

Research on the Prediction of China's Aging Population Based on ARIMA Model

Enlai Li

School of Statistics and Mathematics, Yunnan University of Finance and Economics, Kunming Yunnan
Email: 825634143@qq.com

Received: Jul. 26th, 2016; accepted: Aug. 9th, 2016; published: Aug. 16th, 2016

Copyright © 2016 by author and Hans Publishers Inc.

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

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



Open Access

Abstract

With the Chinese people's living standards achieving a well-off level, the spiritual material is further protected, and the average life expectancy is extended gradually. Hence aging population appears inevitably, and other problems emerge with it. It is quite a challenge for a developing country, which is very difficult to deal with. In this paper, the ARIMA model is used to forecast the China's aging population in a short term. We can find that the current situation of China's population ageing is quite severe, and some policy suggestions are proposed aimed at solving the social problems currently produced from China's aging population.

Keywords

Aging Population, Short-Term Prediction, ARIMA Model, Suggestion

基于ARIMA模型的中国老龄化人口预测研究

李恩来

云南财经大学统计与数学学院, 云南 昆明
Email: 825634143@qq.com

收稿日期: 2016年7月26日; 录用日期: 2016年8月9日; 发布日期: 2016年8月16日

摘要

随着中国人民的生活水平达到小康水平，精神物质得到进一步保障，人口平均寿命逐渐向后延迟。人口老龄化问题必然出现，相继的问题也随之出现。这样的问题对于一个发展中国家来讲是非常有挑战性的，也是非常棘手的。本文利用ARIMA模型来对中国老龄化人口进行短期预测，通过对老龄人口的短期预测，我们可以发现目前我国的老龄化形势相当严峻，同时针对目前我国所面临的人口老龄化所产生的一些社会问题提供了一些政策性的建议。

关键词

人口老龄化，短期预测，ARIMA模型，建议

1. 研究背景

人口老龄化通常是指人口生育率降低和人均寿命延长导致的总人口中因年轻人口数量减少、年长人口数量增加而导致的老年人口比例相应增长的动态。其具有两个含义：一是指老年人口相对增多，在总人口中所占比例不断上升的过程；二是指社会人口结构呈现老年状态，进入老龄化社会。国际上通常看法是，当一个国家或地区 60 岁以上老年人口占人口总数的 10%，或 65 岁以上老年人口占人口总数的 7%，即意味着这个国家或地区的人口处于老龄化社会。本文选用 65 岁以上老年人口进行我国人口老龄化的研究。从新中国成立到现在，我国经济突飞猛进的发展，我国人民的生活水平得到很大的提高，而且我国医疗水平也在不断地发展，导致我国人口平均寿命逐渐向后延迟，因此我国人口老龄化问题必然出现，这将导致我国一些社会问题相继随之出现。这些问题对于一个发展中的国家来说是赋有挑战性，也是非常棘手的问题。能否处理和解决好此问题，是关系我国能否顺利实现下个目标的关键。因此对我国人口老龄化预测就有重要的意义。

国外有许多的学者对人口老龄化的预测模型的研究，相比我国要早一些，国外研究是从 John Graunt 对人口预测模型的研究开始，其中他发布的关于人口预测的文章《死亡表的观察》比较著名，他在死亡表当中引进了一种新的比率，那就是死亡率，据此对以后人口的演变趋势进行预测。从而开创了人口预测的先河。随后日本的 Ueda Masco 以及瑞典的 Grstav Sundbarg [1]等人口学家在 1898 年都对人口老龄化问题做了进一步的研究，找出社会经济发展与人口老龄化之间的关系，从而做了更深入的研究。Grstav Sundbarg 在研究人口预测的时候提出了一个新的假设条件：世界人口的生育率水平首先会随着时间的推移下降至更替水平，然后维持在这个水平长期保持不变。

国内也有许多的学者对人口老龄化的研究方法有很多种。尹春华[2]在 2005 年利用数据挖掘中的 BP 神经网络技术，构建了人口预测模型。安和平[3]在 2005 年研究了中国人人口预测的自回归分布滞后模型。朱兴造[4]在 2009 年将人口预测模型中的自回归模型和经典的 logistic 离散模型进行了对比分析，并且对这两个模型理论进行了推广。甘蓉蓉[5]在 2010 年运用生态足迹法、灰色模型法和回归分析法对 2015 年汉中市总人口进行预测，并对预测结果进行比较分析；蒋远营[6]在 2011 年将人口发展方程模型应用到我国人口的预测中；任强[7]在 2011 年研究了人口预测的随机方法。陈毅华[8]在 2012 年运用径向基函数神经网络(Radial Basis Function, RBF)方法建立了人口老龄化的定量预测模型。

本文基于 ARIMA 模型对我国老龄化人口进行预测研究。主要因为我国老龄化人口可以看作一个不平稳的时间序列，而且 ARIMA 模型是一种精度较高的时序短期预测方法，能够从本质上认识时间序列的结构和特征，可以达到最小均方误差意义下的最有预测。

2. 数据分析

2.1. 数据来源及说明

本文所用的 65 岁及以上的人口数据来源于 2015 年中国统计年鉴。本文基于 2015 年中国统计年鉴上收集了我国 1990~2014 年的我国人口年龄结构数据(具体数据见表 1)。可以通过表 1, 对我国人口老龄化的现状进行简单的了解。本文将 1990~2012 年的数据作为训练集, 根据该数据建立预测模型, 然后将 2013~2014 年的数据作为测试集, 说明所建模型的预测效果。最后, 对我国未来 5 年的人口老龄化情况做预测, 进一步说明其未来发展趋势(数据来源网址: <http://www.stats.gov.cn/tjsj/ndsj/2015/indexch.htm>)。

2.2. 我国老龄化人口的现状分析

根据表 1 可以看出我国人口结构出现两头小中间大, 青少年(0~14 岁)人口数比重逐年在减少, 而老龄人口(65 岁以上)却在逐年在增加。按照国际上通常看法, 我国早在 2000 年, 老龄人口的比重就达到了 7.0%。这说明我国早在 2000 年就进入了老龄化社会。而且我国中间人口(15~64 岁)也逐年在增加, 这势必会在加剧我国今后几十年老龄人口的增长速度, 导致我国老龄化问题加剧。在 1990~2014 年期间, 我国老龄化人口比重增加了 4.6, 平均每年老龄化人口比重增长约 0.184。由此可见, 我国老龄化人口在逐年增长, 老龄化程度越来越严重。

2.3. 老龄化人口数据分析

根据 1990~2014 年我国老龄化人口数, 做出其散点图, 如图 1 所示。由图 1, 可以发现我国老龄化人口数呈逐年递增的趋势, 而从老龄化人口所占比重来看, 我国老龄化人口比重均在 5.6%~6.9%之间, 但是在 2000 年之后, 我国老龄化人口均在 7.0%之上。按照国际上通常看法, 我国从 2000 年就开始进入了老龄化社会, 而且我国老龄化程度在逐年加剧。我国老龄化人口逐年在增加, 老龄化程度不断加剧, 我国所面临的问题也越来越严峻。因此, 对我国老龄化人口预测, 这样有利于我国出台相应的政策来应对我国老龄化问题是很大的帮助。这就使得我国老龄化人口预测有重要的实际意义。

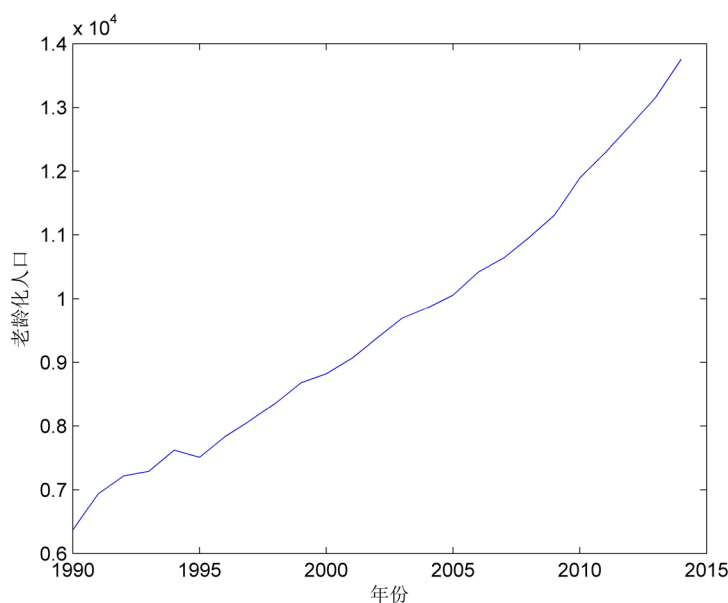


Figure 1. The aging population of our country in 1990-2014

图 1. 我国 1990~2014 年老龄化人口

Table 1. Population structure data of 1990-2014 in China (unit: ten thousand)
表 1. 中国 1990~2014 年历年人口结构数据(单位: 万人)

年份	总人口(年末)	0~14 岁		15~64 岁		65 岁以上	
		人口数	比重(%)	人口数	比重(%)	人口数	比重(%)
1990	114,333	31,659	27.7	76,306	66.7	6368	5.6
1991	115,823	32,095	27.7	76,790	66.3	6938	6
1992	117,171	32,339	27.6	77,614	66.2	7218	6.2
1993	118,517	32,177	27.2	79,051	66.6	7289	6.2
1994	119,850	32,360	27	79,868	66.6	7622	6.4
1995	121,121	32,218	26.6	81,393	67.2	7510	6.2
1996	122,389	32,311	26.4	82,245	67.2	7833	6.4
1997	123,626	32,093	26	83,448	67.5	8085	6.5
1998	124,761	32,064	25.7	84,338	67.6	8359	6.7
1999	125,786	31,950	25.4	85,157	67.7	8679	6.9
2000	126,743	29,012	22.9	88,910	70.1	8821	7
2001	127,627	28,716	22.5	89,849	70.4	9062	7.1
2002	128,453	28,774	22.4	90,302	70.3	9377	7.3
2003	129,227	28,559	22.1	90,976	70.4	9692	7.5
2004	129,988	27,947	21.5	92,184	70.9	9857	7.6
2005	130,756	36,504	20.3	84,197	72	10,055	7.7
2006	131,446	25,961	19.8	95,066	72.3	10,419	7.9
2007	132,129	25,660	19.4	95,833	72.5	10,636	8.1
2008	132,802	25,166	19	96,680	72.7	10,956	8.3
2009	133,450	24,659	18.5	97,484	73	11,307	8.5
2010	134,091	22,259	16.6	99,938	74.5	11,894	8.9
2011	134,735	22,164	16.5	100,283	74.4	12,288	9.1
2012	135,404	22,287	16.5	100,403	74.1	12,714	9.4
2013	136,072	22,329	16.4	100,582	73.9	13,161	9.7
2014	136,782	22,558	16.5	100,469	73.4	13,755	10.1

3. 模型构建

3.1. ARIMA 模型

自回归积分移动平均模型[9]-[11] (Autoregressive Integrated Moving Average Model, 简记 ARIMA)是由 Box 和 Jenkins 提出的一个著名时间序列预测方法, 又称为 Box-Jenkins 模型。它作为一元时间序列分析中的经典模型, 是时间序列分析中较为成熟和应用较为广泛的方法之一。由自回归 AR(p)、差分 I(d)和移动平均 MA(q)三个部分组成, 模型中 p 表示模型的自回归阶数、d 表示非平稳资料转化成平稳资料的差分阶数、q 表示模型移动平均阶数。ARIMA 模型按照季节性波动分为非季节性 ARIMA 模型与季节性 ARIMA 模型。本文主要基于非季节性 ARIMA 模型进行建模。非季节性 ARIMA 模型表达式如下:

$$\nabla^d x_t = c + \frac{\Theta(B)}{\Phi(B)} \varepsilon_t$$

$$\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3 - \dots - \phi_p B^p$$

$$\Theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \theta_3 B^3 - \dots - \theta_q B^q$$

其中 x_t 为所分析的时间序列, ∇ 表示差分算子, B 表示后移算子, c 为常数项, ε 属于白噪声, 此时模型记为 ARIMA(p, d, q)。由上面第一个式子可知, ARIMA 模型的实质就是差分算子与 ARMA 模型的结合。其中上面第二个式子为平稳可逆 ARMA(p, q) 模型的自回归系数多项式, 其中上面第三个式子为平稳可逆 ARMA(p, q) 模型的移动平均系数多项式。这一关系意义重大, 这说明非平稳的序列可以通过适当阶数的差分实现差分后平稳, 就可以对差分后的序列进行 ARMA 模型拟合。而如今 ARMA 模型的分析方法非常成熟, 这意味着对差分后的平稳序列的分析也将是非常简单、极其可靠。需要特别指出的是:

当 $p = 0$ 时, ARIMA(p, d, q) 模型实际上就是 IMA(d, q) 模型。

当 $d = 0$ 时, ARIMA(p, d, q) 模型实际上就是 ARMA(p, q) 模型。

当 $q = 0$ 时, ARIMA(p, d, q) 模型实际上就是 ARI(p, d) 模型。

ARIMA(p, d, q) 模型是指 d 阶差分后自相关最高的阶数为 p , 移动平均最高的阶数为 q , 模型通常包含 $p + q$ 个独立的未知系数。

3.2. 模型的识别

ARIMA 建模是把含有趋势的序列通过差分后转换为平稳的序列, 对差分后平稳的序列拟合 ARMA 模型, 这也就是指时间序列的平稳是模型建立的必要条件。本文通过时序图, 自相关图以及纯随机性检验来确定序列的平稳性与平稳性的纯随机检验。在利用 ARIMA 模型时如何确定其阶数较为关键。常见的方法是 Box-Jenkins 定阶法, 主要是根据样本自相关、偏相关函数的截尾与拖尾的性质初步判断序列的合适阶数(表 2)。

利用自相关、偏相关函数的统计特性来判断模型的阶数, 虽然简单, 但是在实践中却不是一件容易的事, 同时也不能保证精确度。需要与其他模型定阶方法一起使用, 提高精确度。本文选用 AIC, BIC 来提高模型的精确度。

4. 老龄化人口数据分析及预测结果

本次实验的数据是从中国 2015 年统计年鉴上收集而来的, 本文将 1990~2012 年的数据作为训练集, 而 2013~2014 年数据作为测试集。本文首先要对数据的平稳性与其纯随机性进行验证, 在确保数据的平稳性和非纯随机性后, 利用 ARMA 模型来拟合序列, 其次根据 Box-Jenkins 定阶法和 AIC 以及 BIC 来进行模型定阶, 从而确定最优模型。然后利用测试数据来验证模型的优劣, 最后进行我国今后五年的老龄化人口的预测。

4.1. 数据平稳性检验

本文首先通过时序图来看序列是否有明显的趋势, 如果有明显的趋势, 则可以说明该序列是非平稳序列, 如果没有明显的趋势, 那么需要通过自相关图来检验其是否为平稳序列, 如果自相关系数随着延迟期数 k 的增加, 自相关系数会很快地衰减向零, 则是平稳序列, 否则不是平稳序列。通过图 1 可以看出其时序图具有明显的上涨趋势。因此可以知道老龄化人口数据是非平稳序列。我们对该数据做一阶差分, 得到一个新的序列, 对新序列做时序图(见图 2), 我们不能看出其具有明显的趋势, 故需要做其自相关图进行判断其平稳性。通过其自相关图(见图 3), 发现其衰减向零的速度比较快, 故可以判断其为平稳

Table 2. The statistical property of autocorrelation or partial autocorrelation function
表 2. 各个模型的自相关、偏相关函数的统计特性

自相关函数	偏自相关函数	模型的阶数
拖尾	P 阶结尾	AR(p)
q 阶结尾	拖尾	MA(q)
拖尾	拖尾	ARMA(p, q)

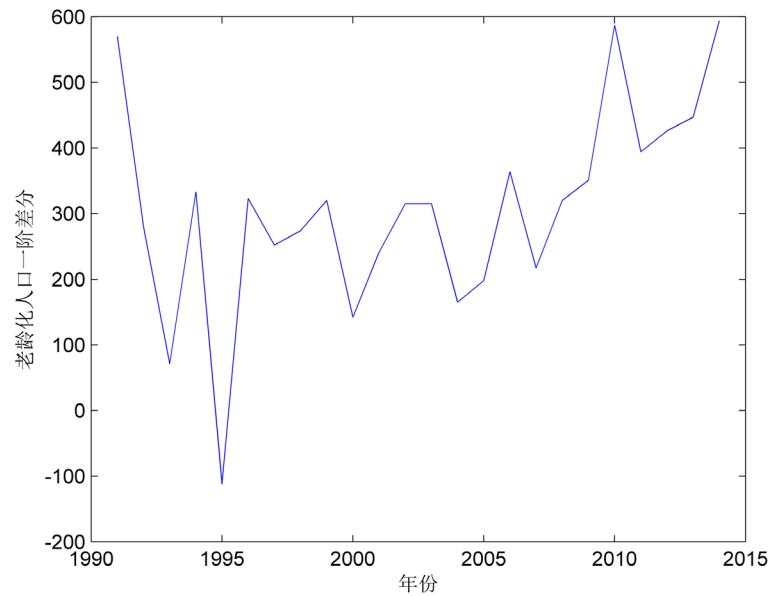


Figure 2. The sequence diagram of the first order difference of the aging population

图 2. 老龄化人口一阶差分的时序图

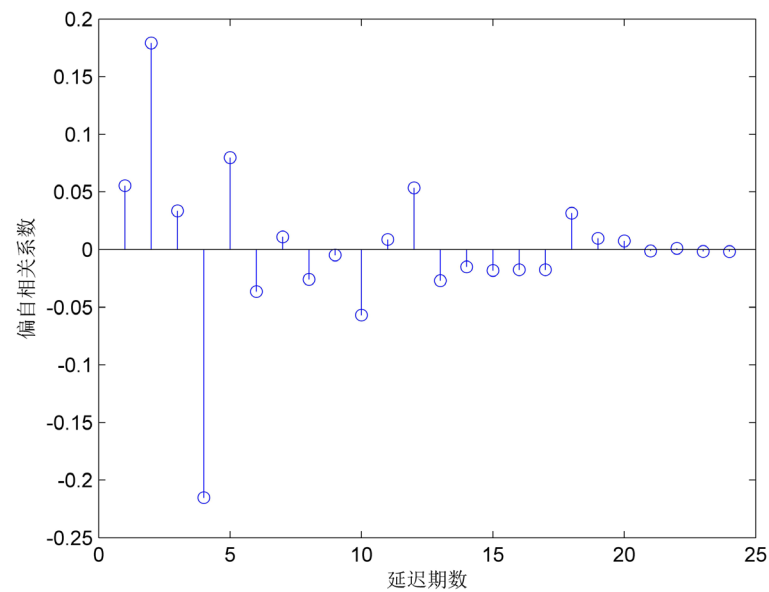


Figure 3. The autocorrelation plot of the first order difference of the aging population

图 3. 老龄化人口一阶差分的自相关图

序列,故可以判断其为平稳序列。最后来验证其是否为纯随机性序列,检验结果见表3($\alpha = 0.05$),可以断定老龄化人口的一阶差分序列属于非白噪声序列。

4.2. ARMA 模型拟合和模型识别

通过上面序列的平稳性检验,我们可以知道,通过做一阶差分序列就平稳了,现在我需要拟合老龄化人口一阶差分的数据,本文是基于极大似然估计方法对 ARMA 模型所涉及的参数进行估计,但是在估计参数之前我们需要对 ARMA 模型进行定阶数,本文基于 Box-Jenkins 定阶法,主要是利用自相关图以及偏自相关图来粗略定一个大概的阶数,然后通过 AIC 以及 BIC 来选择最优的 ARMA 模型,通过上面的自相关图(图 3)可以看出有拖尾现象,通过下面的偏自相关图(图 4),可以看出也拖尾现象。因此本文选择 ARMA 模型,通过 AIC 和 BIC 在 ARMA(p, q), $p = 0, 1, \dots, 5$, $q = 0, 1, \dots, 5$ 中选择最优模型,各个模型的 AIC 和 BIC 值本文就一一给出,仅给出最小五个的值,见表 4,可以看出最优的模型是 ARMA(1, 1),可以得到其相应的表达式为:

$$x_t = 116.1248 + 0.6045x_{t-1} + \varepsilon_t - 0.5066\varepsilon_{t-1}$$

4.3. ARIMA 模型预测

根据上面的分析,可以知道最优的模型为 ARIMA(1, 1, 1),利用 Matlab 程序来预测 2013 年和 2014 年的我国老龄化人口,可知 2013 年和 2014 年的预测结果分别为 13,049 万人和 13,657 万人。我们可以算出其与 2013 年和 2014 年真实值相比都偏小,但是其绝对偏差均低于 1% (分别为 0.851% 和 0.713%),下面是利用 ARIMA(1, 1, 1)模型进行预测我国未来五年的老龄化人口,其结果见表 5。

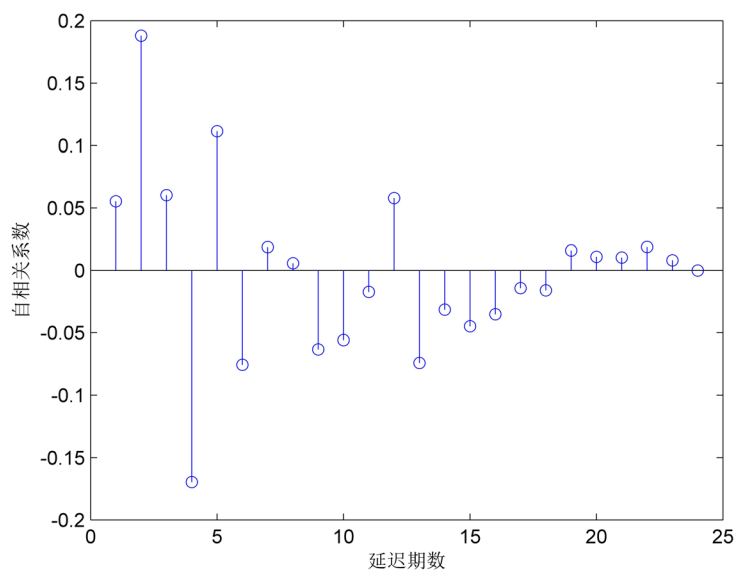


Figure 4. The partial autocorrelation of the first order difference of the aging population
图 4. 老龄化人口一阶差分的偏自相关图

Table 3. LB statistics test results
表 3. LB 统计量检验结果

延迟阶数	LB 检验统计量的值	P 值
6	45.23	<0.01
12	58.47	<0.01

Table 4. AIC and BIC values of the ARMA model
表 4. ARMA 模型的 AIC 和 BIC 值

	$p = 0, q = 0$	$p = 0, q = 1$	$p = 1, q = 0$	$p = 1, q = 1$	$p = 0, q = 2$
AIC	287.7911	288.6307	289.1822	286.1530	289.7047
BIC	302.3374	305.3590	305.9106	301.9323	310.6153

Table 5. Prediction results of aging population in China during 2015-2019 s
表 5. 2015~2019 年我国老龄化人口数预测结果

年份	2015	2016	2017	2018	2019
人口数(单位: 万人)	14,193	14,732	15,297	15,888	16,511

5. 结论与建议

5.1. 结论

经过上面的分析, 可以知道利用 AMIMA(1, 1, 1)模型对我国老龄化人口(65 岁及以上)进行预测, 其预测精度相对比较高(测试集的绝对误差均在 1%)。通过预测我国老龄化人口未来五年的预测, 预测结果表明: 随着时间的推移, 我国老龄化人口的数量越来越多, 呈逐年增加的趋势, 这与我国现在人口结构有很大的关系, 我国 15~64 岁的比重高达 70% 以上。这也我国的医疗越来越发达和我国人民生活水平的提高有关, 从而导致我国人均寿命大幅度提高。但可以发现我国老龄化人口增长速度在下降, 这主要与我国 1971 年实行计划生育政策有关, 使得我国人口得到有效的控制。

5.2. 建议

随着我国人口老龄化加剧, 这会导致许多的社会问题, 比如: 养老问题, 通过上面的分析, 可以发现虽然我国人口老龄化程度还在不断地加剧, 但是我国老龄化人口的增长速度却在下降, 这说明我国实施的计划生育控制我国人口增长的政策是有效的。既然我国人口老龄化是不可避免, 那么就需要我们积极面对, 首先我们应该完善我国各项养老保险制度, 使得我们老有所养; 适当地延迟退休年龄, 以减轻国家的养老负担; 适当地调整人口生育政策, 以延迟老龄化人口峰值期的到来。大力促进老年产业的发展, 这样不仅有利于我国经济的可持续发展, 而且有利提高我国养老保障水平, 从而缓解老年人所带来的养老负担。

参考文献 (References)

- [1] Hansen, P.E. (1989) Leslie Matrix Models. *Mathematical Population Studies*, 2, 209-222.
<http://dx.doi.org/10.1080/08898488909525291>
- [2] 尹春华, 陈雷. 基于 BP 神经网络人口预测模型的研究与应用[J]. 人口学刊, 2005(2): 45-49.
- [3] 安和平. 中国人口预测的自回归分布滞后模型研究[J]. 统计与决策, 2005(16): 6-9.
- [4] 朱兴造, 庞飞宇. 自回归及 logistic 离散模型在中国人口预测中的应用[J]. 统计与决策, 2009, 13(13): 159-161.
- [5] 甘蓉蓉, 陈娜姿. 人口预测的方法比较——以生态足迹法、灰色模型法及回归分析法为例[J]. 西北人口, 2010, 31(1): 57-60.
- [6] 蒋远营, 王想. 人口发展方程模型在我国人口预测中的应用[J]. 统计与决策, 2011(15): 54-56.
- [7] 任强, 侯大道. 人口预测的随机方法: 基于 Leslie 矩阵和 ARMA 模型[J]. 人口研究, 2011(2): 30-44.
- [8] 陈毅华, 李永胜, 苏昌贵, 等. 径向基神经网络模型在人口老龄化预测中的应用——以湖南省为例[J]. 经济地理, 2012, 32(4): 34-39.
- [9] Akhtar, S. and Rozi, S. (2009) An Autoregressive Integrated Moving Average Model for Short-Term Prediction of

Hepatitis C Virus Seropositivity among Male Volunteer Blood Donors in Karachi, Pakistan. *World Journal of Gastroenterology*, **15**, 1607-1612. <http://dx.doi.org/10.3748/wjg.15.1607>

- [10] 吴家兵, 叶临湘, 尤尔科. ARIMA 模型在传染病发病率预测中的应用[J]. 数理医药学杂志, 2007, 20(1): 90-92.
- [11] 沈艳辉, 江初, 敦哲, 等. 北京市城区 1957~2004 年猩红热流行趋势及预测[J]. 现代预防医学, 2008, 35(7): 1224-1226.

期刊投稿者将享受如下服务:

1. 投稿前咨询服务 (QQ、微信、邮箱皆可)
2. 为您匹配最合适的期刊
3. 24 小时以内解答您的所有疑问
4. 友好的在线投稿界面
5. 专业的同行评审
6. 知网检索
7. 全网络覆盖式推广您的研究

投稿请点击: <http://www.hanspub.org/Submission.aspx>