

全国第六届研究生数学建模竞赛



题 目 城镇就业人数影响因素分析

摘 要

就业问题涉及的影响因素众多,仅仅只从数据来判断影响因子是缺乏理论依据的。因此,首先我们对就业理论做了简单的回顾和综述,随后,并没有立即对影响因素进行分析,而是通过对城镇就业人数的时间序列分析,得到“突变点”与重大政策有着很大关系、目标的时间序列具有一定的自我解释能力等结论。

(1)选择具有相关作用的影响因素

通过建立 GARCH(1, 1) 模型,我们得出目标波动受自身特点影响占总原因的 31%,而外部因素的冲击占 69%的结论,因此,对外部因素分析显得很重要。随后我们对之前定性分析总结出的经济指标和劳动指标逐个进行格兰杰因果检验。结果显示有十九个经济指标对目标有解释作用,而目标又对七个经济指标和三个劳动指标有解释作用,即部分宏观经济指标与目标指标间具有相互的因果关系。

(2)建立反映指标与影响因素之间联系的模型

我们以城镇就业人数作为被解释变量,以上述分析中的原因指标作为解释变量,建立多元回归模型,结果显示:经济活动人口与固定资产支出(劳动力供求双方)共同作用后对目标进行了很好的解释,可决系数达到 0.998 以上。考虑到“目标是因素的原因”这部分信息无法被多元回归利用,我们建立就业人数、经济活动人数、居民消费支出、固定资产投资等指标的向量自回归(VAR)模型。

(3)分地区、分行业进一步分析

在分地区的分析中,运用聚类分析方法,将拥有不同就业特点的地区对应应在平面坐标轴上,结果显示各地区就业压力的排布呈现线性变化特征,按照“广东、江苏、山东……”的顺序依次递减。在分行业的分析中,建立 Panel Data 模型对各行业平均就业人数与主营业务收入关系进行分析。随后,我们考虑到分行业指标统计体系具有非线性映射特点,建立 BP 神经网络模型,用来说明目标与指标间的复杂解释关系,并且对 9 个行业进行样本外预测,结果表明仿真效果不错。

参赛队号 k000005

队员姓名 马桢干、陈奕山、伍骏骞

参赛密码 _____

(由组委会填写)

目 录

一、问题重述与分析	3
1.1 问题重述	3
1.2 问题的背景及分析	3
1.2.1 问题提出的背景	3
1.2.2 就业理论的回顾与综述	4
1.2.3 问题的分析	5
二、指标的选择和符号的说明	7
2.1 宏观经济指标和符号说明	7
2.2 劳动力素质指标和符号说明	9
2.3 目标指标和符号说明	10
三、问题的求解	10
3.1 问题一的模型建立与求解	10
3.1.1 问题一的基本假设	10
3.1.2 模型的建立	11
3.1.2.1 对就业人数时间序列的初步分析	11
3.1.2.2 对就业人数外部影响因素的选择	16
3.1.2.3 运用格兰杰因果检验方法对影响因素的选择结果	22
3.2 问题二的模型建立与求解	23
3.2.1 多元回归模型原理	23
3.2.2 运用多元回归模型进行实证分析	23
3.2.3 考虑使用向量自回归(VAR)模型以及该模型的原理	27
3.2.4 运用 VAR 模型对目标与影响因素之间关系进行实证分析	28
3.3 问题三的模型建立与求解	30
3.3.1 对问题三的分析	30
3.3.2 分地区对就业问题进行建模和求解	30
3.3.2.1 分地区对就业问题研究的分析	30
3.3.2.2 运用“星座”图聚类分析法划分地区	31
3.3.2.3 运用聚类分析方法对各地进行划分的结果	34
3.3.3 分行业对就业问题进行建模和求解	37
3.3.3.1 分行业建模的分析	39
3.3.3.2 Panel Data 模型分析原理	39
3.3.3.3 运用 Panel Data 模型进行实证分析	40
3.3.3.4 神经网络模型原理	47
3.3.3.5 使用神经网络模型对分行业系统进行仿真和预测	48
3.3.3.6 将训练好的神经网络模型应用于实际样本外预测	51
3.4 问题四、五的分析	52

3.4.1 对未来就业形势进行的预测	52
3.4.2 对增加就业人口、减少就业压力的政策建议	53
四、本文总结和对该课题研究的进一步思考	53
4.1 本文总结	53
4.2 对该课题研究的进一步思考	54
参考文献	53

一、问题重述与分析

1.1 问题重述

经济增长、充分就业、通货膨胀是宏观经济调控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。

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；从中观层面，不同地区、不同产业也会表现出不同的特征。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。

问题一：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。

问题二：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。

问题三：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。

问题四：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。

问题五：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。

1.2 问题的背景及分析

1.2.1 问题提出的背景

经济增长、充分就业、稳定物价、国际收支平衡是国家宏观经济政策的四大目标，国家之所以将就业问题列入宏观调控的四大目标之一，是因为就业问题是关系到人民切身利益和社会稳定的大事情。实现充分就业，不仅是实现经济可持续发展的必要条件，也是构建和谐社会的重要要求。我国经济的快速增长和劳动力就业情况主要体现为三方面，一、人口的快速增长和年龄结构的转变为我国经济持续发展提供了丰富的劳动力市场；二、过于丰富的劳动力资源是我国最大的

竞争优势，但是同时也带来了巨大的就业压力；三、我国经济的高速增长并没有带来就业的实质增长，引用欧洲经济学家对我国就业现状的看法是“发达国家只需要保持 2%~3% 的经济增长速度就能保持就业的整体稳定，但是在中国，需要将经济增长速度保持在 8%~9% 才能够使满足国内的就业”。^[6]

1.2.2 就业理论的回顾与综述

充分就业是宏观经济调控中四大目标之一，自古就一直受到经济学家的关注，影响就业的因素很多并且作用与传导的机制很复杂，仅仅从数据的相关程度就判断就业与某些宏观经济指标相关联是缺乏理论依据的，不但寻求影响因素的过程有“大海捞针”的感觉，而且做出的判断结果也难以让人信服。因此，在进行数据分析之前，首先需要对经济学史中对就业问题的看法和分析部分进行简单的回顾，为后面进行数据分析奠定一定的理论基础，同时也避免盲目地分析就业影响因素。

西方经济学一般认为：“看不见的手”和“看得见的手”会使经济增长和就业增长两者目标趋于一致。

1. 古典学派观点

亚当·斯密在《国民财富及其增进原因的研究》中最早提出劳动数量和劳动质量的增加会导致国家财富的增长。李嘉图则指出社会财富的增加来源于劳动数量的扩大和劳动生产率的提高。萨伊定律“供给创造需求”表示市场机制下的充分就业可以有效地杜绝大量、长期的非自愿型失业。庇古更进一步认为，如果资本主义条件下有失业的话，也只能是“自愿失业”或“摩擦性失业”，它们与资本主义自由竞争制度下的“充分就业”是不悖的。^{[3][9]}

2. 马克思主义观点

马克思的资本论对就业进行了相应的探讨。首先，技术进步能够为工厂创造更多的利润和剩余产品，有利于扩大再生产和增加就业，个别企业率先运用先进技术来提高劳动生产率获得超额利润，这样会促进全社会的劳动生产率提高，从而降低生产资料的物价水平，增加就业弹性。由于生产率的提高，使生产资料价格便宜，同一可变资本以推动更多的劳动力。其次，科学技术的进步和其设备的普遍使用，会促进相关产业的产生和发展，从而拓宽就业领域。虽然机器应用它的劳动部门排挤工人，但是，它能引起其它劳动部门就业的增加。最后，发展大机器工业能带来更深层的社会分工，扩大工人就业。生产资料和生活资料的增加，较远的将来才能生效的产品(运河、船坞、桥梁)的部门中，劳动扩大了。^[10]

3. 新古典学派观点

凯恩斯认为资本主义经济不具有充分就业均衡的调节机制，他否定了工资率是有充分伸缩性的观点，认为货币工资具有下降刚性。托宾、杜生贝里等人提出“结构性失业问题”，力图用市场结构的变化来解释失业和通胀并发症，认为是微观市场的不完全性和结构变化引起滞胀。弗里德曼提出“自然失业”，自然失业的工人包括：认为实际工资低于劳动力边际效用的自愿失业；源于信息不完全的“寻业的失业”；源于现行劳动力流动性不完全的摩擦性失业；经济结构和技

术发生变化,劳动力流动性和技术水平不相适应的较长期处于失业状态的结构性失业。从古典学派到理性预期学派的失业理论,大多是沿着劳动力市场以外的技术因素和制度因素来探索影响和决定失业的因素。^{[1][5]}新凯恩斯主义在就业理论中维护了凯恩斯主义劳动力市场非出清的信条,克服了凯恩斯主义研究劳动力市场几乎不涉及劳动力市场的致命缺陷。^[11]

4. 劳动力流动观点^{[3][10]}

刘易斯认为在国民经济体系中并存着两大经济部门:一个是传统的自给自足的农业经济部门,另一个是城市现代工业部门。在这种结构中,发展中国家的就业存在特殊的性质。由于发展中国家的传统农业部门庞大,吸收了大量的就业人数,加上资本和土地的有限性,农村必然存在大量的剩余劳动力,即劳动生产率等于或小于零的那部分劳动力。这种情况导致农村劳动力的收入只能维持在生存收入的水平上。刘易斯认为,经济发展依赖于工业部门的扩张,工业通过增加新的资本和扩大生产规模,不断吸收农业剩余劳动力,从而获得更多的利润。依次循环,直至农业剩余劳动力被完全吸收。刘易斯的模型提出以后,拉尼斯对刘易斯的模型进行了修改和扩充。给予了农业部门在经济发展中的合理地位;并比较透彻地分析了农业部门是如何决定和影响工业部门的扩张以及劳动力转移的。乔根森对二元结构模型又做了进一步的发展,它不承认农业有边际生产率为零的剩余劳动的存在,也不认为农业与工业的工资水平是固定不变的。指出人口增长取决于人均粮食供给。如粮食供给是充分的,人口增长率将达到生理最大量。当人均粮食供给增长率大于最大人口增长率时,农业剩余就产生了。农业剩余一出现,农业劳动力就开始向工业部门转移,于是,工业部门就开始增长。托达罗的人口流动模型回答了为什么农村向城市的移民过程会不顾城市失业或隐蔽性失业的存在而继续进行下去的问题,这是主要在于发展中国家中,一个农业劳动者决定他是否迁入城市的决策不仅决定于城乡实际收入差异,而且还取决于就业概率,就业概率在迁移决策中起到十分重要的作用。因此,农民向城市流动,是在市场经济条件下既注意现实又含有预期的理性行为。

1.2.3 问题的分析

在回顾经济学中有关就业的理论的基础之上,我们从就业供给·需求的整体框架出发,总结分析出我国城镇就业压力大体受以下方面因素的影响:

1. 从劳动力供给方来看

(1)人口基数大、增长速度快

由于建国以来我国遗留的人口政策导致了中国在相当长的一段时期内保持了飞快的自然增长速度,造成了就业供给过大的现象。虽然计划生育政策的成功实施成功地遏制了60年代、70年代的人口飙升,但是由于三次生育高峰形成的人口惯性,我国目前处于劳动力供给最为丰富的时期,并且这种趋势还将持续相当长的一段时期。主要表现为:人口增长速度减慢、劳动年龄段的人口占总体人口比重增加、抚养比例下降……尽管丰富的劳动力资源为经济的快速发展创造了良好的条件,但是过重的就业压力为经济的稳定社会的和谐带来了不小隐患。^[6]

(2)城市外来人口的涌入

中国历史中的城乡分割的二元结构管理政策,在很长的一段时间里阻碍了农村剩余劳动人口向城市的转移,使得农村集聚了数量相当大的剩余劳动人口。随

后,伴随着中国改革开放的进展,城市化的进程不断地加快,让农民通过向城镇转移是解决三农问题、促进经济快速增长的一条有效途径,但是城乡之间人力流动管制的放松,导致了农民工进城工作热潮的出现,不仅是受收入的诱导,更重要的是希望定居在城市,一人进城往往会带动亲戚子女的进城,而子女在若干年后又会成为城市劳动力的提供者,但是城市就业岗位供给相对劳动力需求的来说是有限的,另外,我国农村人口比重很大,虽然经过长时间的城市化过程比重减少不少,但是外来人口的涌入会导致城市就业的重大压力。

(3)劳动者技能和偏好

在就业市场上,一方面某些行业存在着劳动力剩余,但是同时另外的不少行业也存在着供不应求的现象。这里面一方面是由于劳动者的技术水平、知识结构或者专业达不到企业的要求,另一方面与劳动者的偏好也有着重要的关系,主要表现为劳动者对行业的歧视和对就业地区的歧视。因此,在劳动市场上整体表现的是供大于求的状况,但是如果能够对劳动者的技能和观念上进行一定的调节,就业压力将有一定的改观。

2. 从劳动力需求方来看

(1)国有企业与民营企业比例

中国处于从计划经济体制向市场经济体制过渡的阶段,国有企业仍然占据着国民经济中的主导地位,并且绝对优势在可以预见的未来不会有实质性的改变,国有企业属于资本密集型企业,吸纳劳动力有限。与之相对应的民营企业大多属于劳动密集型企业,对劳动力数量需求较大,但是规模较小,并且相对于受国家保护的国有企业来说,在税收、融资等政策方面未得到公平的待遇,因此,在经济当中处于弱势,一旦经济动荡首先受到冲击,进而导致劳动力市场的不稳定,这一点在 08 年发生经济危机众多民营企业的倒闭导致失业人员增加的实例中得以体现。

(2)产业结构的不平衡

发达国家产业结构中,第一产业与第二产业只占总体 GDP 的 30%以下,而第三产业作为最大的产业,对劳动力有着巨大的需求。相比之下,如今的中国仍然处于工业化中期,第二产业仍然是劳动力的主要去向,服务业的发展是解决就业压力的一个重要解决途径。如果更进一步来细分行业,第二产业(主要是工业)可以划分为三十九个大行业,面对的地区不同,各个行业的发展情形有着明显的不同,以江苏省为例,各地的主导产业分别为:徐州“工程机械”,连云港“新材料”,盐城“汽车、风电”,泰州“生物医药”,常州“电工电器”……^[5]正是各地(特别是发达地区与不发达地区)的产业结构有着差别,各地就业人群结构也有着较大的差异。

(3)其他要素对劳动力的替代作用

劳动力就业的过程,实质上就是劳动力与其他资源重新配置的过程。^[1]由于在一定技术条件下生产资源相互可替代,劳动力要素在大量投入的情况下,其边际收益降低,而被其他资源排斥出来。中国经济转型和产业结构升级的过程中,生产技术的改进会提高装备的技术水平和劳动生产率,这就会迫使一部分劳动者成为多余。

(4)地域发展差距造成的就业压力

中国劳动力就业问题,明显地呈现出地域性的非均衡特征,大体表现出地区劳动力剩余量与人口的密度呈现负相关的关系。将地域划分为东、中、西三个部

分的话，可以发现东部沿海发达地区的就业紧张情况较为明显，而人口密度较小的中、西部则是剩余劳动力的主要存在区。

(5)行业投资的程度

在我国国民经济统计系统中，各行各业除了会公布工业总产值、工业增加值、利润总额等重要指标之外，还会将该时间段（通常以年或季度为单位）的投资情况公布出来。在短时间内技术进步不大的假设之下，一个行业固定资产投资额的大小直接关联着下一轮经济运行中对劳动力需求的增加程度。

3. 宏观经济环境的影响

除了从就业的供给方和需求方来分析就业平衡系统的影响因素之外，宏观经济环境会同时对以上两方同时产生作用，从而直接影响最终的供需效果。例如，经济的快速发展和有利的宏观政策会增加企业方的投资意愿，随着企业的发展和扩大，又会拉动对劳动力的需求。对外贸易的良好形势会带动企业（特别是外贸企业）业务的扩展，而外贸企业业务的扩张又会增加对内陆生产企业产品的需求，进而带动原材料、采掘业等一系列产业的发展。此外，经济周期和产业的发展规律是客观存在的，当经济周期到达低谷之时，经济的低迷导致企业对劳动力需求减少是必然的，面对一个地区，当产业升级换代之时，被淘汰的产业也会发生同样的现象。

二、指标的选择和符号的说明

通过第一部分中对就业理论的回顾和就业影响因素的定性分析，同时在考虑数据的可获得性和统计单位的一致性后，我们选择如下指标进行分析：

2.1 宏观经济指标和符号说明

表 2-1 可搜寻到数据的宏观经济重要指标一览表

指标类别	选择的统计指标	分析使用的符号
宏观整体	国内生产总值	<i>GDP</i>
	国内生产总值—第一产业	<i>GDP1</i>
	国内生产总值—第二产业	<i>GDP2</i>
	国内生产总值—第三产业	<i>GDP3</i>
	工业总产值	<i>ITO</i>
	工业增加值	<i>IAO</i>
	国民总收入指数	<i>RCE</i>
	居民消费水平指数	<i>GCE</i>
	固定资产投资	<i>FI</i>
	居民消费支出	<i>JXZ</i>
	政府消费支出	<i>ZXZ</i>
	城镇固定资产投资完成额	<i>UFI</i>
	城镇居民家庭平均每人消费支出	<i>CPX</i>
	社会消费品零售总额	<i>SSC</i>
	国际贸易额	<i>IT</i>

	进口额	<i>E</i>
	出口额	<i>O</i>
	外汇储备	<i>ER</i>
	美元汇率	<i>UER</i>
	当年是否具有重大政策——虚拟变量	<i>D</i>
能源生产	发电量	<i>PE</i>
国内存贷	各项贷款余额	<i>LR</i>
	新增贷款	<i>XDK</i>
	新增居民贷款	<i>XJK</i>
	企业存款余额	<i>CK</i>
	储蓄存款余额	<i>CCK</i>
货币供给量	活期存款利率	<i>DR</i>
	短期贷款利率	<i>SLR</i>
	法定准备金率	<i>SRR</i>
	流通中的现金	<i>M0</i>
	狭义货币供给量	<i>M1</i>
经济指数	广义货币供给量	<i>M2</i>
	景气指数—预警指数	<i>EWI</i>
	景气指数—一致指数	<i>COI</i>
	景气指数—先行指数	<i>LII</i>
	景气指数—滞后指数	<i>LAI</i>
	消费价格指数	<i>CPI</i>
	生产价格指数	<i>PPI</i>
	消费者信心指数	<i>CCI</i>
	消费者满意指数	<i>SAI</i>
	消费者预期指数	<i>EXI</i>
	采购经理指数	<i>PMI</i>
	零售物价指数	<i>RPI</i>

部分不常见的经济指标解释如下：

景气指数是用以反映宏观经济运行和企业生产经营状况所处的状态和未来发展变化趋势的指标。一致指数用以反映当前经济的基本走势，由工业生产、就业、社会需求（投资、消费、外贸）、社会收入（国家税收、企业利润、居民收入）等4个方面合成；先行指数是由一组领先于一致指数的先行指标合成，用于对经济未来的走势进行预测；滞后指数是由落后于一致指数的滞后指标合成得到，它主要用于对经济循环的峰与谷的一种确认；预警指数是把经济运行的状态分为5个级别，“红灯”表示经济过热，“黄灯”表示经济偏热，“绿灯”表示经济运行正常，“浅蓝灯”表示经济偏冷，“蓝灯”表示经济过冷。^[9]

CPI 即消费者物价指数(Consumer Price Index)，英文缩写为 CPI。是反映与居民生活有关的产品及劳务价格统计出来的物价变动指标，通常作为观察通货膨胀水平的重要指标。如果消费者物价指数升幅过大，表明通胀已经成为经济不稳定因素，央行会有紧缩货币政策和财政政策的风险，从而造成经济前景不明朗。

PPI 即生产者物价指数(Producer Price Indexes) 或者说是产品价格指数，英文缩写为 PPI。如钢铁、木材、电力、石油之类引致需求的生产要素价格现期

与基期的价格和不变的商品种类和权数乘积之比。人们普遍认为 CPI/PPI 过高，一般是经济过热的表面特征，国家会出台相关政策来抑制经济过快增长（例如：加息，提高准备金率等）。

PMI 即采购经理指数 (Purchase Management Index)，英文缩写为 PMI，它是月度发布的、综合性的经济监测指标体系，分为制造业 PMI、服务业 PMI、建筑业 PMI。它通过对采购经理的月度调查汇总出来的指数，来反映经济的变化趋势，PMI 涵盖了经济活动的多个方面，其综合指数反映了经济总体情况和总的变化趋势，而各项指标又反映了企业供应与采购活动的各个侧面。中国采购经理指数由国家统计局和中国物流与采购联合会共同合作完成，它包括制造业和非制造业采购经理指数，与 GDP 一同构成我国宏观经济的指标体系。

2.2 劳动力素质指标和符号说明

表 2-2 可搜寻到数据的衡量劳动力素质的指标一览表

指标类别	选择的统计指标	分析使用的符号
人口总体	总人口数目	<i>TP</i>
	男女比例	<i>MFR</i>
	就业人数	<i>JYR</i>
	在岗职工人数	<i>ZGR</i>
	职工人数	<i>GR</i>
城乡结构	城镇人口	<i>TP</i>
	乡村人口	<i>CP</i>
	乡村就业人数	<i>CJP</i>
可供劳动力	经济活动人口	<i>EP</i>
	15~64 岁年龄所占比例	<i>AP</i>
教育程度	高中以上学历比例	<i>HE</i>
体制结构	国有单位数目	<i>SN</i>
	城镇集体单位数目	<i>TN</i>
	其他单位数目	<i>ON</i>
就业吸引	职工平均工资	<i>LR</i>
失业情况	城镇登记失业人数	<i>UUP</i>
	城镇登记失业率	<i>UUR</i>

一般认为总人口数量与就业压力之间存在着一定的正相关关系（准确的说，应该按年龄段将人口数量进行细分，例如，如果平均工作年龄为 20 岁，5 年后的就业压力与此时 15 岁的年龄段人群数目有着密切的关系，但是由于数据搜索的问题，细分年龄段工作暂时放下），这一点在“1.2.3”中，我们已经分析过，经济活动人口和 15~64 岁年龄所占比例这两个指标是对它的更精确的表达。男女比例也与就业问题之间存在着一定的联系，就算我们将就业中女性歧视的问题排除在外，但是有些工作由于本身的特性对性别要求，因此我们也初步认为性别是就业的影响因素之一。城镇与乡村的人口，在“1.2.3”中我们分析过，中国历史中的城乡二元分割结构，在很长的一段时间里阻碍了农村剩余劳动人口向城市的转移，使得农村集聚了数量相当大的剩余劳动人口，改革开放之后，特别是随着城市化进程的加快，城市外来人口冲击着有限的就业岗位。国有和城镇集体单位数目，中国处于从计划经济体制向市场经济体制过渡的阶段，国有企业仍然

占据着国民经济中的主导地位，它属于资本密集型企业，吸纳劳动力有限。而与之相对的民营企业属于劳动密集型企业，对劳动力需求大，但是由于处于弱势地位，一旦经济受到震荡冲击，会造成大批的失业人员。劳动的报酬是从劳动者的需求来考虑的，初步认为它与就业之间存在着正向的关系，但是需要与物价指数结合起来考虑。

2.3 目标指标和符号说明

表 2-3 暂定的目标指标		
指标类别	选择的统计指标	分析使用的符号
目标指标	城镇就业人数	EN

三、问题的求解

3.1 问题一的模型建立与求解

3.1.1 问题一的基本假设

在对问题一、二进行分析之前，我们首先得进行一定的假设，这是我们的基本出发点，随着我们对问题的逐步深入分析，假设条件会得到逐步的放松。

我们对问题一、二的基本假设是：决定市场就业人数与失业人数的是劳动力供求双方的供求平衡。劳动力的需求方——宏观经济，将全国看做一个均匀的整体，没有各个地区发展快慢的差异之分，也没有不同行业之分（事实上我们知道这是不可能的，就业问题在不同地区和不同行业间的差异非常明显，但是为了切入问题，有必要进行如此苛刻的假设）；与之相对应，劳动力的供给方——就业人群，我们也不对其进行地区与行业的细致分类，将其视为一统一的整体。因此，在我们的眼前就出现了这样的一个基本的二元模型：宏观经济环境 \longleftrightarrow 就业人群，经济环境作为劳动人群的需求方，它的发展特点（经济总产量、消费量、净出口量等）决定了社会对就业人群的需求量，而劳动人群的条件（总人数、男女比例、年龄分布、教育程度等）决定了它能否顺利地满足社会需求或超出社会需求。正是这两者供需情况的差额导致了劳动力的剩余。

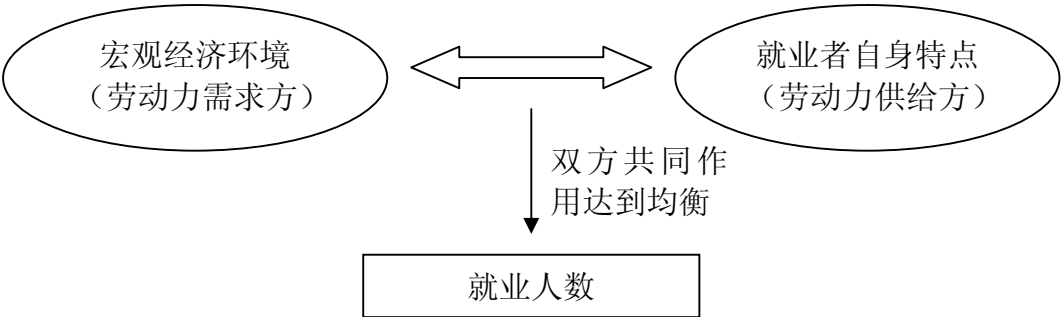


图 3-1 初步分析结构图

3.1.2 模型的建立

3.1.2.1 对就业人数时间序列的初步分析

(1) 对目标时间序列的直观分析

表 3-1		全国历年就业人数（单位：万人）			
年份	就业人数 EN	年份	就业人数 EN	年份	就业人数 EN
1952 年	20729	1971 年	35620	1990 年	64749
1953 年	21364	1972 年	35854	1991 年	65491
1954 年	21832	1973 年	36652	1992 年	66152
1955 年	22328	1974 年	37369	1993 年	66808
1956 年	23018	1975 年	38168	1994 年	67455
1957 年	23771	1976 年	38834	1995 年	68065
1958 年	26600	1977 年	39377	1996 年	68950
1959 年	26173	1978 年	40152	1997 年	69820
1960 年	25880	1979 年	41024	1998 年	70637
1961 年	25590	1980 年	42361	1999 年	71394
1962 年	25910	1981 年	43725	2000 年	72085
1963 年	26640	1982 年	45295	2001 年	73025
1964 年	27736	1983 年	46436	2002 年	73740
1965 年	28670	1984 年	48197	2003 年	74432
1966 年	29805	1985 年	49872.7	2004 年	75200
1967 年	30814	1986 年	51282	2005 年	75825
1968 年	31915	1987 年	52783	2006 年	76400
1969 年	33225	1988 年	54334	2007 年	76990
1970 年	34432	1989 年	55329	2008 年	

数据来源：中国统计局

在对就业人数（问题一、二中暂定的目标指标）进行与其他因素之间的相关分析之前，我们需要对目标的时间序列进行初步的分析，分析其本身具有的波动特点，例如，是否具有一定的周期性波动规律，外部变量与自身特点对整体时间序列各占有的百分比。

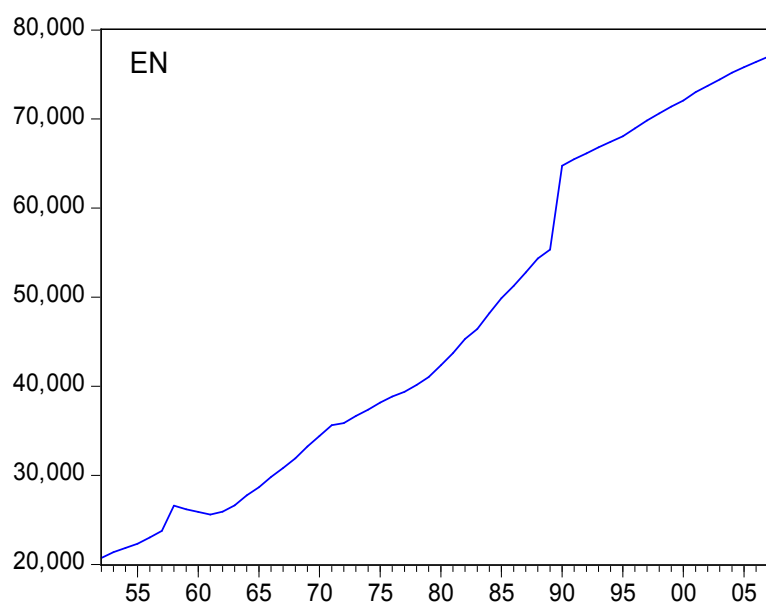


图 3-2 就业人数 EN 的历年走势

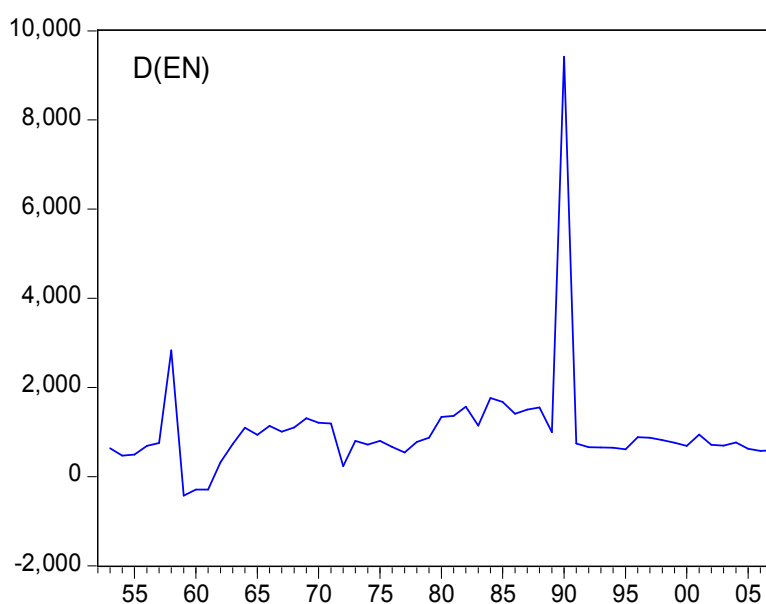


图 3-3 就业人数一阶差分 $D(EN)$ 走势图

从图 2-2 中我们可以看出从就业人数从 1952 年开始保持较快且固定的增长速度向上增长，从图 2-3 中我们可以看出，就业人数的一阶差分 ($D(EN) = EN - EN(-1)$) 在 1963 年~1989 年和 1992 年~2007 年间保持在 500~750 的范围内，在 1958 年与 1990 年有了突破性的增长，而 1959 年就业人数增长突变为负数，通过查找历史资料我们发现 1958 年的大跃进政策和其后此政策的失败是就业人数突变的主要原因，而在 1990 年国务院召开全国经济体制

改革工作会议，宣布了在上海实行经济技术开发区和某些经济特区的政策^[7]，由此我们可以怀着这样的初步猜想——就业人数上的“明显突变”很有可能与特殊的政策有着关联。

(2) 使用退势平稳序列模型拟合

表 3-2 使用 Eviews6 对就业人数时间序列建立退势平稳序列模型的结果

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	10685.23	5204.969	2.052890	0.0451
@TREND(1952)	1234.083	117.1078	10.53801	0.0000
AR(1)	0.902427	0.050814	17.75925	0.0000
R-squared	0.995821	Mean dependent var		47192.47
Adjusted R-squared	0.995661	S.D. dependent var		18769.23
S.E. of regression	1236.414	Akaike info criterion		17.13082
Sum squared resid	79493421	Schwarz criterion		17.24031
Log likelihood	-468.0975	Hannan-Quinn criter.		17.17316
F-statistic	6195.986	Durbin-Watson stat		1.882706
Prob(F-statistic)	0.000000			
Inverted AR Roots	.90			

注释：Eviews6 运行程序请参考电子稿附件“M2-1”

该时间序列函数表达式为：

$$EN_t = 10685.23 + 1234.08 \times T + u_t \quad (\text{令 } T=1952 \text{ 年时, } t=1)$$

$$(2.05) \quad (10.54)$$

其中， $u_t = 0.9024u_{t-1} + v_t$

$$(17.76)$$

整理后得到：

$$EN_t = 10685.23 + 1234.08 \times T + 0.90(EN_{t-1} - 10685.23 - 1234.08(T-1)) + v_t$$

$$\text{即 } EN_t = 2179.20 + 123.41 \times T + 0.90 \times EN_{t-1} + v_t$$

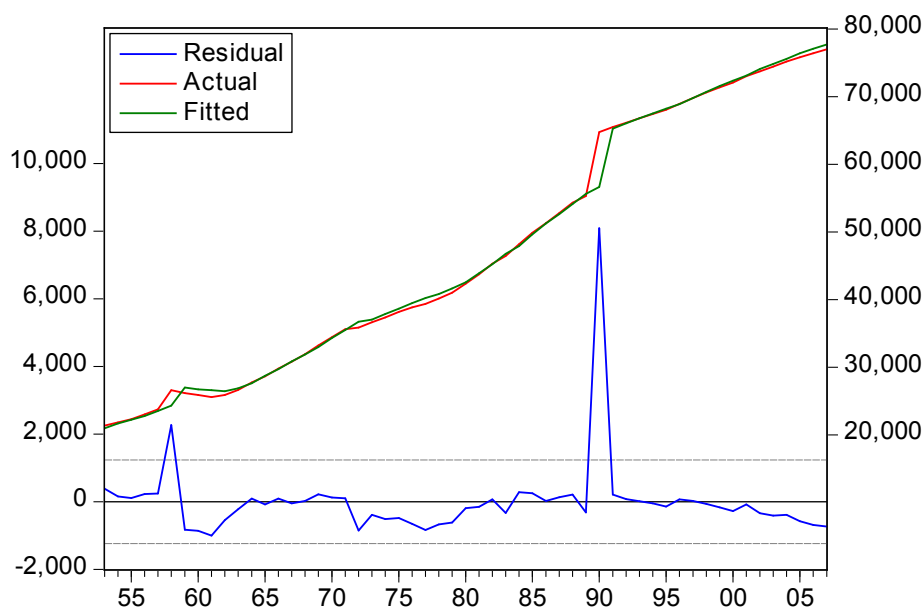


图 3-4 退势平稳序列实际值与拟合值比较图

注释：“Actual”代表实际值，“Fitted”代表拟合值，而“Residual”代表了实际值和拟合值二者之间的差额。

从最终的时间序列模型：

$$EN_t = 2179.20 + 123.41 \times T + 0.90 \times EN_{t-1} + v_t \quad \langle M \ 3-1 \rangle$$

(10.54) (17.76)

函数中的系数均通过 t 检验，且从实际值和拟合值比较图中可以看出，拟合值与实际值曲线非常接近（可决系数 $R^2=0.9958$ ），二者的差额保持在-800~300 的范围之内（除了 1958 年和 1990 年两个“突变年份”之外），换句话说，目标指标——就业人数的历史数据本身即具备了很强的自我预见性，或者说目标指标本身（除了政策突变的外在因素外）是很有规律的，仅仅通过就业人数的历史规律性来判断未来虽然具备一定的道理，但是我们更希望进一步了解哪些外在因素对它具有作用效果，以及外在因素是如何对它产生作用的、作用的大小是多少、持续的时间有多长等等，这些都是我们后面需要分析的问题。

(3) 使用 GARCH 模型对外在因素和内在规律的作用大小进行判断

GARCH 的意思是广义自回归条件异方差 (General ARCH)，旨在对因变量（被解释变量）的方差进行描述和预测，其中，被解释变量的方差按照公式的设定可能依赖于该变量的过去值，或依赖于一些独立的外生变量。^[7]

GARCH 模型的表达式为：

$$y_t = x_t \pi + \varepsilon_t, \quad \sigma_t^2 = \omega + \sum_{i=1}^p \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2$$

其中，第一个方程称为均值方程， y_t 是被解释变量， x 是解释变量，第二个

方程称为方差方程，其中 σ_t^2 是 ε_t 的条件方差。 p 是 ARCH 项的最大滞后阶数， q 是 GARCH 项的最大滞后阶数。

GARCH模型将经济变量的波动来源划分为两部分：变量过去的波动 σ_{t-j}^2 和外部冲击 ε_{t-i}^2 ， α_i 反映经济变量前期外部冲击对本期波动的作用强度， β_j 反映经济变量过去的波动对本期波动的作用强度。因此，GARCH(p, q) 模型可以被看作是被观测系统的一种波动率机制。

应用GARCH(p, q) 模型对于经济波动进行分析时，有下列结论：

变量本期的波动=常数+ $\alpha \times$ 前期的外部冲击+ $\beta \times$ 变量前期的波动率

表3-3 使用Eviews6对时间序列建立GARCH(1, 1)模型的结果

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	48210.28	6474.064	7.446680	0.0000
AR(1)	0.948071	0.016673	56.86153	0.0000
Variance Equation				
C	2.25E+08	84098350	2.673315	0.0075
RESID(-1)^2	-2.218788	0.895092	-2.478837	0.0132
GARCH(-1)	-0.999733	0.001108	-902.3078	0.0000
R-squared	0.989944	Mean dependent var	47192.47	
Adjusted R-squared	0.989139	S.D. dependent var	18769.23	
S.E. of regression	1956.034	Akaike info criterion	19.61625	
Sum squared resid	1.91E+08	Schwarz criterion	19.79874	
Log likelihood	-534.4470	Hannan-Quinn criter.	19.68682	
F-statistic	1230.509	Durbin-Watson stat	0.819265	
Prob(F-statistic)	0.000000			
Inverted AR Roots	.95			

注释：Eviews6 运行程序请参考电子稿附件“M2-2”

$$\text{均值方程为: } EN_t = 48210.28 + 0.948071 \times EN_{t-1} + \hat{u}_t$$

$$(7.45) \quad (56.86)$$

$$R^2 = 0.9899, \text{ DW} = 0.8193$$

$$\text{GARCH(1, 1)模型为: } \sigma_t^2 = -2.2187\hat{u}_{t-1}^2 - 0.9997\sigma_{t-1}^2 \quad \langle \text{M 3-2} \rangle$$

$$(-2.48) \quad (-902.31)$$

GARCH(1, 1)模型系数项 α 为负，且 $\alpha = -2.2187$ 表明就业人数当期的波动受前期外部因素波动的冲击较大，即就业系统以外的因素。系数项 β 为负，且

$\beta = -9997$ 表明就业人数波动受到前期就业波动的影响较小, 因而就业人数序列的波动具有受外部冲击影响较大的特性。外部因素和自身规律对就业人数序列的波动影响各占 69% 和 31% 的比例。因此, 为了对就业人数走势进行更准确的把握, 仅仅对历史数据规律的分析 and 总结是不够的, 对外部影响因素分析是很有必要。

(4) 就业人数时间序列规律分析结论总结

通过 2.1.3.1 的分析我们得出这样的结论:

① 目标指标——就业人数的“明显突变”时间点与特殊政策有着密切联系。

② 如果排除突变时间点, 通过建立退势平稳序列模型来对就业人数历史时间走势进行把握时, 我们发现可绝系数达到 99.58% 以上, 从 1952 年~2007 年长达 56 年的历史数据中, 拟合效果如此之好, 并且所有的指标都通过检验, 我们有理由相信这不是偶然的现象, 即目标指标本身具有很强的规律性, 能够通过历史情况对未来走势进行一定的预测。

③ 通过建立 GARCH(1, 1) 模型对就业人数的时间序列波动特点进行分析时, 我们发现外部因素对序列波动的作用效果远大于本身特点对序列波动的贡献率 ($69\% \gg 31\%$), 因此, 在时间序列分析的基础之上, 我们得出这样的结论: 目标指标的时间序列能够通过历史走势进行把握, 对就业人数的绝对数值有较好的预测小姑, 但是, 如果需要对波动的特点(波动方向、波动大小和波动持续时间)进行把握, 仅依靠目标指标的历史数据是不够的, 我们还需要从另外具有 69% 信息源的外部影响因素进行考虑。

3.1.2.2 对就业人数外部影响因素的选择

通过对目标指标时间序列的规律分析, 我们得出这样的结论: 本身的历史数据具有较强的规律性, 可以对未来就业人数的绝对值进行一定的预测。但是同时, 我们在对波动特点进行分析时发现, 外部因素对序列波动的作用效果远大于本身特点对序列波动的贡献率 ($69\% \gg 31\%$), 因此, 为了波动的特点(波动方向、波动大小和波动持续时间)进行把握, 仅依靠目标指标的历史数据是不够的, 我们还需要从提供 69% 信息的外部影响因素进行分析, 这就需要对众多外部影响因素进行筛选和分析。

(1) 对相关指标进行筛选的数据分析方法——格兰杰因果关系检验^[7]

在计量经济学中, 我们如果要判断两个指标之间是否存在着因果关系, 就会运用到格兰杰因果关系检验方法 (Granger no-causality test), 它的基本思想是: 如果 x 的变化引起 y 的变化, 则 x 应该有助于预测 y , 即在 y 关于 x 的过去值回归中, 增加 x 的过去值作为独立变量应当显著地增加回归的解释能力。检验 x 是否为引起 y 变化的原因基本过程如下:

① 作为原假设 “ x 不是引起 y 变化的原因”

② 带入 y 对 y 的滞后项及 x 的滞后项进行回归, 建立无限制条件的回归模

$$\text{型: } y_t = \sum_{i=1}^m a_i y_{t-i} + \sum_{i=1}^m b_i x_{t-i} + u_t$$

③把 y 只对 y 的滞后值进行回归，建立有限制条件的回归模型：

$$y_t = \sum_{i=1}^m a_i y_{t-i} + u_t$$

④用回归模型的残差平方和计算 F 统计值，检验回归系数 b_1, b_2, \dots, b_m 是否同时显著地不为零。如果是，就拒绝“ x 不是引起 y 变化的原因”的原假设，即 x 是引起 y 变化的原因，说明 x 与 y 存在因果关系。同理可以检验 y 是否为 x 的因。

(2) 使用格兰杰检验方法对目标指标与变量指标之间的因果关系进行判断

在我们本文的分析中，构造如下的模型：

$$\text{设定原假设: } \begin{cases} H_0: EN \text{ Does Not Granger Course } DCZB \\ H_0^1: DCZB \text{ Does Not Granger Course } EN \end{cases} \quad \langle M \ 3-3 \rangle$$

其中 EN 是我们的目标指标——就业人数， $DCZB$ 代表我们在 2.1.2 中二十多个待测指标，在这里我们假设各指标都假设是平稳序列。

①对国内生产总值与目标指标之间的因果关系进行判断

我们以对国内生产总值与就业人数之间是否具有明显的因果关系检验为例：

表 3-4 各季度全国城镇单位就业人员期末人数和国内生产总值一览表

季度	EN	GDP	季度	EN	GDP
2000 年 1 季度	119303	20646.96	2004 年 4 季度	110989	49910.7
2000 年 2 季度	118981	23101.26	2005 年 1 季度	109723.5	38763.6
2000 年 3 季度	118525	24339.28	2005 年 2 季度	110302.6	42443.2
2000 年 4 季度	116125	31127.05	2005 年 3 季度	111027.3	44370.7
2001 年 1 季度	114499	23299.54	2005 年 4 季度	114040.3	57640
2001 年 2 季度	114037	25651.32	2006 年 1 季度	113054.3	44419.8
2001 年 3 季度	113676	26867.34	2006 年 2 季度	114069.8	49191.7
2001 年 4 季度	111658	33836.98	2006 年 3 季度	114945.4	50958.1
2002 年 1 季度	110400	25375.69	2006 年 4 季度	117131.7	67353.8
2002 年 2 季度	110065	27965.32	2007 年 1 季度	115803	53058.286
2002 年 3 季度	110053	29715.69	2007 年 2 季度	116805	59400.007
2002 年 4 季度	109852	37275.98	2007 年 3 季度	117584	61969.338
2003 年 1 季度	108343	28861.8	2007 年 4 季度	120244.3	82877.926
2003 年 2 季度	108407	31007.1	2008 年 1 季度	118760.1	63474.494
2003 年 3 季度	109071	33460.4	2008 年 2 季度	119910.5	71251.289
2003 年 4 季度	109697	42493.4	2008 年 3 季度	120778.4	73299.537

2004 年 1 季度	108250	33420.6	2008 年 4 季度	NA	92644.68
2004 年 2 季度	108959	36985.3	2009 年 1 季度	119965.3	65745.145
2004 年 3 季度	109443	39561.7	2009 年 2 季度	120755.5	74117.055

数据来源：中国统计局

注释：全国城镇单位就业人员期末人数单位是千人，国内生产总值由公布的累积值计算得来，单位是亿元。

表 3-5 使用 Eviews6 对 EN、GDP 因果关系进行一期格兰杰检验结果

Null Hypothesis:	Obs	F-Statistic	Prob.
GDP does not Granger Cause EN	35	3.95503	0.0553
EN does not Granger Cause GDP		0.13635	0.7144

注释：Eviews6 运行程序请参考电子稿附件“M2-3-1”，只需将“GroupM2-3”点击后，将“Granger Causality”选项改为“1”即可，代表一期检验，其后分析同理。

一期检验的初始假设为：

$$\begin{cases} H_0 : EN_t \neq \alpha + \beta_1 GDP_{t-1} + \beta_2 EN_t + \beta_3 EN_{t-1} \\ H_0^1 : GDP_t \neq \alpha + \beta_1 EN_{t-1} + \beta_2 GDP_t + \beta_3 GDP_{t-1} \end{cases}$$

使用 Eviews6 获得结果中的概率定义为：

$P(F > 3.95503) = 0.0553$ （意味着 $F = 3.95503$ 在临界值的右侧）

$P(F > 0.13635) = 0.7144$ （意味着 $F = 0.13635$ 在临界值的左侧）

一期检验得出的结论是：“GDP 变动不是 EN 变动的原因”被拒绝，原假设“EN 价格变动不是 GDP 变动的原因”被接受，即 GDP 变动是 EN 变动的原因，而 EN 变动不是 GDP 变动的原因。

表 3-6 使用 Eviews6 对 EN、GDP 因果关系进行二期格兰杰检验结果

Null Hypothesis:	Obs	F-Statistic	Prob.
GDP does not Granger Cause EN	33	10.1343	0.0005
EN does not Granger Cause GDP		2.61770	0.0907

使用 Eviews6 获得结果中的概率定义为：

$P(F > 10.1343) = 0.0005$ （意味着 $F = 10.1343$ 在临界值的右侧）

$P(F > 2.61770) = 0.0907$ （意味着 $F = 2.61770$ 在临界值的左侧）

二期检验得出的结论是：“GDP 变动不是 EN 变动的原因”被拒绝，原假设“EN 价格变动不是 GDP 变动的原因”被接受，即 GDP 变动是 EN 变动的原因，而 EN 变动不是 GDP 变动的原因。

表 3-7 使用 Eviews6 对 EN、GDP 因果关系进行三期格兰杰检验结果

Null Hypothesis:	Obs	F-Statistic	Prob.
GDP does not Granger Cause EN	32	8.24953	0.0006

使用 Eviews6 获得结果中的概率定义为:

$P(F > 8.24953) = 0.0006$ (意味着 $F = 8.24953$ 在临界值的右侧)

$P(F > 1.96612) = 0.1449$ (意味着 $F = 1.96612$ 在临界值的左侧)

三期检验得出的结论是:“GDP 变动不是 EN 变动的原因”被拒绝,原假设“EN 价格变动不是 GDP 变动的原因”被接受,即 GDP 变动是 EN 变动的原因,而 EN 变动不是 GDP 变动的原因。

四期检验、五期检验……将所有滞后期检验结果列表如下:

表 3-8 对 EN、GDP 因果关系进行格兰杰检验的结果综合

因果关系	Prob. (F>...)					结论
	一期	二期	三期	四期	五期	
GDP—>EN	0.0553	0.0005	0.0006	1.E-08	0.0123	GDP 是 EN 的原因
EN—>GDP	0.7144	0.0907	0.1449	0.5802	0.3059	EN 不是 GDP 的原因

注释:天蓝色底纹的代表拒绝原假设。

②其他经济指标与目标指标之间的格兰杰因果检验分析结果

表 3-9 EN 与国内生产总值(包括一、二、三产业)的格兰杰检验结果

因果关系	Prob. (F>...)					结论
	一期	二期	三期	四期	五期	
GDP—>EN	0.0553	0.0005	0.0006	1.E-08	0.0123	GDP 是 EN 的原因
EN—>GDP	0.7144	0.0907	0.1449	0.5802	0.3059	EN 不是 GDP 的原因
GDP1—>EN	0.8813	0.6025	0.1371	2.E-07	0.0120	GDP1 是 EN 的原因(远期)
EN—>GDP1	0.2492	0.1621	0.0603	0.6591	0.6369	EN 不是 GDP1 的原因
GDP2—>EN	0.0105	3.E-05	9.E-07	6.E-08	0.0754	GDP2 是 EN 的原因
EN—>GDP2	0.8403	0.0082	0.0128	0.3417	0.0482	EN 不是 GDP1 的原因
GDP3—>EN	0.0941	8.E-05	3.E-05	2.E-08	0.0090	GDP3 是 EN 的原因
EN—>GDP3	0.7409	0.0507	0.0095	0.4957	0.8758	EN 不是 GDP3 的原因

数据来源:中国统计局,详细数据在电子稿附件“M2-3-2”中有保留,GDP、GDP1、GDP2、GDP3 的单位均为亿元。

注释: Eviews6 运行程序请参考电子稿附件“M2-3-2”,只需打开“Group”后,将“Granger Causality”选项改为“1”、“2”、“3”、“4”、“5”即可,其后分析同理。

表 3-10 EN 与工业总产值(ITO)、工业增加值(IAO)的格兰杰检验结果

因果关系	Prob. (F>...)					结论
	一期	二期	三期	四期	五期	
ITO—>EN	0.6536	0.7298	0.9144	0.9851	0.9996	ITO 不是 EN 的原因
EN—>ITO	0.1237	0.8038	0.8185	0.7879	0.8609	EN 不是 ITO 的原因
IAO—>EN	0.8283	0.2808	0.6208	0.0128	NA	IAO 不是 EN 的原因
EN—>IAO	0.5270	0.4278	0.4093	0.6483	NA	EN 不是 IAO 的原因

数据来源:中国统计局,详细数据在电子稿附件“M2-3-3”中有保留,其中,工业总产值和工业增加值的单位均为亿元。

表 3-11 EN 与国民总收入指数(RCE)、居民消费水平指数(GCE)的格兰杰检验结果

因果关系	Prob. (F>...)					结论
	一期	二期	三期	四期	五期	
RCE→EN	0.7135	0.7550	0.4352	0.6602	0.6486	RCE 不是 EN 的原因
EN→RCE	0.0237	0.2755	0.3044	0.4167	0.3787	EN 不是 RCE 的原因
GCE→EN	0.4009	0.0314	0.0953	0.0495	0.0415	GCE 是 EN 的原因
EN→GCE	0.5572	0.8242	0.3130	0.2579	0.4581	EN 不是 GCE 的原因

数据来源：中国统计局，详细数据在电子稿附件“M2-3-4”中，原始数据记录的是同比增长（国民总收入指数从 53 年起有记载，居民消费收入水平从 78 年起有记载）。

表 3-12 EN 与居民消费支出(JXZ)、政府消费支出(ZXZ)的格兰杰检验结果

因果关系	Prob. (F>...)					结论
	一期	二期	三期	四期	五期	
JXZ→EN	0.0001	0.0031	0.0155	0.0672	0.2480	JXZ 是 EN 的原因
EN→JXZ	0.8325	0.0106	0.1639	0.3060	0.4722	EN 不是 JXZ 的原因
ZXZ→EN	0.1659	0.0018	0.0298	0.0479	0.2176	ZXZ 是 EN 的原因
EN→ZXZ	0.6243	0.3163	0.0729	0.0772	0.1543	EN 不是 ZXZ 的原因

数据来源：中国统计，详细数据在电子稿附件“M2-3-5”中。

表 3-13 EN 与城镇居民家庭平均每人消费支出(CPX)的格兰杰检验结果

因果关系	Prob. (F>...)					结论
	一期	二期	三期	四期	五期	
CPX→EN	0.3464	NA	NA	NA	NA	CPX 不是 EN 的原因
EN→CPX	0.0002	NA	NA	NA	NA	EN 不是 CPX 的原因

数据来源：中国统计局，详细数据在电子稿附件“M2-3-6”中。

表 3-14 EN 与消费者价格指数(CPI)的格兰杰检验结果

因果关系	Prob. (F>...)					结论
	一期	二期	三期	四期	五期	
CPI→EN	0.7495	0.810	0.0020	0.0035	0.5823	CPI 是 EN 的原因
EN→CPI	0.5259	0.0002	0.0008	0.0573	0.1172	EN 是 CPI 的原因

数据来源：万德数据库，详细数据在电子稿附件“M2-3-8”中。

表 3-15 对外贸易指标（包括国际贸易额(IT)、出口额(EX)、进口额(IM)）与目标之间的格兰杰因果检验分析结果

因果关系	Prob. (F>...)					结论
	一期	二期	三期	四期	五期	
IT→EN	0.0004	5.E-06	3.E-06	3.E-09	0.0223	IT 是 EN 的原因
EN→IT	0.5267	2.E-05	0.0002	0.0002	0.3022	EN 是 IT 的原因
0→EN	0.0006	2.E-05	9.E-06	9.E-10	0.0012	0 是 EN 的原因
EN→0	0.5111	5.E-05	0.0004	2.E-05	0.1934	EN 是 0 的原因
E→EN	0.0003	2.E-06	5.E-06	3.E-08	0.1092	E 是 EN 的原因
EN→E	0.6654	0.0006	0.0151	0.0124	0.2463	EN 是 E 的原因

数据来源：万德数据库，详细数据在电子稿附件“M2-3-8”中。

表 3-16 发电量(PE)与目标指标的格兰杰因果检验分析结果

因果关系	Prob. (F>...)					结论
	一期	二期	三期	四期	五期	
PE→EN	0.0003	2.E-08	2.E-08	4.E-08	0.0295	PE 是 EN 的原因
EN→PE	0.6474	0.0605	0.1908	0.2964	0.7086	EN 是 E 的原因

数据来源：万德数据库，详细数据在电子稿附件“M2-3-8”中。

表 3-17 城镇固定资产投资完成额(UFL)与目标指标的格兰杰因果检验分析结果

因果关系	Prob. (F>...)					结论
	一期	二期	三期	四期	五期	
UFL→EN	0.8527	0.4524	0.1595	2.E-09	0.0184	UFL 是 EN 的原因
EN→UFL	0.4292	0.3351	0.0859	0.0109	0.0667	EN 是 UFL 的原因

数据来源：万德数据库，详细数据在电子稿附件“M2-3-7”中。

表 3-18 国内信贷指标（国内信贷指标包括新增贷款（XDK）、新增居民贷款（XJK）、各项储蓄余款（LR）、企业存款余额（CK）、储蓄存款余额(CCK)）与目标指标之间的格兰杰因果检验分析结果

因果关系	Prob. (F>...)					结论
	一期	二期	三期	四期	五期	
XDK→EN	0.0042	0.0631	0.0577	0.0516	0.2812	XDK 是 EN 的原因
EN→XDK	0.0062	0.0031	0.0027	0.2416	0.0861	EN 是 XDK 的原因
XJK→EN	0.1540	0.6197	0.0101	NA	NA	XJK 不是 EN 的原因
EN→XJK	0.3870	0.2010	0.4479	NA	NA	EN 不是 XJK 的原因
LR→EN	0.0003	8.E-05	8.E-06	1.E-07	0.0489	LR 是 EN 的原因
EN→LR	0.6902	0.1738	0.5694	0.8532	0.5383	EN 不是 LR 的原因
CK→EN	0.0006	0.0002	0.0004	4.E-08	0.0249	CK 是 EN 的原因
EN→CK	0.9301	0.0214	0.0460	0.0870	0.4319	EN 不是 CK 的原因
CCK→EN	0.0002	7.E-05	7.E-05	5.E-11	0.0183	CCK 是 EN 的原因
EN→CCK	0.9928	0.0053	0.0024	0.0001	0.0272	EN 是 CCK 的原因

数据来源：万德数据库，详细数据在电子稿附件“M2-3-8”中。

表 3-19 广义货币供应量(M2)与目标指标之间的格兰杰因果检验分析结果

因果关系	Prob. (F>...)					结论
	一期	二期	三期	四期	五期	
M2→EN	0.0003	0.0003	0.0004	7.E-07	0.0486	M2 是 EN 的原因
EN→M2	0.8762	0.3835	0.2173	0.0542	0.0763	EN 是 M2 的原因

数据来源：万德数据库，详细数据在电子稿附件“M2-3-8”中。

③劳动力指标与目标指标之间的格兰杰因果检验分析结果

表 3-20 EN 与经济活动人数(EP)、就业人数(JYR)、乡村就业人数(CJP)、在岗职工数(ZGR)、城镇登记失业率(UUR)、职工人数(GR)、职工平均工资(LR)的格兰杰检验结果

因果关系	Prob. (F>...)					结论
------	---------------	--	--	--	--	----

	一期	二期	三期	四期	五期	
EP—>EN	0.9967	0.1999	0.4120	0.5817	0.7194	EP 不是 EN 的原因
EN—>EP	0.2820	0.0164	0.0590	0.1486	0.2873	EN 不是 EP 的原因
JYR—>EN	0.3845	0.6675	0.8559	0.8688	0.9549	JYR 不是 EN 的原因
EN—>JYR	0.0895	0.0332	0.0722	0.1433	0.2109	EN 是 JYR 的原因(近期)
CJP—>EN	0.3845	0.6676	0.8559	0.8691	0.9550	CJP 不是 EN 的原因
EN—>CJP	0.0270	0.0089	0.0239	0.0655	0.1425	EN 是 CJP 的原因
ZGR—>EN	0.8815	0.8049	0.9450	0.9737	0.5446	ZGR 不是 EN 的原因
EN—>ZGR	0.0016	0.1982	0.6918	0.9261	0.8950	EN 不是 ZGR 的原因
UUR—>EN	0.7617	0.1387	0.1542	0.2927	0.2016	UUR 不是 EN 的原因
EN—>UUR	0.0029	0.0048	0.0018	0.0675	0.0360	EN 是 UUR 的原因
GR—>EN	0.9796	0.9971	0.9922	0.9924	0.9968	GR 不是 EN 的原因
EN—>GR	0.9876	0.0049	0.6909	0.8479	0.9136	EN 不是 GR 的原因
LR—>EN	0.5544	0.5671	0.5820	0.7690	0.8863	LR 不是 EN 的原因
EN—>LR	0.6568	0.7175	0.8839	0.9425	0.2976	EN 不是 LR 的原因

数据来源：中国统计局，详细数据在电子稿附件“M2-3-8”中。

3.1.2.3 运用格兰杰因果检验方法对影响因素的选择结果

通过前面的格兰杰建议，我们将目标指标与影响因素的因果系列综合。

表 3-21 因果关系分析中结果列表如下

是就业人数原因的影响因素 (DCZB—>EN)	
国内生产总值 (GDP)	发电量 (PE)
第二产业总产值 (GDP2)	消费者价格指数 (CPI)
第三产业总产值 (GDP3)	城镇固定资产投资完成额 (UFL)
第一产业总产值 (GDP1)	国内信贷指标包括新增贷款 (XDK)
居民消费水平指数 (GCE)	新增居民贷款 (XJK)
居民消费支出 (JXZ)	各项储蓄余款 (LR)
政府消费支出 (ZXZ)	企业存款余额 (CK)
国际贸易额 (IT)	储蓄存款余额 (CCK)
出口额 (O)	广义货币供应量 (M2)
进口额 (E)	
是就业人数结果的影响因素 (EN—>DCZB)	
消费者价格指数 (CPI)	国内信贷指标包括新增贷款 (XDK)
包括国际贸易额 (IT)	储蓄存款余额 (CCK)
出口额 (EX)	就业人数 (JYR)
进口额 (IM)	乡村就业人数 (CJP)
城镇固定资产投资完成额 (UFL)	城镇登记失业率 (UUR)

在文献综述的基础上，我们通过定向分析，综合了所有与就业人数有着直接或间接关系的指标。通过前面的因果分析，我们从这些众多的指标当中又筛选出与目标有着因果关系的指标，宏观经济指标当中有 19 项指标对目标有着解释能力，而劳动力素质评价系统中没有对就业人数拥有解释能力的指标，却有 3 项指标——就业人数、乡村就业人数、城镇登记失业率可以通过城镇就业人数指标得

到解释，此外，还有 7 项宏观经济指标可以通过目标指标得以解释。

3.2 问题二的模型建立与求解

3.2.1 多元回归模型原理

(1) 多项式分布滞后模型

分布滞后是指以下模型形式：

$$y_t = \beta_1 x_t + \beta_2 x_{t-1} + \dots + \beta_L x_{t-L+1}$$

系数 β_j 表达了 x_t 及其滞后项对 y_t 的影响。在许多情况下，系数可以直接估计，但是在其他一些情况下， x_t 和滞后项的高度共线性会使直接估计失败。多项式分布滞后的目的是通过在滞后系数上施加一个平滑条件来减少需要估计的参数的个数，所谓平滑也就是要求系数依赖于一个次数相对较低的多项式。

多项式分布滞后的设定有 3 个要素：滞后的长度 L ，多项式的次数（多项式中的最高幂）以及所要采用的约束。^[7]

(2) 非线性最小二乘法

在经济模型中有两种非线性：变量非线性和参数非线性。一个模型可能是变量非线性的，或者是参数非线性的，或者两者都是非线性的。变量非线性但参数为线性可以通过用最小二乘法或两阶段最小二乘法来估计。^[12]例如，模型

$$y_t = \beta_1 + \beta_2 \log L_t + \beta_3 \log K_t$$

其中， L_t 和 K_t 是非线性的，但参数是线性的，因此可以先设定新变量然后再用最小二乘法估计。如果用公式来设定估计方程，则不需要创建任何新变量，它们会暗含在公式中。

下面是一个参数非线性方程的例子

$$y_t = \beta_1 L_t^{\beta_2} K_t^{\beta_3}$$

该方程无法事先设定变量并用普通最小二乘法估计，因此，需要使用非线性最小二乘法。

3.2.2 运用多元回归模型进行实证分析

通过对问题一的分析，我们已经得出了哪些影响因素与目标指标之间具有因果关系，在本次研究的过程中，存在的问题不在技术难度上，而在对指标数据的搜集上，我们需要查询到时间单位较为统一（以月、季度、年都可以，但要求统一），且时间段不能过短，另外，需要将建立的最终模型通过一系列严格的检验。我们将有因果关系的影响因素都进行了综合，将能够最终统一时间单位与时间段的指标总结如下：

表 3-22

目标指标与具有因果关系的影响因素季度数据综合表

季度	城镇就业人数	GDP	GDP1	GDP2	GDP3	CPI
2000 年 1 季度	119303	20646.96	1924.915	9981.424	8740.621	0.10
2000 年 2 季度	118981	23101.26	3169.551	11448.47	8483.242	0.10
2000 年 3 季度	118525	24339.28	4074.812	11199.6	9064.866	0.27
2000 年 4 季度	116125	31127.05	5775.444	12926.38	12425.22	0.93
2001 年 1 季度	114499	23299.54	2035.23	11127.4	10136.9	0.67
2001 年 2 季度	114037	25651.32	3261.871	12685.88	9703.573	1.57
2001 年 3 季度	113676	26867.33	4418.071	12042.86	10406.4	0.80
2001 年 4 季度	111658	33836.98	6066.097	13656.15	14114.73	-0.23
2002 年 1 季度	110400	25375.69	2181.266	11811.32	11383.11	-0.60
2002 年 2 季度	110065	27965.32	3453.948	13751.28	10760.09	-1.07
2002 年 3 季度	110053	29715.7	4690.37	13244.66	11780.67	-0.77
2002 年 4 季度	109852	37275.98	6211.436	15089.51	15975.03	-0.63
2003 年 1 季度	108343	28861.8	2258.1	13776.4	12827.3	0.50
2003 年 2 季度	108407	31007.1	3521.3	15701.6	11784.2	0.67
2003 年 3 季度	109071	33460.4	5087	15375.1	12998.3	0.83
2003 年 4 季度	109697	42493.4	6515.3	17583.2	18394.9	2.67
2004 年 1 季度	108250	33420.6	2663.5	16077.3	14679.8	2.77
2004 年 2 季度	108959	36985.3	4364	18597.2	14024.1	4.40
2004 年 3 季度	109443	39561.7	6357.5	18195.1	15009.1	5.27
2004 年 4 季度	110989	49910.7	8027.7	21034.7	20848.3	3.17
2005 年 1 季度	109723.5	38763.6	2928.6	18968.4	16866.6	2.83
2005 年 2 季度	110302.6	42443.2	4507.7	21934.2	16001.3	1.73
2005 年 3 季度	111027.3	44370.7	6607.5	20639.8	17123.4	1.33
2005 年 4 季度	114040.3	57640	8376.2	25822.2	23441.6	1.37
2006 年 1 季度	113054.3	44419.8	3093	22076.1	19250.7	1.20
2006 年 2 季度	114069.8	49191.7	4880.6	25833.3	18477.8	1.37
2006 年 3 季度	114945.4	50958.1	7085	24098.8	19774.3	1.27
2006 年 4 季度	117131.7	67353.8	8981.4	31153.8	27218.6	2.03
2007 年 1 季度	115803	53058.29	3654.036	26464.62	22939.63	2.73
2007 年 2 季度	116805	59400.01	5629.283	31149.84	22620.89	3.60
2007 年 3 季度	117584	61969.34	8653.478	28790.37	24525.49	6.10
2007 年 4 季度	120244.285	82877.93	10690.2	38394.14	33793.59	6.63
2008 年 1 季度	118760.1	63474.49	4720	31657.77	27096.72	8.03
2008 年 2 季度	119910.5	71251.29	7080	37671.8	26499.49	7.77
2008 年 3 季度	120778.4	73299.54	10262.26	34644.89	28392.39	5.27
2008 年 4 季度		92644.68	11937.74	42208.94	38498	2.533
2009 年 1 季度	119965.332	65745.15	4700	31968.1	29077.05	
2009 年 2 季度	120755.523	74117.06	7325	38102.4	28689.65	

接上表右侧:

国际贸易总额	出口总额	进口总额	发电量	城镇固定资产投资完成额	各项存款余额	M2
327.29	172.33	154.97	1017.89	1063.41	110020.57	123666.1
394.83	209.43	185.40	1077.70	5460.47	115169.75	127677.3

429.21	225.94	203.29	1148.87	11349.23	118443.48	132381.2
429.66	223.01	206.66	1132.67	20033.31	121875.92	136241.5
377.68	195.92	181.76	1087.42	1230.31	126463.25	142293.1
425.41	218.90	206.52	1162.24	6454.30	132180.04	145846.5
451.02	234.45	216.57	1256.00	13447.27	136651.01	150330.3
444.76	237.64	207.12	1231.82	23196.24	141323.21	154629
406.87	215.53	191.29	1189.37	1557.23	144982.52	161556.3
495.57	257.98	237.62	1296.04	8115.86	153835.72	166744.3
581.68	301.74	279.95	1446.86	16705.71	162042.04	173694.6
585.73	310.14	275.57	1435.82	28158.99	168656.07	180679.1
578.89	287.74	291.15	1336.90	2138.19	177430.09	191694.7
674.94	346.44	328.50	1465.14	10971.78	189296.19	200188.9
767.49	391.55	375.95	1661.39	22543.63	198705.14	210117.4
817.25	435.83	381.42	1638.76	36965.42	206084.04	217348
799.59	385.74	413.85	1585.62	3448.50	215146.76	227935.8
944.29	474.96	469.32	1706.62	16109.54	225693.28	235632.6
1016.33	527.36	488.97	1876.43	32443.36	232565.15	239442.9
1088.03	590.79	497.24	1878.11	50483.63	238932.21	248027.8
984.03	519.66	464.38	1818.81	4419.49	249898.02	260565.9
1166.06	621.79	544.27	1941.17	20570.33	263623.76	270669.2
1265.46	680.58	584.89	2131.40	41509.85	275239.76	281892.9
1326.28	719.07	607.21	2099.82	64716.15	284713.31	292899.2
1237.27	657.14	580.13	2062.34	5634.17	298213.81	306192.9
1415.95	771.24	644.70	2197.94	26605.83	314302.95	317722.7
1589.22	875.81	713.41	2484.13	53081.60	323846.31	327936.9
1627.99	926.92	701.07	2420.35	80951.66	333031.22	338609.8
1525.90	840.47	685.44	2317.23	7026.21	346610.13	358087.6
1744.46	982.61	761.84	2586.25	33572.34	362813.89	371625.6
1966.02	1104.91	861.12	2888.52	67201.20	377116.99	388062.9
2011.67	1132.53	879.14	2809.82	102323.95	384454.00	399121.1
1901.70	1019.91	881.79	2689.06	8812.74	404057.37	420646.2
2210.05	1201.49	1008.56	2894.18	42370.08	430821.72	436201.2
2442.07	1359.93	1082.13	3082.82	85650.16	449595.12	449381.8
1982.11	1181.57	800.54	881.67	129656.82	462301.34	462314.9

注释：目标指标与GDP(包括三大产业产值)数据来源于中国统计局，其后的数据来自于万德数据库，是将月度数据通过平均的方式得到的季度数据。

表3-23 目标指标与具有因果关系的影响因素年度数据综合表

年份	就业人数	经济活动人口	居民消费支出	政府消费支出
1989	55329	55707	8812.60	2351.60
1990	64749	65323	9166.73	2560.23
1991	65491	66091	10065.72	3153.03
1992	66152	66782	11461.09	3705.61
1993	66808	67468	12614.79	4218.07

1994	67455	68135	13529.45	4582.03
1995	68065	68855	15005.20	4431.53
1996	68950	69765	16583.41	4866.03
1997	69820	70800	17540.62	5329.95
1998	70637	72087	18787.30	5918.80
1999	71394	72791	20361.15	6662.24
2000	72085	73992	22183.30	7576.59
2001	73025	74432	23642.61	8486.52
2002	73740	75360	25459.55	9259.50
2003	74432	76075	27197.74	9865.22
2004	75200	76823	29400.50	10685.20
2005	75825	77877	32221.44	12037.18
2006	76400	78244	35872.65	13425.30
2007	76990	78645	39325.60	15258.46

年份	固定资产支出	进口额	出口额	平均工资
1989	4419.40	2199.90	1956.10	1935.00
1990	4682.64	2496.90	2896.02	2075.65
1991	5694.18	3188.11	3589.97	2195.01
1992	7505.81	3917.28	4122.70	2390.06
1993	10229.81	4601.16	4062.04	2591.04
1994	10722.81	6168.90	6454.86	2810.66
1995	11046.42	5843.52	6585.96	2909.04
1996	11744.63	5644.41	6142.07	3032.85
1997	12335.42	5609.02	7202.52	3073.76
1998	13681.98	5567.85	7290.73	3581.77
1999	14827.41	6671.90	7848.97	4053.73
2000	16373.07	9016.98	9982.40	4533.45
2001	18137.71	9684.72	10580.78	5222.08
2002	21130.42	11831.26	13050.50	6015.80
2003	25597.63	16364.09	17365.34	6718.75
2004	29991.97	21387.45	22616.05	7380.35
2005	34975.56	24555.44	28344.33	8308.56
2006	40184.84	28250.29	34587.85	9361.21
2007	44755.51	31265.25	39750.64	10604.45

注释：数据由各年度的统计年鉴综合而来，其中居民消费支出、政府支出、固定资产投资、进口额、出口额、平均工资等数据都进行了价格调整(在当年价格的基础之上排除了物价上涨的因素)

经过模型建立的尝试，并通过一系列的检验(最明显的是多重共线性)后，最终建立的多元回归模型如下所示：

表3-24 使用Eviews6建立多元回归模型结果

Variable	Coefficient	Std. Error	t-Statistic	Prob.
----------	-------------	------------	-------------	-------

JJRK	0.913398	0.010235	89.24582	0.0000
TZ	2.493569	0.366499	6.803746	0.0000
R-squared	0.998017	Mean dependent var	70134.05	
Adjusted R-squared	0.997900	S.D. dependent var	5267.409	
S.E. of regression	241.3564	Akaike info criterion	13.90973	
Sum squared resid	990299.7	Schwarz criterion	14.00914	
Log likelihood	-130.1424	Hannan-Quinn criter.	13.92655	
Durbin-Watson stat	1.273900			

注释：详细的数据处理程序请参见电子稿附件中的“M2-3-9”

最终的多元回归方程为：

$$EN = 0.91339755213 * JJRK + 2.49356895529 * TZ \quad \langle M3-4 \rangle$$

(89.24582) (6.803746)

$$R^2 = 0.9980$$

众多指标之间存在着严重的多重共线性，即这些指标的变化具有相同或类似的特征，因此，在使用其中一个指标进行分析时，就代表了其他指标的基本信息，所以最终表现出来的结果就是城镇就业人口与经纪人口总量、固定资产支出的大小有着密切的关系，即就业人数与潜在的就业人数总量(劳动力供给方的供给力量)和城镇投资扩张的力度大小(劳动力需求方的容纳能力)有着密切的关系。

3.2.3 考虑使用向量自回归(VAR)模型以及该模型的原理

由于考虑到多元回归模型是对影响因素——>目标指标的单向分析过程，即只运用了影响因素对目标指标的单方面因果关系这类信息，而不能够在模型中将目标指标——>影响因素，即目标指标对影响因素的因果关系，因此，我们引入向量自回归模型，这种计量分析模型的特点是最终建立的模型是一个系统，它将目标与影响因素较为平等地看待，分析结果不仅体现了影响因素对目标指标的作用效果，也体现了目标指标对影响因素的作用效果。^[7]

向量自回归(VAR)模型通常用于时间序列数据的预测和分析随机扰动对变量系统的动态影响。模型避开了结构建模方法中需要对系统中每个内生变量关于所有内生变量滞后值函数的建模问题。它是指系统内每个方程有相同的等号右侧变量，而这些右侧变量包括所有内生变量的滞后期。当每个变量都对预测其余变量起作用时，这组变量适合用VAR模型表示。向量自回归模型对于相互联系的时间序列变量系统是有效的预测模型。同时，向量自回归模型也被频繁地用于分析不同类型的随机误差项对系统变量的动态影响。

向量自回归模型实际上是向量自回归移动平均(VARMA)模型的简化，后者因参数过多带来很多问题而很少应用。最一般的VAR模型数学表达式为：

$$y_t = A_1 y_{t-1} + \dots + A_N y_{t-N} + B_1 x_t + \dots + B_r x_{t-r} + \varepsilon_t$$

这里 y_t 是 m 维内生变量列向量， x_t 是 d 外生变量向量， A_1, \dots, A_N 和 B_1, \dots, B_r 是待估的系数矩阵，内生变量和外生变量分别有 N 和 r 阶滞后期，

而 ε_t 则是误差向量，它是随机扰动项，其同时刻的元素可以彼此相关，但不能与自身滞后值和模型右边的变量相关。

3.2.4 运用 VAR 模型对目标与影响因素之间关系进行实证分析

我们选择就业人数、经济活动人数、居民消费支出、固定资产投资四个指标建立向量自回归 (VAR) 模型进行分析。

(1) VAR模型初步估计

建立 VAR 模型时，所需要做出的几个选择：无约束 VAR (Unrestricted VAR)，暂且将内生变量滞后期定为 2，样本的范围定为 1989~2007，初步模型结果如下：

表 3-25 使用 Eviews6 建立初步 VAR 模型结果

	X1	X2	X3	X5
X1(-1)	0.962602	1.582982	-0.469273	0.595746
X1(-2)	0.232285	-0.002069	-0.625063	0.106479
X2(-1)	0.002969	-0.015964	0.701534	-0.759568
X2(-2)	-0.235514	-0.046874	0.630104	-0.190805
X3(-1)	0.141474	0.029175	0.571310	0.302854
X3(-2)	-0.049918	-0.236922	-0.021922	0.243486
X5(-1)	-0.063781	-0.069570	0.097945	1.683209
X5(-2)	0.012077	0.085456	0.209898	-0.949487
C	2624.824	-30864.50	-12321.48	12864.27
R-squared	0.999702	0.999133	0.999483	0.996455

注释：建立 VAR 模型的详细过程请参见电子稿附录 “M2-3-10”

从初步建模结果中可以看出 R^1 、 R^2 、 R^3 都大于 99.91%、 R^4 大于 99.65%，，单从拟合效果上来说，这个模型系统达到了很理想的效果。

(2) VAR模型滞后期的选择

为了对 VAR 模型滞后期进行选择，进一步分析：

表 3-26 VAR模型的5个评价指标Eviews6分析结果

Lag	LogL	LR	FPE	AIC	SC	HQ
0	-571.7796	NA	3.08e+24	67.73878	67.93483	67.75826
1	-458.9169	159.3356*	3.72e+19	56.34316	57.32341*	56.44060
2	-437.2326	20.40869	2.87e+19*	55.67443*	57.43888	55.84982*

* indicates lag order selected by the criterion

LR: sequential modified LR test statistic (each test at 5% level)

FPE: Final prediction error

AIC: Akaike information criterion

SC: Schwarz information criterion

HQ: Hannan-Quinn information criterion

注释：其中 “*” 表示评价指标对滞后阶数的选择。

五个评价指标中有三个认为应该建立VAR(2)模型，则确认建立VAR(2)模型。

(3) VAR模型平稳性检验

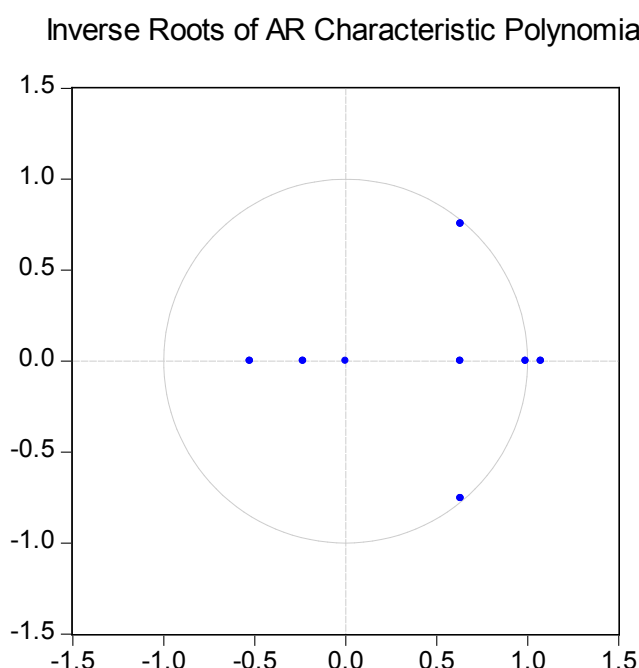


图 3-5 VAR 模型的全部特征根的位置图

从图中我们可以看到，所有的特征根基本都在单位圆内，可以认为这是一个平稳系统。

(4) VAR 模型最终形式

X1、X2、X3、X4 分别代表了就业人数、经济活动人数、居民消费支出、固定资产投资四个指标，它们的函数形式为： <M3-5>

$$X1 = 0.9626 \cdot X1(-1) + 0.2323 \cdot X1(-2) + 0.0030 \cdot X2(-1) - 0.2355 \cdot X2(-2) + 0.1415 \cdot X3(-1) - 0.0500 \cdot X3(-2) - 0.0638 \cdot X5(-1) + 0.0121 \cdot X5(-2) + 2624.8240$$

$$X2 = 1.5830 \cdot X1(-1) - 0.0021 \cdot X1(-2) - 0.0160 \cdot X2(-1) - 0.0469 \cdot X2(-2) + 0.02918 \cdot X3(-1) - 0.2370 \cdot X3(-2) - 0.0696 \cdot X5(-1) + 0.0855 \cdot X5(-2) - 30864.4955$$

$$X3 = -0.4692 \cdot X1(-1) - 0.6251 \cdot X1(-2) + 0.7015 \cdot X2(-1) + 0.6301 \cdot X2(-2) + 0.5713 \cdot X3(-1) - 0.0219 \cdot X3(-2) + 0.0979 \cdot X5(-1) + 0.2099 \cdot X5(-2) - 12321.47780$$

$$X5 = 0.5957 \cdot X1(-1) + 0.1065 \cdot X1(-2) - 0.7596 \cdot X2(-1) - 0.1908 \cdot X2(-2) + 0.3029 \cdot X3(-1) + 0.2435 \cdot X3(-2) + 1.6832 \cdot X5(-1) - 0.9495 \cdot X5(-2) + 12864.2712$$

通过这个函数模型，每个指标长达 19 年的时间序列都可以得到 99.61%以上的说明，且四个指标互相得到了解释，可以运用这个模型对未来进行分析和预测。

3.3 问题三的建立与求解

3.3.1 对问题三的分析

在前面的分析中，我们在劳动力二元供求的假设下，通过对时间序列的分析得出外部因素占整体波动原因的 69% 的结论。随后，我们从宏观经济体系和劳动力本身特点两个方面，运用格兰杰因果模型逐个分析指标与目标之间的关系，判断出哪些指标是对就业波动的原因，而就业波动又是其他哪些指标的变动原因。分析到这里，我们通过对挑选的指标建立回归模型，并且通过了一系列检验，最终确立了合理的形式，一、二问的分析进行到这里实际上已经完成了，但是我们考虑到目标指标对一些指标会产生作用效果，即目标是一些指标的原因，而我们建立的模型是单向回归模型，因此，对于这部分的指标所包含的信息实际上是造成了一种浪费，因此，我们建立向量自回归模型 (VAR 模型) 运用建立的模型实际上是对称模型的特点，充分利用指标 \longleftrightarrow 目标之间的双向的因果关系。

对就业问题进行进一步的思考，我们会发现地区的就业形势存在着很大的差距，按方位来分：中、西、东部的就业形势存在着巨大的差异；按地理特点来分，沿海地区的就业人数大于内陆地区；按级别、政策优惠来分，城市大于城镇、省会大于普通市、经济优惠区大于普通地区。因此，对所有的地区运用相同的模型来对待和分析就业问题必然存在着不合理性，做出的分析结果也难以令人信服。

这种问题不单存在于不同地区之间，同时也存在于不同的行业之间，传统行业、资源消耗型行业、劳动密集型行业对劳动力的需求很大，但是随着技术的改进和时代的进步，一方面，这些行业本身会通过技术改进而减少对劳动者的需求量，或者说提高对劳动者素质的要求从而减少劳动者供给与需求之间的顺利融合，另一方面，根据发展的自然规律，产业结构的合理调整将成为必然的现象（例如，有些行业出于社会的要求而将逐渐的退出，或者随着资源的消耗而进行地区转移），此外，国际间的交流与合作在顺利的开展，并且在可预见的未来，这种形式还将继续扩大，国际间交流的不仅仅只是技术，还有资本的流动与产品的需求，行业将受国际资金的帮助下更快的实现重组和扩张，而国际市场对产品需求也将吸引着企业扩大自己的业务和生产，这样，不同行业由于自身具有的特点将具有不同的就业要求和变化趋势，因此，分行业进行分析很有必要。

3.3.2 分地区对就业问题进行建模和求解

3.3.2.1 分地区对就业问题研究的分析

我们分析一下按地区的不同来分析就业问题所存在的优点和问题。首先，我们得承认这样的一个事实，各个地区的经济发展和就业形势实际上是都不相同的，出于精确分析的角度，我们最好能够地区的单位尽可能的小，对每一个细小的地区单位都根据当地的特点来建立模型，分析当地的就业形势和未来的发展情况。但是，这样又存在着另外一个问题：数据、信息从哪里来？不是所有的地区都有着类似中国统计局那样庞大的统计体系，针对这个问题，有一条解决的方案——调查，通过科学的随机调查能够解决这一问题（我们此次比赛分析中不现实）。另外，“1: 1 的地图是没有意义的”，有些重要指标对细小的地区单位是没有意义的，它们属于国家、大的地区统一的指标，例如利率、汇率等指标是

有特殊的部门确定的，它们的面对对象是全国或较大的地区。

因此，从数据来源的角度出发，我们不能将面对的地区对象设定得过小，至少拥有一套较为完善的统计系统，至少以地级市为限，因为城市拥有的统计局和统计网站能够帮助我们完成此方面的研究。但是，另一方面，地区之间在就业问题上的差异决定了我们很有必要进行一定的地区划分和归类，对拥有就业特征相同或类似的地区统一建立数学模型，用以分析就业问题。我们的思路是：选择与就业问题关联较大的指标作为就业问题地区划分的重要指标，通过合适的聚类分析方法将各地区划分到合适的类别，单独针对各个地区按照问题一、二中的分析方法重新建立一套数学模型分析各类别地区的就业形势和未来发展前景（由于分类后的分析方法与问题一、二类似，出于时间限制，在这里只进行地区的划分）。

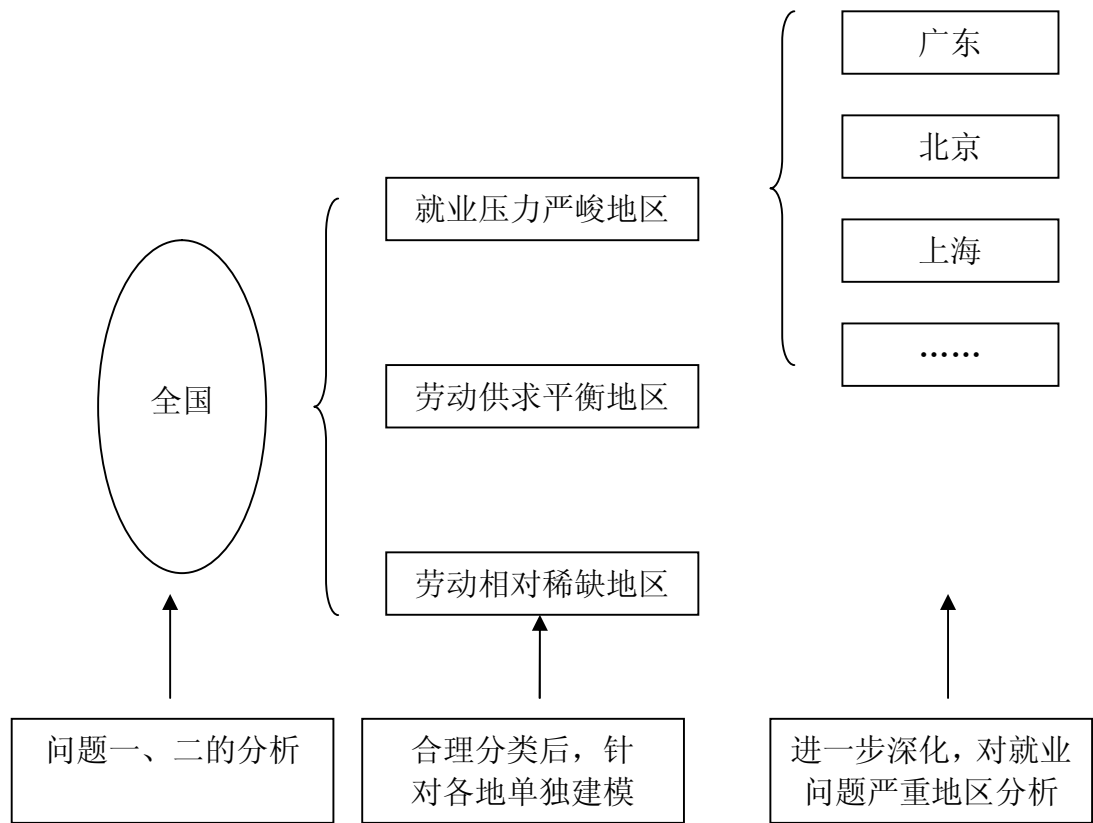


图 3-5 分地区进行分析的流程图

3.3.2.2 运用“星座”图聚类分析法划分地区

聚类分析是研究“物以类聚”规律的一种多元统计方法，以观测数据的样本为对象，根据一定的准则，研究是否可以将其分类，分为几类的问题，它是通过样本数据的相似、相近或相互关联来分类的。我们需要分析出哪些地区在就业压力各指标的特点上相同或类似，即对这些地区进行分类。聚类是一种合适的分类方法。“星座”图聚类分析方法是对每个时间点的各项指标进行权重分配后，将指标的数值合并为平面坐标上的 x 、 y 坐标，从而将各地区按照各自具备的特点定位在平面坐标上，具有相似特点的地区在坐标平面上对应的位置相近，因此，我们可以将聚在一起的时间点定义为一类。

(1) 灰色关联分析法

灰色关联度分析是进行聚类分析的准备工作,通过这种方法得出各指标之间关联度的大小,作为指标之间权重分配的重要依据。

①原始数据变换:

对原始数据进行初值化处理,并求差序列 $\Delta_{0i}(k)$ 及两级差最大差 Δ_{\max} 与最小差 Δ_{\min} ,一般取 $\Delta_{\min}=0$ 。

原始数据 $y(i) = (y_i(1), y_i(2), y_i(3), \dots, y_i(n))(i = 0, 1, \dots, m)$

注: i 代表地区, 范围是 $0 \sim m$, 在这里 $m = 31$ (各省和直辖市的总和为 31)

n 代表指标个数, 在这里 $n = 5$, 指标分别是:

- 1) 就业人数 (用以代表当地可容纳的劳动力数量)
- 2) 劳动报酬 (用以表示各地对劳动力的吸引大小, 通常报酬越大的地区, 潜在劳动力增量越大, 它代表劳动力向该地区流动的力量大小)
- 3) 失业率 (用以代表该地区的就业供求不平衡状况)
- 4) 财政支出 (代表消费与投资)
- 5) 各地生产总值 (用以代表各地的经济发展状况)

表 3-25 地区划分重要指标一览表

地区	就业人数	劳动报酬	失业率	财政支出	生产总值
北京市	5637	28441	1.8	14926380	9353.32
天津市	1976.098	19423	3.6	5404390	5050.40
河北省	4968.664	12481	3.8	7891198	13709.50
山西省	3731.69	12055	3.2	5978870	5730.99
内蒙古	2451.988	13756	4.0	4923615	6091.12
辽宁省	5045.415	13666	4.3	10826948	11023.49
吉林省	2614.36	11816	3.9	3206892	5601.14
黑龙江	4719.688	10982	4.3	4404689	7065.00
上海市	3665.279	26275	4.2	20744792	12188.85
江苏省	6946.193	16192	3.2	22377276	25741.15
浙江省	7150.243	16401	3.3	16494981	18780.44
安徽省	3383.699	12866	4.1	5436973	7364.18
福建省	4520.667	12902	3.9	6994577	9339.51
江西省	2900.266	11059	3.4	3898510	5500.25
山东省	9001.822	13513	3.2	16753980	25965.91
河南省	7154.16	10806	3.4	8620804	15012.46
湖北省	4679.421	11513	4.2	5903552	9550.04
湖南省	4517.741	12093	4.3	6065508	9200.00
广东省	9981.319	16858	2.5	27858007	31084.40
广西	2877.789	12556	3.8	4188265	5955.65
海南省	754.018	10946	3.5	1082935	1223.28
重庆市	2393.813	12921	4.0	4427000	4303.84
四川省	5423.733	12535	4.2	8508606	10505.30
贵州省	2097.973	12133	4.0	2851375	2741.90

云南省	2975.454	11590	4.2	4867146	4741.31
西藏	202.182	21769		201412	342.19
陕西省	3427.83	12113	4.0	4752398	5465.79
甘肃省	1927.12	11569	3.3	1909107	2702.40
青海省	477.326	14533	3.8	567083	783.61
宁夏	612.332	14311	4.3	800312	889.20
新疆	2540.24	11491	3.9	2858600	3523.16

数据来源：2008 中国统计年鉴。

注释：就业人数是 09 年第一季度就业人数，单位是千人；劳动报酬是 09 年第一季度的平均劳动报酬，单位是元；失业率是城镇失业率，单位是%；财政支出单位是万元，生产总值单位是亿元。

初值化得

$$x_i = \left(\frac{y_i(1)}{y_i(1)}, \frac{y_i(2)}{y_i(1)}, \frac{y_i(3)}{y_i(1)}, \dots, \frac{y_i(n)}{y_i(1)} \right) = (x_i(1), x_i(2), x_i(3), \dots, x_i(n)) (i = 0, 1, \dots, m) \quad \langle 1 \rangle$$

$$\Delta_{0i}(k) = |x_0(k) - x_i(k)| (i = 1, 2, \dots, m; k = 1, 2, \dots, n)$$

$$\textcircled{2} \text{ 计算关联系数 } \xi_{0i}(k) = \frac{\Delta_{\min} + \Delta_{\max}}{\Delta_{0i(k)} + \rho \Delta_{\max}}, (0 < \rho < 1) \quad \langle 2 \rangle$$

$$\textcircled{3} \text{ 求关联度 } r_{0i} : r_{0i} = \frac{1}{n} \sum_{k=1}^n \xi_{0i}(k) (i = 1, 2, \dots, m) \quad \langle 3 \rangle$$

④列关联矩阵

⑤优势分析：根据各行或各列关联度的大小判断主要影响因素即优势因素和次要影响因素。

(2) “星座”图聚类法

运用“星座”图聚类法对各年份进行划分。“星座”图聚类的基本原理是：将每个时间点按一定的数量关系定在一个半圆之内，每一个样点用一个星点表示，同类的样点便可以组成一个“星座”，然后归类并区分不同“星座”的界线，从而得到一个“星座”类型图。在本文的研究中以 WTI、油价增量、同比增长、环比增长指标作为参考指标，将每个时间点作为一个样点，在半圆坐标中组成“星座”，聚类法的具体过程如下：

首先，观测数据进行极差标准化处理，按下式：

$$\Phi_{ij} = \frac{x_{ij} - x_{j\min}}{x_{j\max} - x_{j\min}} \times \pi \quad \langle 4 \rangle$$

式中 $i=0, 1, 2, 3, \dots, n$ ，为样点数（在此表示每一个地区）； $j=1, 2, 3, \dots, p$ ，为指标数（在此表示各指标）； x_{ij} 为原始数据； $x_{j\max}, x_{j\min}$ 分别为第 j 个因子的最大值和最小值； Φ_{ij} 为标准化后的数据。

然后，在以上基础上按下列各式计算各样点的坐标值、综合指标值：

$$X_i = \sum_{j=1}^p w_j \cos \phi_{ij}$$

$$Y_i = \sum_{j=1}^p w_j \sin \phi_{ij}$$

$$Z_i = \sum_{j=1}^p \phi_{ij} w_j \quad (0 \leq w \leq 1, \text{且} \sum_{j=1}^p w_j = 1) \quad \langle 5 \rangle$$

式中 X_i 为第 i 个样点的横坐标； Y_i 为第 i 个样点的纵坐标； w_j 为第 j 个指标的权重值； Z_i 为第 i 个样点的综合指标值。

再根据 X_i 、 Y_i 的值确定每个样点在图内的星点位置，并依次作出全部样点位置的坐标平面，性质相似和接近的样点代表星点聚在一起形成一个“星座”型图。

3.3.2.3 运用聚类分析方法对各地进行划分的结果

(1) 各指标的关联分析

将原始数据处理后，运用灰关联分析方法分别计算各指标间的关联度，并列出可关联矩阵，由于 $\langle 2 \rangle$ 式中的 ρ 需要视情况而定，这一步骤可以运用 Eviews6 相关关联度直接代替。结果见下列各表：

表 3-26 各指标相关关联度分析

指标	就业人数	劳动报酬	失业率	财政支出	生产总值
就业人数	1.0000	0.1849	-0.4509	0.8247	0.9372
劳动报酬	0.1849	1.0000	-0.4691	0.5721	0.2454
失业率	-0.4509	-0.4691	1.0000	-0.5005	-0.4410
财政支出	0.8247	0.5721	-0.5005	1.0000	0.9069
生产总值	0.9372	0.2454	-0.4410	0.9069	1.0000

由以上各表各行得各因子与其它因子的平均关联度大小。列于下表：

表 3-27 各指标平均相关关联度表

指标	就业人数	劳动报酬	失业率	财政支出	生产总值
平均关联度	0.4992	0.3066	-0.1723	0.5606	0.5297

注释：失业率与其他指标之间呈现负相关的关系，代表在体现就业压力这一问题上与其他指标呈现相反方向变动的特点，属于正常状况。

(2) 区域划分中因子的权重分配

对上表中因子的平均关联度进行归一化，将其视为最后划分区域时考虑的权重，结果如下表所示：

表 3-28 对指标进行权重分配结果

指标	就业人数	劳动报酬	失业率	财政支出	生产总值
权值分配	0.2896	0.1779	-0.1000	0.3252	0.3073

注释：此权值为划分类别时所考虑的各指标的权重分配，其中失业率的权重分配为负值，代表当考虑失业率这一指标时，应当加上一个负号才能与其他指标的判断方向相同。

(3) 聚类分析分类结果

参照公式<4>、<5>按步骤计算出横纵坐标 X、Y 和综合指标值 Z，并按横纵坐标将各地区列在平面坐标轴上。 <M3-6>

表 3-29 各地聚类分析后的 X、Y 坐标

地区	X	Y	地区	X	Y
北京市	0.0203	-0.1755	湖北省	0.4630	0.7546
天津市	0.6080	0.8556	湖南省	0.4766	0.7716
河北省	0.3077	0.5364	广东省	-0.6441	-0.9030
山西省	0.6605	0.8385	广西	0.6902	0.9896
内蒙古	0.6778	0.9994	海南省	0.8667	1.1451
辽宁省	0.2643	0.5170	重庆市	0.7204	1.0496
吉林省	0.7191	1.0417	四川省	0.3214	0.5853
黑龙江	0.5324	0.8386	贵州省	0.7726	1.1112
上海市	-0.2107	-0.0535	云南省	0.6804	1.0147
江苏省	-0.5011	-0.5731	西藏	0.6878	0.8861
浙江省	-0.2438	-0.2392	陕西省	0.6459	0.9638
安徽省	0.6017	0.9139	甘肃省	0.8373	1.0752
福建省	0.4586	0.7254	青海省	0.8191	1.1380
江西省	0.7407	0.9665	宁夏	0.8023	1.1626
山东省	-0.4184	-0.4638	新疆	0.7525	1.0769
河南省	0.1381	0.2527			

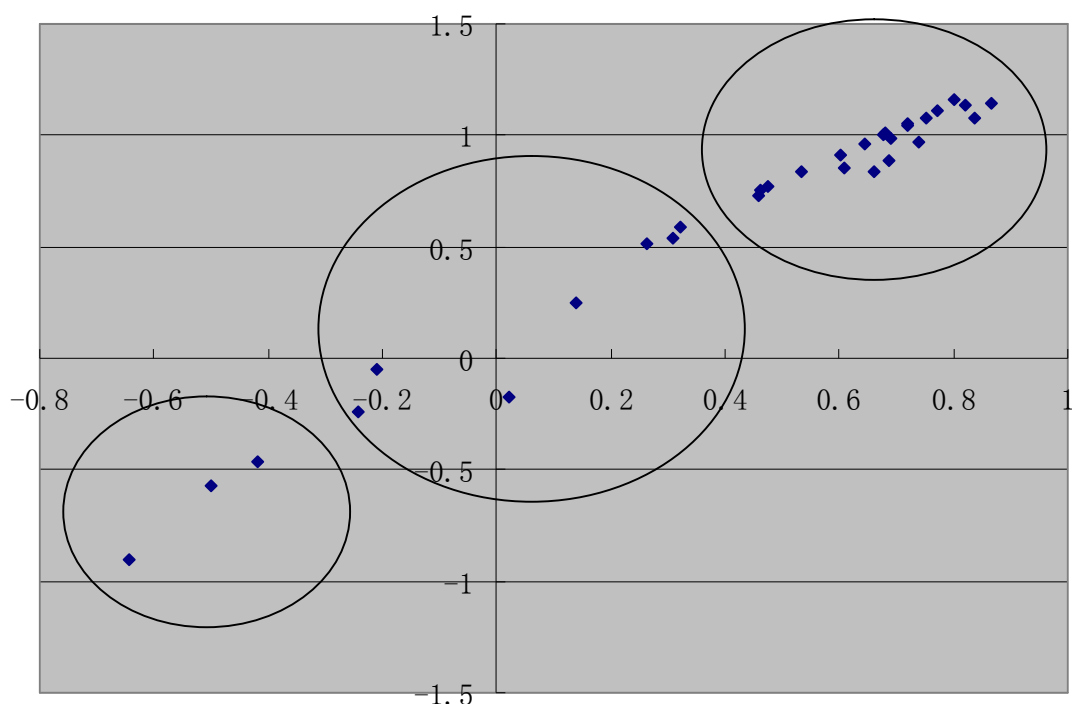


图 3-6 “星座”聚类分析法对各地区聚类后的结果如图所示

从图中我们可以看出，即使是通过各类指标希望对地区就劳动力就业这一问题进行合理的聚类划分，但是我们可以从最终的聚类结果图中可以看出，31 个省份和地区的就业情况并不是很的显现出“就业压力严重区”、“劳动供求平衡区”、“劳动相对稀缺地区”三类，各地区之间呈现出递变的特征，并且具有很明显的线性特征。

我们将所有的按就业特点聚类后在坐标轴上显示的结果按照从左至右列表如下，在地理位置允许的前提下，可以认为就业的压力降呈现梯度降低的形式，而未来人口的流动很有可能仍然将保持如此的特征。

表 3-30 三类地区划分结果

第一类	就业压力严重地区(从左到右就业形势依次递减)
	广东省、江苏省、山东省
第二类	劳动供求平衡区(从左到右就业形势依次递减)
	浙江省、上海市、北京市、河南省、辽宁省、河北省、四川省
第三类	劳动供求平衡区(从左到右就业形势依次递减)
	福建省、湖北省、湖南省、黑龙江、安徽省、天津市、陕西省、山西省、内蒙古、云南省、西藏、广西、吉林省、重庆市、江西省、新疆、贵州省、宁夏、青海省、甘肃省、海南省

地区的划分实际上是根据平面坐标图上点的位置紧密与否来确定的，具有一定的主观性，因此，某个边界点属于上一类还是下一类是可以进一步讨论的。但是，各个地区的左右顺序的排列却是固定的，我们可以看到广东、江苏、浙江、上海、北京等经济发达地区都排在对就业人数具有吸引力的前列，符合我们普遍的看法，但是又与我们通常的观念有所不同，北京、上海等地并没有凸显就业压力过大的问题，这里存在着一定的博弈问题，经济发达地区一方面对劳动力的供

给方有着很强的吸引力,但是同时,由于本地经济的迅速发展企业的扩张将进入的劳动力顺利吸收,使得就业的情况并没有上升到压力过重的级别。

在对地区进行分类的工作结束之后,我们需要将各类地区看做一个整体,对其建立模型分析就业形势与未来走势,这一部分的工作类似于对一、二问题的处理,技术上的难度已经公平,由于时间有限,在此略过。

3.3.3 分行业对就业问题进行建模和求解

首先,我们观察一下分行业主要指标的统计系统所具有的特点,其后神经网络模型建立的数据源就来自于下面这张表。

表 3-31 按行业分,2007 规模以上工业企业主要指标(行业就业指标在最后一列)

行业	企业单位数	工业总产值	工业增加值	资产总计
煤炭开采和洗选业	7537	9201.83	4696.33	13864.21
石油和天然气开采业	184	8300.05	6450.83	9930.78
黑色金属矿采选业	2899	2130.61	928.78	1986.09
有色金属矿采选业	2183	2288.75	973.32	1813.33
非金属矿采选业	3004	1365.63	517.24	969.95
其他采矿业	24	10.97	3.27	3.40
农副食品加工业	18140	17496.08	4642.45	8798.13
食品制造业	6644	6070.96	1861.56	4415.77
饮料制造业	4422	5082.34	1883.66	4902.58
烟草制品业	150	3776.23	2918.82	3767.93
纺织业	27914	18733.31	4913.92	13734.66
纺织服装、鞋、帽制造业	14770	7600.38	2265.11	4559.15
皮革、毛皮及其制品业	7452	5153.49	1480.39	2648.94
木材加工及木、竹、藤、	7852	3520.54	1030.29	2048.13
家具制造业	4110	2424.94	646.76	1655.92
造纸及纸制品业	8376	6325.45	1743.05	6115.13
印刷业和记录媒介的复制	5083	2117.57	691.94	2257.69
文教体育用品制造业	4087	2098.79	554.57	1366.39
石油加工等加工业	2149	17850.88	3096.98	9398.79
化学原料及化学制造业	22981	26798.80	7340.42	22420.71
医药制造业	5748	6361.90	2286.60	6916.55
化学纤维制造业	1556	4120.80	809.43	3462.37
橡胶制品业	3695	3462.41	959.00	2749.66
塑料制品业	15376	8120.41	2137.14	5966.67
非金属矿物制品业	24278	15559.44	4849.19	13971.51
黑色金属冶炼加工业	7161	33703.01	9007.14	29097.44
有色金属冶炼加工业	6701	18031.88	4477.61	11407.18
金属制品业	18008	11447.08	3010.41	7494.01
通用设备制造业	26757	18415.52	5107.54	14868.07
专用设备制造业	13409	10591.98	3067.36	9962.73
交通运输设备制造业	14091	27147.40	6974.48	25189.96
电气机械及器材制造业	19322	24019.07	6053.78	16411.68

通信设备、计算机及其他	11220	39223.77	7924.57	24376.20
仪器仪表及文化、办公用	4526	4307.99	1163.25	3137.87
工艺品及其他制造业	6416	3387.71	917.62	2129.01
废弃资源和废旧加工业	652	680.71	162.05	272.27
电力、热力的生产业	5565	26462.65	8828.89	53484.80
燃气生产和供应业	591	988.72	306.67	1632.61
水的生产和供应业	1735	797.08	365.96	3849.09

接上表右侧：

负债总计	所有者权益	主营业务收入	主营业务成本	利润总额	从业平均人数
8477.42	5386.79	9593.08	6609.03	1022.18	463.69
3835.79	6094.99	8497.14	4029.22	3535.41	90.67
959.72	1026.37	2072.41	1454.39	349.03	49.14
868.84	944.49	2239.39	1567.49	428.12	55.11
490.64	479.31	1295.58	972.21	108.97	46.62
1.94	1.46	9.80	8.08	0.42	0.26
4997.36	3800.77	17131.12	14907.57	893.58	264.80
2331.32	2084.47	5853.09	4585.18	394.22	135.03
2613.06	2289.51	4993.98	3541.20	445.43	101.02
932.01	2835.87	3737.61	1224.23	608.35	18.61
8257.94	5476.68	18164.42	16076.30	765.87	626.26
2517.54	2041.67	7335.75	6228.09	357.13	414.19
1499.71	1149.23	4967.63	4301.28	255.56	256.98
1090.90	957.23	3375.84	2865.02	192.96	106.18
923.31	732.61	2360.52	1977.99	108.96	91.30
3596.31	2518.83	6151.49	5214.95	381.23	138.30
1161.89	1095.81	2039.27	1662.05	155.88	72.38
731.34	635.05	2029.44	1746.11	73.32	119.32
5317.83	4069.27	17952.50	16449.80	216.25	80.64
12387.14	9985.71	26482.84	22010.77	1834.34	380.28
3377.33	3530.36	5967.13	4065.98	581.28	137.34
2106.29	1356.08	3981.47	3595.94	161.02	45.30
1607.69	1141.97	3380.39	2887.90	177.90	87.51
3352.86	2613.81	7895.38	6802.69	402.39	224.05
7931.12	6040.29	14987.46	12390.43	1037.19	448.41
17737.42	11352.84	34930.91	30780.43	2087.48	304.43
6663.75	4737.26	17917.04	15737.56	1177.00	156.27
4413.32	3080.69	11100.92	9638.43	532.60	273.48
8866.39	5999.66	17837.30	14843.15	1172.25	420.71
5885.77	4073.78	10265.54	8378.31	774.58	256.51
15853.95	9335.86	26637.10	22348.57	1685.08	408.59
9991.08	6420.59	23213.79	19759.40	1233.35	449.15
15085.36	9290.50	39014.14	34829.60	1445.89	587.92
1673.71	1464.16	4205.44	3522.87	278.66	106.97
1190.92	938.09	3297.86	2834.27	168.22	136.94

180.66	91.64	682.51	603.07	24.44	6.64
31104.94	22219.86	26236.07	22633.70	1982.22	256.96
918.42	714.19	1137.32	975.42	75.52	15.88
1980.70	1868.39	746.40	539.94	30.89	41.36

数据来源：2008 中国统计年鉴

3.3.3.1 分行业建模的分析

我们从以上该表看出分行业的经济主要指标的统计系统（由于我国第二产业比例高达 50%以上，作为就业人数主要去向，我们主要分析工业中的 39 个行业），行业的划分和指标的选择较为统一，如果数据来源通畅的话，我们可以将历年的分行业统计体系整理完毕，甚至能够在较长的时间段的基础之上，将各个地级市的分行业数据搜集完成，这样，我们就能够进行分地区、分行业对各个身为时间序列的指标之间的关系进行分析，计量经济学中对此这种面板数据的分析方法叫做 Panel Data 模型，我们首先运用该模型来分析行业的面板数据。另外，对于这类没有较好拟合模式的复杂数据源，我们建立神经网络模型来模拟就业人数与其他行业指标之间的非线性映射关系。

3.3.3.2 Panel Data 模型分析原理

Panel Data（面板数据）是指对不同时刻的界面个体进行连续观测所得到的多维时间序列模型。面板数据在 Eviews 中也被称为时序与界面合成数据(pooled time series and cross-section data)。面板数据模型是一类利用混合数据分析变量间相互关系并预测其变化趋势的计量模型。它能够同时反映演技对象在时间和截面单元两个方向上的变化规律及不同时间、不同单元的特性。面板数据模型综合利用样本信息，并且可以减少多重共线性的影响。^[7]

模型的一般形式为：

$$y_{it} = \alpha_{it} + x_{it} \cdot \beta_{it} + \varepsilon_{it}, \quad i = 1, 2, 3 \dots N, \quad t = 1, 2, 3 \dots T$$

其中， x_{it} 为解释变量向量， i 代表了不同个体(地区)， t 代表时间(月、季度或年)， α_{it} 、 β_{it} 分别为截距向量及参数向量。模型中的系数随着时间和个体的不同而改变，因而可以反映模型中被忽略的时间因素和个体差异因素的影响。

(1) 变系数模型

假定时间序列参数齐性，即参数满足时间一致性，也就是参数值不随时间的不同而改变，则一般形式可以写出：

$$y_{it} = \alpha_i + x_{it} \cdot \beta_i + \varepsilon_{it}, \quad i = 1, 2, 3 \dots N, \quad t = 1, 2, 3 \dots T$$

(2) 变截距模型

假定斜率系数 $\beta_1 = \beta_2 = \dots = \beta_N$ ，但截距不同，即区域之间经济差异只表现在截距上，则一般形式可以写出：

$$y_{it} = \alpha_i + x_{it} \cdot \beta + \varepsilon_{it}, \quad i = 1, 2, 3 \dots N, \quad t = 1, 2, 3 \dots T$$

(3) 混合回归模型

假定截距和斜率系数都是常数，则得到：

$$y_{it} = \alpha + x_{it} \cdot \beta + \varepsilon_{it}, \quad i = 1, 2, 3 \dots N, \quad t = 1, 2, 3 \dots T$$

这个模型中被忽略的个体差异因素对截距项 α 和系数向量 β 都无影响，此时相当于将 T 个时期的截面数据融合成一个“混合样本”，利用普通最小二乘法来求出参数 α 和 β 即可。

3.3.3.3 运用 Panel Data 模型进行实证分析

表 3-32 全国按主要行业分从业人员平均人数季度统计（单位：万人）

行业	2003. 1	2003. 2	2003. 3	2003. 4	2004. 1	2004. 2
煤炭开采和洗选业	382. 82	375. 24	376. 97	377. 13	427. 11	378. 84
石油和天然气开采业	59. 83	60. 29	72. 26	65. 05	83. 82	71. 88
黑色金属矿采选业	25. 19	23. 6	24. 08	24. 43	27. 01	31. 89
有色金属矿采选业	56. 71	41. 47	41. 46	41. 66	38. 11	39. 39
非金属矿采选业	42. 77	44. 98	45. 84	45. 83	45. 25	44. 83
其他采矿业	1. 76	1. 76	1. 78	1. 78	0. 13	0. 11
农副食品加工业	167. 59	171. 05	173. 38	178. 96	180. 24	182. 69
食品制造业	90. 52	93. 08	95. 05	97. 15	97. 44	100. 51
饮料制造业	88. 96	89. 59	89. 85	90. 21	85. 2	90. 52
烟草制品业	29. 45	20. 78	20. 25	20. 79	20. 47	19. 95
纺织业	472. 53	483. 79	489. 95	495. 72	491. 29	500. 97
纺织服装、鞋、帽制造业	257. 92	265. 67	275. 59	282. 77	283. 05	300. 06
皮革、毛皮及其制品业	134. 82	145. 64	151. 95	155. 1	156. 88	168. 25
木材加工及木、竹、藤、家具制造业	60. 66	58. 44	59. 08	62. 79	61. 9	64. 22
造纸及纸制品业	35. 1	36. 11	37. 56	38. 99	45. 12	48. 23
印刷业和记录媒介的复制	110. 8	111. 42	113. 34	114. 92	113. 79	114. 76
文教体育用品制造业	55. 58	56. 16	56. 99	58. 17	58. 31	60. 36
石油加工等加工业	71. 26	77. 16	80. 11	81. 12	85. 35	89. 52
化学原料及化学制造业	55. 68	57. 47	57. 76	58. 14	59. 35	59. 31
医药制造业	306. 6	301. 81	305. 47	304. 36	303. 62	308. 94
化学纤维制造业	107. 93	111. 31	113. 56	112. 83	114. 77	121. 69
橡胶制品业	34. 81	34. 62	34. 84	35. 43	35. 42	38. 11
塑料制品业	60. 36	60. 94	62. 29	62. 05	61. 83	63. 03
非金属矿物制品业	126. 34	131. 71	133. 71	137. 01	143. 8	144. 49
黑色金属冶炼加工业	372. 83	381. 61	387. 01	391. 87	396. 95	411. 02
有色金属冶炼加工业	241. 75	243. 49	244. 82	248. 05	251. 7	252. 18
	110. 8	102. 26	102. 51	103. 94	110. 49	111. 51

金属制品业	154.5	160.37	163.56	167.88	172.01	178.89
通用设备制造业	269.49	273.72	278.25	279.88	286.81	294.27
专用设备制造业	205.16	206.86	213.58	216.02	201.87	204.5
交通运输设备制造业	301.99	302.88	305.73	310.73	311.77	319.76
电气机械及器材制造业	233.41	245.16	253.77	258.8	282.17	280.48
通信设备、计算机及其他	236.7	242.65	254.82	266.4	308.1	307.92
仪器仪表及文化、办公用	65.55	69.77	71.51	73.32	75.41	74.29
工艺品及其他制造业	88.53	94.28	97.51	100.18	104.2	104.1
废弃资源和废旧加工业	0.71	0.81	0.84	0.83	1.52	1.57
电力、热力的生产业	234.58	226.72	229.29	231.96	250.2	242.48
燃气生产和供应业	15.16	14.57	14.48	14.42	14.44	14.38
水的生产和供应业	45.15	45.14	45.48	45.73	45.92	46.65

接上表右侧：

2004.3	2004.4	2005.1	2005.2	2005.3	2005.4	2006.1	2006.2	2006.3	2006.4
380.73	387.57	401.8	412.9	419.5	425.7	422.1	433.4	440.1	442.1
71.08	70.87	75.1	81.5	81.9	82.4	136.9	82.4	82.8	83.1
28.67	28.85	32.8	35.6	36.4	37.4	35.9	40.8	41.8	43
38.54	39.24	35.7	37.9	39	39.9	38.8	41.1	42.2	43.3
45.44	45.23	40.2	40.3	40.9	42	39.8	41	42.2	43.6
0.11	0.14	0.1	0.3	0.3	0.3	0.1	0.2	0.1	0.1
182.78	188.74	189.3	198.4	204	213.5	209.4	218.4	225.1	232.4
101.86	104.26	102	106	109.1	113.9	116	121.1	123.5	125.4
87.44	87.31	82.5	84.9	85.1	85.5	84.9	87.2	89	90.1
19.52	19.81	19.3	18.9	18.1	18.7	18.5	18.3	17.8	18.6
507.66	516	505.4	548.3	564.1	572.5	573.4	592.3	602.2	605.8
306.09	314.7	307	317.1	329.4	332.3	333.2	355	363.4	367
173.53	178.23	176.6	203.5	211.8	219.1	211.7	227.4	232.2	236
65.98	68.33	69	74.8	77	79.8	78	81.8	84.7	87.7
49.71	51.56	55.7	64.3	66.1	68.8	71.4	75.6	76.3	78.3
115.3	116.45	115.1	124.9	127.9	129.7	125.8	128.8	130.2	131.7
60.28	61.08	59	61.2	63.3	64.2	64.4	66.4	67.4	68.7
92.73	92.92	93.9	102.4	106.6	106.2	104.7	110.6	114	113.4
60.38	62.07	67.8	68.7	68.6	69.2	72.6	75.8	73.9	74.4
308.96	313.42	309.7	318.9	326.8	329.1	326.7	336.9	343.3	348
116.27	117.29	115.2	117.7	119.6	121	122.1	125.3	126.9	127.7
38.3	38.91	35.5	38.9	40.4	40.6	40.9	40.8	41.3	41.3
64.03	64.45	62.9	74.4	75.5	77.3	75.6	78.9	80.3	81.4
146.94	149.79	151	168.4	172.7	177.2	179.2	186.8	191	193.8
401.2	404.69	375.2	397.3	404.3	409.4	386.7	405.6	414.7	420.7
255.18	259.73	265.6	274.2	278.3	279.8	276.7	284.4	295.2	291
114.48	115.08	115.5	124.4	126.7	128.6	127.5	131.4	134.7	137
182.98	187.32	189.5	201	208.1	213.4	215.9	228.5	233.7	238.5
296.73	303.04	309.8	327.4	336	341.2	347.3	360.7	368	373.5
205.44	206.13	209.5	213	218	222.2	215.2	220.6	225.6	232.5

321.94	325.73	324.4	334.1	338.7	345.4	345.7	357.4	364.1	370
290.58	295.06	303.5	340.3	351.6	360.6	360.8	376.8	386.5	394.5
316	329.7	348.3	392.7	408.6	429.3	441.3	472.7	488.4	497.2
76.1	77.84	78.1	81.6	85.2	87.3	88.8	91.6	93	95.7
118.15	107.68	103.4	114	118.1	120.6	116.4	125.2	129.5	130.2
1.55	1.55	2.2	3	3.2	3.4	3.7	4.2	4.5	4.8
235.01	236.64	240.4	243.8	243.9	247.2	245.9	245.8	248.8	248.3
14.32	14.43	14.4	14.8	14.9	14.9	13.5	14.4	14.5	14.5
46.01	46.17	45.1	45.9	46.1	46	45.6	46.3	46.3	46.4

接上表右侧:

2007.1	2007.2	2007.3	2007.4	2008.1	2008.2	2008.3	2008.4	2009.1
437.5	451.78	457.54	459.73	458.2	462.4	471.9	474.07	465.21
87.94	90.12	89.65	89.44	86.6	85	91.1	93.8	101.7
40.03	43.51	44.71	46.76	43.8	47.8	49.4	50.78	46.34
40.73	47.71	47.76	49.22	48	50.1	50.1	49.61	42.15
41.6	43.1	44.11	45.1	43.6	45.1	46.2	46.63	45.22
0.13	0.14	0.14	0.14	0.3	0.3	0.3	0.31	0.28
228.17	236.25	244.77	253.81	249.5	254.7	264.1	270.65	275.18
123.66	128	130.46	133.84	129.9	133.8	137.3	139.81	136.53
104.84	93.4	95.5	97.65	101.2	104.6	104.8	107.41	103.1
26.27	18.59	17.54	17.98	17.8	18.9	17.9	18.39	18.48
594.15	612.08	618.62	631.54	601.8	607.5	612.2	614.61	578.09
361.07	378.86	389.73	396.82	384.7	401.9	404.3	406.7	398.24
220.02	233.64	242.28	247.22	240.4	251.4	255	254.62	233.5
86.49	92.09	94.63	99.5	101.8	104.1	108.1	110.7	110.25
75.25	79.86	83.79	85.69	89.7	91.6	91.7	92.14	85.81
127.14	131.31	133.76	136.37	132.6	136.3	139.1	140.7	131.54
65.73	66.55	69.48	70.54	69.9	72.2	73	76.84	71.71
102.11	108.79	114.21	115.47	113.3	118.9	120.7	132.87	105.64
74.56	77.35	78.67	78.91	78.5	79.7	81.2	80.89	77.09
342.51	355.92	362.44	369.08	535.2	376.1	387.4	392.88	383.8
128.21	130.35	131.8	133.21	134.5	135.8	138.7	140.4	139.67
42.72	43.94	43.83	44.58	43.9	44.6	45.3	45.19	42.3
79.97	82.71	83.93	86.2	84.5	87.3	88.7	87.96	85.29
189.98	198.93	204.9	209.18	206.8	218.3	220.5	222.78	207.82
405.39	422.77	430.08	440.13	421.9	436.1	443.3	449.16	427.19
288.87	294.5	298.15	311.21	301.7	300.5	307.4	306.58	292.8
136.65	140.71	144.42	148.23	150.2	158.6	163	201.05	157.84
237.35	247.78	254.61	259.71	260.9	269.9	277	278.08	274.98
378.53	389.94	399.41	406.8	411.8	426	434.2	436.2	434.14
227.67	235.01	240.42	246.51	255.7	258.1	264.4	266.99	269.44
375.55	395.29	398.15	406.54	410.7	424.1	434.9	438.37	434.96
382.75	404.45	418.13	431.33	446.7	465.1	473.3	473.75	449.22
489.1	511.38	537.39	566.07	616.7	610.3	627	634.6	541.69

93.39	96.48	99.2	102.38	105.3	110.5	112	112.52	102.26
124.17	130.29	133.42	134.78	126.3	131.2	133.9	132.03	118.62
5.41	5.73	6.19	6.33	5.9	6.6	6.8	6.88	7.66
249.74	271.64	247.14	250.23	261.1	246.8	252.3	255.09	255.52
14.26	14.6	14.63	14.99	15.2	15.8	16.1	16.14	17.1
40.33	40.93	40.94	41.02	41.2	41.6	41.5	41.73	42.46

表 3-33 全国按主要行业分主营业务收入季度统计（单位：亿元）

	2003.1	2003.2	2003.3	2003.4	2004.1	2004.2
煤炭开采和洗选业	302.03	867.09	1490.89	2160.49	420.66	1297.46
石油和天然气开采业	541.98	1376.31	2188.25	3021.6	579.39	1540.79
黑色金属矿采选业	28.08	96.87	185.14	279.44	61.78	216.23
有色金属矿采选业	62.13	195.07	340.97	495.93	74.05	245.52
非金属矿采选业	48.31	158.28	281.08	402.24	73.26	206.08
其他采矿业	0.88	3.59	7.91	11.33	0.2	0.6
农副食品加工业	713.57	1991.92	3428.88	5083.93	917.07	2682.96
食品制造业	287.72	784.93	1298.58	1888.19	346.23	970.64
饮料制造业	320.9	805.38	1384.44	1919.77	368.26	970.68
烟草制品业	417.28	936.87	1469.08	2046.5	466.33	1078.97
纺织业	874.86	2603.61	4507	6608.76	1076.33	3256.46
纺织服装、鞋、帽制造业	388.91	1111.82	1914.67	2806.49	464.92	1334.26
皮革、毛皮及其制品业	224.17	661.03	1197.51	1792.08	287.17	852.7
木材加工及木、竹、藤、	106.79	335.52	576.76	845.91	131.86	421.45
家具制造业	83.64	227.32	386.67	565.15	111.04	326.73
造纸及纸制品业	294.8	864.14	1491.16	2156.61	388.92	1109.73
印刷业和记录媒介的复制	119.81	335.4	561.15	818.76	149.91	417.85
文教体育用品制造业	98.46	291.87	509.01	759.18	129.77	386.1
石油加工等加工业	939.15	2443.27	3968.13	5632.24	1153.46	3132.3
化学原料及化学制造业	1133.23	3294.62	5508.98	7896.88	1479.26	4351.29
医药制造业	372.54	1066.46	1696.66	2440.32	432.28	1223.45
化学纤维制造业	186.46	474.64	820.36	1199.31	215.09	671.99
橡胶制品业	138.67	429.68	747.87	1073.96	180.72	570.24
塑料制品业	357.61	1047.68	1777.72	2566.17	434.67	1327.86
非金属矿物制品业	555.62	1748.98	3101.87	4609.16	750.82	2409.35
黑色金属冶炼加工业	1180.5	3431.13	6038.41	8869.38	2138.09	5867.27
有色金属冶炼加工业	396.73	1188.19	2077.58	3037.33	627.46	1950.43
金属制品业	431.21	1275.61	2204.9	3214.84	555.25	1734.08
通用设备制造业	604.16	1872.76	3224.48	4662.02	843.8	2700.56
专用设备制造业	401.46	1309.88	2274.84	3270	522.89	1698.69
交通运输设备制造业	1329.04	4038.03	6761.4	9633.22	1594.25	5138.42
电气机械及器材制造业	850.85	2642.17	4519.23	6555.02	1156.45	3578.94
通信设备、计算机及其他	1853.74	4980.71	8834.3	13529.52	2664.56	7677.27
仪器仪表及文化、办公用	196.82	543.19	939.17	1426.41	266.97	751.69
工艺品及其他制造业	152.46	434.57	757.76	1086.59	181.45	521.11

废弃资源和废旧加工业	4.73	14.02	24.97	36.96	8.95	27.31
电力、热力的生产业	1485.91	3975.46	6724.89	9477.47	1863.32	5065.8
燃气生产和供应业	61.41	155.38	240.15	334.17	75.44	193.33
水的生产和供应业	55.88	147.2	252.16	365.05	65.38	173.48

接上表右侧：

2004.3	2004.4	2005.1	2005.2	2005.3	2005.4	2006.1	2006.2
2262.08	3369.96	683.3	2052.5	3527.9	5136.8	848.1	2518.3
2577.81	3802.07	717.5	2133.8	3757.1	5416	1080.9	2988.2
368.89	523.75	96.5	322.7	573.1	843.3	112.3	406.7
432.66	663.97	116.4	370.5	653.3	951	172.4	546.6
363.61	528.59	84.9	250.6	447.8	658.2	101.3	323
0.97	1.67	0.3	3.7	5.9	8.2	0.6	1.6
4649.5	6870.73	1228.3	3540.2	6195.8	9119.5	1541.4	4434.6
1628.16	2402.93	451.2	1258.5	2140.3	3149.3	612.2	1679.9
1651.1	2272.09	429.2	1152.9	1985	2746	556.6	1442.6
1712.27	2379.08	565	1224.2	1881	2621	609.3	1361.7
5671.94	8308.09	1346.4	4213.5	7385	10774.3	1764.4	5280.8
2329.18	3426.05	582	1604.1	2810.2	4105.9	719.9	2043.3
1540.42	2294.96	341.5	1068.8	1950	2889.5	486.5	1385.1
726.64	1073.23	163.5	566.4	1019.4	1523.6	231.2	768.9
549.82	793.8	153.2	485.6	808.6	1178	224.4	651.1
1859.67	2672.1	470.1	1411.8	2421.2	3569.5	617	1801.7
693.21	1007.49	178.2	486.4	824.1	1198.1	211.4	592.8
673.76	990.07	165.8	483.4	852.5	1251.1	201.3	602.9
5301.6	7748.82	1570.4	4329.5	7286.5	10560.3	2058	5670.4
7359.5	10757.85	1972.8	5878.8	10016.3	14377.2	2444.8	7305.2
1993.22	2852.78	508.8	1452.7	2451	3537	625.1	1755.1
1143.32	1684.71	308.2	925.1	1596	2286.6	399	1158.1
971.07	1376.83	231.8	768.6	1306.8	1880	324.1	974.2
2284.18	3278.61	542.9	1710.4	2983.1	4306.3	734.4	2183.8
4217.19	6169.61	901.4	2911.8	5225.6	7730.3	1171.3	3797.8
9699.4	14219.57	2848.6	8419.3	13889	19326.7	3062.6	9268.3
3334.7	4813.75	872.4	2735.4	4707.7	6844.8	1335.4	4348.9
3034.51	4443.35	758.9	2238	3857.9	5578.3	915.5	2809.6
4560.28	6592.23	1102.9	3514	6090.5	8744.2	1481.1	4677.2
2808.05	3975.03	665.9	2087.7	3597.1	5234.4	851	2704.4
8306.47	11659.66	1776.8	5511.4	9208.3	13199.4	2372.1	7352.3
6275.72	8950.71	1504	4728.8	8157	11840.2	1947.8	6061.5
12909.66	19149.7	3324	9026.5	15544.8	23109.5	4318.4	11844.5
1281.51	1882.96	371.4	955.4	1650.6	2436.4	396.3	1161.6
897.81	1306.53	216.4	659.8	1148.6	1691.9	287.7	856.2
45	67.12	15	83.2	139	204.9	39.3	134.2
8505.9	12052.79	2465.7	6511.2	11093.6	15652.2	2981.4	7713.7
299	415.22	96.4	261.2	414	574.1	129.5	330

294.09	421.51	74.1	200.8	338.4	488	84.7	226.8	
接上表右侧:								
2007.1	2007.2	2007.3	2007.4	2008.1	2008.2	2008.3	2008.4	2009.1
1105.5	3284.3	5758.6	8365.8	1593	5056.8	9476.4	13621.03	1909.33
1098	3070.9	5225.6	7687.6	1541.9	4187	7337.1	10190.9	844.47
170.8	596.2	1129.3	1779	327.3	1137.8	2129.5	2916.19	281.76
257.5	847.5	1488.4	2160.3	321.5	1016.2	1724	2350.92	253.93
125.2	432.8	800.4	1180.8	179.5	621.5	1143.2	1649.43	211.98
0.6	2.1	4	6.9	1.4	4.5	8.1	10.78	1.76
1958.4	5747.7	10217.4	15139.4	2796.2	8218.8	14402.6	20437.16	3233.6
777.9	2126.1	3637.6	5229.6	1017.5	2811.3	4753.4	6662.1	1127.55
693.4	1853.2	3185.6	4472.8	880.8	2331.8	4020.2	5541.98	996.79
763.1	1626	2536.8	3445.1	902.6	1879.1	2916.4	3933.11	1007.84
2180.1	6476.4	11252	16258.9	2552.8	7626.6	13105.3	18562.24	2561.26
894.2	2528.1	4418.9	6415.8	1070.9	3023.8	5232.3	7536.82	1191.96
568.9	1715.3	3078.3	4477.7	672.1	2025.6	3614.8	5159.8	688.85
327.6	1073.5	1976.4	2933.1	486.2	1534.9	2789.9	4025.41	541.85
297.9	850.6	1442.5	2068.4	367	1051.2	1792.6	2535.37	382.99
764.4	2244	3803.5	5486.8	974.2	2867.6	4801	6647.44	932.01
254.4	719.1	1242.4	1785.9	311.5	877.6	1510.3	2148.37	345.78
240.3	717.6	1253.3	1828.9	287	811.6	1434.6	2059.03	288.02
2440.6	6672	11162.6	16004.2	3259.5	8818.2	15433.5	20801.46	2395.41
3230.4	9733.5	16528.7	23776.6	4190.6	12658.9	21746.6	29624.43	3944.03
736.4	2146.2	3631.4	5207.2	960.6	2717.8	4645.1	6561.3	1117.62
500	1492.3	2525.6	3612.1	568.2	1615.2	2650	3639.15	425.55
409.3	1233.3	2138.6	3047.9	512.8	1543.7	2643.6	3654.47	508.74
913.1	2754.6	4822.1	6942.8	1134.9	3374	5877.7	8306.74	1141.33
1595.1	5050	9141.6	13320.1	2079.4	6726.9	12054.7	17388.09	2400.9
4417.8	12920.5	22057	31607.9	6016.6	18360.4	31817.9	42090.09	5277.23
2066.8	6412.7	11023.2	15909.4	2625.8	8000.7	13697.5	18390.45	2058.19
1247.2	3827.5	6707.7	9682.9	1642.8	4975.3	8821.1	12439.03	1690.51
1961.6	6115.3	10737	15508.7	2618.9	8219.2	14375	20069.59	2866.03
1105.2	3558.2	6148.6	8888.9	1492.2	4771.2	8183.5	11682.1	1722.41
3173.8	9648.2	16261.1	23195.2	4160.9	12505.7	20745.7	28512	4058.57
2612.7	8171.5	14227.8	20560.2	3414.3	10568.1	18026.2	25246.34	3443.21
4988.7	13584.3	23328.4	34877.8	5705.4	16141.5	27208.2	38529.56	4779.77
493.4	1491.7	2550	3760.3	611.2	1815.6	3028.9	4322.26	570.26
373.6	1108.2	1933.2	2797.1	496	1465.2	2502.4	3520.61	499.53
60.9	230.6	393.4	595.4	103.2	357.6	624.5	825.12	116.27
3694.3	9519.2	16123.8	22876.8	4326.8	11142.9	18689.4	25844.75	4394.51
163.4	428	683.2	972.4	231.8	606.8	988.6	1321.31	254.47
102.2	264.1	451.3	650.9	112.5	299.2	502.3	707.07	134.39

(1) 建立分行业的就业人数与主营业务收入的混合回归模型

模型的形式是：

$$y_{it} = \alpha + x_{it} \cdot \beta + \varepsilon_{it}, \quad i = 1, 2, 3 \dots N, \quad t = 1, 2, 3 \dots T$$

混合回归模型假定截距和斜率系数都是常数，利用普通最小二乘法来求出参数 α 和 β 即可。

表 3-34 Eviews5 建立就业人数 (JYRS) 与主营业务收入 (ZYSR) 的混合回归模型

Variable	Coefficient	Std. Error	t-Statistic	Prob.
ZYSR?	0.025483	0.000768	33.17482	0.0000
R-squared	-0.113367	Mean dependent var		172.7227
Adjusted R-squared	-0.113367	S.D. dependent var		147.5666
S.E. of regression	155.7067	Akaike info criterion		12.93485
Sum squared resid	23614222	Schwarz criterion		12.93986
Log likelihood	-6304.740	Durbin-Watson stat		0.647032

从 Eviews5 的分析结果中，我们看到 $R^2 = -0.1137$ ，混合模型将 α 和 β 看成是普通的常数，忽略各个行业特点建立的整体模型效果非常不好，我们建模的过程中需要逐步推进，因此，我们建立个体固定效应回归模型。

(2) 建立分行业的就业人数与主营业务收入的个体固定效应回归模型

表 3-35 使用 Eviews5 建立 JYRS 与 ZYSR 的固定效应模型

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	157.5407	1.331914	118.2815	0.0000
ZYSR?	0.004105	0.000247	16.62683	0.0000
Fixed Effects (Cross)				
X1--C	252.9436	X21--C	-42.80140	
X2--C	-88.42975	X22--C	-123.0112	
X3--C	-123.0352	X23--C	-87.72203	
X4--C	-117.3476	X24--C	7.467899	
X5--C	-115.8292	X25--C	230.2901	
X6--C	-157.1119	X26--C	67.31599	
X7--C	32.99668	X27--C	-48.21096	
X8--C	-50.79319	X28--C	44.89651	
X9--C	-72.99027	X29--C	167.8542	
X10--C	-144.9335	X30--C	54.85177	
X11--C	374.7693	X31--C	163.0707	
X12--C	172.0187	X32--C	167.7963	
X13--C	41.50393	X33--C	218.8604	
X14--C	-80.40138	X34--C	-74.99394	

X15--C	-93.85060	X35--C	-44.75803
X16--C	-41.57432	X36--C	-154.4787
X17--C	-96.00285	X37--C	47.95841
X18--C	-58.48408	X38--C	-144.4224
X19--C	-115.8583	X39--C	-114.5319
X20--C	146.9781		

Effects Specification

Cross-section fixed (dummy variables)

R-squared	0.959587	Mean dependent var	172.7227
Adjusted R-squared	0.957902	S.D. dependent var	147.5666
S.E. of regression	30.27750	Akaike info criterion	9.698847
Sum squared resid	857139.8	Schwarz criterion	9.899152
Log likelihood	-4688.188	F-statistic	569.2660
Durbin-Watson stat	0.538459	Prob(F-statistic)	0.000000

注释：具体的 Eviews5 运行程序请参见电子稿附录“M2-3-11”

相应的数学表达式为：

〈M3-7〉

$$JYZB_{it} = 157.54 + 252.94D_1 - 88.43D_2 + \dots - 114.53D_{39} + 0.0041054 * ZYSR_{it}$$

(118.28)
(16.63)

$$R^2 = 0.959587$$

其中虚拟变量 D_1 、 D_2 、 D_3 …… D_{39} 的定义是：

$$D_i = \begin{cases} 1, & \text{如果属于第 } i \text{ 个行业, } i = 1, 2, 3, \dots, 39 \\ 0, & \text{如果不属于第 } i \text{ 个行业} \end{cases}$$

从最终的模型结果中我们可以看出，尽管各行业方程的截距不相同，但是，具有相同的斜率。

最终数据分析的结果： $R^2 = 0.959587$ ，对目标的解释效果很好，且通过相应的 t 检验、F 检验，是能较好说明分行业指标体系与目标之间相互关系的模型。

3.3.3.4 神经网络模型原理

首先，我们给出行业有关就业人数的统计体系：

(1) 神经网络模型分析非线性映射关系数据的优点^[2]

我们所处的客观世界是复杂的，所接触的绝大多数时间序列往往是由随机信号和由确定性非线性系统产生的信号的混合体，复杂的市场是一个信息量巨大的非线性动力系统，其功率谱具有宽带、连续等特征。传统的线性预测方法因为无法将其与噪声的功率谱加以区分，预测效果往往一般。

神经网络方法是解决这类问题的一种很好的方法，因为，具有非线性隐含神经元的神经网络具备了可以逼近任何非线性连续函数的能力，可通过神经元连接权值的调整，更精确地逼近市场价格波动中反映出来的非线性映射关系，从而实现对混沌时间序列的短期精确预测。神经网络是一种高度自适应的非线性动力

系统，通过神经网络学习可以得到输入与输出之间的高度非线性映射，因此使用神经网络可以建立起输入与输出之间的非线性性。

(2) 人工神经网络模型结构

人工神经元是人工神经网络操作的基本信息处理单位。人工神经元的模型如下图所示，它是人工神经网络的设计基础。

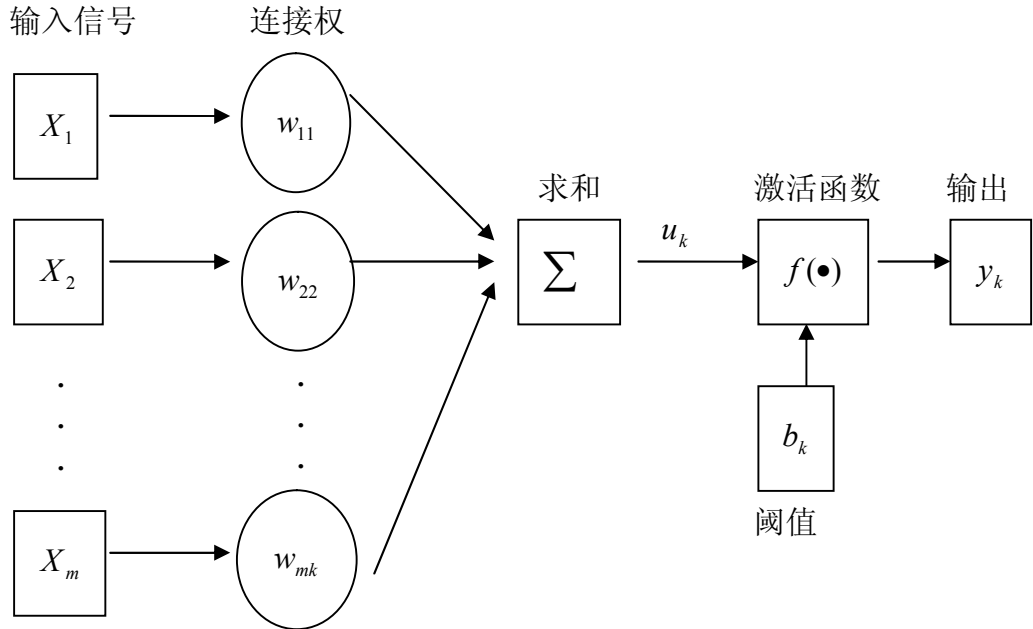


图 3-7 人工神经元模型结构示意图

人工神经网络模型可以被看成由三种元素组成：

- 1) 一组连接：连接强度由各连接上的值表示，权值可以取正值也可以取负值，权值为正表示激活，权值为负表示抑制。
- 2) 一个加法器：用于求输入信号对神经元的相应突触加权之和。
- 3) 一个激活函数：用来限制神经元输出振幅。激活函数也成为压制函数，因为它将输入信号压制(限制)到允许范围之内的一定值。通常，一个神经元输出的正常幅度范围可写成单位闭区间 $[0, 1]$ ，或者另一种区间 $[-1, +1]$ 。

另外，可以给神经元模型加上一个外部偏置 b_k ，它的作用是根据其为正还是为负，相应地增加或降低激活函数的网络输入。

3.3.3.5 使用神经网络模型对分行业系统进行仿真和预测

我们选取企业单位数、工业总产值、工业增加值、负债总计、主营业务收入、主营业务成本等六个指标作为输入指标，而从业平均人数为输出指标，建立 $6 \times ? \times 1$ 的 BP 神经网络模型来分析和预测油价的变动。其中，由于输入样本为 6 维的输入向量(自变量由六个指标)，因此，输入层一共有 6 个神经元；? 代表中间层的神经元，中间层神经元个数的增加，虽然可以提高网络的映射精度，但并不意味着一定会提高网络的性能，所以，在设计 BP 网络时，不能无限地增加中间层神经元的个数，在这里用 ? 表示是因为后面我们会对适宜个数进行判断，最终

才确定。

<M3-8>

(1)输入输出数据预处理

由于BP神经网络训练样本集中，输入、输出样本参数的绝对值如果离散性太大或过于集中，在权值矩阵误差函数逼近过程中，易产生局部误差最优或误差震荡等缺陷。

因此，必须对输入、输出进行归一化处理，此外，归一化处理可以使得具有不同物理意义的输入变量赋以同等重要的地位，避免数值大而不重要的变量掩盖数值小相反却重要的变量，此外，还可以保证网络的收敛性，提高网络的收敛速度。为了进行归一化处理，我们可以利用下面公式将数据压缩在(0, 1)的范围内，

$$X_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}}$$

我们选取六个指标作为输入指标，而从业平均人数为输出指标。这七项指标共 39 组行业数据，但为了对建立的神经网络效果进行预测和检验，在这里将数据分为两部分，一部分数据用来建立模型，而另一部分数据用来做样本外预测和检验，考虑到要避免建立模型所使用数据的经济背景与样本外预测所使用数据的经济背景不同，这两部分的数据的提取是对总时间段随机抽取得来。

表 3-32 训练数据和样本外预测数据范围

模型训练样本(30 个行业)	第 1 行业到第 15 行业，第 25 行业到第 39 行业
样本外预测数据(9 个行业)	第 16 行业到第 24 行业

考虑到不同的指标的数值存在数量级上的差异，故为了增强数据的可比性，我们首先要对数据进行无归一化处理。经过归一化处理后的训练样本如下所示：

表 3-33 各指标进行归一化结果

行业	企业单位数	工业总产值	工业增加值	负债总计	主营业务收入	主营业务成本	从业平均人数
第 1 行业	0.2694	0.2344	0.5212	0.2725	0.2457	0.1896	0.7403
第 2 行业	0.0057	0.2114	0.7161	0.1233	0.2176	0.1155	0.1444
第 3 行业	0.1031	0.0541	0.1028	0.0308	0.0529	0.0415	0.0781
第 4 行业	0.0774	0.0581	0.1077	0.0279	0.0572	0.0448	0.0876
第 5 行业	0.1068	0.0345	0.0571	0.0157	0.0330	0.0277	0.0741
第 6 行业	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
第 7 行业	0.6496	0.4459	0.5152	0.1606	0.4390	0.4279	0.4226
第 8 行业	0.2374	0.1545	0.2064	0.0749	0.1498	0.1314	0.2153
第 9 行业	0.1577	0.1293	0.2088	0.0840	0.1278	0.1015	0.1610
第 10 行业	0.0045	0.0960	0.3238	0.0299	0.0956	0.0349	0.0293
第 11 行业	1.0000	0.4775	0.5454	0.2654	0.4655	0.4614	1.0000
第 12 行业	0.5287	0.1935	0.2512	0.0809	0.1878	0.1786	0.6612
第 13 行业	0.2663	0.1311	0.1641	0.0482	0.1271	0.1233	0.4101
第 14 行业	0.2807	0.0895	0.1141	0.0350	0.0863	0.0820	0.1692
第 15 行业	0.1465	0.0616	0.0715	0.0296	0.0603	0.0566	0.1454
第 16 行业	0.2995	0.1610	0.1932	0.1156	0.1575	0.1495	0.2205

第 17 行业	0.1814	0.0537	0.0765	0.0373	0.0520	0.0475	0.1152
第 18 行业	0.1457	0.0532	0.0612	0.0235	0.0518	0.0499	0.1902
第 19 行业	0.0762	0.4550	0.3436	0.1709	0.4600	0.4722	0.1284
第 20 行业	0.8231	0.6831	0.8149	0.3982	0.6787	0.6319	0.6071
第 21 行业	0.2052	0.1620	0.2536	0.1085	0.1527	0.1165	0.2190
第 22 行业	0.0549	0.1048	0.0895	0.0677	0.1018	0.1030	0.0719
第 23 行业	0.1316	0.0880	0.1061	0.0516	0.0864	0.0827	0.1394
第 24 行业	0.5504	0.2068	0.2370	0.1077	0.2022	0.1951	0.3575
第 25 行业	0.8696	0.3965	0.5382	0.2549	0.3840	0.3556	0.7159
第 26 行业	0.2559	0.8592	1.0000	0.5702	0.8953	0.8837	0.4859
第 27 行业	0.2394	0.4596	0.4969	0.2142	0.4591	0.4517	0.2492
第 28 行业	0.6448	0.2916	0.3340	0.1418	0.2844	0.2766	0.4365
第 29 行业	0.9585	0.4694	0.5669	0.2850	0.4571	0.4260	0.6716
第 30 行业	0.4799	0.2698	0.3403	0.1892	0.2629	0.2404	0.4093
第 31 行业	0.5044	0.6920	0.7742	0.5097	0.6827	0.6416	0.6523
第 32 行业	0.6919	0.6123	0.6720	0.3212	0.5949	0.5672	0.7171
第 33 行业	0.4014	1.0000	0.8798	0.4850	1.0000	1.0000	0.9388
第 34 行业	0.1614	0.1096	0.1288	0.0537	0.1076	0.1009	0.1705
第 35 行业	0.2292	0.0861	0.1016	0.0382	0.0843	0.0812	0.2183
第 36 行业	0.0225	0.0171	0.0176	0.0057	0.0172	0.0171	0.0102
第 37 行业	0.1987	0.6746	0.9802	1.0000	0.6724	0.6498	0.4101
第 38 行业	0.0203	0.0249	0.0337	0.0295	0.0289	0.0278	0.0250
第 39 行业	0.0613	0.0200	0.0403	0.0636	0.0189	0.0153	0.0657

注释：拥有底纹的是训练样本，没有底纹的是样本外预测数据。

(2)学习速率、训练方法、传输函数和期望误差的选择

学习速率决定每一次循环训练所产生的权值的变化量。过大的学习速率可能导致系统的不稳定，但是过小的学习速率将导致训练时间较长，收敛速度很慢，不过能保证网络的误差值趋于最小。一般情况下，学习速率的选取范围在 0.01~0.9 之间。

鉴于自适应学习速率通过保证稳定训练的前提下，达到了合理的高速率，可以减少训练时间。我们采用自适应学习速率训练方法。考虑到数据样本非线性

特点，我们采用对非线性问题解决较好的 S 型正切激活函数 $\tan sig = \frac{1-e^{-2x}}{1+e^{-2x}}$ 作为

隐含层的传递函数，对数激活函数 $\log sig = \frac{1}{1+e^{-x}}$ 作为输出层的传递函数。

期望误差的控制，通过对网络训练参数的设定实现。考虑到具体网络训练时的训练精度和训练时间，同时确保能够对系统进行较好的预测，保证预测的实际精度，我们在网络中使训练次数极值，网络训练的误差允许值

$nettrainParam.goal = 0.005$ 。

BP 神经网络的基本流程图为：

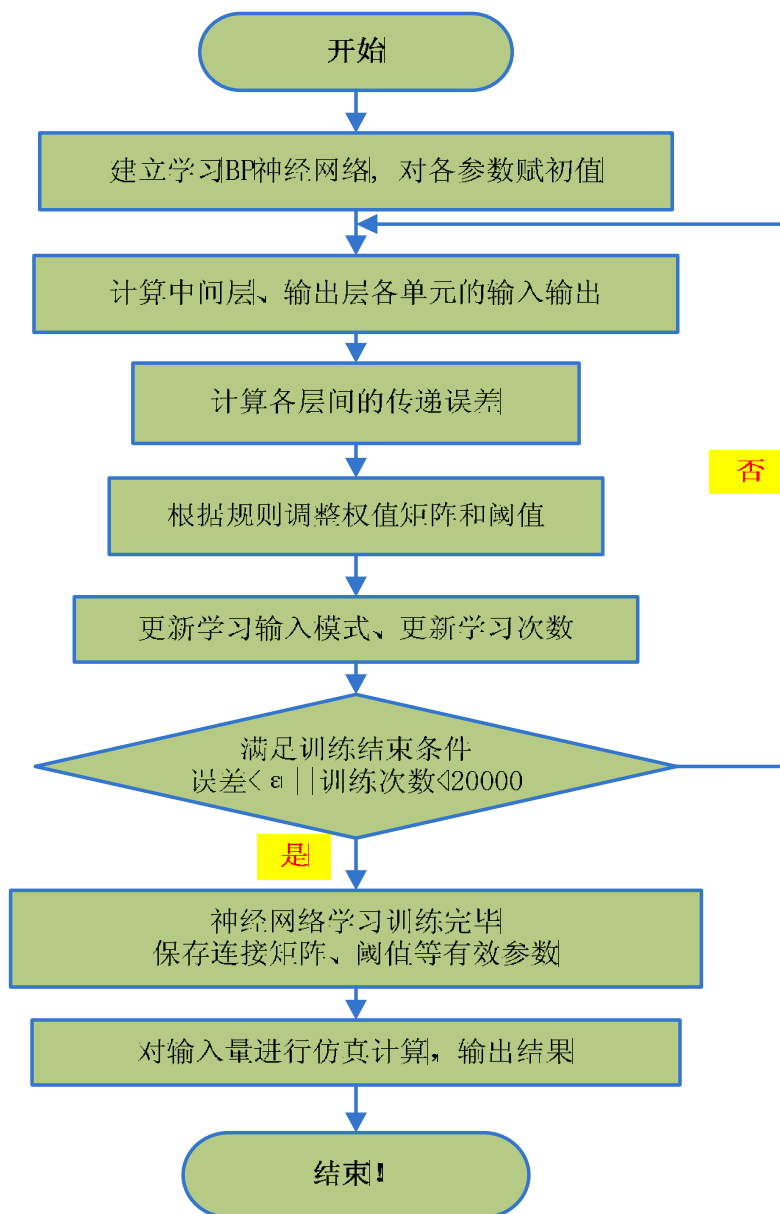


图3-8 BP神经网络训练流程图

3.3.3.6 将训练好的神经网络模型应用于实际样本外预测

中间层神经元个数的增加, 虽然可以提高网络的映射精度, 但并不意味着一定会提高网络的性能, 所以, 在设计BP网络时, 不能无限制地增加中间层神经元的个数, 我们选取11、12、13、14个神经元分别建立模型, 各自样本外预测误差如下图所示:

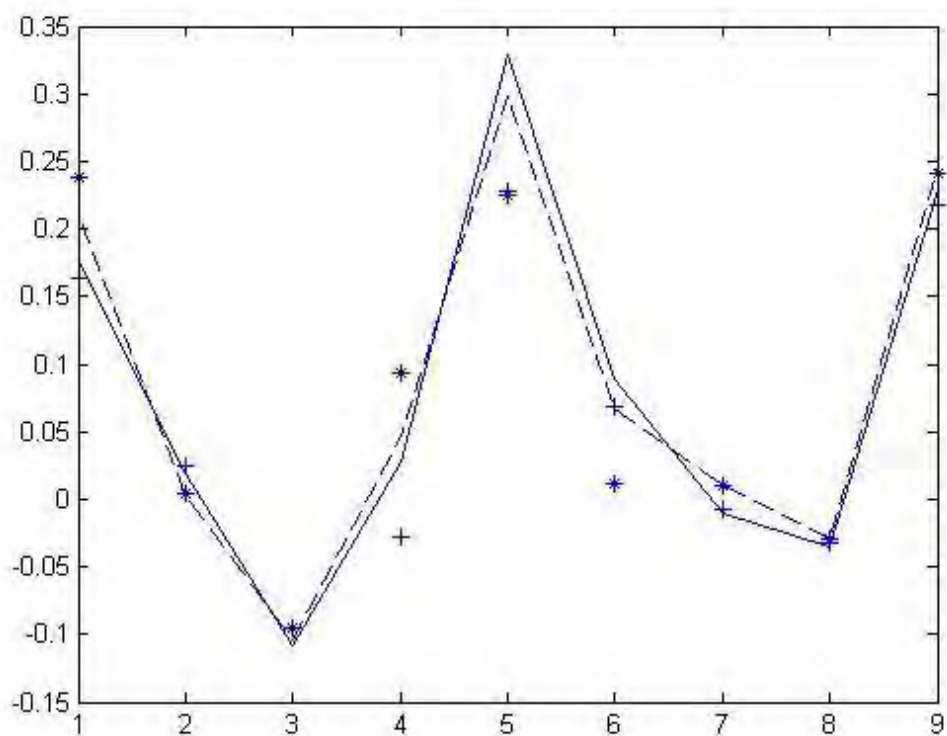


图3-9 不同数目神经元建立模型，各自样本外预测误差示意图
 (“+”中间层神经元为11，“——”中间层神经元为12
 “---”中间层神经元为13，“*”中间层神经元为14)
 注释：详细运行程序请参见电子稿的附件“神经网络模型.M”

从图中可以看到，四种中间层神经元数目的神经网络预测误差相差都不大，通过30个行业样本建立非对称性映射系统的神经网络模型对9个行业进行样本外预测的结果和实际值之间的差距保持在-10%和+25%的范围之内，其中对第17、19、21、22、23的预测效果尤其好，误差控制在±5%以内。

3.4 问题四、五的分析

3.4.1 对未来就业形势进行的预测

在前面分析就业问题与各指标之间关系时，实际上已经对这个问题进行了一定的说明。

在对城镇就业人数进行时间序列的初步分析时，得到城镇就业人数与重大政策有着莫大的关系（从1958年与1990年可以看出），今年四万亿促进经济发展政策以及开放江苏沿海（类似于1990年开发上海浦东地区）的政策对就业人数的增加有着正向的促进作用。但是，不能仅凭重大政策这一个指标，在随后对就业人数的时间序列建立退势平稳序列模型的过程中，我们可以看到未来短期内城镇就业人数将会继续保持较好的势头增加。

随后建立的多元回归模型表示：城镇就业人数可以通过经济活动人数和固定资产支出两个指标来介绍，如果运用此模型对未来短期进行预测，固定资产支出增量较大是毫无疑问的，经济活动人口按照往年5%左右的增长速度，城镇就业

人数的增幅将大于往年 3.52%水平。

后面的模型是反映目标与其他影响因素之间关系的模型，在运用这些模型进行未来预测时，碰到的问题是：这些因素的最新数值搜索较为困难，我们对此采用的解决方式是通过历史数据首先建立各指标的时间序列模型，在对最新时间点进行预测之后，带入事先建立好的关系模型中，用以预测未来的就业形势。

在分地区、分行业的分析中，我们得出这样的结论，三十一个直辖市和省份在就业压力的排布上呈现线性变化特征，按照广东、江苏、山东、浙江、上海、北京……顺序就业压力依次递减。可以结合当前政策预测未来各地的就业形势。

3.4.2 对增加就业人口、减少就业压力的政策建议

在前面的分析中我们得出这样的结论，短期的突变受政策影响较大，目前的扩张型宏观政策有利于扩大企业的发展，对吸收剩余劳动力很有帮助。

通过建立模型，我们发现就业人口与经济发展之间存在着明显的联系，面对劳动力供大于求的局面的长期性，我们需要通过经济的增长拉动就业人数的增加。分地区的分析中，三十一个直辖市和省份在就业压力的排布上呈现线性变化特征，按照广东、江苏、山东、浙江、上海、北京……顺序就业压力依次递减，可以根据各地就业严重的不同程度订立合适的宏观政策来解决就业问题。分行业的分析中，需要把握行业结构的调整，向劳动力的需求方输送数量合适且能满足要求的就业者。

四、本文总结和对该课题研究的进一步思考

4.1 本文总结

在进行这个对就业问题研究时，我们考虑到它属于经济学问题，且涉及的影响因素众多，仅仅只从数据的相关关系来判断影响因子是缺乏理论依据的，因此，我们首先对经济学中有关就业的理论做了一个简单的回顾和综述，然后定性地分析和总结了就业问题产生的原因和过程。在理清了这方面的思路后，我们有目的性的从宏观经济和劳动者特点两个方面搜集与就业有着直接或间接联系的指标数据。不过我们没有立即对影响因素进行分析，在此之前，我们对目标指标——城镇就业人数进行时间序列分析，得到就业人数与重大政策有莫大关系的结论（从 1958 年与 1990 年可以看出），政策指标对于解释突变点有着很好的效果。另外，时间序列对目标指标本身具有一定的解释性。

(1) 选择具有相关作用的影响因素

通过建立 GARCH(1, 1) 模型，我们初步得出目标指标时间序列的波动受自身特点（即可以通过历史数据对未来进行判断）的影响占总原因的 31%，而外部因素的冲击作用占了 69%，因此，不能仅从历史数据的走势对未来进行判断，对外部影响因素的分析将显得非常重要。随后我们在进行文献综述的基础之上，从定性分析的角度，总结出三十多个可能对就业有着直接或间接作用的宏观经济指标与十几个反映劳动力特点的指标，并且逐个运用格兰杰因果检验方法，对目标指标与影响因素之间的因果关系进行判断。最终的结果显示有十九个宏观经济指标对城镇就业人数有着解释作用，而城镇就业人数又对七个宏观经济指标与三个劳

动力特点指标有着解释作用，即部分宏观指标与目标指标具有相互的因果关系。

(2) 建立反映指标与影响因素之间联系的模型

在此基础之上，我们以城镇就业人数作为被解释变量，以对它具有原因关系的指标作为解释变量，建立了多元回归模型。在建模的过程中发现多重共线性的问题较为严重（这主要是由于众多宏观经济指标之间有着类似的波动特点），因此，在消除多重共线性之后，模型中留下了经济活动人口与固定资产支出（一个说明供给方的供给能力，一个说明需求方的扩张能力），拟合效果很好，可决系数达到了 0.998 以上。

考虑到有些影响因素与目标指标之间的因果关系是目标对影响因素有解释作用，而这部分的信息在多元回归模型中无法体现，因此，我们建立向量自回归 (VAR) 模型，在通过一系列的检验后，就业人数、经济活动人数、居民消费支出、固定资产投资四个指标在最终的模型中相互得以说明，并且互相解释的效果很好，达到 99.60% 以上。

(3) 分地区、分行业进一步分析

在前面的分析中我们假定全国是一个整体，即没有行业与地区之间的差别，而我们知道地区或不同行业之间的就业形式是有很大的不同的，因此，在这一部分的分析中，我们分地区、分行业进行对就业的形式进行建模分析。

在分地区的分析中，我们选择与就业有着密切关系的几个指标（在前面格兰杰检验的基础之上）运用聚类分析方法，将拥有不同就业特点的地区对应平面坐标轴上特定的点，结果发现三十一个直辖市和省份在就业压力的排布上呈现线性变化的特征，按照广东、江苏、山东、浙江、上海、北京……的顺序就业压力依次递减，我们可以人为地将聚集在一起的点归为一类，大体分为就业压力严重、就业较为平衡、稀缺劳动力等三类，针对每一类再来进行问题一、二的分析，从而使分析的过程更加合理。

在分行业的分析中，我们对这样同时具有时间序列和截面性质的面板数据的分析中，建立 Panel Data 模型对各行业的平均就业人数与主营业务收入之间的关系进行分析，最终模型的形式是各行业的截距不同而斜率相同，且模型整体的可决系数达到了 0.95 以上，通过 t 检验与 F 建议，是用来分析不同行业就业人数与经济指标关系过程中的一个较好的模型。最后，我们考虑到分行业的指标统计体系中，众多因素与目标指标之间的关系具有非线性映射的特点，因此，建立 BP 神经网络模型，用来说明目标与众多解释变量之间复杂的解释关系，并且对 9 个行业进行了样本外预测，结果表明训练的效果不错。

4.2 对该课题研究的进一步思考

在本文中我们的分析思路：首先将全国看做一个均衡的整体，劳动力的供给方和劳动力的需求的不同导致了最终就业人数的多少，当供大于求时就发生了劳动力的剩余，即失业率的产生，通过分析影响因素与目标指标之间的关系，我们对就业人数的未来走势可以有一个较好的把握。随后，我们意识到各地区和各行业之间的就业情况是有着明显差别的，因此我们分地区、分行业对就业进行分析。

在分析过程中，我们逐渐意识到以下几个问题可以作为继续研究的方向：

(1) 在之前分地区的分析中，我们是将全国按行政单位划分为严重、平衡、稀缺这几类，然后针对每一类别单独进行分析。在其后的分析中我们可以讲这三

类再细分，即严重地区再精确到直辖市或省，但是我们很快又意识到以省为最终的单位进行分析还不够（例如，江苏的苏南、苏北虽在同一省，但是，经济与就业的差别还是很大的），同时我们还需要考虑数据获得的可能，如果能够精确到市级单位最好了，各市有统计局（例如，在这次比赛过程中查询到的南京、苏州等统计年鉴与国家统计年鉴的统计规格与方法类似），因此，对分地区的就业问题，我们可以做到地域的更精确的分析。

(2)对行业的划分，以工业为例，39 个行业的划分还可以根据需要在进一步细分，此外，我们也意识到对地区的划分和对行业的划分并不是单独进行的，例如，就业压力并不严重的地区也有劳动力供大于求的行业，因此，分地区分行业对就业问题的分析可以将这两方面融合起来，如果数据来源通道顺畅的话，理想的研究路线应该如下图所示：

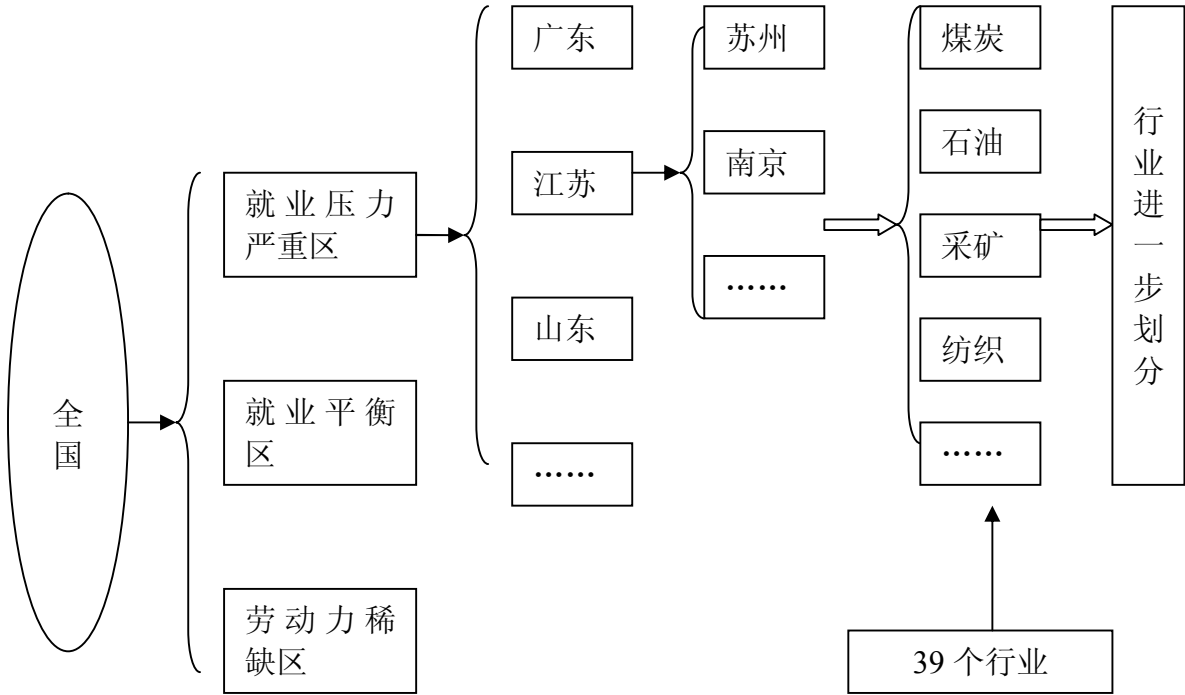


图 3-10 理想的分地区且分行业的就业问题分析图

(3)就业压力指标评价体系的建立，就业压力指标不能仅仅单纯从就业人数、失业率等反映劳动力人数的指标来考虑，例如，现实中存在着的“跳巢”现象，虽然从每时每刻来看，劳动者都处于就业的状态，但是频繁的“跳巢”也表明劳动者对工作的不满意，就其自身而言，他是处于“失业”的状态，因为他还没有找到令自己满意的工作岗位。一旦环境发生变化，这类人群很有可能因为外部冲击的作用而失业。因此，除了从实际失业数据进行分析外，还需要对潜在失业人员进行分析，即还需要进行就业满意度的分析。

(4)有关就业压力预警系统的建立，对就业人数进行研究的实际目标是希望进一步对就业压力的大小进行把握，明白当下或可预见里某个地区的某个行业的就业压力是多少，是处于警报系统的哪个级别，需要进行怎样的宏观调控来解决严重的就业压力。

参考文献

- [1] 曹建云. 我国经济增长与就业增长的关系研究[D]. 兰州大学博士学位论文, 2008.
- [2] 葛哲学, 孙志强. 神经网络理论与 MATLAB2007 实现[M]. 电子工业出版社, 2007. 9.
- [3] 卡尔·马克思. 资本论(中译本), 北京: 北京人民出版社
- [4] 梁保华. 江苏未来发展报告[R]. 2009, 9.
- [5] 李敏. 中国就业问题研究[D]. 华中科技大学博士论文, 2005.
- [6] 探究当前就业压力形成的深层次原因[EB/OL].
http://whb.news365.com.cn/ly/200903/t20090311_2232554.htm
- [7] 于俊年. 计量经济学软件: EViews 的使用[M]. 对外经济贸易大学出版社, 2006.
- [8] 1990 年: 国务院宣布开放浦东新区的九项政策规定[EB/OL].
<http://bbs1.people.com.cn/postDetail.do?boardId=9&treeView=1&view=2&id=94138965>
- [9] 什么叫景气指数? 一般如何计算? [EB/OL].
<http://iask.sina.com.cn/b/3286517.html>
- [10] Lewis, W. A. Economic Development with Unlimited Supply of Labour, The Manchester School, May 1954:139-191
- [11] J. M. Keynes The General Theory of Employment, Interest and Money China Social Sciences Publishing House Chengcheng Books LTD., 1999
- [12] Friedman, Free to Choose: A Personal Statement, 1979. China Social Sciences Publishing House Chengcheng Books LTD., 1999