

人口结构变动对就业形势的影响研究

杨陈陈, 罗瑾萱

天津商业大学理学院, 天津

收稿日期: 2024年2月5日; 录用日期: 2024年3月31日; 发布日期: 2024年4月8日

摘要

纵观历史, 人口问题当属世界关注前列, 人口转变易引发多种社会问题, 就业困难首当其冲。为增加就业人口, 本文将人口结构作为一个影响因素进行研究, 选取中国二十三年人口子结构与就业数据, 通过向量自回归的方法, 对两者做相关分析, 建立了就业人口与人口自然结构、就业与人口社会经济结构的VAR模型以及LSTM模型, 分析人口结构对就业的影响, 对人口子结构与就业之间的关系做实证研究, 并给出关于如何改善人口子结构进而增加就业人口的建议。

关键词

人口结构, 就业人口, 平稳性检验, VAR模型, LSTM模型

Research on the Impact of Changes in Population Structure on Employment Situation

Chenchen Yang, Jinxuan Luo

College of Science, Tianjin University of Commerce, Tianjin

Received: Feb. 5th, 2024; accepted: Mar. 31st, 2024; published: Apr. 8th, 2024

Abstract

Throughout history, population issues have been at the forefront of the world's attention. Demographic transition can easily lead to a variety of social problems, with employment difficulties bearing the brunt. In order to increase the employed population, this article studies the population structure as an influencing factor, selects China's 23-year population substructure and employment data, conducts correlation analysis on the two through the vector autoregression method, and establishes the relationship between the employed population and the natural popula-

tion. The VAR model and LSTM model of structure, employment and population socioeconomic structure analyze the impact of population structure on employment, conduct empirical research on the relationship between population substructure and employment, and give suggestions for improving the population structure.

Keywords

Population Structure, Employment Population, Stationarity Test, VAR Model, LSTM Model

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

1.1. 研究背景与意义

我国目前正处在经济社会发展转型和人口结构变化的关键时期。错综复杂的社会局面, 让我们正处在向劳动力供求关系发展转变的十字路口。一方面, 在“人口红利”和市场经济的快速发展后, 低出生率问题在中国老龄化过程中进一步加剧, 并且很快地就进入了中国劳动年龄人口负增长的历史转折点。另一方面, 进入 21 世纪以来, 国家调整经济转型和产业结构, 这一措施扩大了第三产业发展和劳动力需求关系的发展路径, 但经济的快速增长并没有带来强劲的劳动力需求。

我国经济稳定发展的关键是劳动力供求平衡, 同时人口结构变动也是最重要的因素, 它会引发各种社会问题。当前我国人口结构变化的特点: 1) 劳动年龄人口总量呈下降趋势; 2) 老龄化和少子化; 3) 抚养比进入上升通道; 4) 劳动力素质提高显著; 5) 加速城镇化政策出台对人口城乡结构变化影响显著[1]。就业是民生之本, 研究人口结构转变对就业形势的影响无论对国家, 对人民都具有深刻的意义。本文将人口结构分为三个子结构: 自然、社会经济、地域结构, 对子结构与就业之间的关系进行研究分析。目的在于准确把握人口发展形势和就业人口相互作用的客观规律, 从而促进人口、经济和社会的持续健康发展。

1.2. 研究现状

近年来人口结构发生变化, 人力资源需求的变化和新技术革命使得就业形式日益严峻, 从人口结构来研究其对就业形式的影响, 国内外学者从人口出发针对就业形式问题展开了深入的研究。朱文涛等基于全国 1978~2013 年时序数据构建了 VAR 模型, 并探究人口城镇化、就业非农化与城乡收入差距之间的关系, 结果显示: 前两者与城乡收入差距存在一种长期均衡的关系[2]。莫旋等根据我国流动人口的动态监测数据建立了分层多元选择模型, 以分层和异质性的视角, 深入研究了影响流动人员就业率的各种因素以及流动人员的职业选择, 研究成果表明: 流动人员在不同职业身份的情形下具有要素禀赋具有显著差异, 以及不同的要素禀赋将会对其职业选择产生深远影响, 创业型就业更受老一代欢迎, 而务工型就业更受年轻一代欢迎[3]。林耿等对广东省的 21 个地级市进行就业空间研究, 对两类人口的就业构成、类型区进行研究, 结果表明: 体制内就业更受常住人口喜欢, 常住人口有也比流动人口有更多元化的就业空间分异, 但是两者就业空间并没有造成社会两极分化[4]。

针对人口与就业的时间序列数据, 多采用向量自回归(VAR)模型、长短期记忆网络(LSTM)模型等进行分析与预测。刘宏等基于全国 1985~2010 年外国直接投资、经济增长以及就业这三个时序数据构建 VAR

模型。研究表明: 外国直接投资对我国经济增长和就业产生了加速效应[5]。Yang 等针对多元时间序列数据, 对 VAR 模型进行了改进, 提出了一种新的 VARMA 在线时间序列预测框架, 证明了 VAR 可以在在线设置下模拟基础 VARMA 模型, 在这个框架下, 开发了两种有效的算法 VARMA-OGD 和 VARMA-ONS 来解决时间序列预测问题[6]。Li 等采用聚类分析法对毕业生就业形式进行研究, 预测了毕业生就业情况, 提出了一种基于长短时记忆(LSTM)递归神经网络的毕业生就业状况预测技术(网络结构设计、网络训练等), 验证了 LSTM 及其参数优化算法在毕业生就业形势分析中的适用性和正确性[7]。

2. 模型介绍

2.1. 向量自回归模型

一般的 VAR (p)模型的数学表达式是:

$$y_t = c + A_1 y_{t-1} + \dots + A_p y_{t-p} + B_0 x_t + B_1 x_{t-1} + \dots + B_q x_{t-q} + \mu_t, t \in \{-\infty, +\infty\} \quad (1)$$

其中 $y_t = (y_{1t} \dots y_{Kt})$ 是 $K \times 1$ 阶时间序列列向量。c 是 $K \times 1$ 阶常数项列向量。A₁ 到 A_p 表示 $K \times K$ 阶参数矩阵。x_t 表示 $M \times 1$ 阶外生变量向量。B₁ 到 B_q 是 $K \times M$ 阶待估系数矩阵。μ_t 是 $K \times 1$ 随机误差列向量。

2.2. 长短期记忆模型

长短期记忆网络(LSTM)是一种循环神经网络(RNN)的变体, 这种网络结构最初是由 Hochreiter 等人引入的, 逐渐被完善为处理时间序列预测问题的工具。LSTM 不同于 RNN 网络的内部结构, RNN 存在长期依赖性, 而 LSTM 的网络内部结构更为复杂, 其通过输入门、遗忘门、输出门引入 sigmoid 函数并结合 tanh 函数, 重复的模块不只是有一个简单的 tanh 结构, 从而避免了 RNN 长期依赖性。LSTM 通过门的结构来调控任何时间的状态, 添加求和操作, 将短期记忆和长期记忆相连接从而解决梯度消失的问题, LSTM 具备删除或添加节点的能力, 通过三个门的信息传递给单元状态, 其中遗忘门决定以前的单元格信息是否传递到当前单元格, 输入门从当前的输入中确定单元格的信息更新, 当前单元格的输出基于最新状态、前一时刻输出和当前输入得到, 也称作输出门, 每一个门都由神经网络表示, 都具备输入层、隐藏层和输出层。

LSTM 在 t 时刻有 3 个输入值, 分别是目前网络的输入、上一时刻网络的输出 h_{t-1} 以及上一时刻的单元状态 c_{t-1} ; 同时, LSTM 网络的 2 个输出值, 分别是当前时刻输出结果 h_t 和现在的单元状态 c_t 。状态单元里面的信息由输入门控制留下或者删除, 输出门则负责把状态单元的长期状态信息控制作为目前网络的输出。下列公式表示了长短期记忆网络在 t 时刻的各个状态单元的更新。

$$\begin{aligned} f_t &= \sigma(W_{fx} x_t + W_{fh} h_{t-1} + b_f) \\ i_t &= \sigma(W_{ix} x_t + W_{ih} h_{t-1} + b_i) \\ \tilde{c}_t &= \tanh(W_{cx} x_t + W_{ch} h_{t-1} + b_c) \\ c_t &= f_t * c_{t-1} + i_t * \tilde{c}_t \\ o_t &= \sigma(W_{ox} x_t + W_{oh} h_{t-1} + b_o) \\ h_t &= o_t * \tanh(c_t) \end{aligned} \quad (2)$$

其中 σ 表示 sigmoid 函数, x_t 是 t 时刻的输入, f_t 为遗忘门, i_t 为输入门, o_t 为输出门, W 表示权重, b 表示偏置。

3. 数据来源及指标选取

本文采用中国 1998~2020 年的时序数据, 通过运用 R、Eviews、Python 软件来探究中国人口结构与与

就业形势之间的关系, 数据来源于 RESSET 数据库、《统计年鉴》。在研究过程中, 选取不同的人口子结构与就业之间进行分析, 人口自然结构选取人口性别比、人口抚养比; 人口社会经济结构选取家庭规模、城镇居民人均可支配收入、教育结构; 人口地域结构选取反映城镇化水平的非农人口占比。对于就业形势, 选取就业人口(表 1)。

Table 1. Variable description

表 1. 变量说明

类型	名称	符号	定义
解释变量	人口性别比	FM	男性人口与女性人口的比值
	人口抚养比	PDR	总人口中非劳动人口占劳动人口比值
	家庭规模	FS	家庭户人口数与总家庭户数的比值
	收入结构	PDI	城镇居民人均可支配收入
	教育结构	ES	普通高等教育人口占比
	非农人口占比	NAR	总人口中非农人口的占比
被解释变量	就业人口	LP	总人口中就业人口的数量

4. 实证分析

4.1. 数据预处理

对各个指标作方差齐性变换, 变换之后的序列分别记为 $\ln FM$ 、 $\ln PDR$ 、 $\ln FS$ 、 $\ln PDI$ 、 $\ln ES$ 、 $\ln NAR$ 、 $\ln LP$ (图 1)。

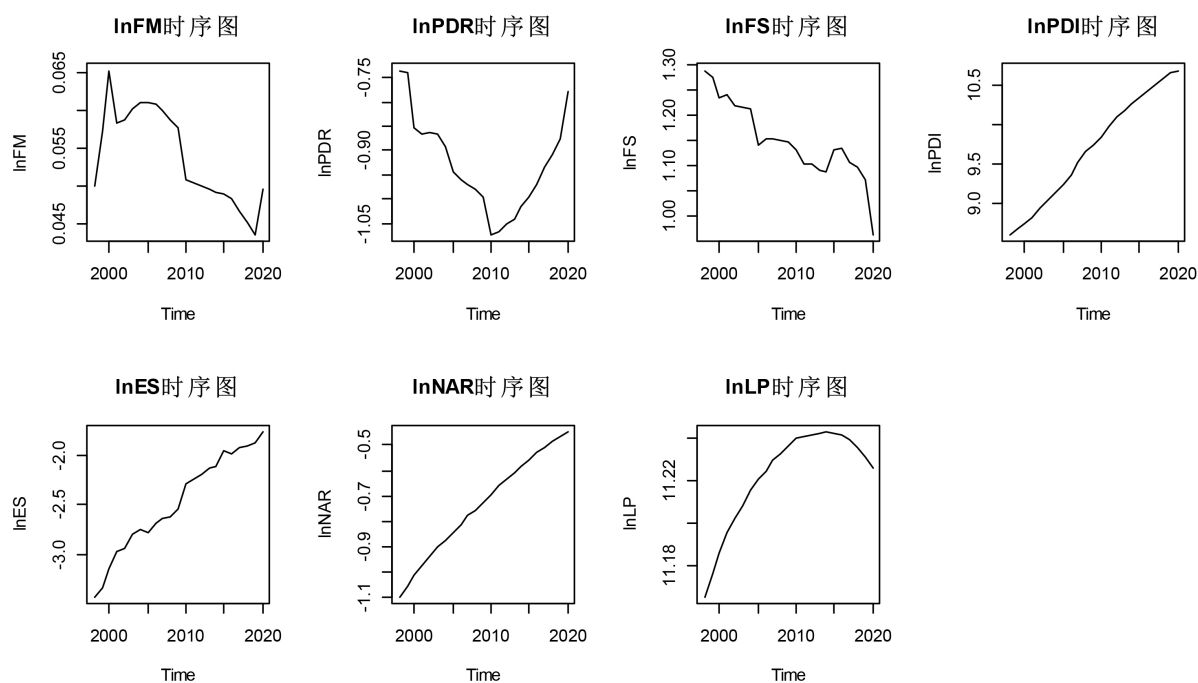


Figure 1. Time series diagram of population structure and employment population series

图 1. 人口结构和就业人口序列的时序图

4.2. 平稳性检验

传统的建立 VAR 模型要求平稳的时间序列。但大多数序列是非平稳的, 如果不处理直接运用可能会出现“伪回归”现象, 所以要对非平稳序列进行差分处理成平稳序列再运用 VAR 模型。本文将对各个序列进行 ADF 检验来确定是否平稳, 结果如表 2。

由表 2 的结果知, 在 0.05 显著水平下, 对数序列 lnFM、lnPDR、lnFS、lnPDI、lnES、lnLP 这 6 个序列为非平稳; 只有 lnNAR 这 1 个序列为平稳。对数一阶差分序列 dlnFM、dlnPDR、dlnFS、dlnPDI、dlnES、dlnNAR、dlnLP 这 7 个为平稳的序列。

Table 2. Results of ADF test

表 2. ADF 检验结果

变量	ADF 统计量		ADF 检验		
	t 值	1%	5%	10%	
lnFM	-0.6111	-2.66	-1.95	-1.60	
dlnFM	-5.4342	-4.38	-3.60	-3.24	
lnPDR	-0.2483	-2.66	-1.95	-1.60	
dlnPDR	-2.0946	-2.66	-1.95	-1.60	
lnFS	-1.8895	-2.66	-1.95	-1.60	
dlnFS	-2.2609	-2.66	-1.95	-1.60	
lnPDI	-0.2483	-2.66	-1.95	-1.60	
dlnPDI	-2.0946	-2.66	-1.95	-1.60	
lnES	-1.8343	-3.75	-3.00	-2.63	
dlnES	-3.4175	-3.75	-3.00	-2.63	
lnNAR	-3.2003	-3.75	-3.00	-2.63	
dlnNAR	-3.7413	-4.38	-3.60	-3.24	
lnLP	-0.3086	-3.75	-3.00	-2.63	
dlnLP	-3.6965	-4.38	-3.60	-3.24	

4.3. 协整检验

本文通过 ADF 检验发现人口地域结构的 lnNAR 为平稳序列, 因此对该变量不做协整检验; 而对数序列 lnFM、lnPDR、lnFS、lnPDI、lnES、lnLP 都是非平稳的时间序列, 而对数一阶差分之后的序列均为平稳序列, 可以进行协整检验。本文将分别对人口结构的两个子结构(人口自然结构、人口社会经济结构)所选变量与就业人口之间进行协整检验。

由表 3 可知, 在 0.05 的显著性水平下, $-2.3355 < -1.95$, 因此人口自然结构与就业人口的残差序列 1 平稳, 即存在协整关系; $-0.1008 > -1.95$, 所以人口社会经济结构与就业人口的残差序列 2 非平稳, 即不存在协整关系。

Table 3. ADF test results of two residual sequences

表 3. 两个残差序列的 ADF 检验结果

	α 显著性水平下的临界值	实际检验值
残差序列 1	0.05: -1.95	-2.3355
残差序列 2	0.05: -1.95	-0.1008

4.4. 实证分析

4.4.1. 格兰杰因果检验

协整检验表明了就业人口与人口自然结构之间存在长期稳定的关系；而与人口社会经济结构没有这种关系，所以将 LNFM、LNPDR、LNLP 进行滞后一期和二期的 Granger 检验。但由表 2 的 ADF 平稳性检验可知，人口社会经济结构和人口地域结构的五个对数一阶差分序列都是平稳的，因此将 DLNFS、DLNPDI、DLNES、DLNNAR、DLNLP 进行滞后一期和二期的 Granger 检验。具体检验结果如下(表 4)。

Table 4. Results of Granger test

表 4. Granger 检验结果

虚拟假设	Lag = 1			Lag = 2		
	Obs	F 统计量	P 值	Obs	F 统计量	P 值
LNLP does not Granger Cause LNFM	22	15.8713	0.0008	21	4.06561	0.0374
LNFM does not Granger Cause LNLP		32.0958	0.00002		10.0094	0.0015
LNLP does not Granger Cause LNPDR	22	4.11577	0.0567	21	9.97246	0.0015
LNPDR does not Granger Cause LNLP		24.1613	0.0001		0.18957	0.8291
DLNLP does not Granger Cause DLNFS	21	0.39802	0.536	20	1.51072	0.2525
DLNFS does not Granger Cause DLNLP		0.10614	0.7483		0.29285	0.7503
DLNLP does not Granger Cause DLNPDI	21	4.24393	0.0541	20	3.95409	0.0418
DLNPDI does not Granger Cause DLNLP		0.10608	0.7484		11.3565	0.001
DLNLP does not Granger Cause DLNES	21	1.99295	0.1751	20	0.71555	0.5049
DLNES does not Granger Cause DLNLP		0.46186	0.5054		3.2943	0.0652
DLNLP does not Granger Cause DLNNAR	21	15.1605	0.0011	20	7.24253	0.0063
DLNNAR does not Granger Cause DLNLP		0.00524	0.9431		0.23107	0.7965

在人口自然结构中，由于人口性别比(lnFM)、人口抚养比(lnPDR)与就业人口(lnLP)之间存在相关性，因此将人口性别比(lnFM)、和人口抚养比(lnPDR)纳入 VAR 模型中的解释变量，就业人口(lnLP)为被解释变量；在人口社会经济结构中，由于家庭规模(dlnFS)与就业人口(dlnLP)不存在相关性，实际纳入 VAR 模型的变量为城镇居民人均可支配收入(dlnPDI)、教育结构(dlnES)，就业人口(dlnLP)；在人口地域结构中，非农业人口比重(dlnNAR)没有通过 Granger 因果检验，因此不采用人口地域结构指标建立 VAR 模型。

4.4.2. VAR 模型构造及稳定性检验

①人口自然结构 VAR 模型

对数序列 lnFM、lnPDR 与 lnLP 存在协整关系，接下来对人口自然结构基于序列协整建立 VAR 模型。根据表 5 结果可知，最优滞后期为 2，则建立 VAR (2)模型如下。

Table 5. Optimal lag period selection for model 1

表 5. 模型 1 的最优滞后期选择

Lag	LogL	LR	FPE	AIC	SC	HQ
0	169.6493	NA	1.16E-11	-16.66493	-16.51557	-16.63577

续表

1	267.5518	156.6441*	1.63E-15	-25.55518	-24.95774*	-25.43856
2	279.6443	15.72027	1.29E-15*	-25.86443*	-24.81892	-25.66034*
3	288.5195	8.875194	1.60E-15	-25.85195	-24.35836	-25.56039

$$\begin{pmatrix} \text{LNLP}_t \\ \text{LNFM}_t \\ \text{LNPDR}_t \end{pmatrix} = \begin{pmatrix} 0.970265 \\ -0.177867 \\ 60.286003 \end{pmatrix} + \begin{pmatrix} 1.451396 & 0.357317 & -0.003709 \\ 0.669687 & 0.529731 & 0.046487 \\ -29.694451 & -2.927217 & 0.395345 \end{pmatrix} \begin{pmatrix} \text{LNLP}_{t-1} \\ \text{LNFM}_{t-1} \\ \text{LNPDR}_{t-1} \end{pmatrix} \\ + \begin{pmatrix} -0.540041 & -0.156168 & -0.011784 \\ -0.650522 & 0.041615 & -0.029401 \\ 24.261092 & 5.792171 & -0.051594 \end{pmatrix} \begin{pmatrix} \text{LNLP}_{t-2} \\ \text{LNFM}_{t-2} \\ \text{LNPDR}_{t-2} \end{pmatrix}$$

如图 2 所示, 向量自回归模型对应的特征根的模长都小于 1, 说明构建的 VAR 模型具有稳定性。

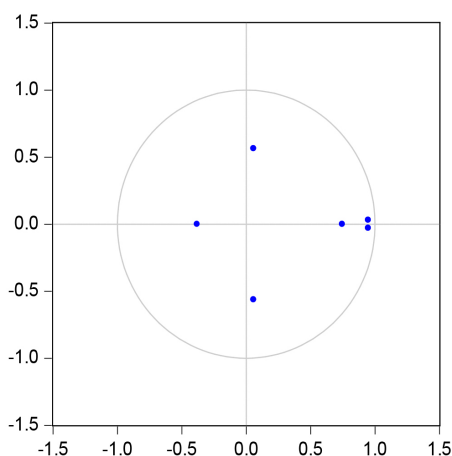


Figure 2. AR root graph of VAR (2) model
图 2. VAR (2)模型 AR 根图

②人口社会经济结构 VAR 模型

对数一阶差分序列 dlnPDI 、 dlnES 、 dlnLP 为平稳序列, 接下来对人口社会经济结构和就业人口基于平稳序列建立 VAR 模型。

根据表 6 结果可知, 最优滞后期为 1, 则建立的 VAR (1)模型如下。

$$\begin{pmatrix} \text{DLNLP}_t \\ \text{DLNPDI}_t \\ \text{DLNES}_t \end{pmatrix} = \begin{pmatrix} -0.000694 \\ 0.013580 \\ 0.186660 \end{pmatrix} + \begin{pmatrix} 0.9874029 & 0.001517 & -0.002222 \\ 1.358529 & 0.709367 & 0.117111 \\ 7.335353 & -1.141832 & -0.309808 \end{pmatrix} \begin{pmatrix} \text{DLNLP}_{t-1} \\ \text{DLNPDI}_{t-1} \\ \text{DLNES}_{t-1} \end{pmatrix}$$

Table 6. Optimal lag period selection for model 2

表 6. 模型 2 的最优滞后期选择

Lag	LogL	LR	FPE	AIC	SC	HQ
0	162.4821	NA	5.08E-11	-15.18877	-15.03956	-15.15639
1	202.1225	64.17970*	2.79E-12*	-18.10691*	-17.51004*	-17.97737*

如图 3 所示, 向量自回归模型对应的特征根的模长都小于 1, 说明构建的 VAR 模型具有稳定性。

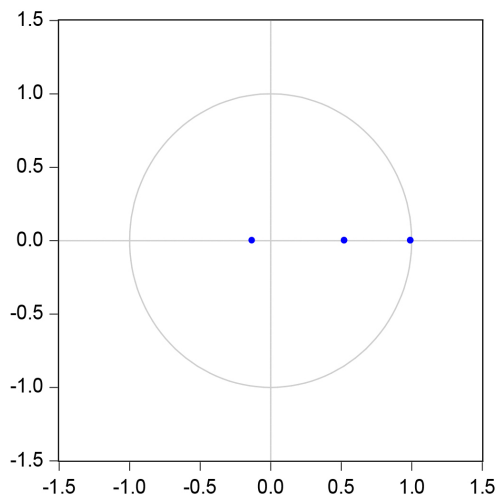


Figure 3. AR root graph of VAR (1) model

图 3. VAR (1)模型 AR 根图

4.4.3. 脉冲响应函数分析

①人口性别比(LNFM)冲击就业人口(LNLP)

图 4 显示, 人口性别比的正向冲击持续正向拉升就业人口, 而且这种拉升效应一直在增强。由此可见, 对于全国而言, 人口性别比的增加在短期和长期都能够促进就业人口的上升, 也可以解释为男性就业率比女性就业率高。

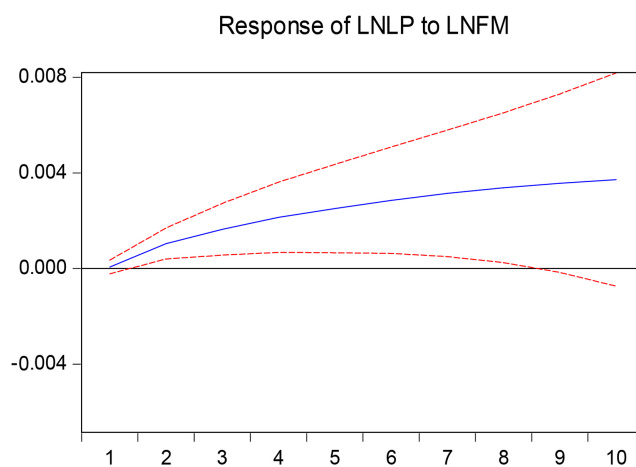


Figure 4. The impulse response function diagram of the employment population generated by the shock to the sex ratio of the population

图 4. 人口性别比冲击产生的就业人口脉冲响应函数图

②人口抚养比(LNPDR)冲击就业人口(LNLP)

图 5 显示, 人口抚养比的正向冲击对就业人口脉冲响应函数的负向影响开始逐渐增强。由此可见, 人口抚养比的增加在短期和长期都会导致就业人口数目的下降, 也可以解释为由于人口老龄化会导致人口抚养比增加, 进而使得就业人口数目减少。

Response of LNLP to LNPDR

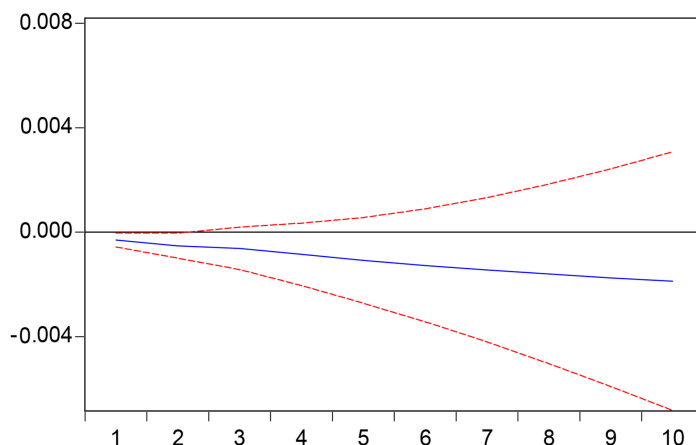


Figure 5. The impulse response function diagram of the employment population generated by the dependency ratio shock

图 5. 人口抚养比冲击产生的就业人口脉冲响应函数图

③人均可支配收入(DLNPD)冲击就业人口(LNLP)

图 6 显示, 人均可支配收入在第一期对于就业人口的冲击几乎没有影响, 在第二期达到最大的负向影响, 从第五期逐渐稳定且为负。由此可见, 长期来看人均可支配收入对于就业人口具有负向作用, 可以解释为人均可支配收入减少, 就业人口就会增加。

Response of DLNLP to DLNES

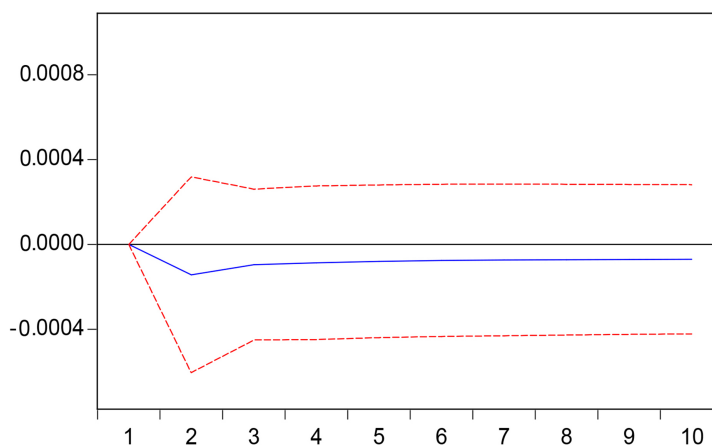


Figure 6. Impulse response function diagram of employment population generated by per capita disposable income shock

图 6. 人均可支配收入冲击产生的就业人口脉冲响应函数图

④教育结构(DLNES)冲击就业人口(LNLP)

图 7 显示, 教育结构施加的正向冲击持续正向拉升就业人口数目, 而且这种拉升效应一直在增强, 在第六期达到最大的正向影响, 此后逐渐稳定并为正。教育结构反映的是接受高等教育及以上的情况。全国的教育结构近年来稳定缓慢上升, 至 2020 年已经达到 17.19%, 也可以解释为受高等教育的人数越多, 就业人口越多。

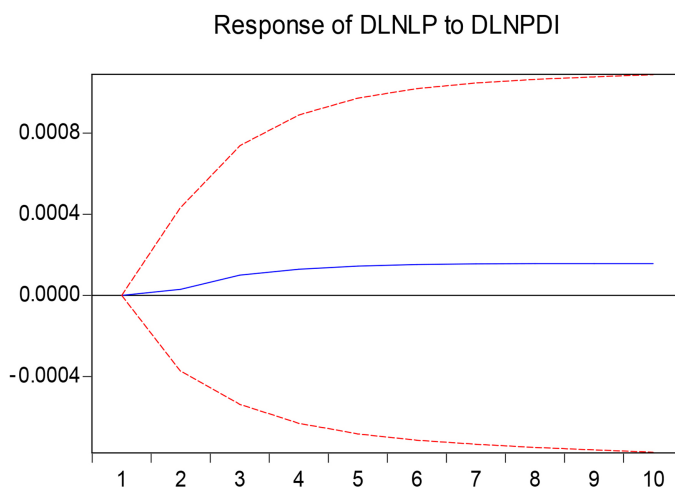


Figure 7. Impulse response function diagram of employment population caused by educational structure shock

图 7. 教育结构冲击产生的就业人口脉冲响应函数图

4.4.4. 方差分解分析

方差分解进一步评价人口性别比、人口抚养比、人均可支配收入、教育结构的贡献度。

①人口自然结构方差分解结果

由表 7 可知, 从短期来看人口性别比对就业人口数目影响不大, 但在长期人口性别比波动对就业人口波动具有重要的影响, 而且人口性别比对就业人口的波动要高于人口抚养比。

Table 7. The contribution rate of population sex ratio and dependency ratio to the employed population

表 7. 人口性别比、抚养比对就业人口的贡献率

Period	S.E.	LNFM	LNPDR	LNLP
1	0.002712	0.957703	20.76371	78.27859
2	0.003341	43.7381	14.45444	41.80746
3	0.003591	55.47597	10.97877	33.54526
4	0.00363	61.40186	10.78734	27.8108
5	0.003651	63.68125	11.46271	24.85604
6	0.003687	64.7748	12.11838	23.10682
7	0.003757	65.33335	12.68425	21.9824
8	0.003866	65.56807	13.24764	21.1843
9	0.004012	65.57713	13.81048	20.61239
10	0.004189	65.4565	14.34597	20.19754

②人口社会经济结构方差分解结果

由表 8 可知, 从短期来看人均可支配收入对就业人口数目影响较小, 教育结构对就业人口影响较大; 在长期人均可支配收入波动对就业人口波动有微小的影响, 同样教育结构对就业人口的波动要高于人均可支配收入。

Table 8. The contribution rate of per capita income and educational structure to the employed population
表 8. 人均可支配收入、教育结构对就业人口的贡献率

Period	S.E.	DLNPDI	DLNES	DLNLP
1	0.068312	0	14.6462	85.3538
2	0.07381	0.042675	10.54433	89.413
3	0.07442	0.361989	9.818888	89.81912
4	0.074632	0.701642	9.537522	89.76084
5	0.074693	1.003925	9.426878	89.5692
6	0.074715	1.252897	9.376587	89.37052
7	0.074726	1.453611	9.351634	89.19476
8	0.074734	1.615178	9.337957	89.04686
9	0.07474	1.746239	9.32968	88.92408
10	0.074747	1.853778	9.324175	88.82205

4.4.5. 拟合

①两个 VAR 模型的拟合

人口自然结构与人口社会经济结构的模型拟合优度分别为 0.998953 和 0.950503, 两个 VAR 模型拟合效果较好, 图 8 和图 9 的拟合图也与之对应。

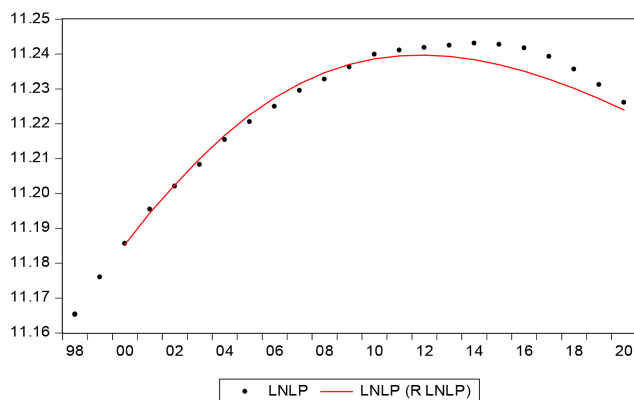


Figure 8. Fitting diagram of natural structure of population-VAR (2) model
图 8. 人口自然结构-VAR (2)模型的拟合图

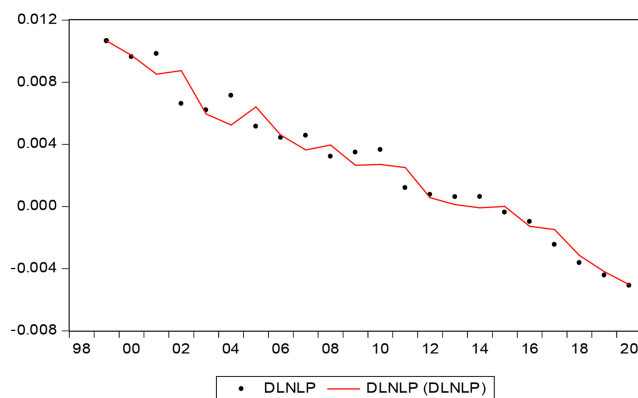


Figure 9. Population socioeconomic structure-fitting diagram of VAR (1) model
图 9. 人口社会经济结构-VAR (1)模型的拟合图

②LSTM 模型的拟合

对原始时间序列作归一化处理, 训练集为前 85% 的数据, 测试集为后 15% 的数据, 输入层为人口结构的三个子结构出发所得 6 个因素, 输出层为预测的 1998 年到 2020 年的就业人数, 输入层的维度为 6, 输出层的维度为 1, 训练次数为 3000, 采用 sigmoid 函数作为激活函数。

运用三个指标对 LSTM 模型拟合结果进行评价: 平均绝对误差、均方误差、均方根误差。平均绝对误差: $MAE = \frac{1}{m} \sum_{i=1}^m |y_i - \hat{y}_i|$, 能反映预测值误差实际情况, 其值越小, 模型的拟合效果就越好。均方误差:

$MSE = \frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2$, 可以评价数据的变化程度, 其值越小说明预测模型精确度更高。均方根误差

$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2}$, 用来衡量观测值和真实值的偏差[8] (图 10)。

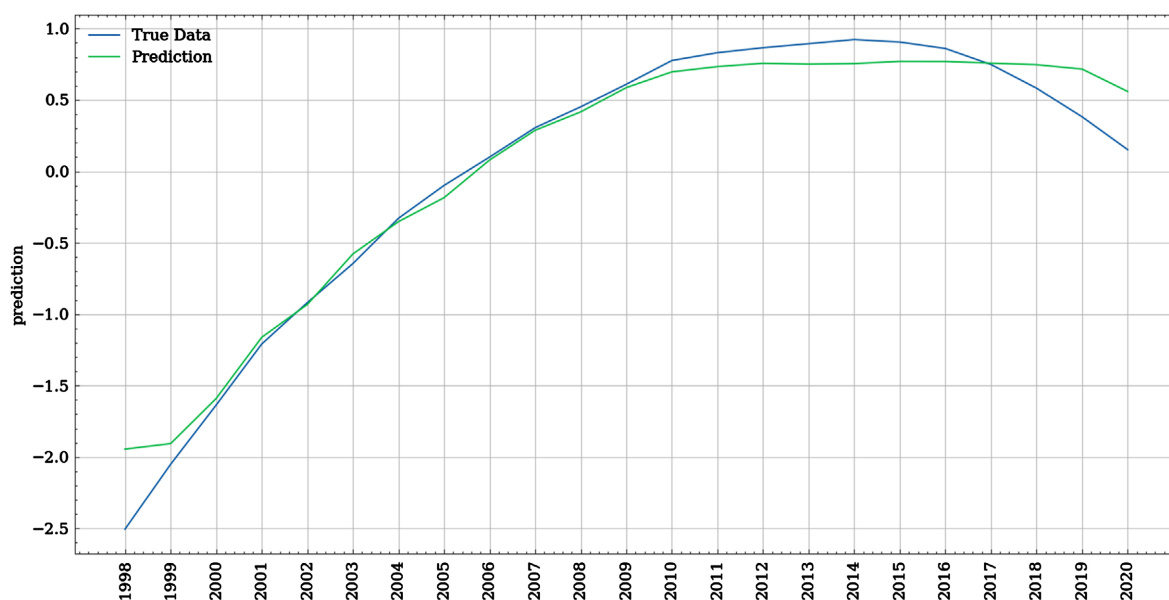


Figure 10. Fitting diagram of LSTM model

图 10. LSTM 模型拟合图

运用 Python 软件计算得到模型评价指标值: MAE: 0.12228; MSE: 0.02218; RMSE: 0.18214, 说明该模型拟合效果很好。

5. 总结与建议

5.1. 总结

本文对人口结构的指标与就业之间的关系进行了实证分析, 得出以下结论: 格兰杰因果检验显示, 人口性别比、人口抚养比、人均可支配收入、教育结构与就业人口之间具有相关性; 脉冲响应函数分析显示, 人口性别比、教育结构对就业人口具有长期的正向影响, 人口抚养比、人均可支配收入对就业人口具有长期的负向影响; 方差分解分析结果显示, 人口自然结构中短期内人口抚养比对就业人口的影响大于人口性别比, 但是长期来看人口性别比影响更大; 人口社会经济结构中人均可支配收入对于就业人口的影响无论是长期还是短期均小于教育结构的影响; VAR 模型与 LSTM 模型拟合效果都很好, 但是 VAR 模型可解释性更强, 是由于 VAR 通过其模型参数直接揭示了变量间的相互作用和联系, 且每个系

数都有明确的经济意义, 可以直接解释为一个变量对另一个变量的影响强度和方向。

5.2. 建议

5.2.1. 建立健全养老保障体系

人口自然增长率下降, 老年群体不断扩大, 人口老龄化现象加重了年轻一代的压力, 他们需要投入工作赚取更多的抚养资本, 这使得相当一部分老年人口沦为“空巢老人”, 而随着“空巢老人”数量的不断增加, 家庭养老功能进一步弱化, 养老难题成为社会有关部门的重点关注对象, 社会负担加重, 因此建立健全养老保障体系才会使社会经济稳定发展。

5.2.2. 大力发展教育事业

根据当前就业形势来看, 毕业的人数逐年递增, 竞争压力越来越大, 要想在众人中脱颖而出, 学历是首先考虑的一个因素。教育结构与就业人口关系较为紧密, 其原因是用人单位更愿意招收高等院校毕业的学生, 从而对就业人口稳定健康发展具有重要意义。每个城市的经济发展水平不同, 部分城市在教育上的投入远低于发达城市。建议在未来政府应该加大对教育的投入、改善办学软件条件和硬件设施, 以提高居民的文化水平、文化素质和劳动技能, 让更多的人获取学历, 营造良好的就业形势。

参考文献

- [1] 中国人事科学研究院. 《人口结构变化对就业的影响研究》项目成果摘要[EB/OL]. http://www.mohrss.gov.cn/xgk2020/fdzdgknr/ghtj/kyxx/202109/t20210908_422627.html, 2022-05-21.
- [2] 朱文涛, 刘秀玲. 人口城镇化、就业非农化与城乡收入差距的动态关系分析[J]. 统计与决策, 2016(12): 141-144.
- [3] 莫旋, 唐成千, 阳玉香. 城镇化进程中流动人口就业影响因素与就业选择——分层异质视角下多元选择模型的实证分析[J]. 商业研究, 2019(7): 36-41.
- [4] 林耿, 王炼军. 阶层化背景下的就业空间——以常住人口与流动人口为例[J]. 地理研究, 2010, 29(6): 1069-1082.
- [5] 刘宏, 李述晟. FDI 对我国经济增长、就业影响研究——基于 VAR 模型[J]. 国际贸易问题, 2013(4): 105-114. <https://doi.org/10.13510/j.cnki.jit.2013.04.002>
- [6] Yang, H., Pan, Z., Tao, Q. and Qiu, J. (2018) Online Learning for Vector Autoregressive Moving-Average Time Series Prediction. *Neurocomputing*, **315**, 9-17. <https://doi.org/10.1016/j.neucom.2018.04.011>
- [7] Xing, L. and Ting, Y. (2021) Forecast of the Employment Situation of College Graduates Based on the LSTM Neural Network. *Computational Intelligence and Neuroscience*, **2021**, Article ID: 5787355.
- [8] 曾安, 聂文俊. 基于深度双向 LSTM 的股票推荐系统[J]. 计算机科学, 2019, 46(10): 84-89.