

全国第六届研究生数学建模竞赛



题 目 我国就业人数的主要影响因素分析及前景预测

摘 要

本文针对社会与国民经济中极其重要的问题——就业问题进行分析研究,通过对有关统计数据进行分析,运用统计软件 *SPSS*,采取主成份分析方法提取了影响就业的主要因素或指标,建立了我国就业人数的数学建模,分别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。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。最终结合模型与仿真结果,对提高我国城镇就业人口数或减少城镇登记失业率提出较好的建议。

针对问题 1,本文结合就业人数折线图,将我国就业分成两个阶段进行研究,综合运用了灰色关联度与相关系数两种方法,建立了指标排序模型,按所得值大小来选取影响就业的主要指标,依照 $\gamma \geq 0.81$ 的原则对主要因素或指标进行提取。

针对问题 2,本文分别利用主成分回归分析法与向量自回归分析法分段构建了城镇就业人数拟合模型,并依据所得到的向量自回归模型做出脉冲响应函数图以及方差分解图,得到每一个指标对城镇就业人数的冲击效应及方差贡献率,综合两阶段模型的结果可知对我国城镇就业人数影响最大的是城市化水平。

针对问题 3,结合上述数学模型从包含主要的经济社会指标,文章分别从东、中、西三大区域,和国有、集体和其他经济行业三方面以及农民工、大学毕业生及其他三种不同人群的角度来构建相关的人口就业模型。

针对问题 4,利用神经网络预测的方法,结合国家的有关决策和规划,对 2009 年及 2010 年上半年我国城镇居民的就业前景进行仿真预测。文章最后根据所建立的数学模型的相关影响因素,我们从经济发展、财政政策、货币政策等 5 个角度出发,并结合社会实际,提出提高城镇就业水平的政策、建议。

关键词: 就业; 影响就业指标体系; 灰色关联度; 相关系数; 神经网络预测

§1 问题的重述

1.1 就业与失业统计法

1.1.1 什么是就业

就业就是一定年龄阶段内的人们所从事的为获取报酬或经营收入所进行的活动。细分则从三个方面进行界定：①就业条件，指一定的年龄；②收入条件，指获得一定的劳动报酬或经营收入；③时间条件，即每周工作时间的长度。按照已有研究，就业也可以定义为一定年龄阶段内的人们在三个月内有稳定的收入或与用人单位有劳动聘用关系的活动。

1.1.2 经济中地位

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

1.1.3 失业统计法

失业的统计方法各国差异较大，我国采用城镇登记失业率，其定义如下：

$$\text{城镇登记失业率} = \frac{\text{城镇登记失业人数}}{\text{城镇从业人数} + \text{城镇登记失业人数}}$$

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。

1.2 就业的背景与对策

1.2.1 影响因素分析

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

1.2.2 我国现实状况

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

1.2.3 政府就业对策

面对挑战，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）。

1.3 数据获取的途径

1.3.1 参考网站

(1) <http://www.stats.gov.cn/tjsj/ndsj/>;

(2) <http://dlib.cnki.net/kns50/>;

(3) <http://www.drcnet.com.cn> 点击宏观经济,再点击运行数据或数据平台或旧库数据平台;

(4)<http://chinese.wsj.com/gb/index.asp> 点击经济,再点击经济数据;

(5)<http://news.hexun.com/> 点击宏观数据。

1.3.2 相关附件

附件中有部分逐月经济运行数据,供参考,具体如以下。

(1)附件一: WORD 文档“A 题经济数据”;

(2)附件二: WORD 文档“A 题新闻”;

1.3.3 统计资料

《新中国 55 年统计资料汇编》、《2005-2008 年中国统计年鉴》、《1996-2008 中国劳动力统计年鉴》、《2005-2008 年中国人口统计年鉴》等统计资料。

1.4 要解决的具体问题

请研究生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(参见上面网站,也可以对比其他国家的统计数据)就我国就业人数或城镇登记失业率研究如下问题。

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;

问题 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;

问题 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);

问题 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);

问题 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。

希望利用上面网站上的数据,可以参考已有文献的数学模型,但更鼓励创新、使用翔实可靠的数据、在多方面对各项政策措施进行对比论证。

§2、问题的分析

2.1 问题的总体分析

就业(失业)是社会、国民经济中极其重要的问题,本文要求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,进行统计分析,寻找影响就业的主要因素或指标。

研究对象为我国近年来经济发展中的一系列时间序列分析,采用方法为主成分分析、关联度分析,找出与就业问题相关的主要因素或指标,并结合各类具体数据,运用灰色预测法对未来一两年内的就业问题进行预测。

本问题的难点是对数据涉及量大,指标涉及太多,且有国家政策等不确定的因素混合在内,因此合理指标体系的构建是本文的重中之重。对于指标体系的构建时,指标的个数不可太少,也不可太多。为此,首先我们结合经济学知识从我国经济相关时间序列数据中查找尽可能多的相关因素;其次我们可运用基本统计方法将相关数据进行分类筛选,采取同类合并、化多为少、取重舍次等方法,最终构建出较为合理、方便操作的“适度”指标以便于研究。

2.2 对具体问题的分析

2.2.1 对问题 1 的分析

问题 1 要求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,这要求我们寻找与就业紧密相关的各类因素,然后将这些因素进行分门别类,形成三级指标体系。然后,我们运用多元统计分析法中的主成分分析法与灰色关联度方法,由此排序可找出影响就业的主要因素。

2.2.2 对问题 2 的分析

首先根据对问题 1 提出来的影响就业的主要指标,结合历年来我国城镇就业人数的变化规律,可考虑运用主成分回归分析法与向量自回归分析法来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。

2.2.3 对问题 3 的分析

前面建立的模型是全盘考虑的多元回归分析,问题 3 要求因素较多,从包含主要的经济社会指标,然后从分行业、分地区、分就业人群三个角度,尝试建立比较精确的数学模型。对分行业,我们考虑到行业过多,由于时间限制,我们选取了几个容易获取数据且影响就业的行业:国有经济企业、乡镇集体经济和其他经济类型等。对分地区,从简洁考虑,我们分东部、中部、西部三个地区进行建模研究,至于更细的可类似处理。对分就业人群,我们选择几个主要就业人群:大中专在校毕业生、农民工数量、其它人群(如高中待业、转业复原军人、下岗工人等)几个主要就业人群。

2.2.4 对问题 4 的分析

为了科学合理对未来城镇就业前景的预测,减小预测模型的误差,本文利用问题 2 中已建立的数学模型预测结果与神经网络预测结果,得出城镇就业的综合预测值。第一种预测方法我们首先利用神经网络方法模拟仿真出不同模型中各影响因素的数值,将其代入我们得到的数学模型,得到预测结果,第二种方法我们直接利用神经网络得到仿真预测结果。最后我们进行细化,结合国家的有关决策和规划对我国不同行业的 2009 年及 2010 年上半年的就业前景进行仿真预测。

2.2.5 对问题 5 的分析

问题 5 要求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。对此,我们要结合问题 2 所建立的模型及问题 4 的仿真结果,分别依照所选的六个主要影响结果的进行相关的政策建议,为有关政府部门提供政策,进而有效应对当前金融危机下的我国高校毕业生以及农民工就业难的困境,在危机中抢抓机遇,从而实现我国经济的稳定可持续的健康快速发展。

§3、模型的假设

1. 我国城市劳动力处于供过于求的状态,没有达到最佳的就业状态;
2. 假设男女就业(获取工作)的机会相同,即不考虑性别因素;
3. 规定年龄在 15—64 岁之间的人为我国的劳动人口,不考虑不具有劳动能力的人;
4. 规定失业人口不仅包括城镇登记失业人口,还包括城镇未登记的、非自愿失业人口和农村剩余劳动力;
5. 不考虑因自然灾害(如四川地震)等原因对就业产生意外影响。

§4、符号说明

序号	符号	符号说明
1	P	表示就业人数
2	GDP	表示国内生产总值
3	K	表示固定资产投资
4	ρ	表示可比价格表示的固定资产投资额的平均增长率
5	δ	表示折旧率
6	mar	市场化指数
7	str	产业结构调整水平
8	zr	人口自然增长率
9	csb	城市化水平
10	lb	表示劳动力资源总数占总人口比例
11	cxc	表示城乡收入差距
12	czs	表示财政收入
13	czz	表示财政支出
14	ne	表示净出口
15	$M2$	表示广义货币供给量
16	tax	表示税收
17	kzf	表示科学研究占财政支出比重
18	eng	表示恩格尔系数
19	com	表示居民消费水平
20	$huma$	表示人力资本
21	d_1	表示小学人口所占比重
22	d_2	表示初中人口所占比重
23	d_3	表示高中人口所占比重
24	d_4	表示高中以上人口所占比重
25	pr	表示平均货币工资水平

§5 模型的准备

5.1 数据的采集与指标的构建

5.1.1 数据的采集

根据问题要求,我们首先通过经济意义的分析,初步选取以下30个因素作为影响就业人数的主要统计数据:就业人数、国内生产总值、固定资产投资总额、财政收入、财政支出、进口总额、出口总额、货币供给量 $M2$ 、税收总收入、科学研究占财政支出比重、职工工资总额、劳动者报酬、城镇非国有单位从业人员、城镇从业人员、外商固定资产投资、第一产业增加值、第二产业增加值、第三产业增加值、人口自然增长率、城镇人口数、总人口数、15-65岁人口数、城镇居民可支配收入、农村居民纯收入、恩格尔系数、居民消费、小学在校学生数、中学在校人数、大专及以上在校人数、平均货币工资等。

数据来源于《新中国55年统计资料汇编》、《2005-2008年中国统计年鉴》、《1996-2008中国劳动力统计年鉴》、《2005-2008年中国人口统计年鉴》等统计资料。

5.1.2 指标体系的说明

根据上述统计数据资料,为了便于进一步建模分析,以下我们将30个因素相关的数据进行同类合并、化多为少、取重舍次等方法,构建以下影响就业的18个主要指标。

(1)**就业人数 (P)**:在本文模型中,我们选取每年就业人数来衡量我国就业情况,考虑到城镇登记失业率在统计上存在一定的系统性误差,没有将既未从业又未参加登记的人群计算在内,故不能反映出真实的就业状况。

(2)**国内生产总值 (GDP)**:从宏观调控角度看, GDP 的增加是创造新增就业机会的主要来源,所以我们初步选取 GDP 作为影响就业人数的主要因素。

(3)**固定资产投资 (K)**:固定资产是社会扩大再生产的基本手段,是实现国民经济持续、快速、健康发展的原动力,它通过拉动经济、促使产业结构升级来扩大就业,因此可选取固定资产投资作为一个影响因素。

对于资本投入量的测算,按照惯例文章采取资本存量并通过永续盘存法来对资本投入量进行测算,即用上一年的资本净存量(资本存量减折旧)加上当年的投资来表示,其计算公式为:

$$K = K_{t-1} + I_t - \delta K_{t-1}。$$

其中,初始年份的资本存量可以采取Hall-Jones(1999)的方法^[1]来估算,其计算公式为:

$$K_0 = I_0 / (\rho + \delta)$$

其中 ρ 为计算期内可比价格表示的固定资产投资额的平均增长率, δ 代表折旧率,一般对于我国固定资产折旧率取值为10%(Hall和Jones在模拟世界上127个国家的资本存量时折旧率取值为6%)。

(4)**财政收入 (czs)**:增加税收会对人民就业产生影响,财政收入从某种程度上减少了人们的收入,更多地就业,增加就业人数,同时也可能减少就业,因此我们引入财政收入作为影响就业人数的主要因素。

(5)**财政支出 (czz)**:能够给予社会补贴,实施特别培训计划,支持企业吸纳就业、支持相关机构提供就业服务,扶持职业中介机构,并通过大规模增加政府投资,吸纳更多的人就业,实现就业增长等。

(6)**净出口(ne)**:净出口额是出口总额与进口总额的差值,其计算公式为:

$$\text{净出口额} = \text{出口总额} - \text{进口总额}$$

净出口额的增加，在某种程度上会刺激国内经济发展，增加就业机会。

(7)货币供给量($M2$): 货币供给量是指一国在某一时期内为社会经济运转服务的货币存量，它由包括中央银行在内的金融机构供应的存款货币和现金货币两部分构成。我们一般用广义货币量 $M2$ 来代替。

$M2$ =流通中现金+企业活期存款+机关团体部队存款+农村存款+个人持有的信用卡类存款+城乡居民储蓄存款+企业存款中具有定期性质的存款+信托类

(8)税收(tax): 税收是国家为实现其职能，凭借政治权力，按照法律规定，通过税收工具强制地、无偿地征参与国民收入和社会产品的分配和再分配取得财政收入的一种形式。税收对就业的增加，可以分两方面考虑：一方面，税收增加，人们的可支配收入减少，人们为了维持原有的收入水平，从而多劳动，劳动供给增加。另一方面，税收增加，劳动力报酬减少，打击劳动者的工作积极性，劳动供给减少。

(9)科研支出率 (kzf): 科技经费支出额占 GDP 比例（反映对科技活动的投入强度和重视程度）。由于政府对科技活动的投入强度和重视程度的加大，将会直接导致全社会科技创新水平的显著提高，进而带动产业结构升级等一系列现象的发生。因此，本文构建科学研究占财政支出比重指标来反映这种现象，其计算公式为：

科学研究占财政支出比重=科技活动的投入/财政支出

(10)市场化指数 (mar): 对于市场化指数的测量，由于观察角度及使用的标准不同，许多学者提出了各自的方法论。国家计委课题组(1996年)是从商品市场(包括生产环节和流通环节)的市场化和要素市场(包括劳动力市场和资金市场)的市场化程度入手进行测算的。陈宗胜(1999年)认为，对经济体制市场化程度的测度，要按企业、政府、市场三方面来考虑。徐明华(1999年)则从所有制结构、政府职能转变和政府效率、投资的市场化、商品市场发育、要素市场发育、对外开放、经济活动频度、人的观念，8个方面进行了测算。樊纲、王小鲁（2003年）则以政府与市场的关系、非国有经济的发展、产品市场的发育程度、要素市场的发育程度、市场中介组织发育和法律制度环境5个方面，23个指标为基础，运用“主因素分析法”进行市场化指数的测量。

对于市场化指数的影响，可以分为正向指标和逆向指标。正向指标越大，逆向指标越小，市场化指数越大；反之市场化指数越小。本文根据以往研究经验，经过正逆指标的调整（劳动力市场中城镇国有单位从业人员/城镇从业人员为逆指标，而城镇非国有单位从业人员/城镇从业人员则为正指标），从正向指标的方向考虑，选取分别代表收入分配、劳动力市场、和资本市场的各一项指标，运用 $SPSS$ 的主成份分析方法对市场化指数进行计算。

(11)产业结构调整水平 (str):对于影响就业结构的产业而言，第二三产业行业多、门类广，劳动密集、资本密集、技术密集行业并存，具有吸纳各类劳动力就业的独特优势和作用。因此，为了更好地研究产业结构演进和就业数量之间的关系，文章参照张雷(2008)，设定用来反映区域产业结构多元化演进程度的指标 Str ，其计算公式为：

$$str = \sum (P/P, S/P, T/P) \quad (1 \rightarrow \infty)$$

其中， P 为第一产业产出， S 为第二产业产出， T 为第三产业产出，并且 str 的值域可以从1到无穷大，其值越大代表产业结构的多元化程度越高，否则则反之。

(12)人口自然增长率 (zr): 人口自然增长率，是反映人口发展速度和制定人口计划的重要指标，也是计划生育统计中的一个重要指标，它表明人口自然增长的程度和趋势。它反映了人口再生产活动的的能力。在数值上，它是人口出生率与人口死亡率的差值，其计算公式为：

人口自然增长率=人口出生率-人口死亡率

根据分析认为人口自然增长率的增加也是就业人数增加的一个原因。

(13)城市化水平 (csh): 又叫城市化率, 是衡量城市化发展程度的数量指标, 一般用一定地域内城市人口占总人口比例来表示。

城市化水平=城镇人口数/农村居民人口数。

(14)劳动资源率 (lb):

劳动资源率即劳动力资源总数占总人口比例。劳动力的定义分广义和狭义, 广义上的劳动力指全部人口, 狭义上的劳动力则指具有劳动能力的人口。本文中取统计年鉴中15岁至65岁的人口为广义劳动力人口。

劳动资源率=广义劳动力人口/总人口数。

(15)恩格尔系数 (eng): 恩格尔系数是根据恩格尔定律得出的比例数, 是表示生活水平高低的一个指标。主要表述的是食品支出占总消费支出的比例随收入变化而变化的一定趋势, 其计算公式为:

恩格尔系数=食品支出总额/家庭或个人消费支出总额×100%

其中恩格尔系数越大, 代表该地区或家庭生活越贫困, 越小则反之。

(16)居民消费水平 (com): 居民消费水平是指居民在物质产品和劳务的消费过程中, 对满足人们生存、发展和享受需要方面所达到的程度。通过消费的物质产品和劳务的数量和质量反映出来。

(17)人力资本 (huma): 人力资本是指人们花费在教育、健康、训练、迁移和信息获取等方面的开支所形成的资本, 之所以称为人力资本是因为无法将其同它的载体分离开。目前, 关于人力资本的测量有四种方法: 一是以人们已接受学校的教育年限来分类确定; 二是用平均每万人在校中学生人数来衡量 (Barro, Lee, 1993); 三是用入学率作为替代变量来衡量人力资本的教育程度 (Barro, Lee, 1996; Barro, 1997, 2001); 四是用教育经费占GDP或财政支出的比重来衡量 (沈利生, 朱运法, 1997)。

目前学者们普遍使用人均受教育年数来度量, 因此, 文章对于人力资本的测算, 通过考虑小学、中学、大学等不同层次的教育对劳动者素质的提高起到的作用, 即平均受教育年数来代表影响能源使用效率的人力资本, 其计算公式为:

$$H = 6d_1 + 9d_2 + 12d_3 + 16d_4,$$

其中 $d_i (i = 1, 2, 3, 4)$ 表示在6岁及6岁以上人口中文化程度是小学、初中、高中及高中以上人口所占的比重。

(18)城乡收入差距 (cxc): 城乡收入差距指的是我国城镇居民收入与农村居民收入差值。其计算公式为:

$$cxc = Ic - In$$

其中, Ic 代表城市居民收入水平, In 代表农村居民收入水平。

城乡收入差距越来越大, 使得大量农村居民外出务工, 在一定程度上导致就业人数增加。城乡收入差距变化的背后原因是, 城市经济部门对劳动力需求的增加速度与农村劳动力供给的速度的不对称现象, 并且当前者大于后者时即出现了所谓的“刘易斯转折点”。

5.1.3 指标体系模型的建立

下面我们将以上18个影响就业指标体系分成四大类, 宏观经济 (国内生产总值、固定资产投资、净出口、市场化指数、产业结构调整水平), 调控政策 (财政收入、财政支出、税收、货币供给量、科研支出率), 人口状况 (就业人数、人口自然增长率、城市化水平、劳动资源率、人力资本), 生活水平 (城乡收入差距、恩格尔系数、居民消费水平)。为了便于直观理解, 我们将以上18个影响就业指标体系结合分类, 可构建出影响就业指标体系模型 (如图1)。

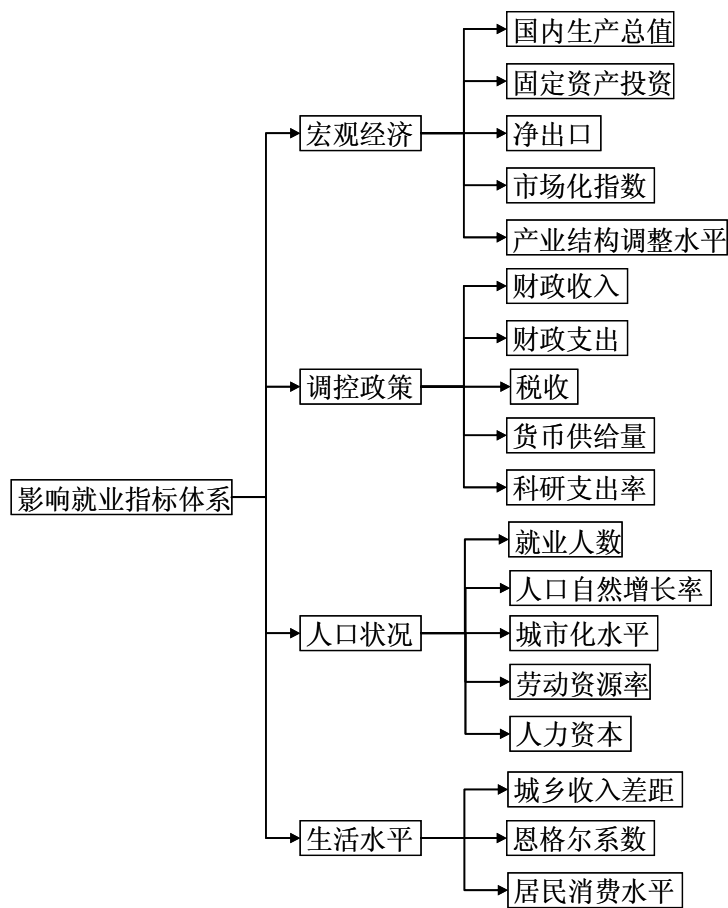


图 1 影响就业指标体系模型

5.2 初始数据提取与分析

5.2.1 就业问题研究阶段的划分

以下利用EXCEL软件对我国历年来城镇就业人数作折线图（见图2），从折线图可以发现：图形大致呈折线形变化，其中城镇就业人数在1989年至1990年发生较大幅度的浮动。因此，就模型的准确性与合理性方面考虑，本文可将我国城镇就业人数的增长模型分两个阶段构建：即第一阶段1978~1989年，我国改革开放初级阶段；第二阶段1990~2007年，是我国在经济上进一步深化改革带来就业需要大幅增加的结果。

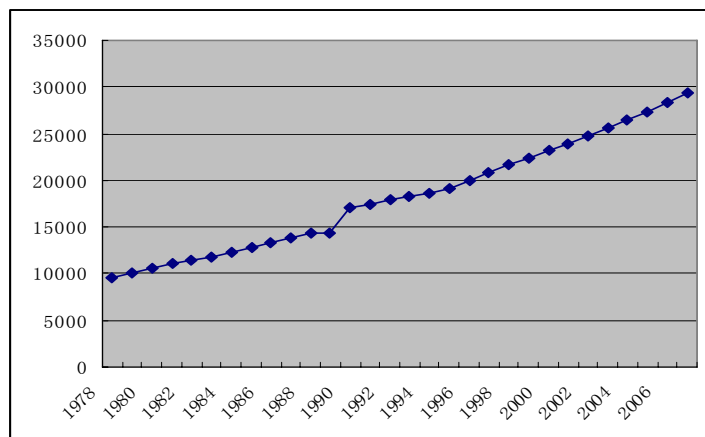


图 2 全国城镇就业人数折线图

5.2.2 数据的补充及算法

由于对数据来源的可获得性以及统计资料的完备性均受到限制，甚至某些指标出现少许缺省值，出于对模型准确与合理方面的考虑，我们以下利用Newton插值法^[2]对原始数据进行插值，补全缺省数据。

(1)牛顿插值法基本思路：

给定插值点序列 $(x_i, f(x_i))$, $i = 0, 1, \dots, n$ 。构造牛顿插值多项式 $N_n(u)$ 。输入要计算的函数点 x ，并计算 $N_n(x)$ 的值，利用牛顿插值公式，当增加一个节点时，只需在后面多计算一项，而前面的计算仍有用；另一方面 $N_n(x)$ 的各项系数恰好又是各阶差商，而各阶差商可用差商公式来计算。

(2)牛顿插值法计算步骤：

①输入 n 值及 $(x_i, f(x_i))$, $i = 0, 1, \dots, n$ ；要计算的函数点 x 。

②对给定的 x ，由

$$N_n(x) = f(x_0) + (x - x_0)f[x_0, x_1] + (x - x_0)(x - x_1)f[x_0, x_1, x_2] + \dots + (x - x_0)(x - x_1) \dots (x - x_{n-1})f[x_0, x_1, \dots, x_n]$$

计算 $N_n(x)$ 的值。

③输出 $N_n(x)$ 。

5.2.3 实际价格的计算

为了消除我们收集的原始数据中价格因素的影响，对于全国统计数据，我们利用全国居民消费价格指数对原始数据进行价格平减，即用实际价格除以以1978年为基期的居民消费价格指数来近似代替。对于各省份之间的价格指数计算我们利用各省份的居民价格指数进行平减。

§6 模型的建立与求解

6.1 问题 1：指标排序模型

6.1.1 建模思路

问题一要求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，我们接下来对§5第一部分所构建的18个指标分别与就业人数进行关联性分析，我们分别运用灰色关联度与相关系数两种方法进行分析，然后将它们综合成排序指标模型，按所得值大小来选取影响就业的主要指标。

6.1.2 理论准备

(1) 灰色关联度^[3]

灰色系统理论是一种用灰色关联度顺序来描述因素间关系的强弱，大小，次序的。其基本思想是：以因素的数据列为依据，用数学的方法研究因素间的几何对应关系。*GRA* 实际上是动态指标的量化分析，充分体现了动态意义。

$$\gamma_{\text{关联度}} = \frac{\min_{i,k} |x_{0(k)} - x_i(k)| + \xi \cdot \max_{i,k} |x_{0(k)} - x_i(k)|}{\Delta_{0j(k)} + \xi \cdot \max_{i,k} |x_{0(k)} - x_i(k)|}$$

若关联度 $\gamma_{\text{关联度}}$ 最大, 说明 $x_i(k)$ 与最优指标 $x_0(k)$ 最接近, 即第 i 个被评价对象优于其他被评价对象, 据此可以排出各被评价对象的优劣次序。

(2) 相关系数

相关系数是描述变量之间相关程度的指标, 用 $\gamma_{\text{相关系数}}$ 表示, 取值范围为 $[-1, 1]$ 。 $|\gamma_{\text{相关系数}}|$ 值越大, 误差 Q 越小, 变量之间的线性相关程度越高; $|\gamma_{\text{相关系数}}|$ 值越接近 0, Q 越大, 变量之间的线性相关程度越低。

$$\gamma_{\text{相关系数}} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot \sum_{i=1}^n (y_i - \bar{y})^2}}$$

6.1.3 模型建立

灰色系统理论的关联度分析与数理统计学的相关系数是不同的, 两者的区别主要在于以下三方面:

(1) 它们的理论基础不同。前者是基于灰色系统的灰色过程, 而后者则是基于概率论的随机过程;

(2) 分析法不同。前者是进行因素间时间序列的比较, 而后者是因素间数组的比较;

(3) 数据量要求不同。前者不要求数据太多, 而后者则需有足够的数量。

因此, 本文为了更加全面的反映各项指标与就业人数之间的相关性, 我们综合考虑灰色关联度 $\gamma_{\text{关联度}}$ 及相关系数 $\gamma_{\text{相关系数}}$ 这两方面, 构建排序指标模型如下:

$$\gamma = \frac{1}{2} \times (\gamma_{\text{关联度}} + \gamma_{\text{相关系数}})$$

可根据此排序指标值的大小对各个指标进行排序, 依此选取影响就业的主要指标。

6.1.4 模型求解

根据对初始就业人数数据的分析, 本文认为对该问题的分析应分为两个阶段: 第一阶段为 1978 - 1989 年, 第二阶段为 1990 - 2007 年。分别对这两个阶段影响我国就业人数的指标进行排序, 具体结果如下表所示:

表 1 1978-1989 年影响就业人数因素的 r 值

名次	1	2	3	4	5	6	7	8	9
指标	<i>M2</i>	<i>csH</i>	<i>czZ</i>	<i>GDP</i>	<i>com</i>	<i>K</i>	<i>pr</i>	<i>inc</i>	<i>tax</i>
r 值	0.85	0.84	0.83	0.82	0.80	0.80	0.79	0.74	0.73
名次	10	11	12	13	14	15	16	17	18
指标	<i>lb</i>	<i>czS</i>	<i>str</i>	<i>zr</i>	<i>eng</i>	<i>ne</i>	<i>mar</i>	<i>kzf</i>	<i>huma</i>
r 值	0.69	0.67	0.62	0.61	0.50	0.40	0.31	0.30	0.23

表 2 1990-2007 年影响就业人数因素的 r 值

名次	1	2	3	4	5	6	7	8	9
指标	<i>K</i>	<i>csH</i>	<i>inc</i>	<i>com</i>	<i>M2</i>	<i>GDP</i>	<i>pr</i>	<i>czZ</i>	<i>czS</i>
r 值	0.86	0.83	0.83	0.82	0.82	0.81	0.80	0.78	0.76
名次	10	11	12	13	14	15	16	17	18
指标	<i>huma</i>	<i>tax</i>	<i>str</i>	<i>mar</i>	<i>ne</i>	<i>zr</i>	<i>eng</i>	<i>lb</i>	<i>kzf</i>
r 值	0.75	0.74	0.71	0.69	0.60	0.60	0.59	0.55	0.52

6.1.5 指标选取

根据表 1 及表 2 我们可以清楚的观察到,不同时期不同指标对就业人数的关联性大小及排序,可以发现各指标与就业人数都有一定的关联性。然而对主要指标的选取原则,却具有一定程度的主观性。此处,结合数据分析所受到的限制及建模结果,本文按照 $\gamma \geq 0.81$ 的原则对主要因素或指标进行提取。1978—1989 年,影响就业人数的有货币供给量 M2、城市化水平、财政支出、国内生产总值这四个指标。而在 1990-2007 年这个阶段,影响就业人数的有固定资产投资总额、城市化水平、城乡收入差距、居民消费水平、货币供给量 M2、国内生产总值这六个指标。

6.2 问题 2: 城镇就业人数拟合模型

6.2.1 建模思路

从上面的分析得知,各个阶段影响城镇就业人数的指标是不同的,运用多元线性回归方法衡量这些因素对我国城镇就业人数影响的大小,为了避免回归时产生多重共线性,我们对不同阶段所选取出来的影响指标提取主成分,利用主成分进行回归,同时,用各阶段影响指标建立 VAR 模型,通过观察其脉冲响应函数及方差贡献率图研究各阶段的影响指标对我国城镇就业人数变动的长期影响及短期影响。

6.2.2 理论准备

(1)主成分回归分析^[5]

主成分回归分析是将回归模型中有严重多重共线性的变量进行因子提取,得到正交的因子变量,然后对因子变量进行回归模型的建立,从而可以解决由于共线性而造成伪回归方程的问题。其中主成分是一种通过降维技术把多个指标约化为少数的几个综合指标,这些称为主成分的综合指标能够反映出原始指标的绝大部分信息。

(2)向量自回归分析

向量自回归(VAR)是基于数据的统计性质建立模型,VAR 模型把系统中每一个内生变量作为系统中所有内生变量的滞后值的函数来构造模型,从而将单变量自回归模型推广到由多元时间序列变量组成的“向量”自回归模型。脉冲响应函数描述的是 VAR 模型中的一个内生变量的冲击给其他内生变量所带来的影响,它是指是随着时间的推移,观察模型中的各变量对于冲击是如何反应的,然而对于只是要简单地说明变量间的影响关系又稍稍过细了一些。方差分解是通过分析每一个结构冲击对内生变量变化(通常用方差来度量)的贡献度,进一步评价不同结构冲击的重要性。因此,方差分解给出对 VAR 模型中的变量产生影响的每个随机扰动的相对重要性的信息。也就是说相对方差贡献率是根据第 j 个变量基于冲击的方差对 y_i 的方差的相对贡献度来观测第 j 个变量对第 i 个变量的影响。

6.2.3 变量检验

建立 VAR 模型的前提条件是各个变量同阶单整,在建立 VAR 模型前,必须对各个变量进行平稳性检验。检验结果如下:

表 3 相关变量的平稳性检验 (1978-1989)

变量名	检验值	临界值	滞后期数	显著性水平
$\Delta \ln p$	-2.852	-2.748	0.000	0.100
$\Delta^2 \ln GDP$	-1.951	-1.600	0.000	0.100
$\Delta \ln czz$	-4.095	-3.321	2.000	0.010
$\Delta \ln M2$	-3.690	-3.213	0.000	0.050
$\Delta \ln csh$	-7.798	-5.835	1.000	0.010

表 4 相关变量的平稳性检验 (1990-2007)				
	检验值	临界值	滞后期	显著性水平
$\Delta^2 \ln P$	-5.622	-3.959	2.000	0.010
$\Delta \ln GDP$	-5.584	-4.800	2.000	0.010
$\Delta \ln K$	-2.277	-1.966	1.000	0.050
$\Delta \ln M2$	-4.566	-3.920	0.000	0.010
$\Delta^2 \ln csh$	-3.806	-3.081	0.000	0.050
$\Delta \ln cxc$	-2.945	-2.681	0.000	0.100
$\Delta \ln com$	-4.078	-3.920	0.000	0.010

由表中可知, 在 1978—1989 年间, $\Delta \ln P$, $\Delta^2 \ln GDP$, $\Delta \ln czz$, $\Delta \ln M2$, $\Delta \ln csh$ 是平稳的; 在 1990—2007 年这一阶段, $\Delta^2 \ln P$ 、 $\Delta \ln GDP$ 、 $\Delta \ln K$ 、 $\Delta^2 \ln csh$ 、 $\Delta \ln cxc$ 、 $\Delta \ln M2$ 、 $\Delta \ln com$ 是平稳的。

6.2.4 模型的建立与求解

(1) 主成分回归模型

① 1978~1989 年阶段

对于 1978—1989 年这一阶段, 我们利用已提取出四个因素: 国内生产总值 GDP , 财政支出 czz , 广义货币供给量 $M2$, 城市化水平 csh , 来对就业人数进行拟合。若直接利用最小二乘法对数据进行拟合, 容易产生多重共线性等问题, 为了避免这种情况的发生, 我们首先对各个指标提取主成分: 第一个主成分的方差贡献率为 97.844%, $\lambda_1 = 3.915$ 这里我们提取一个主成分, 表达式为

$$F = 0.4938 \times \ln GDP + 0.4933 \times \ln czz + 0.5003 \times \ln M2 + 0.5024 \times \ln csh$$

这里得到一个关于主成分的时间序列, 用这个序列对我国 1978—1989 年城镇就业人数进行拟合, 结果为:

$$\ln P_{78} = 8.5518 + 0.1636 \times F$$

其中 $R^2 = 0.9885$ 拟合效果较好, 且综合指标的系数显著。

因此我们可以进一步求出就业人数拟合模型如下:

$$\ln P_{78} = 8.5518 + 0.0808 \times \ln GDP + 0.0808 \times \ln czz + 0.0819 \times \ln M2 + 0.0822 \times \ln csh$$

② 1990~2007 年阶段

对于 1990-2007 年这一阶段, 根据前面所说的指标选取原则, 这里, 我们主要选取国内生产总值 GDP 、固定资产投资 K 、广义货币供给量 $M2$ 、城市化水平 csh 、城乡收入差距 cxc 以及居民消费水平 com , 就业人数记为 P , 为了便于分析, 我们先对各项指标取对数, 分别记为 $\ln GDP$ 、 $\ln czz$ 、 $\ln M2$ 、 $\ln csh$ 、 $\ln cxc$ 、 $\ln com$ 、 $\ln P$ 。

首先来分析我们选取出来的此阶段的各项指标与就业人数的计量模型。由于选取的是六个指标, 运用最小二乘法对各个指标进行拟合会存在多重共线性的问题, 为了避免这种情况的发生, 我们首先运用主成分分析法提取这六个指标的主成分, 然后用所提取出来的主成分对我国城镇就业人数进行回归, 然后再展开, 可以得到各个指标的系数情况。

主成分回归的相关结果: 此时累计方差贡献率大于 85%, 特征值 $\lambda = 2.07494$ 选取一个主成分, 表达式为:

$$F = 0.688 \times \ln GDP + 0.684 \times \ln K + 0.690 \times \ln M2 + 0.679 \times \ln csh + 0.683 \times \ln cxc + 0.686 \times \ln com$$

这里得到一个关于主成分的时间序列, 用这个序列对我国 1990—2007 年城镇就业

人数进行拟合，结果为：

$$\ln P_{90} = 10.3305 + 0.0253 \times F$$

其中 $R^2 = 0.9899$ 拟合效果较好，且综合指标的系数显著。

因此我们可以进一步求出城镇就业人数拟合模型如下：

$$\begin{aligned} \ln P_{90} = & 10.331 + 0.0174 \times \ln GDP + 0.0171 \times \ln czz \\ & + 0.018 \times \ln M2 + 0.017 \times \ln csh + 0.017 \times \ln cxc + 0.017 \times \ln com \end{aligned}$$

(2) 向量自回归模型

① 1978~1989 年阶段

下面我们利用 1978—1990 年的主要影响指标变动对我国城镇就业人数变动的影响建立向量自回归模型。根据表 3 的平稳性检验结果， $\Delta \ln p$ ， $\Delta^2 \ln GDP$ ， $\Delta \ln czz$ ， $\Delta \ln M2$ ， $\Delta \ln csh$ 是平稳的，可以建立 VAR 模型。

$$\begin{pmatrix} \Delta \ln p_t \\ \Delta^2 \ln GDP_t \\ \Delta \ln cee_t \\ \Delta \ln M2_t \\ \Delta \ln csh_t \end{pmatrix} = \begin{pmatrix} 3.3417 \\ 16.6871 \\ -47.7658 \\ -35.3568 \\ -6.8870 \end{pmatrix} + \begin{pmatrix} 0.0115 & -0.0167 & 0.0229 & 0.2536 & 0.6193 \\ 0.7052 & -0.1774 & 0.3928 & 0.0801 & -1.7377 \\ 1.6617 & 0.5889 & 0.0905 & -4.2356 & 5.8674 \\ 1.3537 & -0.1541 & 0.2413 & -0.0706 & 3.9683 \\ 0.2120 & 0.2197 & 0.1238 & 0.2563 & 0.9512 \end{pmatrix} \times \begin{pmatrix} \Delta \ln p_{t-1} \\ \Delta^2 \ln GDP_{t-1} \\ \Delta \ln cee_{t-1} \\ \Delta \ln M2_{t-1} \\ \Delta \ln csh_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \\ \varepsilon_{4t} \\ \varepsilon_{5t} \end{pmatrix}$$

三个方程的拟合优度分别是 $R_p^2 = 0.998$ ， $R_{GDP}^2 = 0.785$ ， $R_{cee}^2 = 0.979$ ， $R_{M2}^2 = 0.979$ ， $R_{csh}^2 = 0.996$ ，由此可以看出，拟合的效果比较好。

② 1990~2007 年阶段

而对于 1978—1990 年的主要影响指标变动对我国城镇就业人数变动的影响，建立向量自回归模型。从表 4 可以看出，于 $\Delta^2 \ln P$ 、 $\Delta \ln GDP$ 、 $\Delta \ln K$ 、 $\Delta^2 \ln csh$ 、 $\Delta \ln cxc$ 、 $\Delta \ln M2$ 、 $\Delta \ln com$ 是平稳的，可以建立 VAR 模型。

$$\begin{pmatrix} \Delta^2 \ln p_t \\ \Delta^2 \ln csh_t \\ \Delta \ln com_t \\ \Delta \ln cxc_t \\ \Delta \ln GDP_t \\ \Delta \ln K_t \\ \Delta \ln M2_t \end{pmatrix} = \begin{pmatrix} -0.6912 & 0.0592 & -0.0094 & -0.0183 & 0.021 & 0.0011 & 0.0127 \\ 0.1016 & -0.3686 & -0.0147 & -0.0781 & -0.010 & 0.0294 & 0.1201 \\ 2.3825 & -2.629 & 0.03117 & -0.3953 & 0.4783 & 0.037 & -0.1028 \\ 12.5463 & -2.1424 & 0.3287 & 0.6705 & -0.0796 & -0.0572 & 0.1025 \\ 0.2454 & -0.258 & 0.1290 & 0.102 & 0.5097 & -0.048 & 0.1885 \\ -0.1379 & -0.0369 & 0.0252 & -0.0117 & 0.0561 & 0.6554 & 0.0221 \\ -0.2666 & -0.7129 & 0.0415 & -0.0481 & -102246 & 0.1227 & -0.5189 \end{pmatrix} \times \begin{pmatrix} \Delta^2 \ln p_{t-1} \\ \Delta^2 \ln csh_{t-1} \\ \Delta \ln com_{t-1} \\ \Delta \ln cxc_{t-1} \\ \Delta \ln GDP_{t-1} \\ \Delta \ln K_{t-1} \\ \Delta \ln M2_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{t1} \\ \varepsilon_{t2} \\ \varepsilon_{t3} \\ \varepsilon_{t4} \\ \varepsilon_{t5} \\ \varepsilon_{t6} \\ \varepsilon_{t7} \end{pmatrix}$$

6.2.5 模型结果的分析

(1) 主成分回归的结果

由 1978—1989 年的主成分回归模型可以看出，国内生产总值、财政支出、广义货币供给量、城市化水平对我国城镇就业人数的弹性的大小相差不大；同样，对于 1990—2007 年的主成分回归模型，国内生产总值、城市化水平、固定资产投资、广义货币供给量或城乡收入差距以及居民消费水平没变动 1%，都，相比之下，1978—1989 年的弹性系数大一些，国内生产总值、财政支出、货币供给量或城市化水平变动 1%，将会引起城镇就业人数约 0.081% 的变动。

(2) 向量自回归模型的结果分析

① 1978~1989 年阶段

观察 1978—1989 年这一阶段每一指数对我国城镇就业人数的脉冲响应函数图形以

及方差分析图。

[GDP 对就业的影响]

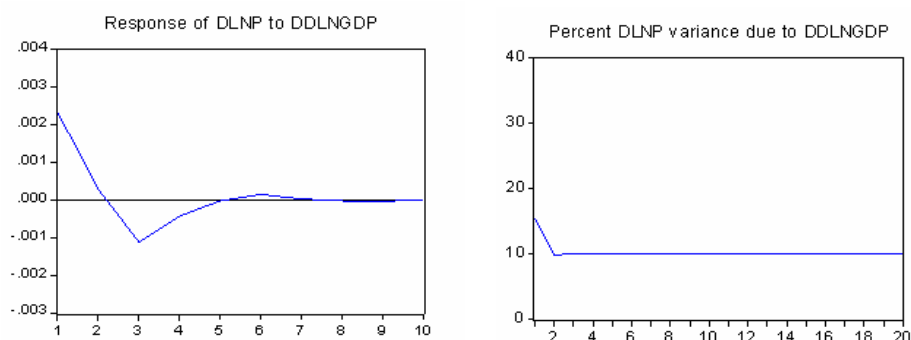


图3 GDP 对就业人数的脉冲响应函数图形及方差分析图

图3左图是GDP增长率变动的冲击引起的我国城镇就业人数的增长率变化的脉冲响应函数。从图中可以看出，当在本期GDP增长率变动受到一个冲击后，在前两期对我国城镇就业人数增长有正向作用，但其滞后作用力度呈下降趋势，在第二年达到最小，短暂恢复到原有水平，在第五年左右恢复到原有的均衡水平，此时达到稳定。这说明，给GDP增长变动率一个冲击，可以引起我国城镇就业人数增长率增加的滞后期为两年，在第二年末，正向冲击很小，几乎为零。从方差贡献率图中也可以看到这种情况。不考虑就业人数本身的贡献率，GDP增长速度对城镇就业人数增长的贡献在当期最大，达15%，之后呈下降趋势，两期后稳定在10%左右的水平。

[城市化水平对就业的影响]

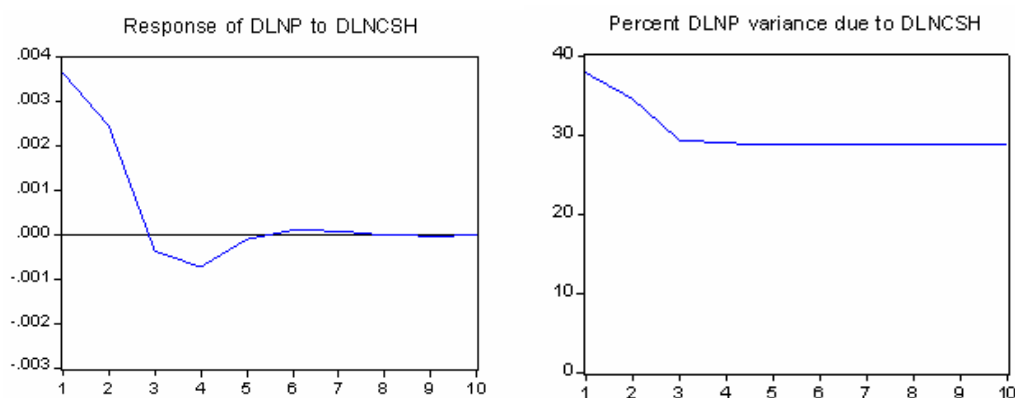


图4 城市化水平对就业人数的脉冲响应函数图形及方差分析图

图4左图是城市化的变化冲击对我国城镇就业人数增长率的影响，从图中可以看出，本期给城市化一个正冲击后，前两期对我国城镇就业人数的增加有正向冲击，但逐渐减少，受到冲击后我国城镇就业人数增长率在五期左右恢复到原来水平。从城市化对就业人数增加的方差贡献率来看，当期最高，达38%，之后虽有下降，但下降幅度并不大，且在三期后达到均衡，维持在30%的水平。

[货币供给量对就业的影响]

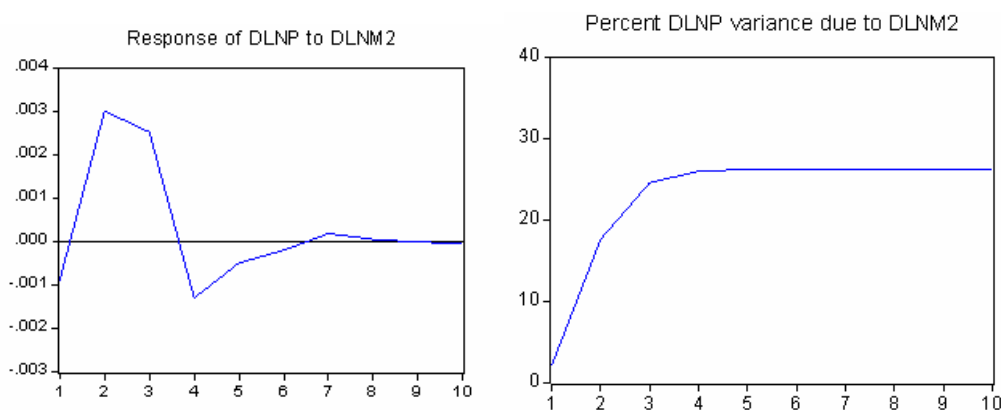


图5 货币供给量对就业人数的脉冲响应函数图形及方差分析图

从左边的脉冲响应函数图中可以看出，给货币供给量增长率一个冲击后，当期对就业的影响不大，其作用在两期后达到最大，这一点从货币供给量对对我国城镇就业人数变动的方差贡献率中也可以看出，当期的方差贡献率小于 5%。但其滞后效应比较大，可以引起我国城镇就业人数增长率增加的滞后期为未来三年且增长幅度比较大。从其方差贡献率中可以看出，货币供应量对城镇就业人数的贡献率是呈指数函数形式递增的，在第三期达到稳定状态，约为 27%。

[财政支出对就业的影响]

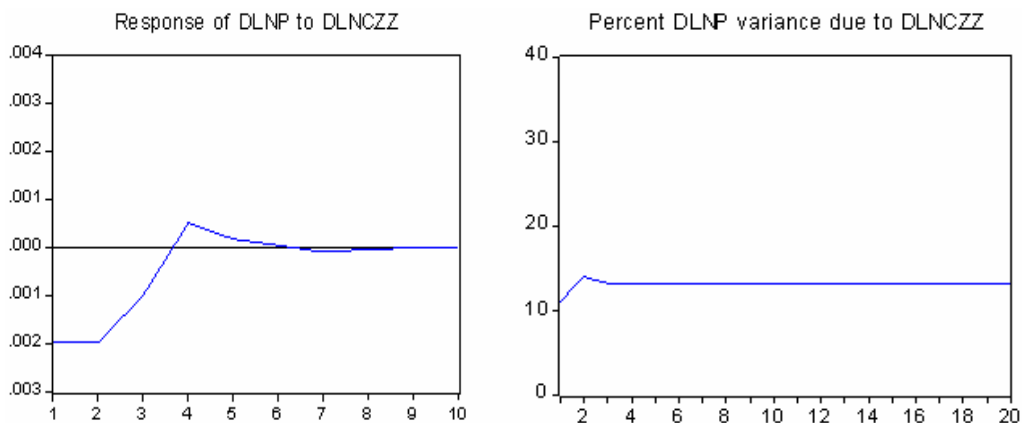


图6 财政支出对就业人数的脉冲响应函数图形及方差分析图

从左图财政支出增长率对城镇就业人数增长率的脉冲图形中看到，当期给财政支出增长率一个反向冲击后，两期之后通过各路径逐步反映到就业市场，同时，它对城镇就业人数变动的贡献率波动不大，在 13% 左右。

从以上图中可以看出，当期给各个影响指标一个冲击，其对城镇就业人数的滞后效应一般在三期左右，影响波动最大的是城市化指数。不考虑城镇就业人数自身的贡献率，各个指标对城镇就业人数变化的贡献率情况是：*GDP* 的贡献率最小，约为 10%，其次是财政支出，城市化指数与货币供给量对城镇就业人数变动的方差贡献率稳定时相差不多，但总体上说，城市化指数的方差贡献率要大些。

②1990~2007 年阶段

观察 1990—2007 年这一阶段的脉冲响应函数图形以方差分解图。

[城市化水平对就业的影响]

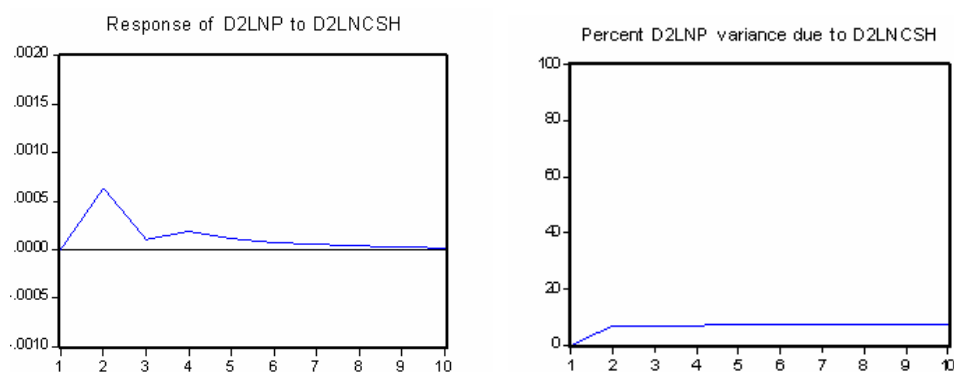


图7 城市化水平对就业人数的脉冲响应函数图形及方差分析图

图7左图是对城市化的冲击引起的城镇就业人数增加率变化的脉冲响应函数图。从图中我们可以看出，当在本期给城市化一个正冲击后，城镇就业人数在第2期达到最高点；从第4期以后开始趋于稳定。这表明城市化受外部条件的某一冲击后，经过市场传递给就业人数，给就业带来同向的冲击，冲击效应在第2期时达到最大，即城市化的正向冲击对就业具有显著的促进作用，并且这一显著促进作用具有较长的持续效应。

[居民消费水平对就业的影响]

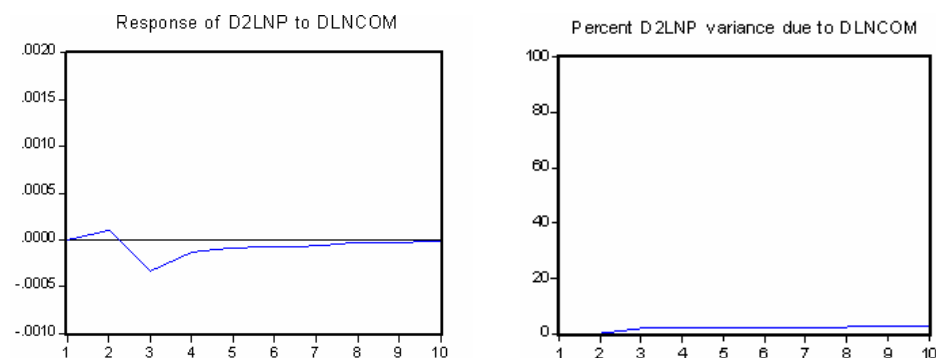


图8 居民消费水平对就业人数的脉冲响应函数图形及方差分析图

从左图的脉冲响应曲线可以看到，当期居民消费水平的变动受到一个单位的冲击后，对就业的作用不是很明显的，而且滞后影响持续时间不长，且对就业人数变动的方差贡献率很低。

[GDP对就业的影响]

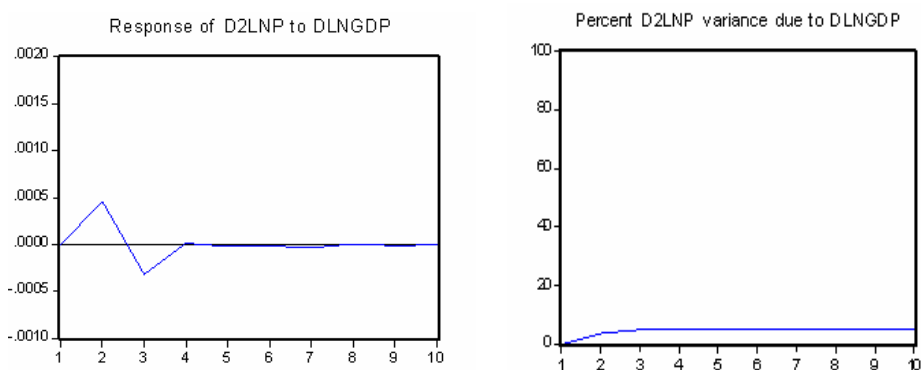


图9 GDP对就业人数的脉冲响应函数图形及方差分析图

从左图的脉冲响应图形可以看出，国内生产总值变动率当期受到一个冲击后，对我国就业人数变动的滞后影响在前两期其作用力度呈上升趋势，在第二年达到最大，然

后滞后影响开始下降，经过两期恢复到原有水平。从方差贡献率来看，国内生产总值变动率最多能解释大约 8% 的方差。

[货币供给量对就业的影响]

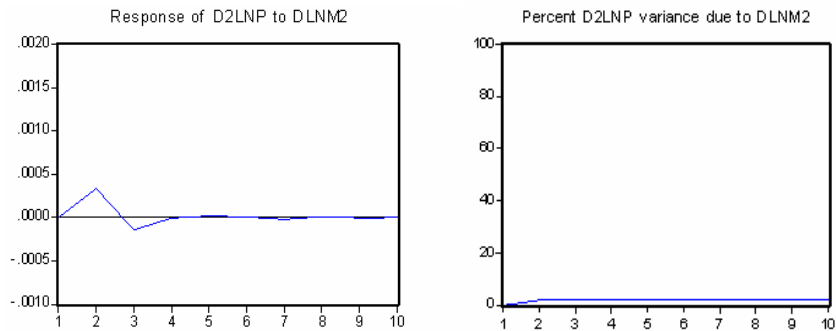


图 10 货币供给量对就业人数的脉冲响应函数图形及方差分析图

结合图 8、图 9、图 10，可以看出这三者的脉冲响应函数图具有一定的相似性。从图形上看，三者对影响就业人数变动具有相似的作用模式，滞后期不长，且有波动趋势相似，其中影响幅度波动最大的是国民生产总值，其中货币供给量的方差贡献率最小。

[城乡收入差距对就业的影响]

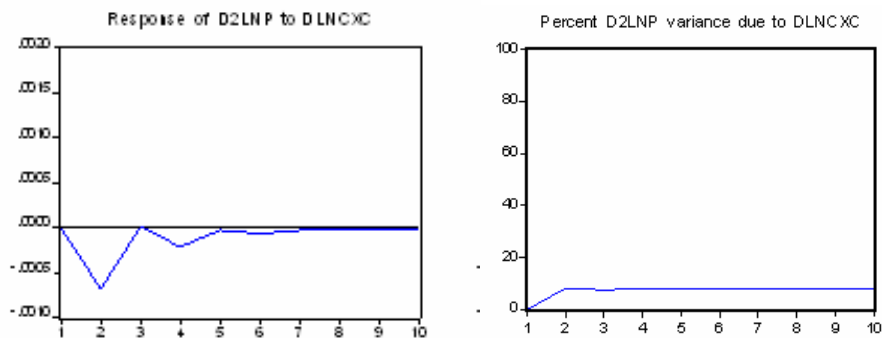


图 11 城乡收入差距对就业人数的脉冲响应函数图形及方差分析图

看左图当期城乡收入差距受到一个单位的冲击后，对城镇就业人数的影响，当期城乡收入差距受到一个单位的正向冲击后，在一定时间内对城镇就业人数的增加有阻碍作用，这种影响将持续大约 5 期，其方差贡献率在两期后维持在 10% 左右的水平。

[固定资产投资总额对就业的影响]

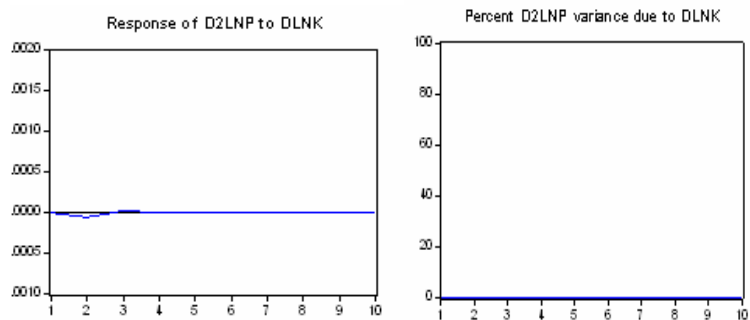


图 12 固定资产投资总额对就业人数的脉冲响应函数图形及方差分析图

从图 12 中可以看出，固定资产投资总额的变动受到一个冲击后，滞后作用不明显，

且其方差贡献率很低，几乎为零。

从上面的几个图中我们可以看出，不考虑城镇就业人数自身的贡献率，城乡收入差距及城市化对就业人数的贡献率较大，在 8% 左右，其次是 *GDP*，固定资产对城镇就业人数的贡献率最小。从中还可以看出以下几个方面特点：居民消费虽然对城镇就业人数的贡献较小，但它对城镇就业的拉动作用将逐渐加强，这一判断来自于居民消费结构、消费观念正发生着相当大的变化。对家庭而言，消费结构的变化可能会引起以下两种后果：第一，家庭负债额可能增大；第二，家庭将寻找增加收入的门路，如兼职、加班、主妇就业等，这两种后果都会在一定程度上是城镇就业人数上升。

综合两个阶段的特点，我们发现，城市化水平在 1978—1989 年及 1990—2007 年这两个阶段对我国城镇就业人数的变化有较大影响，究其原因，城市化水平的提高将急剧扩大消费需求，有力地拉动经济的快速增长，扩张城镇就业总量。城市化使城市人口规模不断地扩大，同时也会使生活消费和生产消费的规模不断扩大，这使许多行业能够获取一定的规模收益，而且也将催生一些新的产业和行业。城市化促进社会分工和职业细化，从而广开就业门路，扩张城镇就业总量，同时社会分工的深化和职业的细化，有力地带动了各行各业的产生和发展，从而提供大量的就业机会。此外，城市化程度的进一步提高还可以提高就业质量、优化产业结构、增强就业弹性、完善就业体系等，这些也间接的促进了城镇就业。

6.3 问题 3：具体分类就业人口模型

6.3.1 分区域的城镇就业人口模型

(1) 建模思路

本文参考上面的城镇就业人口估计模型，并由于受到数据的限制，选取各个省份影响城镇就业人数的固定资产投资总额、城市化水平、城乡收入差距、居民消费水平、货币供给量 *M2*、国内生产总值这六个指标，进行面板数据分析，进而从中观角度中的不同区域，构建分区域的就业人口模型。

(2) 建模准备

① 我国区域的划分

对于我国区域的划分，本文按照国家统计局的划分方法，我国 31 个省、直辖市、自治区（香港、澳门、台湾除外）从地理位置上可分为东、中、西部三个经济带，其中东部经济带包括北京、天津、辽宁、河北、山东、江苏、浙江、福建、广东、海南、广西等 12 个省区；中部经济带包括黑龙江、吉林、山西、安徽、江西、河南、湖北、湖南、内蒙古等 9 个省区；西部经济带包括新疆、四川、重庆、西藏、云南、青海、甘肃、宁夏、陕西、贵州等 10 个省区。而西部大开发战略中所定义的西部地区包括内蒙古、广西、宁夏、西藏、新疆等 5 个少数民族自治区和陕西、甘肃、青海、云南、贵州、四川、重庆等 7 个省区。并且为了便于分析，我们将重庆数据一概并入四川来进行计算。

② 指标数据的选取

由于受到数据的限制，本文主要对 1995-2007 年我国的省际面板数据进行分析，并且这段时间属于我国就业人口变化 1990-2007 的这个阶段。因此，本文认为在这个阶段的基础上，文章应选取各个省份固定资产投资总额、城市化水平、城乡收入差距、居民消费水平、货币供给量 *M2*、国内生产总值这六个指标，来对城镇就业人口进行分析。

(3) 模型的建立与求解

① 模型的初步建立与分析

要根据我国省际面板数据来构建精确的分区域就业人口模型。我们先根据固定资产

投资总额、城市化水平、城乡收入差距、居民消费水平、货币供给量 $M2$ 、国内生产总值这六个指标，对我国东、中、西三大区域进行初步的估算，进而分析各变量的影响程度。此处我们采用固定效应来对面板数据进行分析，三大区域的初步估算模型如表 5 所示：

表 5 三大区域面板数据的初步回归建立结果

	东部	中部	西部
$M2$	0.3412*	0.0751*	0.2725*
	(7.8514)	(2.1323)	(8.2583)
czz	-0.0578	-0.2556*	0.0600
	(-1.4576)	(-3.7122)	(1.5247)
com	-0.8868*	-0.3384*	-1.6073*
	(-8.3712)	(-2.2251)	(-7.2795)
csH	-0.2509*	-0.8774*	-0.6758*
	(-2.4797)	(-10.3981)	(-7.4980)
cxc	-0.0797	0.0422	-0.1960*
	(-0.8073)	(0.4130)	(-1.7159)
K	0.7258*	0.3775*	0.8452*
	(12.0981)	(5.8208)	(14.9040)
常数	7.9146*	9.3568*	11.9288*
	(12.4248)	(10.7923)	(12.5449)
$Adj-R^2$	0.9121	0.7933	0.9146
N	12	9	9

注：表中括号内的数字为 t 检验值。*代表显著性水平为 10%

② 模型的最终建立与求解

根据上表三大区域城镇就业人口模型的初步估算，我们剔除不显著的变量，进而可以得到更加精确的估计模型，如表 6 所示：

表 6 三大区域面板数据的最终回归建立结果

	东部	中部	西部
$\ln M2$	0.3825*	0.0712*	0.2895*
	(9.4504)	(2.1075)	(7.2012)
$\ln czz$		-0.2409*	
		(-4.1017)	
$\ln com$	-0.9759*	-0.33824*	-1.5055*
	(-9.9181)	(-2.2052)	(-7.7505)
$\ln csH$	-0.1672*	-0.8766*	-0.5949*
	(-1.7326)	(-10.4379)	(-4.4199)
$\ln cxc$			-0.2231*
			(-1.8406)
$\ln K$	0.6893*	0.3841*	0.8549*
	(11.7709)	(6.1374)	(13.4679)
常数	7.3064*	9.4089*	12.2991*
	(12.2829)	(11.0178)	(11.0702)
$Adj-R^2$	0.9095	0.7951	0.9304
N	12	9	9

注：表中括号内的数字为 t 检验值。*代表显著性水平为 10%

其中具体分区域的城镇就业人口模型如下：

东部地区： $LnP = 7.3064 + 0.3825 \times \ln M2 - 0.9759 \times \ln com - 0.1672 \times \ln csh + 0.6893 \times \ln K$ ；
 中部地区： $LnP = 9.4089 + 0.0712 \times \ln M2 - 0.2409 \times \ln czc - 0.3382 \times \ln com - 0.8766 \times \ln csh$ ；
 西部地区： $LnP = 12.2991 + 0.2895 \times \ln M2 - 1.5055 \times \ln com - 0.5949 \times \ln csh - 0.2231 \times \ln inc + 0.8549 \times \ln K$ 。

6.3.2 分行业的城镇就业人口模型

(1) 建模思路

对于分行业建立城镇就业人口模型，由于我国行业划分较多，且不同年份有关指标变量的统计口径不一致等原因，我们从国有经济行业、集体经济行业和其他经济行业三方面，并参照之前的有关指标来进行简要的分析，然后运用主成分回归的方法来构建分行业的城镇就业人口模型。

(2) 建模准备

本文着重从国有经济行业、集体经济行业和其他经济行业这三方面来进行分析，而对于影响所选取行业的指标的，我们进行了以下的划分：

对于国有经济行业，本文认为影响国有经济行业就业人数的因素主要有以下几种：国有经济行业职工工资 x_1 、国有经济行业固定资产投资 x_2 、广义货币供给量 $M2$ 、城市化水平 csh 、城乡收入差距 cxc 以及居民消费水平 com 。

对于集体经济行业，本文认为影响集体经济行业就业人数的因素主要有以下几种：集体经济行业职工工资 y_1 、国有经济行业固定资产投资 y_2 、广义货币供给量 $M2$ 、城市化水平 csh 、城乡收入差距 cxc 以及居民消费水平 com 。

对于其他经济行业，本文认为影响其他经济行业就业人数的因素主要有以下几种：其他经济行业职工工资 z_1 、国有经济行业固定资产投资 z_2 、广义货币供给量 $M2$ 、城市化水平 csh 、城乡收入差距 cxc 以及居民消费水平 com 。

而对于时间区间范围的考虑，本文根据指标数据的可得性以及准确性将其指定为 1984-2007 年。

(3) 模型的建立与求解

为了研究分行业对就业数量的影响，我先对各类企业人员的时间序列进行分析。

表 7 历年来各企业就业人数

年份	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995
国有	8637	8990	9333	9651	9984	10108	10346	10664	10889	10920	10890	10995
集体	811	967	1092	1207	1426	1557	1681	1866	2109	2592	3245	3931
其他	37	41	55	72	97	132	164	216	282	536	748	877
年份	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
国有	10949	10766	8809	8262	7715	7168	6621	6621	6438	6232	6170	6148
集体	4302	4512	5331	5774	6262	6867	7667	8678	9814	11283	13014	15595
其他	942	1085	1628	1951	2274	2597	2920	2920	3287	3849	4264	4595

注：数据来源于国家统计局网站

① 国有经济行业就业人口模型

通过观察国有经济行业就业人数趋势图（见图 13），我们以 1995 年为界把国有经济行业就业人口时期划分为 1984-1995 年和 1996-2007 年两个时期，再分别进行主成分回归分析，得出相应模型：

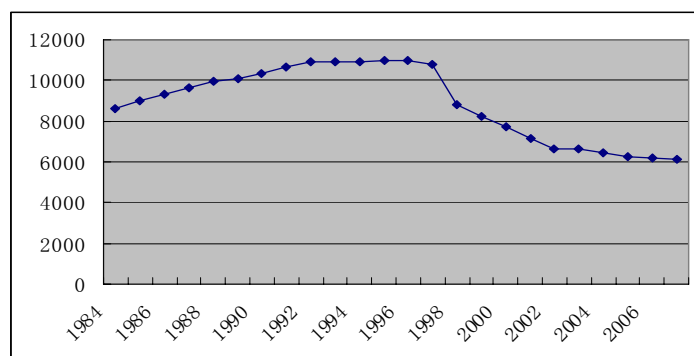


图 13 国有经济行业就业人口趋势图

本文对国有经济行业选取的六个指标提取主成分：第一个主成分的方差贡献率为 98.55%， $\lambda_1 = 5.913$ 这里我们提取一个主成分，即为影响国有经济行业就业人口数量因素的综合值，表达式为

$$F = 0.416 \times \ln x_1 + 0.404 \times \ln x_2 + 0.410 \times \ln x_3 + 0.407 \times \ln x_4 + 0.409 \times \ln x_5 + 0.409 \times \ln x_6$$

我们运用这个综合值序列分别对国有经济行业所划分不同阶段的就业人数进行拟合，其结果为：

$$\ln P_{84} = 7.8857 + 0.0815 \times F$$

$$\ln P_{95} = 13.4496 - 0.2338 \times F$$

其中 $R_{84}^2 = 0.869$ ， $R_{95}^2 = 0.878$ 拟合效果较好，且综合指标的系数显著。

因此我们可以进一步求出不同时期国有经济行业就业人口模型如下：

$$\begin{aligned} \ln P_{84} = & 7.8857 + 0.0335 \times \ln x_1 + 0.0329 \times \ln x_2 + 0.0334 \times \ln x_3 + 0.0332 \times \ln x_4 \\ & + 0.0334 \times \ln x_5 + 0.0333 \times \ln x_6 \end{aligned}$$

$$\begin{aligned} \ln P_{95} = & 13.4496 + 0.096 \times \ln x_1 + 0.0945 \times \ln x_2 + 0.0959 \times \ln x_3 + 0.0951 \times \ln x_4 \\ & + 0.0957 \times \ln x_5 + 0.0955 \times \ln x_6 \end{aligned}$$

② 集体经济行业就业人口模型

通过观察集体经济行业就业人数时序图，我们可以发现其大体上呈现出逐年递减的趋势。因此，我们可以直接进行主成分回归分析，并得出相应模型。

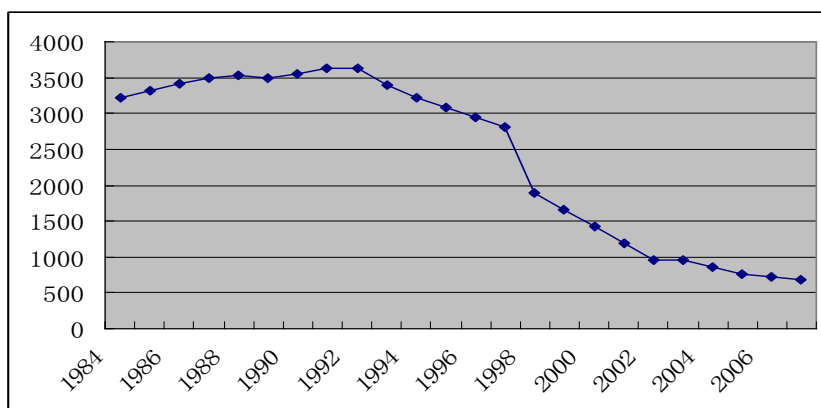


图 14 集体经济行业就业人口趋势图

本文对国有经济行业选取的六个指标提取主成分：第一个主成分的方差贡献率为 96.42%， $\lambda_1 = 5.785$ 这里我们提取一个主成分，即为影响集体经济行业就业人口数量因素的综合值，表达式为：

$$F = -0.389 \times \ln x_1 + 0.413 \times \ln x_2 + 0.406 \times \ln x_3 + 414 \times \ln x_4 + 0.414 \times \ln x_5 + 0.413 \times \ln x_6$$

我们运用这个综合值序列对我国 1984—2007 年集体经济就业人数进行拟合，结果为：

$$\ln P = 13.0229 - 0.0516 \times F$$

其中 $R_{84}^2 = 0.803$ ，拟合效果较好，且综合指标的系数显著。

因此我们可以进一步求出不同时期集体经济行业就业人口模型如下：

$$\begin{aligned} \ln P = & 13.0229 + 0.02 \times \ln x_1 - 0.02 \times \ln x_2 - 0.02 \times \ln x_3 - 0.021 \times \ln x_4 \\ & - 0.021 \times \ln x_5 - 0.021 \times \ln x_6 \end{aligned}$$

③其他经济行业就业人口模型

通过观察其他经济行业就业人数时序图，我们可以发现其大体上呈现出逐年递增的趋势。因此，我们可以直接进行主成分回归分析，并得出相应模型。

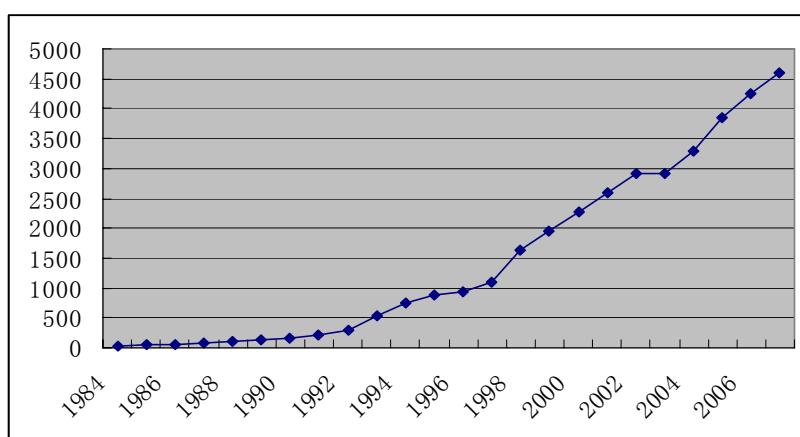


图 15 其他经济行业就业人口趋势图

本文对其他经济行业选取的六个指标提取主成分：第一个主成分的方差贡献率为 98.63%， $\lambda_1 = 5.918$ 这里我们提取一个主成分，即为影响其他经济行业就业人口数量因素的综合值，表达式亦为：

$$F = -0.389 \times \ln x_1 + 0.413 \times \ln x_2 + 0.406 \times \ln x_3 + 414 \times \ln x_4 + 0.414 \times \ln x_5 + 0.413 \times \ln x_6$$

我们运用这个综合值序列对我国 1984—2007 年其他经济就业人数进行拟合，结果为：

$$\ln P = 7.8884 + 0.8137 \times F$$

其中 $R_{84}^2 = 0.96$ ，拟合效果较好，且综合指标的系数显著。

因此我们可以进一步求出不同时期集体经济行业就业人口模型如下：

$$\begin{aligned} \ln P = & 7.8884 + 0.328 \times \ln x_1 + 0.331 \times \ln x_2 + 0.332 \times \ln x_3 + 0.329 \times \ln x_4 \\ & + 0.331 \times \ln x_5 + 0.331 \times \ln x_6 \end{aligned}$$

6.3.3 分人群的就业人口模型

(1)模型建立的思路

对于分人群建立就业人口模型，由于我国就业人群有多种，且不同的分类有着不同的标准，现阶段对于就业的关注主要体现在农民工与大学毕业生这两大人群上。这里我们讨论影响这两大人群的就业情况的指标，我们将人群分为三类：农民工、大学毕业生及其他。

(2) 模型的准备

根据产业就业互动理论,产业结构调整必然会使劳动力结构和技术结构出现一系列变化,促使劳动力就业产生新组合,由于农民工在很大程度上受产业结构的影响,分析影响农民工人数的因素就要考虑产业结构这一指标。选取 2001—2007 年外出务工的农民工人数,在之前所选取的主要影响因素的基础上引入产业结构、城乡收入差距两个指标,计算各项指标与我国就业人数的灰色关联度 $\gamma_{\text{关联度}}$ 及相关系数 $\gamma_{\text{相关系数}}$,根据灰色关联度及相关系数构建排序指标 γ ,然后根据这个排序指标的大小对各个指标进行排序,依此选取影响我国就业人数的指标。

(3) 模型的建立与求解

①农民工就业模型

根据排序结果,选取 5 个指标:净出口 $\ln net$ 、产业结构 str 、城市化水平 csh 、城乡收入差距 $\ln cxc$ 、居民消费 $\ln com$ 等。对选取的 5 个指标提取主成分:第一个主成分的方差贡献率为 96.275%,且 $\lambda = 4.184$,提取一个主成分,表达式为

$$F = 0.439 \times \ln net_1 + 0.445 \times str + 0.448 \times csh + 0.449 \times \ln cxc + 0.454 \times \ln com$$

运用这个综合值序列对我国 2001—2007 年农民工人数进行拟合,结果为:

$$\ln p_{\text{农民工}} = 10.899 + 0.01 \times F$$

其中 $R^2 = 0.9744$ 拟合效果较好,且综合指标的系数显著。

因此我们可以进一步求出城镇就业人口模型如下:

$$\ln p_{\text{农民工}} = 10.899 + 0.0044 \times \ln net_1 + 0.0045 \times str + 0.0045 \times csh + 0.0045 \times \ln cxc + 0.0045 \times \ln com$$

②大学毕业生就业人数模型

根据排序结果,选取 4 个指标:净出口 $\ln net$ 、产业结构 str 、城市化水平 csh 、城乡收入差距 $\ln cxc$ 等。

本文对影响大学毕业生就业人数的四个指标提取主成分:第一个主成分的方差贡献率为 95.602%, $\lambda_1 = 3.842$ 这里我们提取一个主成分,表达式为

$$F = 1.878 \times \ln net + 1.917 \times str + 0.921 \times csh + 1.931 \times \ln cxc$$

我们运用这个综合值序列对我国 2001—2007 年大学毕业生就业人数进行拟合,结果为:

$$\ln P_{\text{毕业生}} = 1.525 + 0.0035 \times F$$

其中 $R^2 = 0.988$,拟合效果较好,且综合指标的系数显著。

因此我们可以进一步求大学毕业生就业人数模型如下:

$$\ln P_{\text{毕业生}} = 1.525 + 0.0066 \times \ln net + 0.0067 \times str + 0.0067 \times csh + 0.0068 \times \ln cxc$$

③其他人群就业人数模型

根据排序结果,选取 6 个指标:国内生产总值 GDP 、城乡收入差距 $\ln cxc$ 、居民消费 $\ln com$ 、净出口 $\ln net$ 、广义货币供给量 $M2$ 以及固定资产投资 K 等。对选取的 6 个指标提取主成分:第一个主成分的方差贡献率为 85.768%,且 $\lambda = 5.164$,提取一个主成分,表达式为:

$$F = 0.436 \times \ln GDP + 0.429 \times \ln cxc + 0.431 \times \ln M2 + 0.436 \times \ln K - 0.249 \times str + 0.428 \times \ln net$$

运用这个综合序列对其他人群就业人数进行拟合,得到:

$$\ln p_{\text{其他}} = 8.748 + 0.049 \times F$$

其中 $R^2 = 0.784$,拟合效果较好,且综合指标的系数显著。

因此我们可以进一步求其他类就业人数模型如下:

$$\ln P_{\text{其他}} = 8.748 + 0.021 \times \ln GDP + 0.021 \times \ln cxc + 0.021 \times \ln M2 \\ + 0.021 \times \ln K - 0.012 \times \ln str + 0.021 \times \ln net$$

6.4 问题 4：就业前景仿真模型

6.4.1 建模思路

为了科学合理对未来城镇就业前景的预测，减小预测模型的误差，本文利用问题二中已建立的数学模型预测结果与神经网络预测结果，得出城镇就业的综合预测值。第一种预测方法我们首先利用神经网络方法模拟仿真出不同模型中各影响因素的数值，将其代入我们得到的数学模型，得到预测结果，第二种方法我们直接利用神经网络得到仿真预测结果。最后我们进行细化，结合国家的有关决策和规划对我国不同行业的 2009 年及 2010 年上半年的就业前景进行仿真预测。

6.4.2 理论的准备

BP网络^[4]，是 1974 年P.Werbos在其博士论文中提出第一个适合多层网络的学习算法，但当时该算法并未受到足够的重视和广泛地应用，直到 20 世纪 80 年代中期，美国加利福尼亚的PDP小组(Parallel Distributed Proccession)于 1986 年发表了专著- *Parallel Distributed Processing*，将该算法应用于神经网络的应用，才使之成为迄今为止最著名的多层网络学习算法—BP算法。由此算法训练的网络称为BP神经网络^[6]。其网络拓扑结构见图 16：

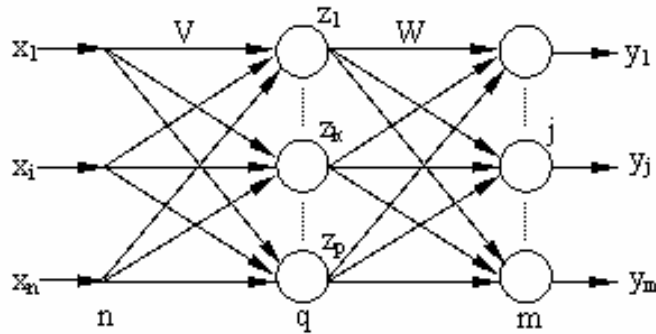


图 16 具有单隐层的 BP 网络结构

设 BP 网络的输入层有 n 个节点，隐层有 q 个节点，输出层有 m 个节点，输入层与隐层之间的权值为 v_{kj} ，隐层与输出层之间的权值为 w_{jk} ，如图 2 所示。隐层的传递函数为 $f_1(\cdot)$ ，输出层的传递函数为 $f_2(\cdot)$ ，则隐层节点的输出为（将阈值写入求和项中）：

$$z_k = f_1\left(\sum_{i=1}^n v_{ki} x_i\right) \quad k = 1, 2, \dots, q$$

输出层节点的输出为：

$$y_j = f_2\left(\sum_{k=1}^q w_{jk} z_k\right) \quad j = 1, 2, \dots, m$$

BP 网络完成 n 维空间向量对 m 维空间向量的映射。

根据以上原理，可以将人工神经网络计算过程归纳为如下几步：

- (1) 初始值选择 $w(0)$
- (2) 前向计算，求出所有神经元的输出： $a^k(t)$
- (3) 对输出层计算 δ ： $\delta_j = (t_j - a_j) a_j (1 - a_j)$

- (4) 从后向前计算各隐层 δ : $\delta_j = a_j(1 - a_j) \sum_i w_{ji} \delta_i$
- (5) 计算并保存各权值修正量: $\Delta w_{ij} = -\eta \delta_j a_i$
- (6) 修正权值: $w_{ij}(t+1) = w_{ij}(t) + \Delta w_{ij}$
- (7) 判断是否收敛, 如果收敛则结束, 不收敛则转至步骤 (2)。

6.4.3 仿真模型的建立与求解

(1) 间接仿真法

首先, 我们利用神经网络方法预测出影响城镇就业人口的 6 个因素的数值, 结果见表 8:

表 8 影响城镇就业人口的因素预测值

年份	<i>GDP</i>	<i>czz</i>	<i>M2</i>	<i>csh</i>	<i>cxc</i>	<i>com</i>
2008	55238.3	58400.89	102638.97	47.69	2010.97	1493.83
2009	61195.3	68847.27	118415.53	49.42	2194.05	1612.51
2010	67794.71	81162.25	136617.09	51.22	2393.8	1740.63

然后根据上述预测值, 并结合问题二中得到的城镇就业人口模型:

$$\ln \hat{p} = 0.3305 + 0.0174 \times \ln GDP + 0.0173 \times \ln czz + 0.0175 \times \ln M2 \\ + 0.0172 \times \ln csh + 0.0173 \times \ln cxc + 0.0174 \times \ln com$$

得到城镇就业人口的预测值, 见表 9:

表 9 城镇就业人口预测值 (万人)

年份	2008	2009	2010
$P_{\text{城}}$	34525.72	34910.53	35299.57

(2) 直接仿真法

利用神经网络方法对城镇就业人口进行预测, 得到预测结果见表 10:

表 10 神经网络城镇就业人口预测值 (万人)

年份	2008	2009	2010
$P_{\text{城}}$	30299.72	31336.73	32409.24

为了更加准确得对就业人口进行预测, 减小预测误差, 增加预测精确度, 以下利用组合预测原理综合考虑上述两种预测方法, 将两种方法的预测结果进行加权平均, 得到组合预测结果, 见表 11:

表 11 组合预测城镇就业人口预测值 (万人)

年份	2008	2009	2010
$P_{\text{城}}$	32412.72	33123.63	33854.41

6.5 问题 5: 相关建议方案

从本文的分析中可知, 影响我国城镇就业人口数量的指标主要是城市化水平、城乡居民收入差距、居民消费、货币供给量、国内生产总值以及固定资产投资额等六项指标。以下我们针对这几方面进行逐一的讨论。

6.5.1 经济发展与就业

首先, 根据上面问题分析得出的模型, *GDP*对就业人数变动的弹性是0.0174, 我们

认为经济增长是解决就业的根本出路，努力实现促进经济增长与扩大就业的良性互动。

农民工与我国的产业结构与区域结构发展结构有关。在珠三角长三角以及一些沿海城市，它们是富裕的城市同时又是农民工聚集的城市，所以因为经济的发展促进了企业的成长，因为劳动密集型企业的成长吸纳了劳动力的增加，因为劳动力的增加成就了城市的发展，所以在经济高速发展中农民工就业不是问题。正如深圳为代表的珠三角地区在中国经济高速发展的时候在深圳的外向型企业劳动密集型企业中缺少的最多的就是劳动力。但当经济发展形式开始出现转变与放缓后，中国的农民工就业问题再次浮出水面，如2008年我国出现农民工提前返乡潮现象，主要就是受到国际金融危机对我国经济发展的冲击影响。

从经济发展的角度出发，我们提出如下建议：经济危机发生之后，农民工的就业方向应该在党中央提出的中部崛起西部大开发和现在的4万亿经济刺激方案下逐步向中西部欠发达地区转移。因为在内陆欠发达地区，建设和发展经济所带动GDP1个百分点比沿海发达地区带动1个百分点所增加的劳动力要多20%左右，同时在内陆城市可以减少企业的用工成本和工业用地成本还可以得到当地政府更多的支持。当在这些城市的企业开始吸纳周边的劳动力的时候就开始带动当地的经济发展与建设，这样就逐步带动了现在中国经济中最困难的内需不足问题，同时在带动内需的时候又增加了就业的机会。

经济发展对大学生就业也有重大影响，经济的发展要求企业需要更多的高度专业化和高度技术型人才，我们认为建议学校应该在了解分析市场和行业后有计划的开设一些相应的课程。这既有利于高校与学生的就业率同时也有利于企业的技术进步和产能提升。这样可以进一步扩大企业规模、增加就业。

6.5.2 财政政策与就业

积极的财政政策对完善和实施积极就业政策具有重要意义，从我国国情看，财政支出政策在很大程度上缓解了就业压力，我们认为可以通过财政拨款兴修水利、进行基础设施等工程、项目的投资，创造大量的就业机会；增加再就业培训、职业介绍服务等方面的财政支出，缩短工人寻找工作的时间，增强工人在不同岗位的就业适应能力等等。

6.5.3 货币政策与就业

根据我国实际状况，当前最佳战略选择是发展中小企业来吸纳劳动力，而这就需要制度上保证中小企业的营商环境，这就需要货币政策和就业政策共同发挥作用的最好组合空间内寻求到较好的解决途径。我们认为实现低失业率不能单纯依靠货币供应量的增加和财政支出的扩大，我们提出如下建议：开辟适合中小企业发展的专门融资渠道；开拓新的融资手段，为中小企业提供必要的信贷支持；加快中小企业直接融资步伐，让中小企业进入资本市场。

6.5.4 城市化与就业

城市化进程是城市在国家经济和社会生活中的主导作用不断增强的过程，它使人口进一步向大城市集中，这一规律不仅表现在大城市数量和规模的更快增长上。它对就业的影响一方面使人们的消费观念发生变化，使一些新兴产业得到进一步发展的机会，从而提供更多的就业机会；另一方面，它会使农村剩余劳动力发生转移，从而增加了就业人数。

我国的城市化水平不仅低于世界平均水平，而且与我国的经济发展和工业化水平不一致。造成这一现象的原因主要是片面发展重工业的产业政策、落后的农业生产率以及我国城市发展政策对城市化得制约。解决这个问题相关建议是，在政策上，要为农村人口向城市人口的转移扫除障碍，同时，在增加城市数量，扩大城市规模的同时，要注意提高城市化的质量水平，在模式上，要使大中城市和小城镇协调发展。

6.5.5 城乡收入差距与就业

城乡收入差距对就业的影响一方面是通过影响消费从而影响就业，另一方面是通过影响城市化进度来影响就业。我国城乡居民收入差距呈现出扩大的趋势，导致这中现象重要原因是二元经济结构、人力资本及物质资本的差异。二元经济结果的核心问题是农村剩余劳动力的转移问题，通过向城市工业转移来消除农村劳动力剩余，并引起产业结构的调整而提高农村的劳动生产率已达到减小城乡收入差距；提高农业生产的技术水平，缩小城乡生产的技术水平差异；加强农村人力资本投资，缩小城乡的人力资本差异；加强物质投入，提高农村居民的平均产出。

§7 模型的误差分析、检验及进一步讨论

7.1 误差分析

7.1.1 考虑影响因素的选取方面

在指标选取模型中，我们在国家财政政策影响因素的确定上，考虑到在数据上，利率因素的确定并非是按照年度时间制定。因此并未将此因素引入模型中。但是我们通过定性分析可知利率的影响对就业人数是有一定作用的，因此在模型的建立过程中会产生一定的误差。

当利率发生变动时，可能会间接导致就业人数变化，因为利率与人们手中握有的实际财富量密切相关，会引起人们就业取向变化。对于不同行业、不同就业人群等，其作用机制与作用大小是不同的。

7.1.2 考虑缺省值

在数据的选取上，因为某些指标部分年份数据缺省，我们利用牛顿插值法进行差值补全数据，我们可以通过相应时刻拟合模型的拟合值来判断误差大小。

7.1.3 考虑我国区域的划分

在问题三中不同地区的就业人口模型中，我们分为中部、东部和西部三地区，没有考虑东北三省老工业基地，模型的精度可能会受影响。

7.2 模型的检验

本文要解决的四个问题中多元线性回归模型、主成分回归模型、向量自回归模型、灰色预测模型和神经网络预测模型在建立的过程中通过了相应软件的检验，具有一定的合理性。

7.3 模型的进一步讨论

本文模型指标的选取只是通过关联系数选取了固定资产投资总额、城市化水平、城乡收入差距、居民消费水平、货币供给量 M2、国内生产总值这六个指标，然而影响就业水平的因素还有很多，本文所构建的影响就业指标体系就包括了 18 个相关指标，诸如：产业结构、市场化指数、知识层次等等，都会对城镇就业人口产生影响。此外，利率、房地产价格、股票等一系列金融指标也同样会对城镇就业人口的变化起到一定的作用。同时，不同行业、地区、人群都存在着不同的特征，其相关影响因素也不同，从而使其相应的就业人口数量也产生不同。

§8 模型的评价与推广

8.1 模型的优缺点

8.1.1 模型的优点：

(1)对于全国就业量的研究，本文以 1990 年为界，对就业人口的影响因素进行了分段考虑。因而，能够从不同时间阶段考虑出对就业量的主要影响因素。

(2)本文利用主成分回归方法，有效避免变量间的多重共线性问题，进而提高了模型的估计精度；

(3)对于影响全国就业量的因素，本文从不同角度，选取了大量指标进行了分析。

(4)本文运用 SPSS、DPS、EViews 等多种数学软件进行计算，取长补短，使计算结果更加准确；

8.1.2 模型的缺点：

(1)在对影响就业量主要因素的选取过程中，主观性较大，极易漏掉其他重要变量，从而不能全面地分析就业量。

(2)建立就业量模型时，对影响就业量的指标我们设定的是一种对数形式，然而现实生活中，其与就业量的相关性并不一定遵守这个规律，因而会产生一定的偏差。

8.2 模型的推广

依据主成分回归的方法我们可以进一步从不同的角度来对就业量的估计进行推广。此处，文章选取农业、工业、建筑业、房地产业相关数据进行进一步分析。

8.2.1 农业行业就业人口模型

本文对农业行业选取的六个指标：支农财政支出 x_1 、农业各税 x_2 、农业总产值 x_3 、农村居民消费水平 x_4 、农村居民恩格尔系数 x_5 、农村居民纯收入 x_6 六项指标，提取主成分：第一个主成分的方差贡献率为 94.41%， $\lambda_1 = 5.665$ 这里我们提取一个主成分，即为影响农业行业就业人口数量因素的综合值，表达式为

$$F = 0.398 \times \ln x_1 + 0.409 \times \ln x_2 + 0.410 \times \ln x_3 + 0.416 \times \ln x_4 - 0.399 \times \ln x_5 + 0.417 \times \ln x_6$$

我们运用这个综合值序列对我国 1990—2007 年就业人数进行拟合，结果为：

$$\ln P_{\text{农}} = 11.907 + 0.189 \times F$$

其中， $R^2 = 0.746$ 拟合效果较好，且综合指标的系数显著，因此我们可以进一步求出农业行业就业人口模型如下：

$$\begin{aligned} \ln P_{\text{农}} = & 11.907 + 0.0752 \times \ln x_1 + 0.0773 \times \ln x_2 + 0.0775 \times \ln x_3 + 0.0786 \times \ln x_4 \\ & - 0.0754 \times \ln x_5 + 0.0788 \times \ln x_6 \end{aligned}$$

8.2.2 工业行业就业人口模型

本文对影响工业行业就业人数的因素：工业企业个数 y_1 、工业总产值 y_2 、利润总额 y_3 、工业贷款 y_4 四项指标提取主成分：第一个主成分的方差贡献率为 95.185%， $\lambda_1 = 3.807$ 这里我们提取一个主成分，即为影响工业行业就业人口数量因素的综合值，表达式为：

$$F = 0.5012 \ln y_1 + 0.5079 \ln y_2 + 0.4941 \ln y_3 + 0.4945 \ln y_4$$

我们运用这个综合值序列对我国 1998—2007 年就业人数进行拟合，结果为：

$$\ln P_{\text{工}} = 4.4489 + 0.2693 \times F$$

其中， $R^2 = 0.669$ 拟合效果较好，且综合指标的系数显著，因此我们可以进一步求出工业行业就业人口模型如下：

$$\ln P_{\text{工}} = 4.4489 + 0.135 \ln y_1 + 0.1368 \ln y_2 + 0.1331 \ln y_3 + 0.1332 \ln y_4$$

8.2.3 房地产行业就业人口模型

本文对房地产业选取的四个指标：房地产企业个数 z_2 、职工平均工资 z_2 、商品房销售额 z_2 、房地产业贷款 z_2 ，提取主成分：第一个主成分的方差贡献率为 97.75%， $\lambda_1 = 3.91$ 这里我们提取一个主成分，即为影响房地产行业就业人口数量因素的综合值，表达式为：

$$F = 0.5193 \ln z_1 + 0.5017 \ln z_2 + 0.5037 \ln z_3 + 0.5023 \ln z_4$$

我们运用这个综合值序列对我国 1997—2007 年就业人数进行拟合，结果为：

$$\ln P_{\text{房}} = 4.2618 + 0.0621 \times F$$

其中， $R^2 = 0.9642$ 拟合效果较好，且综合指标的系数显著，因此我们可以进一步求出房地产业行业就业人口模型如下：

$$\ln P_{\text{房}} = 4.2618 + 0.0322 \ln z_1 + 0.0312 \ln z_2 + 0.0313 \ln z_3 + 0.0312 \ln z_4$$

8.2.1 建筑业行业就业人口模型

本文对建筑业行业选取的四个指标：建筑业生产总值 e_1 、建筑业企业单位数 e_2 、建筑业技术装备率 e_3 、建筑业贷款 e_4 ，提取主成分：第一个主成分的方差贡献率为 85.97%， $\lambda_1 = 3.439$ 这里我们提取一个主成分，表达式为

$$F = 0.5193 \times \ln e_1 + 0.4287 \times \ln e_2 + 0.5134 \times \ln e_3 + 0.5322 \times \ln e_4$$

我们运用这个综合值序列对我国 1995—2007 年就业人数进行拟合，结果为：

$$\ln P_{\text{建}} = 4.751 + 0.1874 \times F$$

其中， $R^2 = 0.781$ 拟合效果较好，且综合指标的系数显著，因此我们可以进一步求出建筑业行业就业人口模型如下：

$$\ln P_{\text{建}} = 4.751 + 0.0973 \times \ln e_1 + 0.0803 \times \ln e_2 + 0.0962 \times \ln e_3 + 0.0997 \times \ln e_4$$

根据以上模型，我们可以进一步对未来几年的数值进行仿真预测，其结果如表 12：

表 12 各行业就业人数预测值

年份	2008	2009	2010
$P_{\text{农}}$	30186.09	31564.21	33004.97
$P_{\text{工}}$	13394.76	14736.04	16211.82
$P_{\text{房}}$	458.0703	468.9532	480.0946
$P_{\text{建}}$	4755.009	4957.898	5169.442

由以上分析，此处仅是对模型的初步推广，更深层次更精确的研究有待进一步讨论，希望能够为政府相关部门提供一定的借鉴。

参考文献

- [1] Hall R, Jone C, Why do some countries produce so much more output per worker than others?[J], Quarterly Journal of Econormics, 1999, 114:83~116;
- [2] 易大义, 沈云宝, 李有法, 计算方法[M], 杭州: 浙江大学出版社, 2002 年 06 月第二版;
- [3] 邓聚龙, 灰色系统理论教程[M]; 武汉: 华中科技大学出版社, 1990;
- [4] 韩力群, 人工神经网络理论、设计及应用[M]; 北京: 化学工业出版社, 2001;
- [5] 高铁梅, 计量经济分析方法与建模, 北京: 清华大学出版社, 2006;
- [6] 吴建国, 数学建模案例精编, 北京: 中国水利水电出版社, 2005;
- [7] 范金, 朱强, 王艳, 中级宏观经济学, 北京: 经济管理出版社, 2007;
- [8] 王振龙, 应用时间序列分析, 北京: 科学出版社, 2007.
- [9] 许毓坤, 股指及宏观经济变量的神经网络仿真建模和预测, 宁波职业技术学院学报, 11 (5): 91-94, 2007.
- [10] 朱加凤, 我国财政货币政策就业效应的实证分析, 学术交流, 6 (6): 91-94, 2009.
- [11] 石为人, 冯治恒, 给予灰色理论与 BP 算法的宏观经济预测模型研究, 计算机与数字工程, 35: 8-10, 2006.

附 录

附件 1：神经网络预测程序

```
a=[3601.993551
3878.653445
4164.90333
4411.28574
4791.098384
5287.171268
6199.307456
7173.238205
7661.689961
8395.433608
9138.826358
8845.917387
9335.404795
10085.64115
11532.43941
13441.36993
14586.26504
15711.52231
16985.10409
18332.29085
19750.60555
21282.8929
23452.59999
25740.39782
28474.63006
31758.97623
35980.56944
40503.94695
46157.65166
51859.21008];
pn=[a(1:10),a(4:12),a(5:14),a(7:16),a(9:18),a(11:20),a(13:22),a(15:24),a(17:26),a(19:28)]
tn=[a(3:12),a(5:14),a(7:16),a(9:18),a(11:20),a(13:22),a(15:24),a(17:26),a(19:28),a(21:30)]
;
[pn1,minpn,maxpn]=premnmx(pn);
[Tn1,minTn,maxTn]=premnmx(tn);
net=newcf(minmax(pn1),[20,10],{'tansig','purelin'},'trainlm','learngdm');
net.trainParam.epochs=3000;
net.trainParam.goal=0.001;
net.trainParam.show=10;
net.trainParam.lr=0.8;
```



```

net.trainParam.mc=0.6;
net=train(net,pn1,Tn1);
rn1=sim(net,pn1);
y=postmnmx(rn1,minTn,maxTn)
rn2=sim(net,Tn1);
yu=postmnmx(rn2,minTn,maxTn)

```

附件 2:

年份	M2 (亿元)	市场化指数	产业结构	城市化水平 (%)	城乡收入差距	恩格尔系数
1978	1121.0474	0.608489	3.558621	17.91521	207.3123	57.5
1979	1278.4747	0.58625	3.207721	18.96106	233.7171	38.13
1980	1522.0655	0.568747	3.323378	19.39111	262.3208	56.9
1981	1828.3408	0.568273	3.145963	20.15649	249.8029	56.7
1982	2132.9467	0.558039	3.00562	21.1305	238.6841	58.6
1983	2472.6928	0.577195	3.026571	21.62356	225.9349	59.2
1984	3082.3444	0.616924	3.123938	23.01427	255.2638	58
1985	3393.192	0.627227	3.52707	23.70691	271.7004	53.3
1986	4738.3836	0.58809	3.691233	24.52491	355.7497	52.4
1987	4537.2574	0.575477	3.733265	25.3193	375.61	53.5
1988	4511.3272	0.561351	3.896711	25.81467	385.9579	51.4
1989	5615.1155	0.531836	3.999338	26.21025	402.0985	54.5
1990	7007.4211	0.59691	3.69701	26.4097	412.0159	54.2
1991	8371.2468	0.592865	4.087622	26.94025	459.3327	53.8
1992	10052.316	0.580209	4.592776	27.45987	532.257	53
1993	11270.4	0.548166	5.032534	27.99008	629.8825	50.3
1994	12257.244	0.484859	4.944318	28.5098	688.5508	50
1995	15700.35	0.472673	4.876019	29.04038	699.1573	50.1
1996	18158.776	0.485981	4.903469	30.47986	695.091	48.8
1997	21123.062	0.525713	5.239712	31.90995	712.6964	46.6
1998	24453.233	0.631539	5.383662	33.35017	763.5837	44.7
1999	28455.151	0.699542	5.670778	34.77971	864.7527	42.1
2000	31819.564	0.750135	6.116139	36.21975	951.8184	39.4
2001	37159.706	0.782765	6.314305	37.65974	1054.731	38.2
2002	43778.677	0.794712	6.525429	39.08978	1236.926	37.7
2003	51727.78	0.810223	6.934635	40.53023	1367.886	37.1
2004	57186.7	0.823029	6.59068	41.76001	1459.492	37.7
2005	66046.047	0.812636	6.112511	42.99	1600.13	36.7
2006	75273.625	0.840636	6.144679	43.9	1779.998	35.8
2007	83846.44	0.855109	6.282623	44.94	2004.581	36.3

