应用成为现实。在本文对城镇登记失业率的影响指标进行分析的过程中，不同的影响因素之间具有一定的相关性，而这样势必增加分析问题的复杂性，因此需要有一种进行简化的方法。

6.1.1 基本原理

主成分分析法是一种降维的统计方法，它借助于一个正交变换，将其分量相关的原随机向量转化成其分量不相关的新随机向量，这在代数上表现为将原随机向量的协方差阵变换成对角形阵，在几何上表现为将原坐标系变换成新的正交坐标系，使之指向样本点散布最开的 p 个正交方向，然后对多维变量系统进行降维处理，使之能以一个较高的精度转换成低维变量系统，再通过构造适当的价值函数，进一步把低维系统转化成一维系统。^[6]

6.1.2 计算步骤

1. 原始指标数据的标准化

由于统计数据各指标数值之间相差悬殊，为了避免计算中数据溢出，我们在预测前对于数据进行标准化处理，使原始数据均转换为无量纲化指标测评值。即各指标值都处于同一个数量级别上，可以进行综合测评分析。首先将采集到的 P 维随机向量 $X = (x_1, x_2, x_3, \dots, x_p)^T$ 的 n 个样品 $X_i = (x_{i1}, x_{i2}, x_{i3}, \dots, x_{ip})^T$ ， $i = 1, 2, 3, \dots, n$ ， $n > p$ ，构造样本阵，对样本阵元进行如下标准化变换：

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}, i = 1, 2, 3, \dots, n; j = 1, 2, 3, \dots, p \quad (6-1)$$

$$\text{其中, } \bar{x}_j = \frac{\sum_{i=1}^n x_{ij}}{n}, s_j^2 = \frac{\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}{n-1}, \text{ 得标准化阵 } Z。$$

对于本题目，将表 5-1 中的数据矩阵进行标准化处理得到如下标准化阵 Z 。

$$Z = \begin{bmatrix} -1.135 & -0.930 & -1.181 & -1.032 & -1.078 & -1.540 & 0.488 & -0.930 \\ -0.989 & -0.891 & -1.030 & -0.970 & -0.983 & -1.188 & 0.484 & -0.916 \\ -0.766 & -0.810 & -0.852 & -0.781 & -0.812 & -0.832 & 0.485 & -0.839 \\ -0.581 & -0.696 & -0.592 & -0.723 & -0.624 & -0.476 & 0.478 & -0.774 \\ -0.396 & -0.529 & -0.323 & -0.537 & -0.432 & -0.122 & 0.478 & -0.546 \\ -0.158 & -0.208 & -0.042 & -0.152 & -0.154 & 0.235 & 0.478 & -0.138 \\ 0.244 & 0.188 & 0.302 & 0.353 & 0.279 & 0.550 & 0.477 & 0.251 \\ 0.664 & 0.674 & 0.708 & 0.784 & 0.698 & 0.868 & 0.093 & 0.763 \\ 1.190 & 1.238 & 1.165 & 1.269 & 1.214 & 1.114 & -0.900 & 1.298 \\ 1.927 & 1.964 & 1.847 & 1.789 & 1.890 & 1.390 & -2.561 & 1.830 \end{bmatrix}$$

2. 对标准化阵 Z 求相关系数矩阵 R

$$R = [r_{ij}]_{p \times p} = \frac{Z^T Z}{n-1} \quad (6-2)$$

其中, $r_{ij} = \frac{\sum z_{kj} \bullet z_{ki}}{n-1}$, $i, j = 1, 2, \dots, p$ 。

依据此公式得出城镇登记失业率影响因素的相关系数矩阵 R 为:

$$R = \begin{bmatrix} 1.000 & 0.995 & 0.998 & 0.994 & 0.999 & 0.951 & -0.864 & 0.992 \\ 0.995 & 1.000 & 0.989 & 0.995 & 0.997 & 0.923 & -0.884 & 0.998 \\ 0.998 & 0.989 & 1.000 & 0.993 & 0.997 & 0.968 & -0.835 & 0.989 \\ 0.994 & 0.995 & 0.993 & 1.000 & 0.997 & 0.949 & -0.834 & 0.998 \\ 0.999 & 0.997 & 0.997 & 0.997 & 1.000 & 0.951 & -0.857 & 0.996 \\ 0.951 & 0.923 & 0.968 & 0.949 & 0.951 & 1.000 & -0.680 & 0.935 \\ -0.864 & -0.884 & -0.835 & -0.834 & -0.857 & -0.680 & 1.000 & -0.852 \\ 0.992 & 0.998 & 0.989 & 0.998 & 0.996 & 0.935 & -0.852 & 1.000 \end{bmatrix}$$

3. 解样本相关矩阵 R 的特征方程 $|R - \lambda I_p| = 0$ 得 P 个特征根, 确定主成分

按 $\frac{\sum_{j=1}^m \lambda_j}{\sum_{j=1}^p \lambda_j} \geq 0.85$ 确定 m 值, 使信息的利用率达 85% 以上, 对每个 $\lambda_j (j=1, 2, \dots, m)$,

解方程组 $Rb = \lambda_j b$, 得单位特征向量 b_j^0 。

利用 MATLAB 求解得出各特征值及其贡献百分率如表 6-1 所示。

表 6-1 相关矩阵 R 的特征值及其贡献百分率

No	特征值	百分率%	累计百分率%
1	7.61663	95.20784	95.20784
2	0.34312	4.28894	99.49678
3	0.03784	0.47296	99.96974
4	0.00157	0.01957	99.98931
5	0.00067	0.00843	99.99774
6	0.00013	0.00167	99.99941
7	0.00003	0.00038	99.99979
8	0.00002	0.00021	100.00000

由表可知，当 $m=1$ 时即有贡献百分率为 $\frac{\sum_{j=1}^m \lambda_j}{\sum_{j=1}^p \lambda_j} = 95.2\%$ ，因此在本题目中取最大

特征根 λ_1 及其对应单位特征向量 b_1^0 即可。

4. 将标准化后的指标变量转换为主成分

利用公式：

$$U_j = z_j^T b_j^0, j=1,2,\dots,m \quad (6-3)$$

可将指标变量转换为主成分。其中， U_1 称为第一主成分， U_2 称为第二主成分， \dots ， U_p 称为第 p 主成分。

由以上分析，用第一主成分 $U_1 = z_1^T b_1^0$ 进行问题分析将会概括大部分影响城镇登记失业率的因素，利用 MATLAB 编程（主程序见附录一）计算得出以下第一主成分：
 $U_1 = 0.3621z_1 + 0.3614z_2 + 0.3611z_3 + 0.3608z_4 + 0.3622z_5 + 0.3424z_6 - 0.3151z_7 + 0.3606z_8$

6.1.3 模型的检验

以上建立的基于主成分分析法的城镇登记失业率影响模型，简单易行，但我们事先并不能断定随机变量 y 与变量 x_1, x_2, \dots, x_p 之间确有线性关系。在进行回归参数的估计前，我们用多元线性回归方程去拟合随机变量 y 与变量 x_1, x_2, \dots, x_p 之间的关系，只是根据一些定性分析所作的一种假设。因此，当求出线性回归方程后，还需要对回归方程进行显著性检验。多元线性回归方程的显著性检验有两方面，一是回归方程显著性的 F 检验，另一个是回归系数显著性的 t 检验。

1. F 检验

对多元线性回归方程的显著性检验就是要看自变量 x_1, x_2, \dots, x_p 从整体上对随机变量 y 是否有明显的影响。为此提出原假设：

$$H_0: \beta_1 = \beta_2 = \dots = \beta_p = 0$$

如果 H_0 被接受，则表明随机变量 y 与 x_1, x_2, \dots, x_p 之间的关系由线性回归模型表示不合适。构造 F 检验统计量：

$$F = \frac{SSR / p}{SSE / (n - p - 1)} \quad (6-4)$$

SPSS 软件会输出 F 值与显著性 P 值。当 P 值 $< \alpha$ 时，拒绝原假设 H_0 。

2. t 检验

在多元线性回归中，回归方程显著并不意味着每个自变量对 y 的影响都显著，所以就需要我们对每个自变量进行显著性检验。检验变量 x_j 是否显著，等价于检验假设

$$H_{0j}: \beta_j = 0, \quad j=1,2,\dots,p$$

如果接受原假设 H_{0j} ，则 x_j 不显著；如果拒绝原假设 H_{0j} ，则 x_j 是显著的。构造 t 统计量：

$$t_j = \frac{\hat{\beta}_j}{\sqrt{\hat{c}_{jj}} \sigma} \quad (6-5)$$

SPSS 软件会输出 β_j 的 t_j 统计量的值及相应的 P 值，当 P 值 $< \alpha$ 时，拒绝原假设 H_0 。

3. 拟合优度

多元线性回归模型除了上述两个检验外，还有拟合优度检验。拟合优度是用于检验回归方程对样本观测值的拟合程度。定义样本决定系数为：

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \quad (6-6)$$

样本决定系数 R^2 的取值在 $[0,1]$ 区间内， R^2 越接近 1，表明回归拟合的效果越好。

以上建立的基于主成分分析法的失业率影响模型经过多元线性回归显著性检验后，发现其线性拟合度较好。因此，此模型可作为影响城镇登记失业率的初级模型来应用。

6.2 基于序列相关性的误差修正模型

6.2.1 序列相关性

多元线性回归模型的基本假设之一是模型的随机干扰项相互独立或不相关。如果模型的随机干扰项违背了相互独立的基本假定，称为存在序列相关性。

对于模型

$$Y_i = \alpha_0 + \alpha_1 X_{1i} + \alpha_2 X_{2i} + \cdots + \alpha_k X_{ki} + \mu_i \quad i = 1, 2, \cdots, n \quad (6-7)$$

在其他假设仍然成立的条件下，随机干扰项序列相关即意味着

$$\text{Cov}(\mu_i, \mu_j) = E(\mu_i \mu_j) \neq 0 \quad (6-8)$$

如果仅存在

$$E(\mu_i \mu_{i+1}) \neq 0 \quad i = 1, 2, \cdots, n-1 \quad (6-9)$$

则称为一阶序列相关或自相关，这是最常见的一种序列相关问题。自相关往往可写成如下形式：

$$\mu_i = \rho \mu_{i-1} + \varepsilon_i \quad -1 < \rho < 1 \quad (6-10)$$

其中 ρ 称为自协方差系数或一阶自相关系数， ε_i 是满足以下标准 OLS 假定的随机干扰项：

$$E(\varepsilon_i) = 0 \quad \text{Var}(\varepsilon_i) = \sigma^2 \quad \text{Cov}(\varepsilon_i, \varepsilon_{i-s}) = 0 \quad (s \neq 0) \quad (6-11)$$

序列相关性的检验方法有很多种，如冯诺曼比检验法、回归检验法、图示检验法等。这些检验方法的共同思路是：首先采用普通最小二乘法估计模型，以求得随机干

扰项的“近似估计量”，用 \tilde{e}_i 表示：

$$\tilde{e}_i = Y_i - (\hat{Y}_i)_{OLS} \quad (6-12)$$

然后通过分析这些“近似估计量”之间的相关性以达到判断随机干扰项是否具有序列相关性的目的。下面介绍几种常用的检验方法：

1. 回归检验法

以 μ_i 为被解释变量，以各种可能的相关量，诸如 \tilde{e}_{i-1} ， \tilde{e}_{i-2} ， \tilde{e}_i^2 等为解释变量，建立各种方程：

$$\tilde{e}_i = \rho_1 \tilde{e}_{i-1} + \varepsilon_i \quad i = 2, \dots, n \quad (6-13)$$

$$\tilde{e}_i = \rho_1 \tilde{e}_{i-1} + \rho_2 \tilde{e}_{i-2} + \varepsilon_i \quad i = 3, \dots, n \quad (6-14)$$

.....

对方程进行显著性检验，如果存在某一种函数形式，使得方程显著成立，则说明元模型存在序列相关性。回归检验法的优点是一旦确定了模型存在序列相关性，也就同时知道了相关的形式，而且它适用于任何类型的序列相关性问题的检验。

2. 杜宾——瓦森(Durbin——Watson)

D-W 检验是杜宾(J.Durbin)和瓦森(G.S.Watson)于 1951 年提出的一种检验序列相关性的方法，该方法的假定条件是：

(1) 解释变量 X 非随机

(2) 随机干扰项 μ_i ，为一阶自回归形式：

$$\mu_i = \rho \mu_{i-1} + \varepsilon_i \quad (6-15)$$

(3) 回归模型中不应该含有滞后应变量作为解释变量，即不应出现下列形式：

$$Y = \alpha_0 + \alpha_1 X_{1i} + \dots + \alpha_k X_{ki} + \lambda Y_{i-1} + \mu_i \quad (6-16)$$

(4) 回归模型含有截距项

杜宾和瓦森针对原假设 $H_0: \rho = 0$ ，即产 μ_i 不存在一阶自回归，构造如下统计

$$D.W. = \frac{\sum_{i=2}^n (\tilde{e}_i - \tilde{e}_{i-1})}{\sum_{i=1}^n \tilde{e}_i^2} \quad (6-17)$$

该统计量的分布和给定样本中的 X 值有复杂的关系，因此其精确的分布很难得到。但他们成功的导出了临界值的上限 d_u 和下限 d_l ，且这些上下限只与样本容量和解释变量的个数有关，而与解释变量的取值无关。因此，在检验时，只须计算该统计量的值，再根据样本容量 n 和解释变量数目 k 查 D.W.分布表，得到临界值 d_u 和 d_l ，然后按照下列准则考虑计算得到的 D.W.值，以判断模型的自相关状态：

若 $0 < D.W. < d_l$ ，则存在正自相关；

若 $d_l < D.W. < d_u$ ，则不能确定；
若 $d_u < D.W. < 4 - d_u$ ，则无自相关；
若 $4 - d_u < D.W. < 4 - d_l$ ，则不能确定；
若 $4 - d_l < D.W. < 4$ ，则存在负自相关。

也就是说，当 D.W. 值在 2 附近时，模型不存在一阶自相关。

D-W 检验因其原理简单、检验方便，目前已成为最常用的自相关性检验方法。EViews 软件在回归分析的输出结果中也给出了 D.W 统计值。在报告回归分析的计算结果时，人们也习惯于将 D.W. 值作为常规的检验统计量，连同 R^2 ， t 值等一起标明。但是，使用 D-W 检验时应注意以下几个问题：

(1) D-W 检验只能判断是否存在一阶线性自相关，对于高阶自相关或非自相关皆不适用。当 D.W. 值接近于 2 时，只能说明 μ_i 与 μ_{i-1} 不相关，并不同时意味着模型不存在高阶自相关性，即不能得出“不存在自相关性”的结论。非一阶自相关形式的时间序列不能用 D-W 检验。

(2) D-W 检验有两个无法判定的区域。虽然杜宾和瓦森对这两个区域的检验又提出了修正的方法，但因计算公式复杂，当 D.W. 统计量落入这两个区域时，人们还是不愿改用其他的检验方法来判断自相关性。一般是调整样本容量的大小(增大样本容量)，或者改变模型的函数形式，采用其他的检验方法。

(3) 这一方法不适用于对联立方程组模型中单一方程随机误差项序列相关的检验。

(4) D-W 检验不适用于模型中含滞后的被解释变量的情况。

3. 偏相关系数检验法

在多个变量 Y, X_1, X_2, \dots, X_m 之间，如果只考虑 Y 与 X_i ($i=1, 2, \dots, m$) 之间的相关关系，其他变量固定不变，这种相关性称为偏相关。用来衡量偏相关程度的数量指标叫做偏相关系数。偏相关系数是衡量多个变量之间的相关程度的重要指标，可以用它来判断自相关性的类型。本文中我们利用 EViews 软件计算偏相关系数并对模型进行检验。

6.2.2 误差修正模型的建立

在处理以上时间序列数据时，本文所拥有的时间序列是非平稳的。对于非平稳的数据，如果采用传统的估计方法，很可能会造成的错误推断，即伪回归。因而在模型建立之前需要对非平稳序列进行一阶差分或二阶差分使之变为平稳序列。对于一组非平稳但具有同阶的序列而言，若它们的线性组合为平稳序列，则称该组序列具有协整关系。对具有协整关系的序列，算出误差修正项，并将误差修正项的滞后一期看做一个解释变量，连同其他反映波动关系的变量一起。

建立误差修正模型 (ECM: Error Correction Model) 的一般步骤如下：首先，对单个序列进行单根检验，进行单根检验有两种：ADF (Augment Dickey-Fuller) 和 DF (Dickey-Fuller) 检验法。在此选择 ADF 单根检验。若序列都是同阶单稳，我们就可以对其进行协整分析，在此只介绍单个向量的检验方法。我们可以先求出误差项，再建立误差修正模型，也可以先求出向量误差修正模型，然后算出误差修正项。补充一点的是，误差修正模型反映的是变量短期的相互关系，而误差修正项反映出变量长期的关系。

以下是 ECM 模型的建立过程：

对模型

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 y_{t-1} + \beta_3 x_{t-1} + \varepsilon_t \quad (6-18)$$

移项后，整理可得

$$\Delta y_t = \beta_0 + \beta_1 \Delta x_t + (\beta_2 - 1)(y - \frac{\beta_1 + \beta_3}{1 - \beta_2} x)_{t-1} + \varepsilon_t \quad (6-19)$$

上方程即为ECM，其中 $y - \frac{\beta_1 + \beta_3}{1 - \beta_2} x$ 是误差修正项，记为 ecm 。ECM 模型解释了因变量 y_t 的短期波动 Δy_t 是如何被决定的。一方面，它受到自变量短期波动 Δx_t 的影响，另一方面，取决于 ecm 。如果变量 x_t 和 y_t 间存在着长期的均衡关系，即有下面我们就问题二建立的 ECM 模型。

首先根据近十年的居民消费水平数据，依据公式 6-19，建立以下居民消费水平与登记失业率 ECM 模型

$$\Delta s_t = \beta_0 + \beta_1 \Delta j_t + (\beta_2 - 1)(s - \frac{\beta_1 + \beta_3}{1 - \beta_2} j)_{t-1} + \varepsilon_t \quad (6-20)$$

结果如表 6-2 所示：

表 6-2 居民消费水平与登记失业率协整结果

Dependent Variable: s

Method: Least Squares

Date: 09/20/09 Time: 16:28

Sample (adjusted): 1999 2007

Included observations: 9 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.569440	0.589551	-0.965888	0.3785
s(-1)	0.753131	0.178064	4.229544	0.0083
j	-0.002635	0.000858	-3.070460	0.0278
j(-1)	0.003267	0.001094	2.985646	0.0306
R-squared	0.934042	Mean dependent var	3.844444	
Adjusted R-squared	0.894466	S.D. dependent var	0.466667	
S.E. of regression	0.151601	Akaike info criterion	-0.634027	
Sum squared resid	0.114914	Schwarz criterion	-0.546372	
Log likelihood	6.853121	F-statistic	23.60175	
Durbin-Watson stat	1.567952	Prob(F-statistic)	0.002222	

由上表结果可得：

$$s_t = -0.56944 - 0.002635 j_t + 0.753131 s_{t-1} + 0.003267 j_{t-1} \quad (6-21)$$

同理，可得其他因素与登记失业率 ECM 模型如下（具体协整结果见附录二）：
全社会固定资产投资与登记失业率 ECM 模型

$$s_t = 2.316769 + 8.98 \times 10^{-5} c_t + 0.486091 s_{t-1} - 0.000115 c_{t-1} \quad (6-22)$$

货物进出口总额与登记失业率 ECM 模型

$$s_t = -1.396947 - 4.52 \times 10^{-5} e_t + 1.593163 s_{t-1} - 4.49 \times 10^{-5} e_{t-1} \quad (6-23)$$

职工平均工资与登记失业率 ECM 模型

$$s_t = 0.168384 + 0.000391 z_t + 1.199909 s_{t-1} - 0.000511 z_{t-1} \quad (6-24)$$

国内生产总值 GDP 与登记失业率 ECM 模型

$$s_t = 0.706865 + 6.33 \times 10^{-6} g_t + 0.944589 s_{t-1} - 1.03 \times 10^{-5} g_{t-1} \quad (6-25)$$

城镇人口与登记失业率 ECM 模型

$$s_t = -3.576434 + 0.001145 c_t + 0.601523 s_{t-1} - 0.001087 c_{t-1} \quad (6-26)$$

人民币汇率与登记失业率 ECM 模型

$$s_t = 8.142618 + 0.013844 r_t + 0.905151 s_{t-1} - 0.023054 r_{t-1} \quad (6-27)$$

当年高校毕业生数与登记失业率 ECM 模型

$$s_t = 2.263509 + 0.016226 b_t + 0.375127 s_{t-1} - 0.018915 b_{t-1} \quad (6-28)$$

6.2.3 模型的序列相关性检验

误差修正模型建立后，还需根据序列相关性的知识对模型进行偏相关系数检验以确定模型是否具有可用性。其中，居民消费水平与登记失业率 ECM 模型残差序列的偏相关系数检验结果如图 6-1 所示：

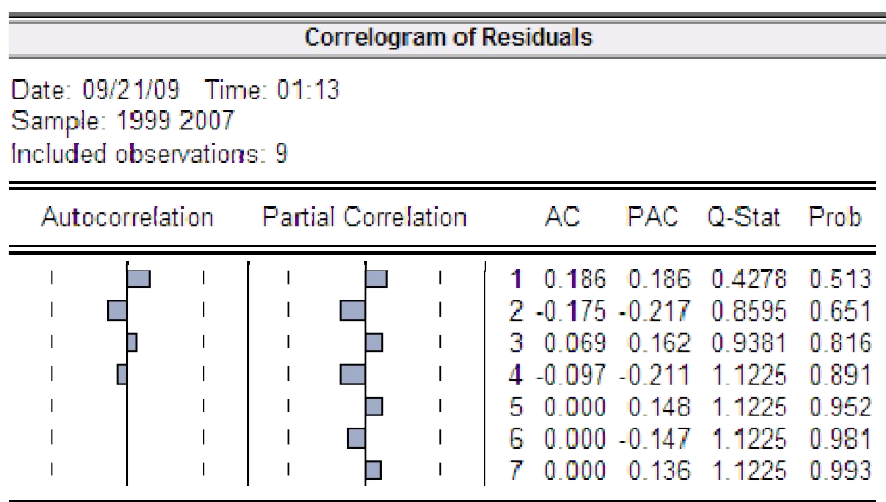


图 6-1 偏相关系数检验图

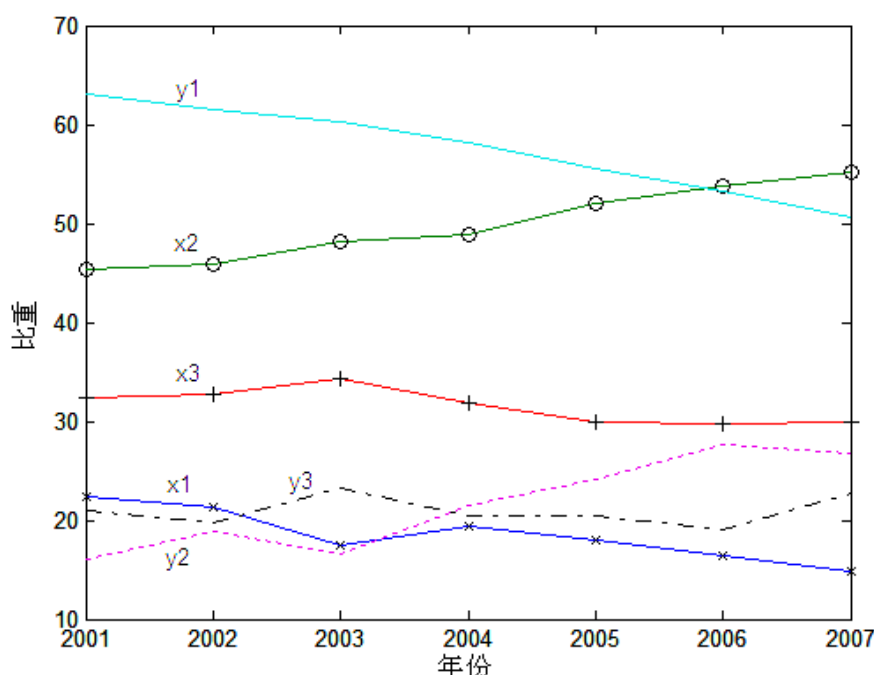
图示结果表明居民消费水平与登记失业率ECM模型的残差序列可被接受，模型通过检验。

其它各因素与登记失业率ECM模型检验同理，且均通过检验。

七、河南省产业结构和就业结构的向量误差修正模型

由于我国地区发展的不平衡性，以及对于不同产业，不同就业人群的情况各不相同，比如沿海地区和西部欠发达地区，在各种因素对失业率的影响上显然风格各异，各有侧重。因此，有必要针对特殊的需要，建立更为合理精确的数学模型。

以下以河南省为例，就当地的产业结构和就业结构的关系问题建立数学模型（具体数据见附录三），进而揭示两者之间在一定期间的协整关系。



注：x₁ 表第一产业占 GDP 比重；x₂ 表第二产业占 GDP 比重；
x₃ 表第三产业占 GDP 比重；y₁ 表第一产业劳动比重；
y₂ 表第二产业劳动比重；y₃ 表第三产业劳动比重

图 7-1 产业比重和就业比重拆线图

从图 7-1 我们还可以看出，第一、二、三产业的就业比重与产业比重基本具有相同的变动趋势，为了验证我们以上得到的直观判断，以下我们采用 kendall 检验对上面的判断进行验证，得到结果如表 7-1 所示：

表 7-1 三次产业 GDP 比重与相应从业人员比重的 Kendall τ 检验（2001-2007 年）

	第一产业	第二产业	第三产业
Kendall's	0.952	0.907	0.978
Prob> z	0.003	0.007	0.002

从表 7-1 可以得出与我们的直观判断是相符合的结果。为进一步分析河南产业结构与就业结构的关系，我们引入各产业的劳动生产率和结构偏离度。计算公式如下：

$$L_i = \frac{Z_i}{C_i} \quad (7-1)$$

其中, L_i : 第*i*产业劳动生产率; Z_i : 第*i*产业实际增加值(按 2001=100, 剔除价格因素); C_i : 第*i*产业从业人员数, $i=1,2,3$ 。

$$P_i = \frac{\hat{Z}_i}{\hat{C}_i} - 1 \quad (7-2)$$

其中, P_i : 第*i*产业结构偏离度; \hat{Z}_i : 第*i*产业计算期增加值占 GDP 的比重; \hat{C}_i : 同期该产业从业人员比重; $i=1,2,3$ 。

表 7-2 各主要年份结构偏离度和劳动生产率

年份	结构偏离度			劳动生产率 (元/人)		
	第一产业	第二产业	第三产业	第一产业	第二产业	第三产业
2001	-0.5018	2.224006	0.315209	1345.491	8856.092	5047.277
2002	-0.40714	1.668464	0.246702	1460.097	9793.93	5284.337
2003	-0.40187	1.253788	0.297595	1827.116	11326.506	5959.975
2004	-0.43365	0.759632	0.626054	1951.119	13435.487	7334.572
2005	-0.48624	0.899837	0.470467	2167.137	14161.718	7912.883
2006	-0.63419	1.866473	0.380519	2374.916	15902.91	8544.971
2007	-0.63135	1.882012	0.101669	2596.176	17747.94	9402.937

如表 7-2 所示, 从三次产业的结构偏离度的变动趋势来看, 第一产业的结构偏离度呈现负值, 偏离度从 2001 年的-0.5018 上升到 2007 年的-0.63135, 偏离零值愈来愈远, 可见第一产业不仅不能吸收更多的劳动力就业反而成了劳动力净流出的部门, 而且随着工业化进程的加快, 第一产业劳动力向其他产业转移的趋势还会增强。第二产业的结构偏离度为正值且较大, 是三个产业中偏离度绝对值最大的, 从 2001—2007 年, 该值呈现变小的趋势, 表明第二产业吸纳了第一产业流出的剩余劳动力, 但随着工业化进程的加快, 尤其河南把石油、化工、煤炭等重工业作为第二产业的重点产业, 资本与技术逐渐替代劳动力, 一方面, 河南第二产业劳动生产率显著提高, 明显高于第一、三产业, 另一方面, 河南第二产业对劳动力的吸纳能力却出现下降的趋势。第三产业的结构偏离度为正且数值相对较小, 并有缓慢趋于零值的趋势, 可以说明第三产业的产业结构和就业结构向均衡状态靠近。从其他部门转移出的劳动力大部分为第三产业吸收, 第三产业今后仍然是吸纳劳动力的主要就业部门。

通过上面对河南产业结构和劳动结构之间结构偏离度的分析我们可以看出, 河南三次产业结构与就业结构存在很强的相关关系。但产业结构与就业结构内在的定量关系还不是很清楚, 下面借用 VEC 模型来揭示它们之间存在的深刻联系。

在介绍 VEC 模型之前, 首先来了解 VAR 模型。

(1) VAR 模型。一般的 VAR 模型为:

$$Y_i = \alpha + A_1 Y_{i-1} + A_2 Y_{i-2} + \cdots + A_k Y_{i-k} + \varepsilon_i \quad (7-3)$$

式中, Y_i 是由内生变量组成的向量, 即 $Y_i = (\text{就业结构}_i, \text{产业结构}_i)$, k 为滞后项数, A_1, A_2, \cdots, A_k 为参数矩阵, ε_i 为扰动向量。

(2) VEC 模型。Engle 和 Granger 将协整与误差修正模型结合起来, 得到了向量误差修正模型(VEC)。只要变量之间存在长期的均衡关系, 我们就称变量之间存在协整

关系。在 VAR 模型中每个方程都是一个自回归分布滞后模型，而 VEC 模型可以看作是含有协整约束的 VAR 模型。向量误差修正模型(VEC)表述如下：

$$\Delta Y_t = \alpha \cdot ecm_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta Y_{t-i} + \varepsilon_t, \quad t = 1, 2, \dots, T \tag{7-4}$$

其中， $ecm_{t-1} = \beta' Y_{t-1}$

式中 Y_t 包含 k 个存在协整关系的一阶单整(即 $I(1)$)变量，但不包含外生变量，其中每个方程的误差项都具有平稳性，误差修正项反映变量之间的长期均衡的关系。

我们首先对六个变量 x_i ：第 i 产业的产业占 GDP 的比重， y_i ：第 i 产业的就业比重进行单整检验得到：第一产业 $x_1 \sim I(1)$ ， $y_1 \sim I(1)$ ，第二产业 $x_2 \sim I(1)$ ， $y_2 \sim I(1)$ ，第三产业 $x_3 \sim I(1)$ ， $y_3 \sim I(1)$ ，即各产业就业比重与产业比重都是一阶单整的，在此基础上做协整检验，采用 JJ 检验可得结果表 7-3 至表 7-5 所示：

表 7-3 x_1 与 y_1 协整检验结果

Unrestricted Cointegration Rank Test (Trace)				
Hypothesized		Trace	0.05	
No. of CE(s)	Eigenvalue	Statistic	Critical Value	Prob.**
None	0.539613	19.15448	15.49471	0.0134
At most 1	0.055515	1.31352	3.841466	0.2517
Unrestricted Cointegration Rank Test (Maximum Eigenvalue)				
Hypothesized		Max-Eigen	0.05	
No. of CE(s)	Eigenvalue	Statistic	Critical Value	Prob.**
None *	0.539613	17.84083	14.26460	0.0130
At most 1	0.055515	1.31352	3.841466	0.2517

表 7-4 x_2 与 y_2 协整检验结果

Unrestricted Cointegration Rank Test (Trace)				
Hypothesized		Trace	0.05	
No. of CE(s)	Eigenvalue	Statistic	Critical Value	Prob.**
None	0.432271	16.02498	15.49471	0.0416
At most 1	0.122452	3.004332	3.841466	0.0830
Unrestricted Cointegration Rank Test (Maximum Eigenvalue)				
Hypothesized		Max-Eigen	0.05	

No. of CE(s)	Eigenvalue	Statistic	Critical Value	Prob.**
None *	0.432271	13.02056	14.26460	0.0779
At most 1	0.122452	3.004332	3.841466	0.0830

表 7-5 x_3 与 y_3 协整检验结果

Unrestricted Cointegration Rank Test (Trace)

Hypothesized		Trace	0.05	
No. of CE(s)	Eigenvalue	Statistic	Critical Value	Prob.**
None	0.530184	17.37964	15.49471	0.0257
At most 1	0.033979	0.760524	3.841466	0.3832

Unrestricted Cointegration Rank Test (Maximum Eigenvalue)

Hypothesized		Max-Eigen	0.05	
No. of CE(s)	Eigenvalue	Statistic	Critical Value	Prob.**
None *	0.530184	16.61911	14.26460	0.0208
At most 1	0.033979	0.760524	3.841466	0.3832

从表 7-3 至表 7-5 可以看出，第一产业，该检验选择的滞后阶数为 4 时， y_1 与 x_1 存在协整关系；对于第二产业，该检验选择的滞后阶数为 4 时， y_2 与 x_2 存在协整关系；对于第三产业，该检验选择的滞后阶数为 5 时， y_3 与 x_3 存在协整关系。据此我们分别建立三次产业就业结构与产业结构的 VEC 模型。

第一产业 VEC 模型：

$$\begin{bmatrix} \Delta y_t \\ \Delta x_t \end{bmatrix} = \begin{bmatrix} -0.4 \\ -3.5 \end{bmatrix} + \begin{bmatrix} -0.1 \\ 0.36 \end{bmatrix} \cdot ecm_{t-1} + \begin{bmatrix} 0.03 & -0.08 \\ -0.36 & 0.51 \end{bmatrix} \begin{bmatrix} \Delta y_{t-1} \\ \Delta x_{t-1} \end{bmatrix} + \begin{bmatrix} 0.09 & -0.07 \\ -1.39 & -0.6 \end{bmatrix} \begin{bmatrix} \Delta y_{t-2} \\ \Delta x_{t-2} \end{bmatrix} \\ + \begin{bmatrix} 0.46 & 0.09 \\ -0.49 & 0.32 \end{bmatrix} \begin{bmatrix} \Delta y_{t-3} \\ \Delta x_{t-3} \end{bmatrix} + \begin{bmatrix} 0.19 & -0.25 \\ -1.66 & -0.15 \end{bmatrix} \begin{bmatrix} \Delta y_{t-4} \\ \Delta x_{t-4} \end{bmatrix}$$

其中， $ecm_{t-1} = y_{t-1} - 1.01x_{t-1} - 29.27$ ， $t = -5.99$ $R^2 = 0.41$

第二产业 VEC 模型：

$$\begin{bmatrix} \Delta y_t \\ \Delta x_t \end{bmatrix} = \begin{bmatrix} 0.04 \\ 0.38 \end{bmatrix} + \begin{bmatrix} -0.45 \\ -0.31 \end{bmatrix} \cdot ecm_{t-1} + \begin{bmatrix} 0.04 & 0.18 \\ 0.6 & 0.23 \end{bmatrix} \begin{bmatrix} \Delta y_{t-1} \\ \Delta x_{t-1} \end{bmatrix} + \begin{bmatrix} 0.38 & 0.16 \\ -0.23 & -0.35 \end{bmatrix} \begin{bmatrix} \Delta y_{t-2} \\ \Delta x_{t-2} \end{bmatrix} \\ + \begin{bmatrix} 0.74 & 0.1 \\ -0.76 & -0.16 \end{bmatrix} \begin{bmatrix} \Delta y_{t-3} \\ \Delta x_{t-3} \end{bmatrix} + \begin{bmatrix} 0.21 & -0.13 \\ 0.02 & 0.31 \end{bmatrix} \begin{bmatrix} \Delta y_{t-4} \\ \Delta x_{t-4} \end{bmatrix}$$

其中， $ecm_{t-1} = y_{t-1} - 0.28x_{t-1} - 25.95$ ， $t = -1.44$ $R^2 = 0.69$

第三产业 VEC 模型:

$$\begin{aligned} \begin{bmatrix} \Delta y_t \\ \Delta x_t \end{bmatrix} &= \begin{bmatrix} 1.42 \\ -0.002 \end{bmatrix} + \begin{bmatrix} -0.1 \\ -1 \end{bmatrix} \cdot ecm_{t-1} + \begin{bmatrix} 0.42 & -0.16 \\ -0.3 & -0.75 \end{bmatrix} \begin{bmatrix} \Delta y_{t-1} \\ \Delta x_{t-1} \end{bmatrix} + \begin{bmatrix} 0.035 & -0.21 \\ 0.12 & -0.82 \end{bmatrix} \begin{bmatrix} \Delta y_{t-2} \\ \Delta x_{t-2} \end{bmatrix} \\ &+ \begin{bmatrix} 0.12 & -0.21 \\ 2.03 & -0.74 \end{bmatrix} \begin{bmatrix} \Delta y_{t-3} \\ \Delta x_{t-3} \end{bmatrix} + \begin{bmatrix} 0.33 & -0.29 \\ 2.07 & -0.51 \end{bmatrix} \begin{bmatrix} \Delta y_{t-4} \\ \Delta x_{t-4} \end{bmatrix} + \begin{bmatrix} 0.41 & -0.17 \\ 1.49 & -0.37 \end{bmatrix} \begin{bmatrix} \Delta y_{t-5} \\ \Delta x_{t-5} \end{bmatrix} \end{aligned}$$

其中, $ecm_{t-1} = y_{t-1} - 0.43x_{t-1} - 9.73$, $t = -6.84$ $R^2 = 0.41$

上面的误差修正模型中, 差分项反映了短期波动的影响。短期变动可以分为两部分: 一部分是短期的内生变量的波动的影响; 一部分是偏离长期均衡所致。误差修正项大小反映了对偏离长期均衡的调整力度。从误差项可以很明显地看出, 第二产业与第一、第三产业相比, 右边第二项的符号相反, 为负。具体来说: 从第一产业看, 河南第一产业就业比重与产业比重成正比, 考虑到 2001—2007 年其产业比重一直处于下降趋势, 可以解释为, 第一产业比重每降低一个百分点, 其就业比重将降低 1.01 个百分点。从第二产业看, 河南第二产业就业比重与产业比重也成正比, 即第二产业比重每增加一个百分点, 其就业比重就提高 0.28 个百分点。从第三产业来看, 河南第三产业比重一直呈现上升趋势, 产业比重每提高一个百分点, 其就业比重相应提高 0.43 个百分点, 看来第三产业确实是吸纳劳动力的主要产业部门, 在 2001—2007 年这个期间, 第三产业承接了第一产业转移的大部分剩余劳动力。

八、就业前景的仿真

以上我们建立了关于城镇登记失业率的数学模型, 在利用此模型对 2009 年及 2010 年上半年的我国就业前景进行仿真前, 我们首先用灰色预测模型对城镇登记失业率进行预测。

8.1 GM(1, 1)灰色预测

灰色预测是指利用 GM(1, 1)模型对系统行为特征的发展变化规律进行估计预测, 同时也可以对行为特征的异常情况发生的时刻进行估计计算, 以及对在特定时区内发生事件的未来时间分布情况做出研究等等。这些工作实质上是将“随机过程”当作“灰色过程”, “随机变量”当作“灰变量”, 并主要以灰色系统理论中的 GM(1, 1)模型来进行处理。

8.1.1 GM(1, 1)模型求解

对于已知的数据参考序列 X_0 进行一次累加(AGO)为:

$$X^{(1)}(k) = \sum_{m=1}^k X_0(m), \quad (k = 1, 2, \dots, n) \quad (8-1)$$

从而得到一组生成序列为： $X^{(1)} = \{X^{(1)}(1), X^{(1)}(2), \dots, X^{(1)}(n)\}$ 。

相应得白化微分方程(影子方程)为：

$$\frac{dX^{(1)}}{dt} + aX^{(1)} = u \quad (8-2)$$

其中， a 为发展灰数， u 为内生控制灰数
引入记号

$$\hat{a} = \begin{bmatrix} a \\ u \end{bmatrix}, \quad y_1 = \begin{bmatrix} X_0(2) \\ \vdots \\ X_0(n) \end{bmatrix}, \quad B = \begin{bmatrix} -\frac{1}{2}(X_1^{(1)}(2) + X_1^{(1)}(1)) & 1 \\ \vdots & \\ -\frac{1}{2}(X_1^{(1)}(n) + X_1^{(1)}(n-1)) & 1 \end{bmatrix}$$

利用最小二乘法求解得到： $\hat{a} = (a, u)^T = [B^T \cdot B]^{-1} \cdot B^T \cdot y_1$

综合以上得微分方程的解为： $\hat{X}^{(1)}(k) = \left(X_0(1) - \frac{u}{a} \right) e^{-a(k-1)} + \frac{u}{a}$

最后将预测累加值还原为预测值： $\hat{X}^{(0)}(k) = \hat{X}^{(1)}(k) - \hat{X}^{(1)}(k-1)$

表 8-1 1998-2007 年城镇登记失业率

年份	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
失业率(%)	3.1	3.1	3.1	3.6	4.0	4.3	4.2	4.2	4.1	4.0

我们将 1998 至 2007 年十年间的城镇登记失业率序列作为数据参考序列（如表 8-1 所示），即令

$$X_0 = [3.1 \ 3.1 \ 3.1 \ 3.6 \ 4.0 \ 4.3 \ 4.2 \ 4.2 \ 4.1 \ 4.0]$$

利用 *MATLAB* 编程（主程序见附录四）求解得到：

$$a = -0.0328, \quad u = 3.2054$$

8.1.2 GM(1, 1)模型检验

1. 残差检验

按预测模型进行计算 $\hat{X}^{(1)}(k)$ ，并进行累减生成 $\hat{X}^{(0)}(k)$ ，然后计算原始序列 $X^{(0)}(k)$ 与 $\hat{X}^{(0)}(k)$ 的绝对误差序列和相对误差序列：

$$\Delta^{(0)}(i) = |X^{(0)}(k) - \hat{X}^{(0)}(k)|, \quad (i = 1, 2, \dots, n) \quad (8-3)$$

$$\Phi(i) = \frac{\Delta^{(0)}(i)}{X^{(0)}(k)} \times 100\%, \quad (i=1,2,\dots,n) \quad (8-4)$$

2. 后验差检验

原始序列标准差: $s_1 = \sqrt{\frac{\sum [X^{(0)}(k) - \bar{X}^{(0)}]^2}{n-1}}$

绝对误差标准差: $s_2 = \sqrt{\frac{\sum [\Delta^{(0)}(k) - \bar{\Delta}^{(0)}]^2}{n-1}}$

方差为绝对误差标准差与原始序列标准差之比: $C = \frac{s_1}{s_2}$

小误差概率: $P = p\{|\Delta^{(0)}(i) - \bar{\Delta}^{(0)}| < 0.6745s_1\}$, 评判标准如表 8-2 所示:

表 8-2

精度等级	好	合格	勉强合格	不合格
要	P	>0.95	>0.80	>0.70
求	C	<0.35	<0.60	<0.65

若此模型经残差检验、后验差检验都能通过, 则可以用所建模型进行预测, 否则, 需要进行残差修正。

8.1.3 GM(1, 1)预测结果及分析

利用此模型对 2009 年及 2010 年我国城镇登记失业率进行预测得出结果分别为 4.5136% 和 4.6639%。而其预测的 2008 年我国城镇登记失业率为 4.3682%, 这一结果比温家宝总理在政府工作报告中作出的 2008 年城镇登记失业率 (4.2%) 要高一些。此外, 由相对误差值 (如图 8-1) 可知, 灰色预测所得的短时间城镇登记失业率预测值的相对误差也都达到了 10% 以上, 此模型并不能通过残差检验。因此, 用简单的 GM(1, 1) 灰色预测模型来预测未来我国城镇登记失业率并不理想。

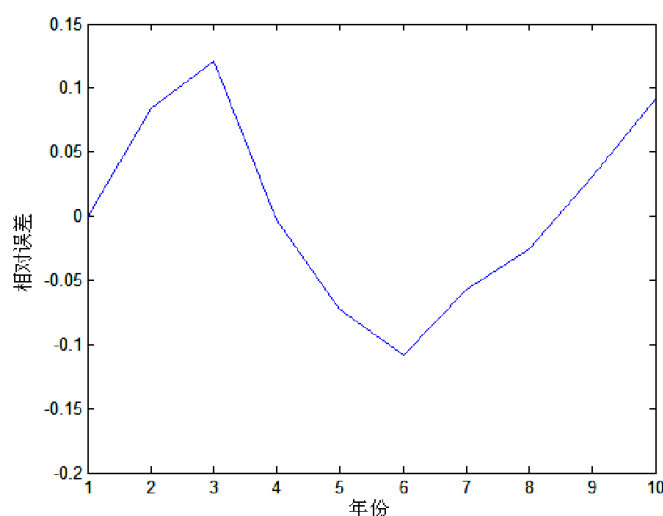


图 8-1 城镇登记失业率预测值的相对误差

8.2 ECM 模型的应用

通过以上计算可知,利用灰色预测模型对城镇登记失业率进行预测所得结果误差较大,国家的决策和规划对预测结果的影响在此模型中也未得到体现。因此,要准确对 2009 年及 2010 年上半年的我国就业前景进行仿真,还要借助我们在问题二和问题三中所建立的误差修正模型,并以此为基础对国家的决策和规划作出合理假设,预测我国未来就业前景。

在解决问题二时,我们首先应用主成分分析法建立了简单的失业率影响模型。将居民消费水平、全社会固定资产投资、职工平均工资、货物进出口总额、国内生产总值、城镇人口、人民币汇率和当年高校毕业生数等影响因素近几年数据代入模型中进行仿真,得出失业率在 2009 年和 2010 年分别为 4.33%和 4.42%左右。

在综合各种因素进行预测后,我们再利用单因素的影响对就业前景进行预测。在解决问题二时我们得到了如公式 6-28 所示的当年高校毕业生数 b 与城镇登记失业率 s 的 ECM 模型,即:

$$s_t = 2.263509 + 0.016226b_t + 0.375127s_{t-1} - 0.018915b_{t-1}$$

我们将利用此 ECM 模型得出就业前景的仿真结果。

根据近几年国家的高校招生政策,我国的高校招生人数逐步增加,表 8-3 中的数据显示了我国在近十年来的高校招生人数。

表 8-3 我国近十年高校招生及毕业人数

年份	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
高校招生人数(万人)	108.4	159.7	220.6	268.3	320.5	382.2	447.3	504.5	546.1	565.9
高校毕业人数(万人)	83.0	84.8	95.0	103.6	133.7	187.7	239.1	306.8	377.5	447.8

由表 8-3 可以看出,我国的高校毕业生人数的在 2009 年及 2010 年仍会有所提高。资料显示 2008 年我国高校毕业生人数为 559 万人,利用灰色预测模型对 2009 年和 2010 年的高校毕业生数量进行预测,得出其数量分别为 597 万人和 644 万人。以此数据为依据,代入公式 6-28 中,得出我国城镇登记失业率在 2009 年和 2010 年分别为 4.43%和 4.61%左右。由此可以看出高校毕业生数量的增加对就业产生了很大的压力,要解决失业问题,高校的招生政策也应当有所变动。

以上我们就高校毕业生数对城镇登记失业率的影响进行了仿真,得出了我国城镇登记失业率在 2009 年和 2010 年将会有所增加的结论。以此为例,利用其它各因素对 2009 年和 2010 年我国城镇登记失业率进行仿真也得出了失业率将会增加的结论。

咨询建议

对于失业问题,许多发达国家都将治理失业问题作为宏观就业政策的一个重要部分。不同的国家和地区在治理失业方面都有其独特的方法,在一定程度上缓解了就业的压力。比如美国的“高新技术的发展不断创造新的工作岗位”,英国的“青年培训

计划”，德国的“双元制”职业教育和新加坡的“精英与大众兼顾”的教育制度。

根据本文所建立的数学模型和仿真结果，我们认为应主要从以下三个方面提高我国城镇就业人口数。

高等教育必须遵循自身的发展规律，与经济发展水平相适应，必须与市场的需求相适应。由文中的模型结果可知，中国高校学生毕业生数冲击对城镇登记失业率波动的影响是很明显的，且二者之间有显著的单向因果关系。在发展高等教育的过程中，不能超越社会法则和客观规律，那种单纯以高校规模来刺激经济发展的思路除恰违背了高等教育的发展规律，不仅浪费了巨大的社会资源，反而使毕业生就业陷入恶性循环的泥沼。中国经济要实现又好又快发展，必须从根本上改变现有的高校招生政策，将学生规模保持在一个科学、合理的范围之内，实现高校学生规模与适度失业率之间协调发展。

转变经济增长模式，注重扩大就业。文中显示片面追求经济数据的增长也会造成失业率的提高，我国在经济发展战略上应采取就业优先、效率与就业并重的增长模式。效率与就业这一尖锐的矛盾，决定着中国未来的政策选择和经济发展，若用牺牲效率和效益来扩大就业机会，将丧失国家竞争力。解决效益与就业的唯一出路，是采取效率与就业兼顾并重的政策。一方而有选择地发展高新技术产业，另一方而同时发展劳动密集型的传统产业。

加快产业结构和就业结构的同步优化。文中通过对河南地区分产业结构的研究发现，产业结构与就业密切相关。我们要在加快产业结构调整的基础上，促进就业结构的改善。首先要积极扶持第三产业的发展。其中，应特别注意发展劳动密集型的第三产业，如商业零售、交通运输、各种信息咨询、社区服务、物业管理、家庭服务业等投资少、见效快、就业潜力大的第三产业；二要运用适当的技术选择政策，按照市场方式，重组衰退产业和扶持新兴产业，提高经济增长的竞争能力，促进就业岗位的增加。通过优化资本结构，优化劳动力的配置。要在增加国家对技术改造和技术更新投资的同时厂采取多种办法扩大企业的筹、融资渠道。对有发展前途、有市场需求的产业和产品，特别是高附加值的产品，支持他们加强技术改造的力度，在投资、贷款、技术和设备引进等方面，给予适当优惠。

参考文献

- [1] 飞思科技产品研发中心, MATLAB7 基础与提高, 北京, 电子工业出版社, 2007.05。
- [2] 徐朝晖, 浙江省城镇失业影响因素分析, 浙江统计, 2006 年第五期。
- [3] 中华人民共和国国家统计局, 中国统计年鉴 2008, 北京, 中国统计出版社, 2009。
- [4] 王婧, 基于因果关系理论与 BP 神经网络整合模型的就业研究, 西安建筑科技大学学位论文, 2008。
- [5] 易丹辉, 数据分析与 Eviews 应用, 北京, 中国统计出版社。
- [6] 秦寿康, 综合评价原理与应用, 北京, 电子工业出版社, 2003。
- [7] 刘璇, 影响上海市青年就业的关键因素分析及对策研究, 东华大学学位论文, 2007。
- [8] 邓聚龙, 灰色系统理论教程, 武汉, 华中理工大学出版社, 1990。

附录

附录一：主成分分析程序

```
dataset=dataset';
stdr=std(dataset);
[n,m]=size(dataset);
sddata=dataset./stdr(ones(n,1),:);
[p,princ,eigenvalue,t2]=princomp(sddata);
p3=p(:,1:2);
p3
sc=princ(:,1:2);
sc
eigenvalue
e=eigenvalue(1:2)'
M=e(ones(m,1),:).^0.5;
compmat=p3.*M
per=100*eigenvalue/sum(eigenvalue);
per
cumsum(per)
pareto(per)
```

附录二：各影响因素与城镇登记失业率协整结果

全社会固定资产投资与登记失业率协整结果

Dependent Variable: SYL

Method: Least Squares

Date: 09/20/09 Time: 16:27

Sample (adjusted): 1999 2007

Included observations: 9 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.316769	2.050130	1.130059	0.3097
SYL(-1)	0.486091	0.546945	0.888738	0.4148
I	8.98E-05	9.22E-05	0.973320	0.3751
I(-1)	-0.000115	0.000114	-1.009718	0.3590
R-squared	0.856200	Mean dependent var		3.844444
Adjusted R-squared	0.769919	S.D. dependent var		0.466667
S.E. of regression	0.223845	Akaike info criterion		0.145374
Sum squared resid	0.250532	Schwarz criterion		0.233030
Log likelihood	3.345816	F-statistic		9.923468
Durbin-Watson stat	1.473135	Prob(F-statistic)		0.015130

货物进出口总额与登记失业率协整结果

Dependent Variable: SYL

Method: Least Squares

Date: 09/20/09 Time: 16:27

Sample (adjusted): 1999 2007

Included observations: 9 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1.396947	1.354865	-1.031060	0.3498
SYL(-1)	1.593163	0.459060	3.470490	0.0178
e	-4.52E-05	2.95E-05	-1.534594	0.1855
e(-1)	4.49E-05	3.13E-05	1.436771	0.2103
R-squared	0.880637	Mean dependent var	3.844444	
Adjusted R-squared	0.809019	S.D. dependent var	0.466667	
S.E. of regression	0.203940	Akaike info criterion	-0.040882	
Sum squared resid	0.207957	Schwarz criterion	0.046773	
Log likelihood	4.183970	F-statistic	12.29633	
Durbin-Watson stat	1.277153	Prob(F-statistic)	0.009590	

职工平均工资与登记失业率协整结果

Dependent Variable: SYL

Method: Least Squares

Date: 09/20/09 Time: 16:29

Sample (adjusted): 1999 2007

Included observations: 9 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.168384	0.932672	0.180540	0.8638
SYL(-1)	1.199909	0.436678	2.747811	0.0404
ZGGZ	0.000391	0.000507	0.770392	0.4759
ZGGZ(-1)	-0.000511	0.000636	-0.802524	0.4587
R-squared	0.831539	Mean dependent var	3.844444	
Adjusted R-squared	0.730463	S.D. dependent var	0.466667	
S.E. of regression	0.242279	Akaike info criterion	0.303651	
Sum squared resid	0.293496	Schwarz criterion	0.391306	
Log likelihood	2.633571	F-statistic	8.226829	
Durbin-Watson stat	1.510507	Prob(F-statistic)	0.022251	

国内生产总值 GDP 与登记失业率协整结果

Dependent Variable: SYL

Method: Least Squares

Date: 09/20/09 Time: 16:12

Sample (adjusted): 1999 2007

Included observations: 9 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.706865	1.025611	0.689214	0.5214
SYL(-1)	0.944589	0.282742	3.340813	0.0205
GDP	6.33E-06	4.17E-05	0.151944	0.8852
GDP(-1)	-1.03E-05	5.09E-05	-0.201426	0.8483
R-squared	0.818183	Mean dependent var		3.844444
Adjusted R-squared	0.709092	S.D. dependent var		0.466667
S.E. of regression	0.251701	Akaike info criterion		0.379949
Sum squared resid	0.316766	Schwarz criterion		0.467605
Log likelihood	2.290227	F-statistic		7.500048
Durbin-Watson stat	1.264978	Prob(F-statistic)		0.026780

城镇人口与登记失业率协整结果

Dependent Variable: SYL

Method: Least Squares

Date: 09/20/09 Time: 16:26

Sample (adjusted): 1999 2007

Included observations: 9 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-3.576434	2.067684	-1.729681	0.1443
SYL(-1)	0.601523	0.313973	1.915841	0.1135
CZRK	0.001145	0.000507	2.259304	0.0734
CZRK(-1)	-0.001087	0.000481	-2.260794	0.0733
R-squared	0.900309	Mean dependent var		3.844444
Adjusted R-squared	0.840495	S.D. dependent var		0.466667
S.E. of regression	0.186378	Akaike info criterion		-0.220981
Sum squared resid	0.173683	Schwarz criterion		-0.133325
Log likelihood	4.994414	F-statistic		15.05172
Durbin-Watson stat	2.179514	Prob(F-statistic)		0.006160

人民币汇率与登记失业率协整结果

Dependent Variable: SYL

Method: Least Squares

Date: 09/20/09 Time: 16:29

Sample (adjusted): 1999 2007

Included observations: 9 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	8.142618	21.80593	0.373413	0.7241
SYL(-1)	0.905151	0.201666	4.488370	0.0065
RMBHL	0.013844	0.022029	0.628449	0.5573
RMBHL(-1)	-0.023054	0.047922	-0.481077	0.6508
R-squared	0.829795	Mean dependent var	3.844444	
Adjusted R-squared	0.727673	S.D. dependent var	0.466667	
S.E. of regression	0.243530	Akaike info criterion	0.313948	
Sum squared resid	0.296534	Schwarz criterion	0.401604	
Log likelihood	2.587232	F-statistic	8.125474	
Durbin-Watson stat	1.484865	Prob(F-statistic)	0.022815	

当年高校毕业生数与登记失业率协整结果

Dependent Variable: SYL

Method: Least Squares

Date: 09/20/09 Time: 16:25

Sample (adjusted): 1999 2007

Included observations: 9 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.263509	1.714954	1.319866	0.2441
SYL(-1)	0.375127	0.552745	0.678662	0.5275
BYS	0.016226	0.013321	1.218057	0.2775
BYS(-1)	-0.018915	0.014520	-1.302686	0.2495
R-squared	0.873465	Mean dependent var	3.844444	
Adjusted R-squared	0.797544	S.D. dependent var	0.466667	
S.E. of regression	0.209977	Akaike info criterion	0.017467	
Sum squared resid	0.220452	Schwarz criterion	0.105122	
Log likelihood	3.921401	F-statistic	11.50492	
Durbin-Watson stat	1.762674	Prob(F-statistic)	0.011065	

附录三：河南省 2001-2007 年产业比重和就业比重时序表

	x_1 (%)	x_2 (%)	x_3 (%)	y_1 (%)	y_2 (%)	y_3 (%)
2001 年	22.3	45.4	32.3	63.0415	16.071	20.8871
2002 年	21.3	45.9	32.8	61.5357	18.798	19.6668
2003 年	17.5	48.2	34.3	60.1879	16.581	23.2312
2004 年	19.3	48.9	31.8	58.0888	21.437	20.4746
2005 年	17.9	52.1	30.0	55.4398	24.095	20.4656
2006 年	16.4	53.8	29.8	53.331	27.623	19.046
2007 年	14.8	55.2	30	50.5803	26.758	22.6619

产业比重和就业比重时序表

附录四：灰色预测程序

```

function A=huise(x)
a=length(x);
y=zeros(1,a); %生成数列
Y=zeros(1,a-1); %数据向量
z=zeros(1,a-1);
B=zeros(a-1,2); %数据矩阵
A=zeros(2,1); %待定参数列
y(1)=x(1);
for i=2:a
    y(i)=x(i)+y(i-1);
end
for i=1:a-1
    z(i)=1/2*(y(i)+y(i+1));
end
for i=1:a-1
    B(i,1)=-z(i);
    B(i,2)=1;
end
for i=1:a-1
    Y(i)=x(i+1);
end
A=inv((B'*B))*B'*Y';

```