

基于ARIMA对养老床位数量的预测

胡天惠, 程佳丽, 陈纪元, 董思妤

上海理工大学, 管理学院, 上海

收稿日期: 2021年10月16日; 录用日期: 2021年11月17日; 发布日期: 2021年11月24日

摘要

随着人口总数的不断增加和人口老龄化的加剧, 如何预测养老床位的需求以及应对随之而来的各种养老问题, 对于个人、企业、政府都至关重要。因此, 可依据各类信息、数据对养老床位的需求进行预测, 并给出合理的建议与运营策略。本文综合选取了全国总人口、老年人比例、人口死亡率、卫生机构床位数、全国GDP、离退人员参加养老保险人数这六个元素作为影响因素, 针对不同数据的特征, 使用Brown线性趋势模型、时间序列分析模型对这些指标进行了十年的数据预测, 再用ARIMA对养老床位需求规模进行预测, 使用神经网络对结果进行修正, 最终得到养老床位需求的分类预测结果。最后使用主成分分析法对各指标进行降维, 将主成分归类为三大指标, 提出了关于如何建设养老服务的建议。

关键词

时间序列, ARIMA, 神经网络, 主成分降维

Prediction of Time Series of Nursing Bed Number Based on ARIMA

Tianhui Hu, Jiali Cheng, Jiyuan Chen, Siyu Dong

School of Management, University of Shanghai for Science and Technology, Shanghai

Received: Oct. 16th, 2021; accepted: Nov. 17th, 2021; published: Nov. 24th, 2021

Abstract

With the continuous increase of the total population and the aggravation of the aging of the population, how to predict the demand for pension beds and deal with the ensuing various pension problems are of great importance to individuals, enterprises and the government. Therefore, the demand for old-age care beds can be predicted based on all kinds of information and data, and reasonable suggestions and operation strategies can be given. The paper selected the population proportion of the population, the elderly, mortality, and health institutions of beds, the national

GDP, the number of personnel to attend endowment insurance applicant retreating the six elements as influence factors, according to the characteristics of the different data, using Brown linear trend model, time series analysis model for data to predict these indexes for ten years. Then ARIMA was used to predict the demand scale of elderly care beds, and neural network was used to modify the results, finally the classification prediction results of elderly care bed demand were obtained. Finally, the principal component analysis method is used to reduce the dimension of each index, the principal component is classified into three indicators, and the suggestions on how to build old-age service are put forward.

Keywords

Time Series, ARIMA, Neural Network, Principal Component Analysis

Copyright © 2021 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

随着国家经济的发展, 全国 GDP 稳步增长; 随着科技进步, 医疗健康的发展, 全国总人口数在逐年递增, 人口死亡率逐渐下降, 老年人比例逐渐升高, 人口老龄化加剧, 社会的进步带来社会福利的增加, 离退人员参加养老保险人数增加[1]。在社会政策支持、人民生活水平提升、老龄化加剧等情况下, 全国公办与民办的养老机构床位数也在随之增加。由此, 如何预测养老床位的需求以及应对随之而来的各种养老问题, 对于个人、企业、政府都至关重要。

2. 养老床位预测模型的准备

2.1. 符号说明

文中出现的符号说明如表 1。

Table 1. Symbol description

表 1. 符号说明

符号	含义
Y	养老床位数(万张)
Total (T)	全国总人口(万人)
Proportion (P)	老年人比例(%)
Death (D)	人口死亡率(%)
Hygiene (H)	卫生机构床位数(万张)
GDP (G)	全国 GDP (亿元)
Insurance (I)	离退人员参加养老保险人数(万人)

2.2. 养老床位自变量数据准备

由于养老服务床位数量受到多因素的影响, 本文查阅参考文献后, 取全国总人口、老年人比例、人

口死亡率、卫生机构床位数、全国 GDP、离退人员参加养老保险人数这六个元素作为自变量(表 2)，数据时间为 2001~2019 年，跨度为 20 年，探究其对养老服务床位数量的影响。

Table 2. Descriptive data of independent variables

表 2. 自变量数据描述性

变量	平均值	标准误	中位数	标准差	最小值	最大值
Y	292.3947	28.14431	316.1	122.6782	114.6	493.7
Total	134065.7	886.948	134091	3866.117	127627	140005
Proportion	11.22759	0.434949	11.27384	1.895901	7.953474	14.0837
Death	6.913158	0.070325	7.09	0.306541	6.4	7.16
Hygiene	525.4547	44.86401	478.68	195.5577	313.6	880.7
GDP	457673.3	64950.43	410354.1	283112.3	109276.2	984179
Insurance	6926.774	661.7059	6305	2884.309	3380.6	12310.4

3. 养老床位自变量数据预测

由于现有各项指标的数据为 2001~2019 年间，所以通过分析，本文对六个自变量分别进行了十年的预测，得到 2020~2030 年的结果。我们绘制出了六个变量随时间变化的折线图来观察其趋势，如图 1 所示。

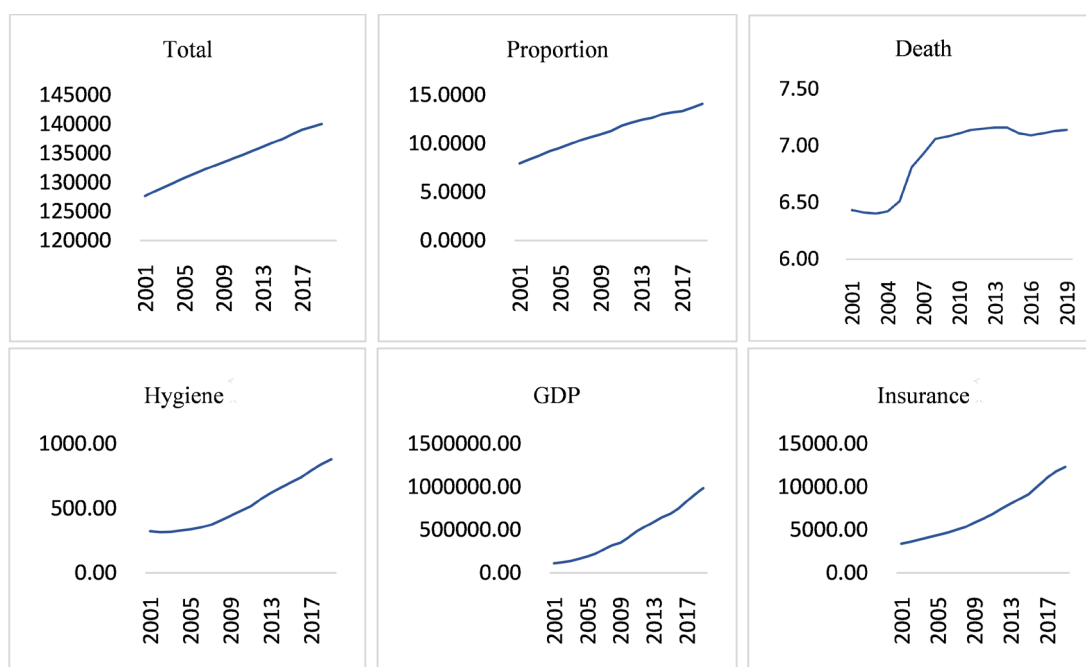


Figure 1. Line chart of the changes of various variables over time

图 1. 各变量随时间变化折线图

指数平滑法是基于移动平均的时间序列分析预测法，主要步骤是观察前期的数值，对其设定合理的权重，根据权重和前期数据推测计算出后期的数据。

从图 1 中可看出, T、H、G、I 四者关于时间没有太大波动, 较为平稳, 有线性趋势且不含季节成分, 所以我们对 T、H、G、I 这四个变量用 Brown 指数平滑法进行预测, 计算方式如下:

$$\begin{cases} l_t = \alpha x_t + (1-\alpha)(l_{t-1} + b_{t-1}) & (1) \\ b_t = \alpha (l_t - l_{t-1}) + (1-\alpha)b_{t-1} & (2) \\ \hat{x}_{t+h} = l_t + hb_t, h=1, 2, \dots & (3) \end{cases}$$

其中, 式(1)为水平平滑方程, 式(2)为趋势平滑方程, 式(3)为预测方程; t 为当前年份; h 为预测的后期年份; x_t 为第 t 年的全国总人口数量的实际观测值; b_t 为第 t 年的全国总人口数量的预测趋势; α 为水平平滑参数, 在 Brown 线性趋势模型中, 认为趋势平滑参数等于水平平滑参数。对于该四个变量, 这里展示变量 T 的年份预测结果, 如图 2 所示。

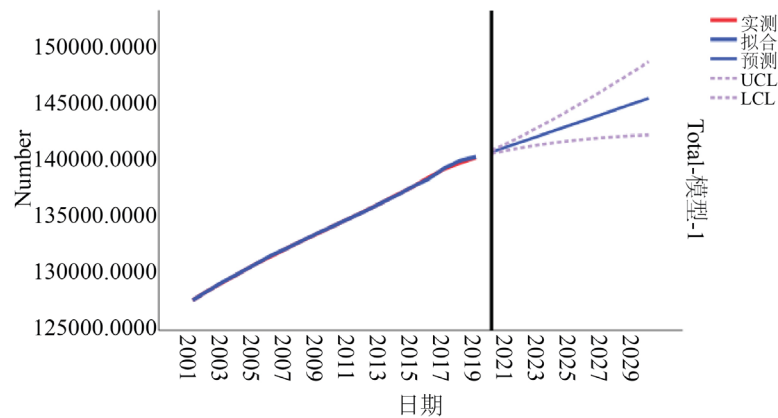


Figure 2. Total exponential smoothing forecast
图 2. Total 指数平滑法预测

从 Proportion 残差的 ACF 和 PACF 图形, 即图 3 中可以看出, 选择 ARIMA (1,1,0)较为合适。我们对 Proportion 进行 ARIMA (0,1,0)预测, 得到结果如图 4 所示。

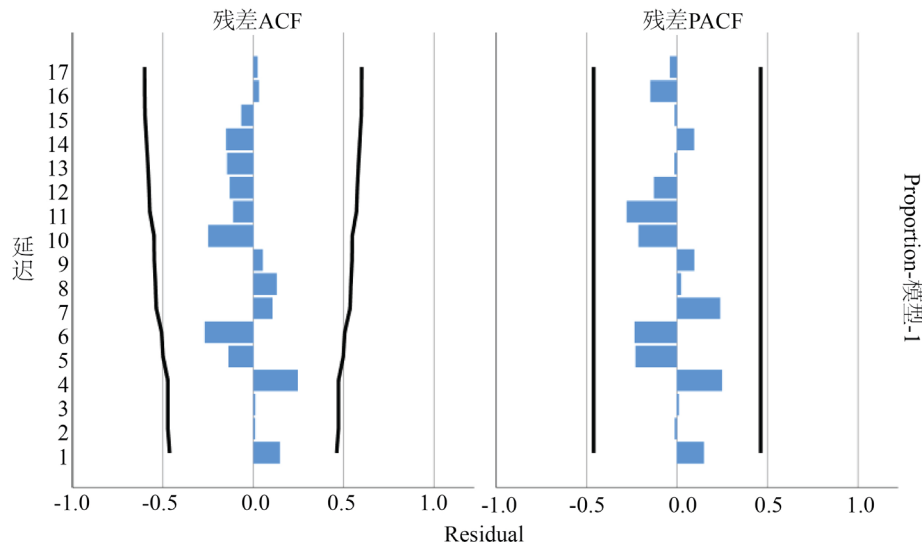


Figure 3. Proportion residual ACF, PACF plot
图 3. Proportion 残差 ACF、PACF 图

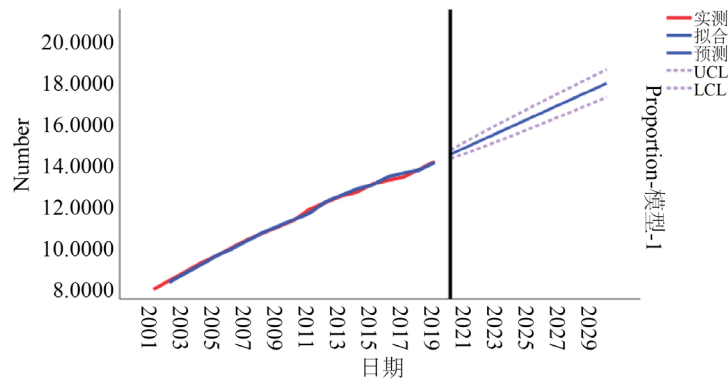


Figure 4. Proportion ARIMA forecast
图 4. Proportion ARIMA 预测

从 Death 残差的 ACF 和 PACF 图形,即图 5 中可以看出,选择 ARIMA (1,1,0)较为合适。我们对 Death 进行 ARIMA (1,1,0)预测,得到结果如图 6 所示。

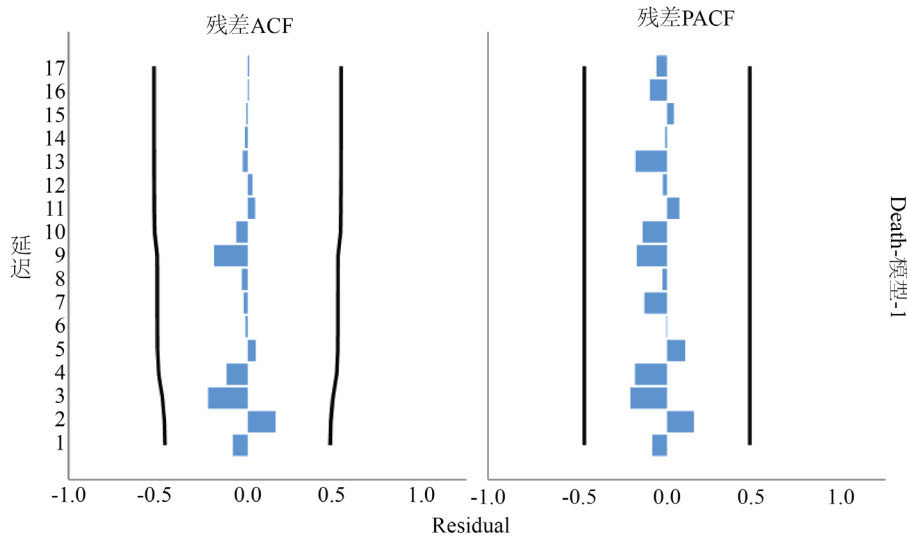


Figure 5. Death residual ACF, PACF diagram
图 5. Death 残差 ACF、PACF 图

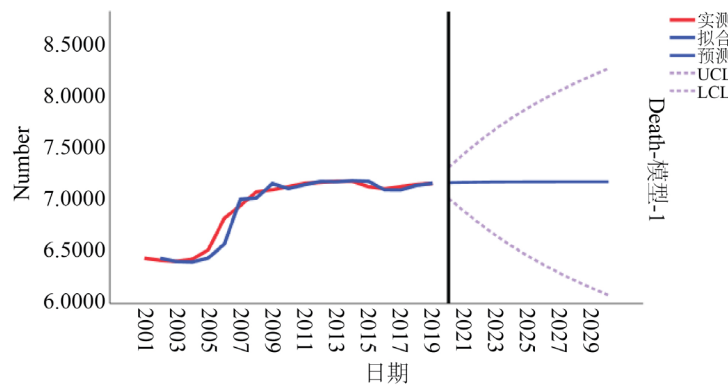


Figure 6. Death ARIMA forecast
图 6. Death ARIMA 预测

最终将得到的各项自变量指标预测结果汇总，如表 3 所示。

Table 3. Forecast results of various factors from 2020 to 2025

表 3. 2020~2025 年各因素预测结果

年份	Total	Proportion	Death	Hygiene	GDP	Insurance
2020	140472	14.42433	7.146	921.1201	1056987	12823.1
2021	140939	14.7649	7.1496	961.5395	1129630	13335.81
2022	141406	15.10547	7.1518	1001.959	1202273	13848.51
2023	141873	15.44604	7.1531	1042.378	1274916	14361.22
2024	142340	15.78661	7.1539	1082.798	1347559	14873.92
2025	142807	16.12718	7.1544	1123.217	1420202	15386.63

4. 历年养老床位数的预处理

对于养老床位数 Y，从图 7 中可看出 2001~2013 年间呈现稳定迅速增长趋势，2013 年到 2015 年迅速回落，2016 年起又平稳增长。三年回落说明床位数并不完全等同于需求量，2013 年前期由于养老院的迅速扩建导致了供过于求，导致养老院入住率低，企业处于亏损状态，大量养老床位被收回，直到 2016 年该现象才有所缓解。

考虑到多数自变量的稳定增长，老年人人数的逐年增加，养老床位需求也应呈现增长趋势，除此之外，考虑到 2013 年的回落来自于前三年的大量扩张，因此本文将 2010~2015 年视为缺失值，根据剩余数据，用线性拟合对缺失年份进行补齐，得到新的数据，调整后的 Y 及其折线图如图所示。

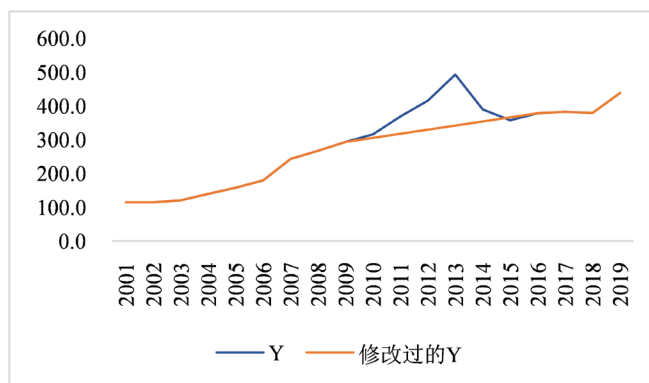


Figure 7. Line chart of Y and adjusted Y over time
图 7. Y 与调整后的 Y 随时间变化折线图

5. 模型的建立与求解

5.1. 神经网络时间序列预测床位

BP 神经网络是一种按照误差逆向传播算法训练的多层前馈神经网络[2]，主要由输入层隐藏层和输出层组成，在前向传递中，输入信号从输入层经过隐藏层处理从输出层得出结果，如果输出层得不到期望输出则进行反向传播通过最速下降算法根据预测误差调整网络权值与阈值从而使 BP 神经网络的预测输出不断逼近期望输出。

具体步骤: (1) 初始化: 为所有连接权值 $M_{ij}^k(t)$ 和阈值 $R_i^k(t)$ ($R_i^k(t)$; $i=1, \dots, P_k$; $t=0$) 赋予随机任意小值。(2) 从 N 组样本中取其中一组样本 $x = [x_1, x_2, \dots, x_{p_1}]^T$, $d = [d_1, d_2, \dots, d_{p_m}]^T$, 把输入信息 x 输入 BP 神经网络。(3) 正向传播: 计算各层节点的输出值 y_j^k 。(4) 计算网络的实际输出与期望输出的误差 $e_i = y_i - y_i^m$ 。(5) 反向传播: 从输出层方向计算第一个隐层, 依照连接权值修正公式分别向减小误差方向调整网络的各个连接权值 $\Delta M_{ij}^{k-1}, d_i^m, d_i^k$ 。(6) 使 $t+1 \rightarrow t$, 取出另一组样本并重复步骤(2)~(5), 直至 N 组输入输出样本的误差达到要求时为止。

公式: $y_i^k = f_k(u_i^k)$, $u_i^k = \sum_j M_{ij}^{k-1} y_j^{k-1}$, $k=2, 3, \dots, m$, $f_k(k=2, 3, \dots, m)$: 隐藏层 k 与输出层中各神经元的非线性输入输出关系, M_{ij}^k : 第 $k-1$ 层的第 j 个神经元与第 k 层的第 i 个神经元的连接权值, u_i^k :

假设为第 k 层的第 i 个神经元输入的总和, y_i^k : 假设为第 k 层的第 i 个神经元输出的输出。

在对数据的综合分析下, 本文基于神经网络时间序列模型对养老床位进行预测。设定 70% 的训练集, 15% 的验证集, 15% 的测试集, 设置隐层神经元的个数为 10, 使用 Levenberg-Marquardt 算法对数据集进行训练, 训练结果如图 8 所示, 从图中可以看出训练结果精确度、拟合度较高。

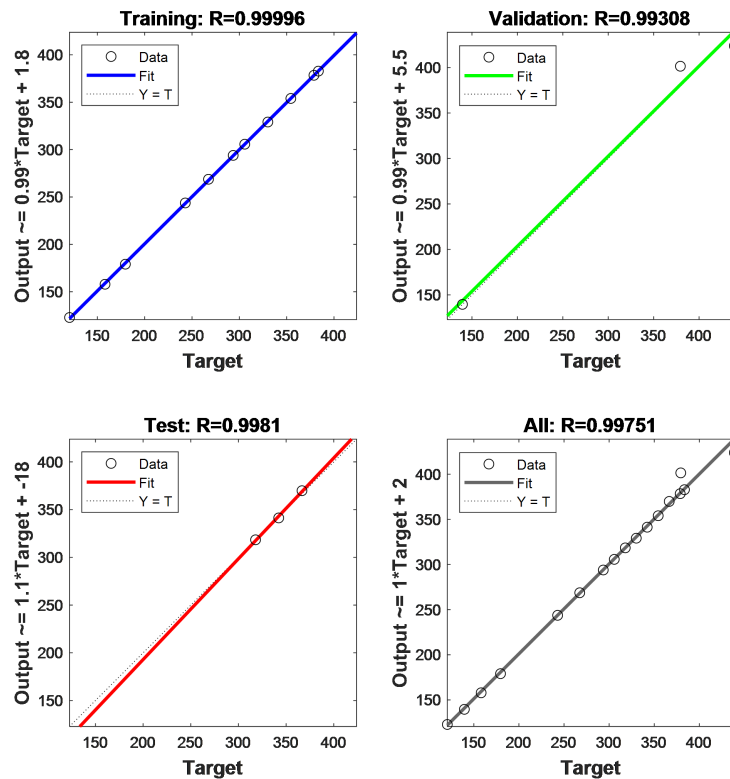


Figure 8. Fitting results after neural network training
图 8. 神经网络训练后拟合结果

在训练结束后, 将表3预测数据作为输入矩阵, 进行神经网络时间序列预测, 得到部分养老床位结果如表4所示。

5.2. 时间序列回归预测床位

ARIMA(0, 1, 0)对 2001~2019 年处理后的年床位数进行回归预测, 得到部分养老床位结果如表 5 所示。

Table 4. The prediction results of neural network time series

表4. 神经网络时间序列预测结果

年份	2020	2021	2022	2023	2024	2025
预测值	367.6045	363.8031	364.4091	369.2679	377.3185	387.5117

Table 5. Time series regression forecast results

表5. 时间序列回归预测结果

年份	2020	2021	2022	2023	2024	2025
预测值	456.8	474.8	492.8	510.8	528.9	546.9

5.3. 养老床位需求量

由于自由市场，我们认为供给量基本等于需求量，加之前文以对年度养老床位数进行了调整、修正，所以我们以修正后的养老床位数量作为养老床位需求量，再对 2020~2030 年的养老床位需求量进行预测。

本文综合神经网络时间序列模型和时间序列回归模型对床位数进行了分析和预测，将两者的加权平均值作为养老床位需求量的结果，如表 6 所示。

Table 6. Forecast results of demand for elderly care beds

表6. 养老床位需求量预测结果

年份	2020	2021	2022	2023	2024	2025
预测值	412.2023	419.3016	428.6046	440.034	453.1093	467.2059

6. 建议

基于上述对于养老服务床位数量的预测，在未来的 10 年内，其需求是稳步增长的。企业在养老服务业扮演的角色是服务的供给方，养老服务主要包括日常生活照料，提供娱乐活动服务、医疗保健等服务。我国社会经济稳步发展，老年人在开放包容的社会环境里更加容易接受多样化、个性化的养老服务，现代养老也更加注重精神上的服务需求。

主成分分析作为一种降维算法，可以将多个变量转换为几个主要的成分，使得数据能被更好的使用。我们先对原有数据进行标准化的处理，去除量纲的影响，再对相关系数矩阵求其特征值和特征向量。进一步计算贡献率，得到三个主成分基本能解释接近 90% 的原始数据信息的结果，可以接受。

$$\begin{cases} F_1 = 1.8597T - 2.4709P - 2.4710D - 2.4572H + 7.8303G - 2.2910I & (4) \\ F_2 = -2.3116T + 0.3487P + 0.3487D + 0.3444H + 0.9660G + 0.3038I & (5) \\ F_3 = -0.0001T - 0.0013P - 0.0013D - 0.0012H - 0.0001G + 0.0038I & (6) \end{cases}$$

结合主成分式子中系数情况和 6 个自变量的意义，对于式(4)，将第一个主成分归纳为全国人口及老年人占比指标，对于式(5)，将第二个主成分归纳为国民消费水平指标，对于式(6)将第三个主成分归纳政府的补贴程度指标。因此，选取的主要因子为老年人人口数量、收入及消费和社会福利，从中挖掘到 3 个角度及对应建议。

政府在我国养老服务体系中扮演重要角色，为促进我国养老服务体系持续、健康、科学的发展，根据以上本文养老服务床位的分析和预测，向政府管理机构提出以下建议：

6.1. 完善监管制度

养老服务床位在未来的需求是持续且稳步增长的，而我国目前的养老服务床位的管理中存在“一床难求”与“床位空置”并存的现象，政府管理机构应努力健全相关法律法规[3]，设立第三方监管机构以及完善公众参与管理机制，统筹养老服务信息与养老服务资源，针对区域分布不平衡以及供需矛盾[4]，合理配置有效的资源，在多方参与的管理下促进我国养老服务的高水平高质量发展。

6.2. 支持层次化多样化的养老服务

由于老年人数的增加，在为企业找寻“商机”中我们发现多样化的养老服务需求日增，为满足老年人的层次化需求政府应支持企业在服务更新上的工作，激发养老行业的潜在消费，为经济市场增添活力。

6.3. 平衡养老服务市场化和社会化

应致力于通过政府将老年人、社区、企业机构、医院等体系中的重要角色有效调配交互，这是养老服务市场化和社会化的有机融合。养老服务的市场化和社会化相辅相成，如果二者出现不协调不稳定对各方会引起不可估计的损害，而政府可以在其中担任调控、指导等工作，我们的养老服务才能更好的为老年人提供保障。

7. 结语

目前我国养老服务仍然面临供给结构不尽合理、市场潜力未充分释放、服务质量有待提高等问题。基于我国养老服务供给的福利性、社会性、市场化的特点，为解决我国目前养老服务的困境，我们认为，要以居家为基础，社区为依托，机构为补充的养老服务体系，计划更加深入地借助互联网技术努力打造更加和谐稳定的“医养康护”养老服务。把“互联网+”信息功能应用到养老老年服务中，开发老年人与社区组织、养老机构的互动交互平台。首要的目的是解决老年人突发疾病状况能够在最短时间内得到救治；其次是确保老年人的养老服务需求信息与服务供给方对等；最后是可以更好地保障服务质量和有效服务。通过网络数据可以将养老服务信息反馈给相关政府管理机构，从而可以在资源配置和管理方面做出及时调整，进而使社会养老服务体系更加高效高质量的运转，用科学的方法提升老年人的生活质量。应用互联网技术使养老服务更高效，得到政府的支持，促进社会的就业，实现合理配置有效资源，数据库可为政府、企业等提供信息指导，同时与医院合作共赢。

养老服务业的前景市场是可以预见的，随着我国老龄人口的不断增加，政府以及社会也更加关注养老服务业的发展。年老是我们每个人的必经阶段，我们年轻时畅想未来，很多人往往忽略了这一重要的阶段。老年不只有步履蹒跚，不只有满身疾病，老年生活依旧可以是多姿多彩的。

参考文献

- [1] 曹稀哲,王海鑫,孙馨愉. 养老服务床位规模预测及其“商机”研究[J]. 营销界, 2020(8): 87-88.
- [2] 王慧健,刘峥,李云,李涛. 基于神经网络语言模型的时间序列趋势预测方法[J]. 计算机工程, 2019, 45(7): 8.
- [3] 张思锋. 中国养老服务体系建设中的政府行为与市场机制[J]. 社会保障评论, 2021, 5(1): 129-145.
- [4] 黄石松. 更好发挥政府作用,推动养老产业高质量发展[N]. 21世纪经济报道, 2021-03-17(004).