

基于结构方程与一阶滞后分位数回归的中国人口老龄化问题分析

摘要

随着中国经济的发展，低生育率与人口老龄化问题所带来的经济问题与社会问题开始逐渐凸显。而在近年来国家出台“三胎”政策后，这一政策对人口结构与市场经济会带来何种程度的调节还有待研究。基于此，本文基于一系列人口社会学与计量经济学方法对问题做出了初步的探讨。

对于问题一，需要预测未来人口结构。人口结构研究可以主要分为三个研究部分，总体人口数量变化，不同年龄层人口结构变化和城乡人口结构变化，而总体人口数量和不同年龄层结构，建立 **Leslie 模型** 对人口情况进行合理预测，效果良好；而对于城乡人口结构，发现城市人口比例随时间在逐步递增，故建立 **OLS 模型**，模型 R^2 分数达到 0.999，说明效果良好，能够进行合理预测。

对于问题二，研究“双减政策”对出生率的影响，本质上是研究教育压力与出生率的影响。通过查找相关实证**面板数据**，将多个变量抽象为升学压力、教育质量、教育投入三类变量，经过**因子分析**抽象出五个因子构建**结构方程**模型。通过对结构图的分析与“双减政策”带来的社会影响得到一系列预判。

对于问题三，探究医疗政策对人口老龄化的影响，使用**一阶滞后分位数回归**对面板数据进行分析能够有效缓解内生性问题。通过对多个变量的回归，发现老年抚养比、卫生人员数、医疗机构数、人均医疗支出、政府医疗投入等多个变量对问题均有显著影响，需要在加强医疗福利保障的同时培养出更多高素质高水平的专业卫生技术人员。

对于问题四，沿用问题三的方案，探究房价、医疗、税收、劳动力、就业率等变量对出生率的影响，认为需要在调整商品房房价与税收结构的同时促进市场，提高就业率，创造更多就业机会的同时提高工资水平从而减轻青年人生活压力。

整体来看，模型方法综合人口社会学和计量经济学手段，简洁直观的同时排除一系列干扰，能够合理对问题进行分析与解释且量化结果与现实社会热点问题相映证，所提出的建议对当下热点问题有一定参考价值。

关键词：Leslie 模型，因子分析，面板数据，结构方程，一阶滞后分位数回归

一、问题重述

1.1 问题背景

中国三孩政策下的低生育形势以及人口政策未来走向,不仅是当前社会各界普遍关注的热点议题,也是关系到中国经济、社会和家庭可持续发展的重大问题。允许一对夫妇生育三个孩子的政策是对我国生育率持续走低、育龄妇女数量减少、育龄人群生育意愿不高、少子化和老龄化问题加剧的积极回应。要落实三孩政策并彰显其实际效应,应针对不同人群配以不同的支持政策,以解决生育意愿、医疗保障等问题,来化解我国老龄化进程。对于生育后的配套支持措施如何实施是当前育龄人群最为关心的问题之一,在此背景下,分析我国的人口学特征,从而预测人口走向十分重要。

1.2 问题重述

本题需要我们结合统计数据,建立数学模型,从而构建人口学的有关模型,并回答以下问题:

- (1) 建立人口学数学模型,预测未来 10 年我国在施行三孩政策下的人口状况。
- (2) 通过考虑相关因素,建立有关“双减政策”的数学模型以预测对新生人口的影响。本质上需要探究教育对人口出生率的影响。
- (3) 针对我国人口老龄化问题分析如何施行医疗政策。本质上是探究医疗福利政策与人口老龄化之间的关系。
- (4) 从多个因素入手,探讨生育意愿偏低的问题,从而为促进生育意愿提供方案。

二、问题分析

2.1 问题一分析

问题一需要结合相关数据对未来十年内中国人口有关情况进行预测,而中国人口有关情况是比较泛的一个概念。这里为了具体化,本文从人口总数预测、年龄结构预测、城乡人口比例三点进行分析。而传统的 Malthus 模型和 Logisitc 模型等考虑因素较少,仅考虑种群增长率和环境阻力,因此本文使用更加深入的 Leslie 模型探究不同年龄结构之间状态的转移。另外对于城乡人口比例,构建 OLS 进行回归分析即可。

2.2 问题二分析

问题二需要分析双减政策对出生率的影响。双减政策本质上影响到的是国家教育,这里需要选择教育有关的变量,若变量过多则需要通过因子分析等手段进行降维与因子提取,然后对变量与出生率之间构建结构方程模型可以清洗揭示教育与人口之间影响的内部机理。由于双减政策 2021 年出台没有足够的面板,这里并不适合作为断点进行回归,进行分析这一政策会对因子有怎样的影响即可。

2.3 问题三分析

问题三需要分析医疗对人口老龄化的影响并提出相应的政策建议，对人口老龄化的描述可以用老年人口在群体中的比例描述。由于统计局的数据为面板数据，为了削弱内生性因素，选择滞后 1 阶的分位数回归对变量影响效果进行分析，最终得出一系列有参考价值的意见。

2.4 问题四分析

而对于问题四，教育与医疗问题问题二三中已经讨论过，可以采用问题三中同样的方法，将平均商品房房价、税收、劳动力、就业率等因素对出生率进行 1 阶滞后的分位数回归，可得到一系列有参考价值的意见。

三、模型假设

- 1、不考虑移民对人口总数的影响。这一假设是为了保证中国内部自身系统稳定性。
- 2、社会稳定，未发生重大自然灾害和战争，出生率和死亡率不会发生跳跃性变化。
- 3、假设男女出生比率较稳定。
- 4、假设所得数据真实有效。

四、符号说明

符号	说明
$f(t)$	女性人口
b_i	第 i 年龄组女性生育女孩的生育率
α_i	修正的生育率
s_i	第 i 年龄组女性生育孩子的存活率
$N(t)$	t 时刻的全国总人口
c_t	城市人口比例
t	时间，以年为单位
o_t	全国老龄化占比
X_t	与医疗有关的控制变量
A	因子载荷
F	分解出的因子
u	变量的均值
ε	特殊的波动因子

五、模型建立与求解

5.1 问题一模型建立与求解

5.1.1 Leslie 模型

在正常社会条件或自然条件下，生育率与死亡率是与群体的年龄构成息息相关的。我们需要对整个群体按年龄进行层次划分，构建与年龄相联系的人口模型。典型的例子就是 Leslie 矩阵模型。Leslie 模型是一类基于不同年龄段生育率不同构建变化矩阵进行分析的模型，能够充分考虑到种群内性别比和年龄段之间的差异。

建立 Leslie 模型，假定各年龄段性别比恒定且相同，通过分析 2020 年分年龄段妇女人数及生育率，依据 2020 年三胎妇女人数预测三胎政策下生育率的变化，以此预测开放三孩之后我国未来十年的人口状况。由于要考虑不同年龄组女性修正后的生育率，我们决定建立 Leslie 模型来预测未来十年的人口数量。通过计算女性人口分布 $f(t)$ 随 t 的变化规律，来研究总人口的变化规律。

将女性人就按年龄大小等间隔划分成 n 个年龄组，并离散化时间，与年龄组间隔相同。设各个年龄组的生育率、存活率不随时间变化，则：

$$\begin{cases} f_i(t+1) = \sum_{i=1}^n b_i f_i(t) \\ f_{i+1}(t+1) = s_i f_i(t) \end{cases} \quad (1)$$

其中 $i = 1, 2, \dots, n-1$ 。

在上式中，假设 b_i 中已扣除了在 t 时段以后出生而活不到 $t+1$ 的婴儿，记 Leslie 矩阵：

$$L = \begin{bmatrix} b_1 & b_2 & \cdots & b_n \\ s_1 & 0 & \cdots & 0 \\ 0 & s_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & s_n \end{bmatrix} \quad (2)$$

记 $f(t) = [f_1(t), f_2(t), \dots, f_n(t)]^T$ ，则上式可以写作：

$$f(t+1) = Lf(t) \quad (3)$$

当 L 、 $f(0)$ 已知时，对任意 $t = 1, 2, \dots$ 有

$$f(t) = L^t f(0) \quad (4)$$

只要求出矩阵 L 并根据人口分布初始变量 $f(0)$ ，就可以求出 t 时刻的女性人口分布向量 $f(t)$ 。再将所得到的女性人口数据预测除以女性占比就可以得到 t 时刻的全国总人口 $N(t)$ 。

5.1.2 生育率的修正

考虑新出台的三胎政策，对不同年龄的女性生育率进行修正，得到表 1：

表 1. 修正后的生育率

年龄（单位：岁）	修正后生育率（单位：‰）	年龄（单位：岁）	修正后生育率（单位：‰）
15	1.06	33	70.45
16	3.48	34	52.78
17	9.53	35	45.91
18	18.62	36	40.48
19	30.03	37	29.02
20	40.42	38	23.04
21	59.1	39	18.91
22	75.54	40	13
23	85.93	41	9.74
24	119.26	42	5.21
25	123.69	43	2.94
26	118.48	44	0.99
27	127.68	45	0.88
28	117	46	0.71
29	108.99	47	0.56
30	97.7	48	0.25
31	83.05	49	0.28
32	80.03		

设女性占比 p 基本不变，且取 2021 年的数据 $p = 0.4876$ ，将所得的修正的生育率 α_i 代入方程

$$b_i = p\alpha_i \quad (5)$$

得到第 i 年龄组女性生育女孩的生育率 b_i 。

5.1.3 模型求解

结合数据对其进行求解得到 2021 年到 2035 年我国总人口的变化趋势如图 1：

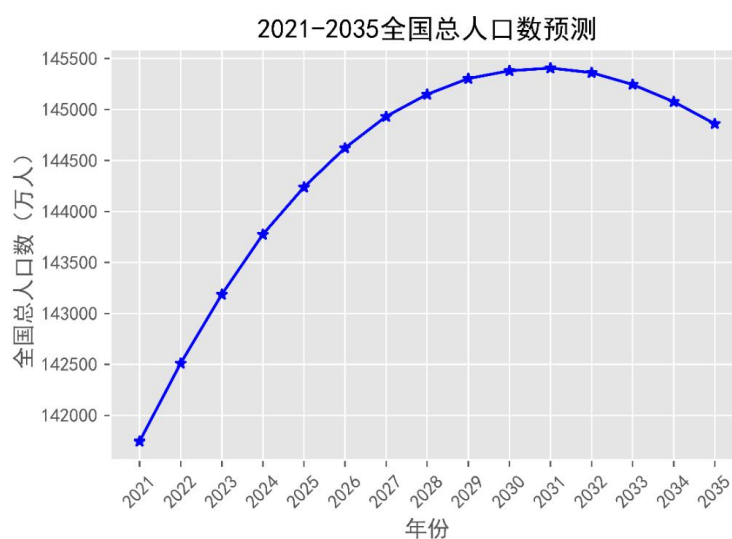


图 1. 全国总人口随时间的变化趋势预测

并得到未来十五年预测数据如表 2:

表 2. 全国人口总数随时间变化预测

年份	2021	2022	2023	2024	2025
人口数量(单位: 万人)	141745	142510	143185	143774	144238
年份	2026	2027	2028	2029	2030
人口数量(单位: 万人)	144622	144931	145148	145303	145379
年份	2031	2032	2033	2034	2035
人口数量(单位: 万人)	145406	145361	145245	145076	144860

由预测数据可知三胎政策发布后我国人口增长率较之前明显升高,但是由于妇女在“黄金”育龄段的人数占比不断下降,人口增长率将保持减速,未来十五年里人口数量在 2031 年会达到峰值 14.54 亿人左右,之后人口数量会逐渐下降。

5.1.4 人口结构情况

在原始 Leslie 模型基础上,将 $f_i(t)$ 按年龄组成分为 3 组: 0-14 岁的总人数, 15-64 岁的总人数, 65 岁及以上的总人数,再将三组数据分别除以女性占比 p ,得到三个年龄段内 2021-2035 年的各组年龄分布情况如图 2 和表 3 所示:

表 3. 不同年龄段人群占比变化

年份	0-14 岁总人数 (单位: 万人)	15-64 岁总人数 (单位: 万人)	65 岁及以上总人数 (单位: 万人)
2021	22547	99261	19938
2022	22644	98921	20945
2023	22689	98477	22019
2024	22657	98239	22878
2025	22573	98369	23296
2026	22614	98183	23825
2027	22410	98392	24128
2028	22168	97621	25359
2029	21892	96413	26998
2030	21531	95575	28273
2031	21175	94685	29546
2032	20659	93913	30788
2033	20131	93340	31773
2034	19797	92045	33234
2035	19517	90947	34396

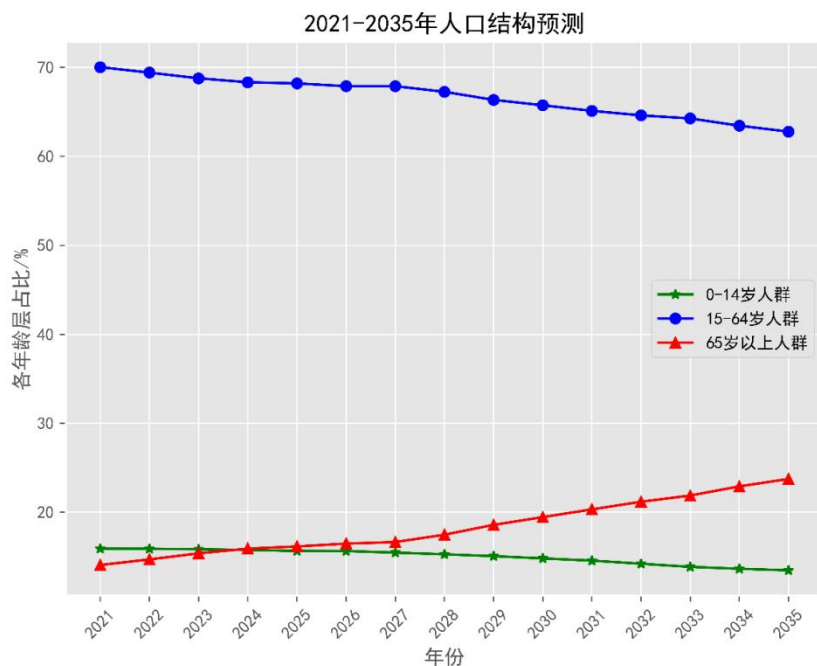


图 2. 不同年龄段人群占比变化

可以看到，在三胎政策执行后未来十五年内 15-64 岁人口数量及所占比重仍在减小，老龄化问题仍然难以得到有效改善。到 3035 年 65 周岁以上的人口数量占总人口的 23.7%。这是由于本身人口老龄化趋势已经比较严重，加上生活压力大，即使开放三胎对丁克人群和一胎群体影响非常小，只对二胎人群有一定影响。

另外，我们考虑城市人口与农村人口在全国人口中的占比。我们将 2002 到 2021 年城市人口和农村人口占比随时间的变化绘制如图 3 所示：

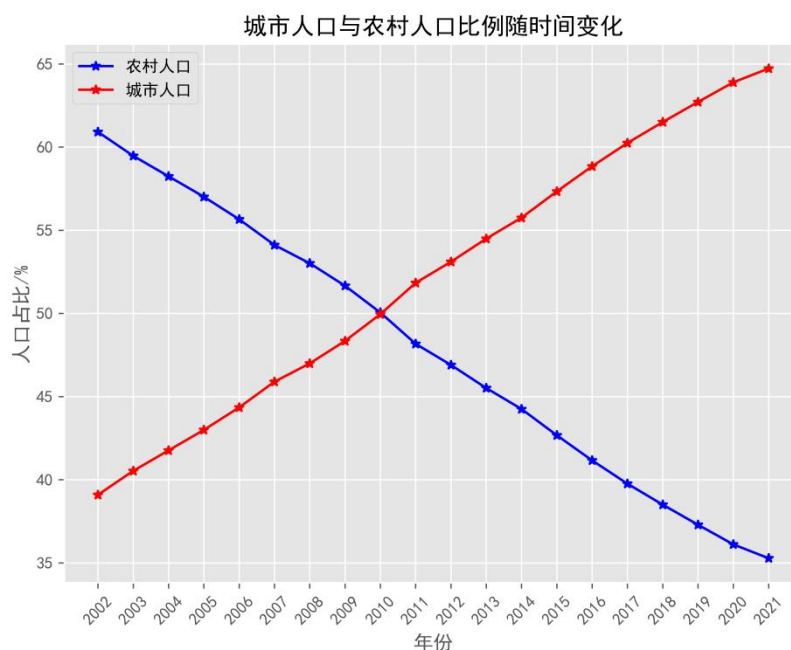


图 3. 城市人口和农村人口比例随时间变化

可以看到，城市人口的比例在逐年稳步增长，这说明中国的城市化进程正在稳步进行，也有越来越多的农村人口向城市迁移。考虑城市人口比例对时间做 OLS 线性回归，可以得到方程：

$$c_t = 0.0139t - 27.5289 \quad (6)$$

该模型在 1%显著性水平显著且 $R^2=0.999$ ，说明该模型拟合效果极佳，可以用此模型来模拟 2021-2035 年城市人口和乡村人口占比。方程的斜率为正说明城市化进程仍然将保持稳定速度快步进行。根据方程，到 2030 年，城市人口比例将达到 68.8%。

5.2 问题二模型建立与求解

因子研究法是对影响对象的各个因素进行研究，以确定它们之间的关系，它是用来探究怎样将许多真实的变量转化成几个典型的变量，并且使代价最小，和如何使变量因子更有说服力的多元统计探究方法。因子研究方法能够根据贡献率或典型特征值，得到几乎包含所有信息普遍适用的因子，能够防止因子中信息的重叠性，并且精简和细化数据集[6]。

因子分析的主要目的是为了将原始变量 X 抽象为因素 F ，即：

$$\begin{cases} x_1 = u_1 + a_{11}f_1 + a_{12}f_2 + \dots + a_{1m}f_m + \varepsilon_1 \\ x_2 = u_2 + a_{21}f_1 + a_{22}f_2 + \dots + a_{2m}f_m + \varepsilon_2 \\ \dots \\ x_n = u_n + a_{n1}f_1 + a_{n2}f_2 + \dots + a_{nm}f_m + \varepsilon_n \end{cases} \quad (7)$$

将其写作矩阵形式，即：

$$X = u + AF + \varepsilon \quad (8)$$

其中，向量 u 为各个变量的均值因子； ε 为特殊的波动因子， A 为载荷矩阵，而 F 则是抽象出的目标因子。

运用因子分析得到的主要因素，对影响新出生人口的各个因素进行归纳，运用少数的几个主要矛盾因素来归纳和分析大部分的影响因素。然后衡量各个因素的分数，从而可以明确影响人口出生率的主要情况，对实现人口出生率增长的相关举措资源的配置和分配具有极大的指导意义。故而，我们首先通过因子分析，不考虑双减政策，筛选出升学压力、教育投入、教育质量三类变量来探究教育对出生率的影响，对每一类变量我们筛选了多个二级变量：

对于升学压力，我们通过升学率来描述，而升学率则为目标学段的招生人数与生源学段的毕业人数比值。这里我们选取小升初升学率、初升高升学率、职校升学率、本科升学率、专科升学率五个层面进行描述。对于教育投入，选择变量包括家庭平均教育开销、国家实际教育投入、国家财政预算内教育投入、政府教育投入、社会教育投入、总学杂费等变量进行分析。对于教育质量，我们在国家统计局找到了不同学段的生师比，我们认为，这一类变量能够有效描述教育质量水平。

每一类变量内包括不同因子，经过最终分解，将因子命名如表 4：

表 4. 因子命名与变量抽象

变量	包含因子
升学压力	小升初升学率
	其他升学率（如初升高、高升大学）
教育投入	国家财政性预算经费（万元）
教育质量	基础教育质量
	职业教育质量
	高等教育质量

（1）升学压力。根据毕业人数和招生人数的比例可以看出不同阶段的升学率，将升学压力分解成小升初升学率和其他升学率，其中由于九年义务教育的普及，小升初升学率接近 100%，可认为该变量影响很小，主要影响在其他阶段升学率上。

（2）教育投入。该变量包含国家对教育的财政投入。

（3）教育质量。该变量主要是根据不同阶段教育生师比来反映教学质量，并按照不同的权重分为三个因子，分别是基础教育质量（主要针对小学，初中，高中）；职业教育质量（中职、高职）；高等教育质量（大学）。这一因子分析的载荷热力图如图 4：

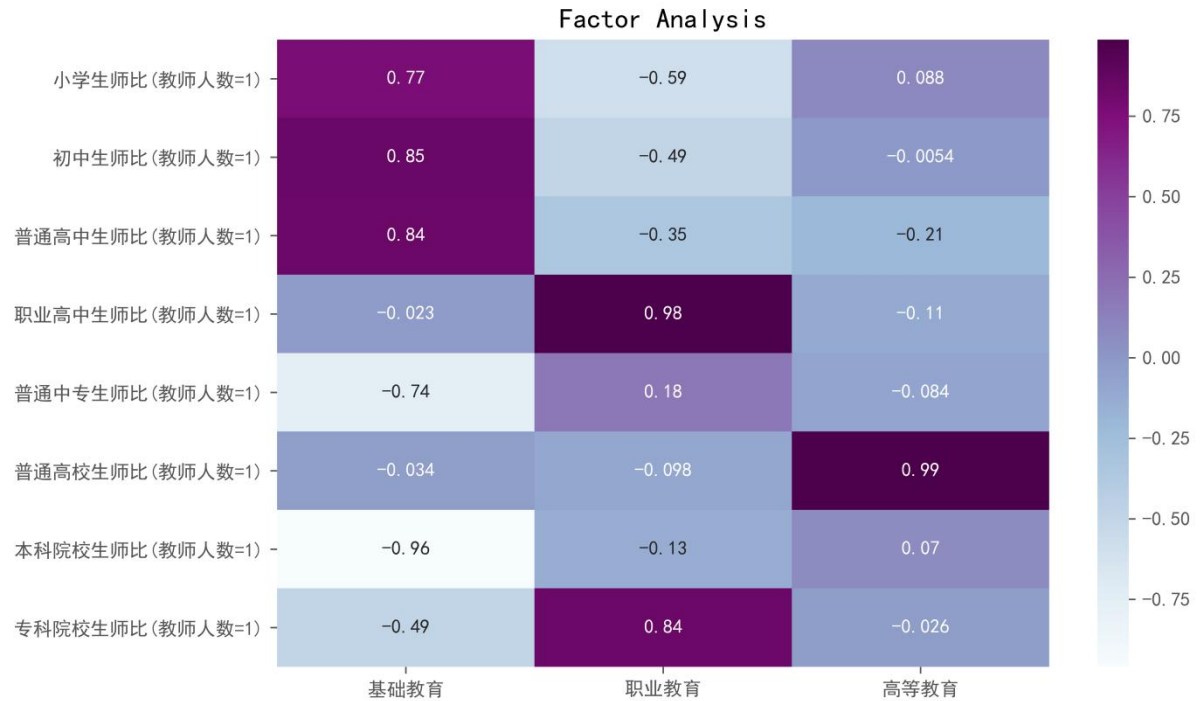


图 4. 教育质量因子分析热力图

其中教育对出生率属于间接影响，一般可认为，教育程度与居民人均可支配收入成正相关关系，而居民人均可支配收入对出生率具有直接影响，因此需要引入可支配

收入这一个变量做调节。

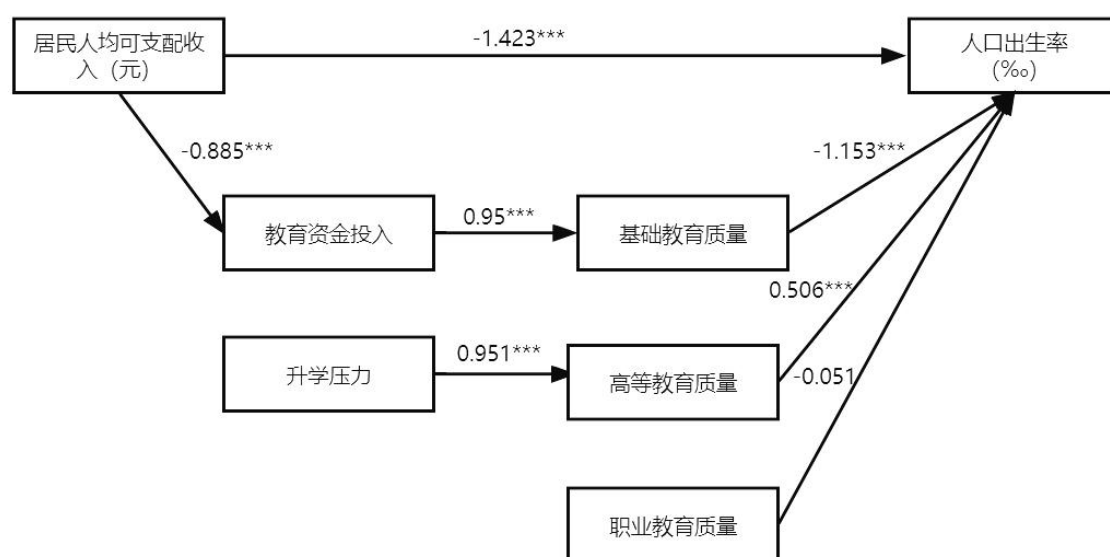
结构方程模型 (Structural Equation Model, SEM) 作为一种多元统计技术, 产生后迅速得到普遍应用。20 世纪 70 年代初一些学者将因子分析、路径分析等统计方法整合, 提出结构方程模型的初步概念。随后, 有研究者进一步发展了矩阵模型的分析技术来处理共变结构的分析问题, 提出测量模型与结构模型的概念, 促成 SEM 的发展。Ullman 定义结构方程模型为“一种验证一个或多个自变量与一个或多个因变量之间一组相互关系的多元分析程式, 其中自变量和因变量既可是连续的, 也可是离散的”, 突出其验证多个自变量与多个因变量之间关系的特点, 该定义具体一定的代表性。

SEM 假定一组隐变量之间存在因果关系, 隐变量可以分别用一组显变量表示, 是某几个显变量中的线性组合。通过验证显变量之间的协方差, 可以估计出线性回归模型的系数, 从而在统计上检验所假设的模型对所研究的过程是否合适, 如果证实所假设的模型合适, 就可以说假设隐变量之间的关系是合理的。SEM 与一些新近的分析方法相比, 也有其独特优势。SEM 将不可直接观察的概念, 通过隐变量的形式, 利用显变量的模型化分析来加以估计, 不仅可以估计测量过程中的误差, 还评估测量的信度与效度。探讨变量关系的同时, 把测量过程产生的误差包含于分析过程之中, 把测量信度的概念整合到路径分析等统计推断决策过程[1]。故而在本题中我们采取 SEM 的方式来进行模型建立与分析, 从而找到各变量之间的回归路径, 从而得出对于出生人口的影响因素及其影响方式。

研究证明, 教育对人口出生率的影响表现在两个方面: 一方面是教育可以影响社会风气进而对人口出生率产生影响, 一方面是人口自身的受教育程度影响人口的出生率。另外, 人口的死亡率与受教育程度呈明显的负相关。由于教育可以影响人口的出生率和死亡率, 出生率的降低, 使儿童少年人口在整个人口中的比重降低; 通过教育, 可以降低死亡率, 延长人口平均寿命。这二者都可以影响人口的年龄构成, 增大人口中老年人口的比重, 加速人口老龄化的进程, 同时也会进一步影响职业构成。

在以往关于教育对经济贡献的研究中, 主要是着眼于教育作为一种人力投资、作为改善劳动力素质的活动对经济增长产生的长远促进效果。而近年关于“教育产业”的经济贡献的讨论, 同以往教育经济学研究的视角有所不同, 主要强调教育作为一种个人或社会“消费”对经济增长产生的短期或直接的拉动效果。一国或一地区教育投入的多少在一定程度上反映了该国或该地区对教育的重视程度。。以舒尔茨为代表的人力资本理论认为, 人力资本是国家的一种重要的生产要素, 教育投资提高了国民的人力资本, 提高了劳动生产率, 促进了国民收入和经济增长[2]。另外, 高等教育以科学技术为纽带, 与社会生产力的发展有着直接的和必然的联系, 随着生产力发展中科技进步因素的日益重要, 高等教育将成为科技进步的直接动力, 对经济发展做出积极的贡献。

将居民可支配收入、出生率与提取出的五个因子做路径分析, 如图 5 所示:



(注: *** $p < 0.001$)

图 5. 结构方程路径

从上述回归路径图中我们可以得知, 居民人均可支配收入与人口出生率有显著的负向预测作用, 且教育资金投入和基础教育质量在居民人均可支配收入对人口出生率的影响中起到了链式中介作用。与此同时, 高等教育质量中介了升学压力对于人口出生率的影响。当经济发达到一定程度时, 出生率会较低, 这在很多发达国家中都观察到了类似的现象。与此同时, 高等教育的发展使得这一部分人群具有更好地生育子女的条件, 但是基础教育目前对于生育及其教育的保障不足, 因而对人口出生率有不同方向的显著影响。

另外值得注意的是, 职业教育质量对人口出生率无显著影响 ($p > 0.05$)。这可能是因为, 目前存在着工人素质不高以及高素质劳动者稀缺的问题, 导致我国劳动生产力不高。但是目前, 职业教育仍然存在着社会污名, 没有得到广泛的认可, 因而其发展无法转变为人口红利, 导致相关路径不显著。职业教育是国家多层次教育体系中, 与劳动力市场需求对接最为紧密的教育类型, 是把“人口红利”变为“技能人才红利”的转换器。这提示我们要注意职业教育的发展及普及, 这样才能够解决技能人口的就业问题, 从而提高生育意愿[7]。

那么考虑到“双减”政策的颁布, 以上变量可能产生跳跃性的变化。如由于区域教育发展不平衡, 人数占比较少的、社会经济地位较高的父母能够相对而言提供更好的教育资源, 因而可能在短时间内导致教育质量的方差增大, 平均水平降低; 另外, 由于后进生在师生比暂时不会有很大变化的条件下, 缺少了课外辅导的机会, 学业上会面临更大阻力, 因而导致总体升学压力增大, 但对高等教育来说, “双减”政策则无显著影响。因而在这些方面, 由于区域发展的不平衡性, 导致欠发达地区的适龄男

女感到缺乏保障，从而生育意愿降低，人口出生率减少。但从长远来看，“双减政策”的推行会使得父母的教育资金会由于不再需要额外报补习班而有所减少，从而在一定程度上降低了生育成本，另外伴随着教育公平的推进，类似的担忧将会减少，因而会对人口出生率的上升有着积极作用。

5.3 问题三模型建立与求解

在计量经济学领域，面板数据是极其重要的一类数据类型。随着面板数据的广泛使用，面板分位数回归也随之出现。结合分位数回归与面板数据，采用分位数回归的方法对面板数据变量的参数进行估计，不仅能够更好的控制个体的异质性，而且能够分析在特定的分位数处自变量对因变量的边际效应，所以采用面板分位数回归可以使各个变量参数结果更加显著，具有更高的实际意义。为了刻画医疗政策对我国老龄化进程的影响，我们对解释变量取一阶滞后之后再进行回归。这样做的另外一个好处是，通过将解释变量取一阶滞后，可以降低内生性。故而我们首先采用全国样本的面板分位数回归分析方法来检验自变量与目标变量的关系，采用不同的控制变量来保证结果的可靠和稳健，之后在总样本解释变量一阶滞后面板分位数回归分析中，我们仍然采用上述回归分析方法和类似的回归方程，来进一步肯定我们的结论[3]。

人口老龄化程度的提高将引致医疗卫生服务需求的变化，通常来说，老龄人口的健康程度开始逐渐下降，对医疗卫生服务的需求更大。因此，一般来说，随着老龄化程度的提高，财政医疗卫生支出也会提高。关于医疗保健支出与老龄化的关系问题，有部分研究指出，对于一些发达国家，由于其健全的社会保障制度，发达国家人口老龄化与居民医疗保健支出并没有表现出明显的影响关系；除某些发达国家之外，有更多的研究证明医疗保健支出影响着人口老龄化，并呈现出正相关关系，这是相关研究中的主流结论[4]。从需求上看，在新的经济状况下，人们对生活有了更高的要求，越来越注重自身的医疗保健，对未来的考虑使得人们对自身寿命做出了越来越多的投资。因此对于养老保险，医疗保健等产业的优化，将大幅地降低中国的人口死亡率，从而加剧了社会的老龄化程度，从这个意义上说，医疗保健支出增长是人口老龄化的一大保障。

老龄化加速额使得老年人在总人口中所占比例越来越大，增加的这部分老年人口，由于机体抵抗力的减弱和城市污染的增加，就会滋生疾病，从而增加城市的医疗消费。因此，老龄化会促使医疗系统的升级和改进。基于此，首先采用全国样本做分位数回归分析来验证，解释变量均未取滞后项。采用以下基本方程：

$$o_t = \varphi X_t + \varepsilon_t \quad (9)$$

o_t 代表全国老龄化占比， X_t 表示各个与医疗有关的控制变量， ε_t 表示残差项。 T 为时间标识。 φ 为回归系数。

研究医疗与人口老龄化的问题，将老年人口（65 岁以上）占比作为因变量，选取 11 个自变量，分别做无滞后和有滞后一阶的分位数回归。表 5 为无滞后的分位数回归，

可以看不同分位数的拟合效果均大于 0.9，说明拟合效果较好。

表 5. 未进行滞后的分位数回归

	分位数 0.10	分位数 0.20	分位数 0.30	分位数 0.40	分位数 0.50	分位数 0.60	分位数 0.70	分位数 0.80	分位数 0.90
常数项	0 (/)	0 (/)	0 (0.062*)	0 (0.687)	0 (0.628)	0 (0.021**)	0 (0.298)	0 (/)	0 (/)
卫生技术人员数 (万人)	0 (/)	0 (/)	0 (0.033**)	0 (0.000***)	0 (0.000***)	0 (0.003***)	0 (0.239)	0 (/)	0 (/)
政府卫生支出 (亿元)	0 (/)	0 (/)	0 (0.688)	0 (0.619)	0 (0.373)	0 (0.13)	0 (0.385)	0 (/)	0 (/)
社会卫生支出 (亿元)	0 (/)	0 (/)	0 (0.016**)	0 (0.000***)	0 (0.000***)	0 (0.000***)	0 (0.286)	0 (/)	0 (/)
每万人拥有执业 (助理)医师数 (人)	-0.001 (/)	-0.001 (/)	-0.001 (0.214)	0 (0.011**)	0 (0.007***)	0.001 (0.003***)	0.001 (0.354)	0.001 (/)	0.001 (/)
人口死亡率(‰)	0.002 (/)	0.002 (/)	0.002 (0.497)	0 (0.568)	0 (0.497)	0 (0.56)	0 (0.988)	0.001 (/)	0.001 (/)
医院数(个)	0 (/)	0 (/)	0 (0.184)	0 (0.062*)	0 (0.019**)	0 (0.474)	0 (0.784)	0 (/)	0 (/)
卫生总费用 (亿元)	0 (/)	0 (/)	0 (0.511)	0 (0.000***)	0 (0.000***)	0 (0.000***)	0 (0.102)	0 (/)	0 (/)
医疗卫生机构数 (个)	0 (/)	0 (/)	0 (0.075*)	0 (0.000***)	0 (0.000***)	0 (0.014**)	0 (0.491)	0 (/)	0 (/)
老年抚养比(%)	0.006 (/)	0.006 (/)	0.006 (0.071*)	0.006 (0.000***)	0.006 (0.000***)	0.007 (0.000***)	0.007 (0.012**)	0.004 (/)	0.004 (/)
人均卫生费用 (元)	0 (/)	0 (/)	0 (0.674)	0 (0.000***)	0 (0.000***)	0 (0.000***)	0 (0.124)	0 (/)	0 (/)
卫生人员数 (万人)	0 (/)	0 (/)	0 (0.151)	0 (0.001***)	0 (0.000***)	0 (0.038**)	0 (0.548)	0 (/)	0 (/)
R2	0.987	0.987	0.986	0.986	0.987	0.989	0.989	0.991	0.993

观察每一行的回归结果可知，对老龄化有影响的变量为每万人拥有执业(助理)医师数(人)、老年抚养比(%)。有如下结论：

(1) 老年抚养比增加对老龄化有积极作用，由 0.1—0.9 分位数水平下，老年抚养比系数显著为正可知，老年抚养比增加对老龄化有积极作用。说明应该出台相关政策，加大子女对于其父母的抚养意愿以及建立健全的养老机构。

(2) 每万人拥有执业(助理)医师数(人)对老龄化有差异化影响。0.1—0.3 分位数的系数为负，0.4—0.5 分位数的系数为零，0.6—0.9 分位数系数为正。说明老年人口占比达到一定比例后，执业医师数会对老龄化产生正向影响。应该加大对有关人员的培训。

根据以上分析，我们得到了一般化的结论，为得到更加稳健的结论，需要更加深入的探究。为了更好刻画医疗对老龄化的影响，我们对解释变量取一阶滞后之后再行回归。这样做的另外一个好处是，通过将解释变量取一阶滞后，可以降低内生性，

这就相当于为上节的结论做了稳健性检验。

但不进行一阶滞后的分位数回归在回归时容易受到某些政策变量以及内生性变量干扰，所得结果可能产生伪因果关系。而对面板数据的内生性问题处理，使用差分和一阶滞后是两种典型的手段。这里我们选用一阶滞后的分位数回归模型再进行测试。

表 6 为一阶滞后的分位数回归，在因变量仍然为老龄化的情况下，我们还引入了死亡率等其他自变量进行探究。

表 6. 一阶滞后的分位数回归

	分位数 0.10	分位数 0.20	分位数 0.30	分位数 0.40	分位数 0.50	分位数 0.60	分位数 0.70	分位数 0.80	分位数 0.90
常数	-0.039 (/)	-0.037 (/)	-0.022 (0.258)	-0.007 (0.000***)	-0.012 (0.000***)	0.004 (0.000***)	0.036 (0.001***)	0.032 (/)	0.053 (/)
人口死亡率 (%)	0.001 (/)	-0.019 (/)	-0.006 (0.99)	0.133 (0.000***)	-0.002 (0.000***)	0.181 (0.000***)	-0.006 (0.96)	0.27 (/)	0.006 (/)
政府卫生支出 (亿元)	0 (/)	0 (/)	0 (0.981)	0 (0.000***)	0 (0.000***)	0.001 (0.000***)	0 (0.841)	-0.002 (/)	0 (/)
社会卫生支出 (亿元)	0 (/)	0 (/)	0 (0.965)	0 (0.000***)	0 (0.000***)	0.001 (0.000***)	0 (0.869)	-0.002 (/)	0 (/)
医疗卫生机构数(个)	0 (/)	0 (/)	0 (0.218)	0 (0.000***)	0 (0.000***)	0 (0.000***)	0 (0.010***)	0 (/)	0 (/)
卫生人员数 (万人)	-0.007 (/)	-0.012 (/)	-0.012 (0.233)	-0.008 (0.000***)	-0.01 (0.000***)	-0.009 (0.000***)	-0.013 (0.004***)	0.001 (/)	-0.014 (/)
人口出生率 (%)	-0.016 (/)	0.017 (/)	0.028 (0.506)	0.048 (0.000***)	0.015 (0.000***)	0.046 (0.000***)	0.01 (0.495)	0.06 (/)	0.03 (/)
每万人拥有 执业(助理) 医师数(人)	0.076 (/)	0.099 (/)	0.077 (0.451)	0.107 (0.000***)	0.069 (0.000***)	0.108 (0.000***)	0.162 (0.001***)	0.217 (/)	-0.001 (/)
人均卫生费用 (元)	0.021 (/)	0.021 (/)	0.021 (0.159)	0.028 (0.000***)	0.021 (0.000***)	0.03 (0.000***)	0.021 (0.001***)	0.031 (/)	0.02 (/)
居民人均可 支配收入 (元)	0 (/)	0 (/)	0 (0.462)	0 (0.000***)	0 (0.000***)	0 (0.000***)	0 (0.214)	0 (/)	0 (/)
老年抚养比 (%)	-0.005 (/)	-0.015 (/)	-0.014 (0.972)	0.174 (0.000***)	-0.005 (0.000***)	0.182 (0.000***)	-0.003 (0.976)	0.409 (/)	0.007 (/)
医院数(个)	0 (/)	0 (/)	0 (0.186)	0 (0.000***)	0 (0.000***)	0 (0.000***)	0 (0.001***)	0 (/)	0 (/)
个人现金卫 生支出 (亿元)	0 (/)	0 (/)	0 (0.967)	0 (0.000***)	0 (0.000***)	0.001 (0.000***)	0 (0.814)	-0.002 (/)	0 (/)
卫生总费用 (亿元)	-0.002 (/)	-0.001 (/)	-0.002 (0.705)	-0.002 (0.000***)	-0.001 (0.000***)	-0.003 (0.000***)	-0.001 (0.231)	0 (/)	-0.001 (/)
卫生技术人 员数(万人)	0.004 (/)	0.01 (/)	0.015 (0.363)	0.011 (0.000***)	0.009 (0.000***)	0.011 (0.000***)	0.016 (0.016**)	0.004 (/)	0.019 (/)
R2	0.923	0.841	0.744	0.696	0.565	0.602	0.343	0.619	0.417

由上表可以明显看出，滞后一项之后，控制变量对老龄化有影响的变量明显变多。结果可以得到如下结论：

(1) 人口死亡率(‰)对老龄化在 0.4-0.6 分位点有显著影响，且影响为负。

(2) 卫生人员数(万人)、卫生总费用(亿元)与老龄化为普遍为负相关关系。也就是说对医疗卫生人员投入越多反而容易让居民平均寿命延长，导致人口老龄化问题。

(3) 人口出生率(‰)、每万人拥有执业(助理)医师数(人)、人均卫生费用(元)、卫生技术人员数(万人)对老龄化的影响为正相关，但是影响较小。

(4) 老年抚养比(‰)对老龄化的影响有显著差异。

医疗人员数量的增多、医疗福利待遇的提高都对老龄化有着显著影响。在通过一阶滞后消除问题内生性的同时，可以看出滞后一项后的分位数回归更加准确，也可以对比滞后一项之后分位数回归与普通 OLS 回归的拟合图形，如图 6 所示。

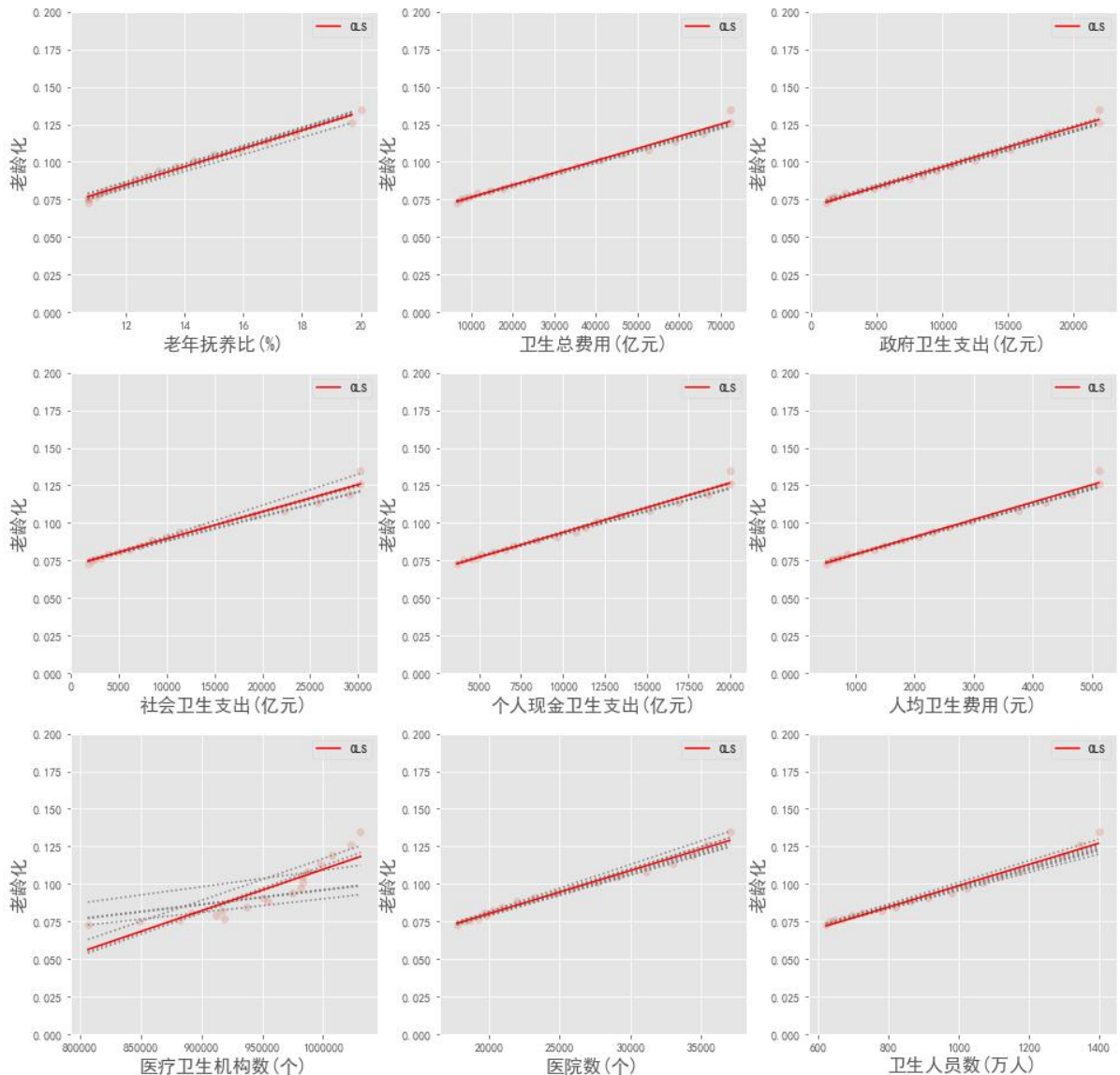


图 6. 医疗政策与人口老龄化之间的回归

可以看出分位数回归与 OLS 回归的拟合效果较吻合，说明一阶滞后结果较好。而从图 5 的曲线趋势中可以看出，这些趋势线均为正相关，说明对医疗卫生领域的经济、人力投入会让居民寿命更加延长从而导致更强烈的老龄化问题。但“老有所养”、

“老有所依”是一个社会稳定长久发展所需要的成果，尽管使得老年人居民占比更高，但这有利于社会稳定和和谐发展。

因此，在医疗新政策方面，我们认为：

- (1) 需要加大医疗财政投入，进行医疗体系的完善和人才培养模式的健全，大力推进医院、卫生所等基本医疗设施建设，以促进居家养老和社区养老，从而为面对人口老龄化提供有力保障。
- (2) 有关机构出台相关政策，降低新一代青年人的就业压力，加大子女对于其父母的抚养意愿以及建立健全的养老机构，鼓励更多人从事养老服务行业。
- (3) 建立健全的女性生育福利政策，降低女性生育压力，促进生育率的提高，延缓人口老龄化。
- (4) 而面对越来越重的养老财政投入，需要促进生产力与市场，切实提高就业率充盈国家财政水平的同时，加大对养老财政的投入比例。
- (5) 对于“就医贵”、“就医难”等问题，需要在给予更多医疗保险福利、药价、就医价格下调的同时，培养更多的高素质高水平医疗卫生技术人才。

5.4 问题四模型建立与求解

从“教育”、“医疗”、“住房”、“用人单位减负”、“个人税专项附加扣除”、“延长男性育儿假”多方面综合出发，以出生率为研究因变量。其中由于“延长男性育儿假”提出时间较短，难以量化，在此处暂不考虑。

首先从教育角度出发，图 7 描述 2012 年-2021 年我国教育资金超出财政预期的份额。可以说明教育实际的投出超出国家预期财政，如 2014 年国家实际教育资金超出预算的 450%，说明前期国家对教育投入太少，财政预算原低于实际需求。近些年财政投入在逐年递增，但是超出预期份额在逐渐递减，说明国家对财政预期越来越高，在未来十年以内，国家教育投入可能会在财政预算之内，说明国家对教育投入在增长。

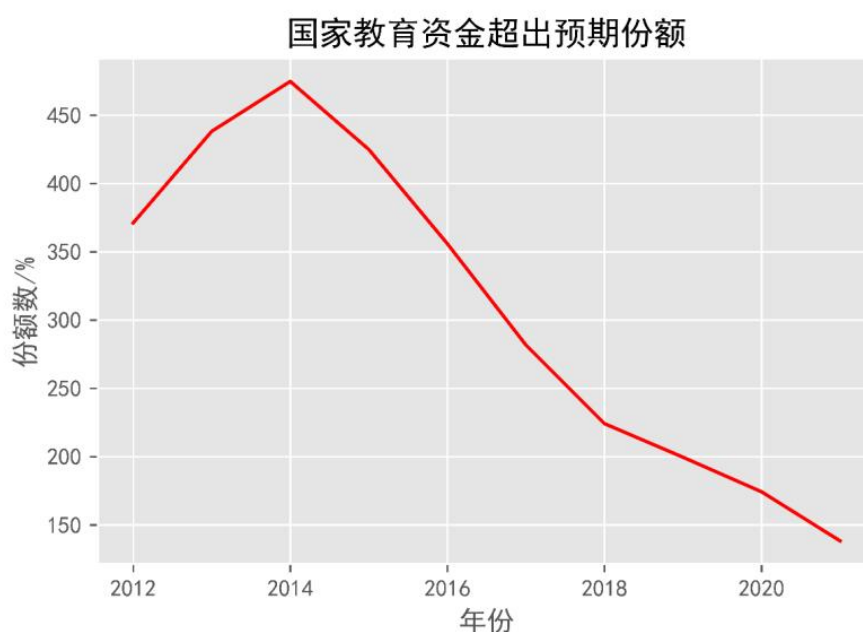


图 7. 教育资金超出财政预期份额百分比变化

近年来，高房价对生育行为的影响引发了学者的探讨。宏观数据方面，住房支付能力减弱直接作用于生育率使其下降，间接影响人口年龄结构导致人口老龄化加剧；有研究者运用动态面板 GMM 估计方法进行研究，发现房价对出生率具有显著的负向影响。微观数据方面，根据家庭生育效用理论模型，房价每上升 1%，家庭初次生育时间约推迟 1.05 年；有研究者基于 2014 年中国家庭追踪调查数据研究了房价对家庭生育决策的影响，发现房价每上涨 1000 元，生育一孩的概率将降低 1.8~2.9%，生育二孩的概率将降低 2.4~8.8%，女性的初次生育年龄将提高 0.14~0.26 年。尤其是在住房面积较大或房价较便宜的城市，生育率较高，高房价意味着沉重的生活成本，因此高房价减少了可用于抚养孩子的资源和生育意愿[5]。

住房方面则从老年人口占比、各项税收（亿元）、就业人员（万人）、居民消费价格指数、劳动力（万人）、居民人均可支配收入(元)六方面做滞后一阶的分位数回归考虑对生育率的影响。

表 7. 对生育率的一阶滞后分位数回归

	分位数 0.10	分位数 0.20	分位数 0.30	分位数 0.40	分位数 0.50	分位数 0.60	分位数 0.70	分位数 0.80	分位数 0.90
const	-2.416 (/)	-1.98 (/)	-1.654 (0.029**)	0.648 (0.000***)	0.653 (0.000***)	0.859 (0.000***)	1.086 (0.001***)	1.404 (/)	1.43 (/)
老龄化	0.001 (/)	0.001 (/)	0.001 (0.000***)	-0.001 (0.000***)	-0.001 (0.000***)	0 (0.000***)	-0.001 (0.000***)	-0.001 (/)	0 (/)
各项税收（亿元）	0 (/)	0 (/)	0 (90.371)	0 (0.000***)	0 (0.000***)	0 (0.032**)	0 (0.083*)	0 (/)	0 (/)
就业人员（万人）	0.005 (/)	0.007 (/)	0.007 (0.073*)	-0.001 (0.040**)	-0.001 (0.025**)	-0.001 (0.124)	-0.002 (0.018**)	-0.001 (/)	-0.001 (/)
居民消费价格指数(上年=100)	0.347 (/)	0.187 (/)	0.161 (0.784)	-0.04 (0.591)	-0.041 (0.567)	-0.067 (0.563)	-0.007 (0.962)	-0.034 (/)	-0.033 (/)
劳动力（万人）	-0.001 (/)	-0.001 (/)	-0.002 (0.496)	0 (0.531)	0 (0.526)	0 (0.533)	0.001 (0.070*)	0 (/)	0 (/)
居民人均可支配收入(元)	0.001 (/)	0.001 (/)	0.001 (0.238)	-0.001 (0.000***)	-0.001 (0.000***)	-0.001 (0.009***)	-0.001 (0.016**)	-0.001 (/)	-0.001 (/)
R2	0.691	0.484	0.25	0.189	0.219	0.252	0.298	0.36	0.452

从表 7 中可以很明显地看到，老龄化、税收、就业人员、居民人均可支配收入是影响出生率的显著因素，而居民消费价格指数和劳动力却并不显著。人口老龄化与出生率之间的影响是可以理解的，因为养老压力重的情况下居民平均生育年龄也在提高，所以生育欲望会比较低；而税收、商品房等则是居民生活压力的来源之一，需要对结构进行合理调整；居民收入水平反映人民生活水平与财务自由程度，对生育率也有一

定影响是可以理解的。

本质上，需要对市场进行合理调整，减轻青年人生活压力的同时为青年人就业工作创造更多机会和收益，切实保障青年劳动力的各项社会福利。

再从其他方面考虑对出生率的影响，如下图所示，其中“卫生人员数”、“卫生技术人员”、“每万人拥有执业（助理）医师数”可以分析医疗情况对出生率影响；“就业人数”、“劳动力”、“商品房平均销售价格”可以用来分析劳动力对出生率影响；“居民消费价格指数”、“居民人均可支配收入”、“各项税收”可以用来分析个人所得税方面对出生率的影响。

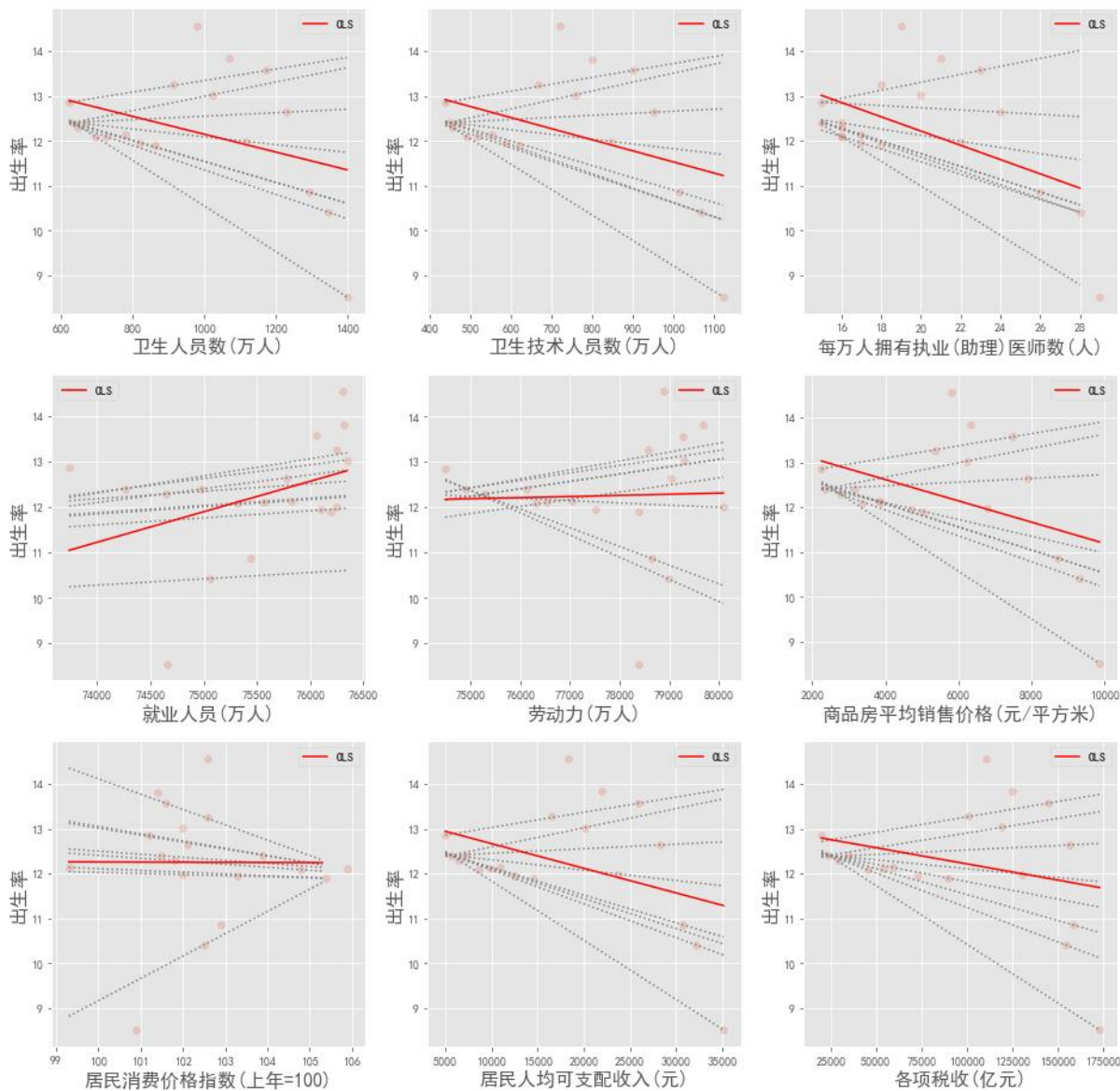


图 8. 不同变量对生育率回归效果

从图 8 给出的曲线中可以看出，变量的分布相对较为零散，对于卫生人员、卫生机构数的增多生育率反而呈现下降的趋势。但是这三个变量的回归都是不显著的，因此我们认为这三项的回归结论存在一定程度的内生性问题。

而就业人员和劳动力的增加，尤其是就业人员的增加，能有效促进出生率。这是由于就业人员与市场的发展会给家庭创造更多收入，从而减轻生育压力。一个牵制出生率最主要的因素就是生活压力与收入水平不够，而国家在就业市场上还需要创造更多就业机会与提高基本工资水平。另外，商品房、税收会对出生率产生负相关关系，因为商品房本身是居民经济负担的重要源头之一。需要对房价结构、税收结构进行合理调整，降低青年人生活负担促进生育意愿。

一个有意思的现象是居民人均可支配收入与出生率之间呈负相关关系，这是由于可支配收入的增高居民本身自由感和满足感较强，也会降低生育意愿。

因此，我们根据模型给出的结果给出以下促进生育意愿的方案：

（1）延长男性育儿假。研究表明，当丈夫承担家庭成员照料与家务劳动的时间增加,女性面临的工作与家庭间的冲突会减少,可能正向激励生育意愿。与之相反,当家庭性别分工不平等时，会对女性生育意愿起到负向抑制作用。还有研究认为家庭性别关系与生育率呈U型关系，在性别角色传统和性别角色平等的家庭中生育率更高[8]。

（2）营造对于女性更加宽松的劳动市场政策，减轻家庭经济负担，创造有利于女性兼顾事业与家庭的制度环境，从而降低生育成本。

（3）完善教育医疗条件，如促进区域平衡发展，完善医疗保险，完善育儿设施建设等，从而降低生育成本，保证生育意愿。

（4）房价上涨已严重制约家庭的生育行为，因此政府必须加强对房地产市场的宏观调控，出台更多利民便民的购房政策，严格管控房价的大幅度上涨；完善公共租赁住房体系，规范租房机制，杜绝恶性高价、非法租赁行为。

（5）采取据实扣除与定额扣除相结合的扣除制度，依据年龄、地域、家庭结构差异细化扣除标准，并将教育支出以外的其他生计费用纳入扣除范围，以期实现个税专项附加扣除政策与“三孩政策”的配套衔接[9]。

六、模型的分析与评价

6.1 模型的优点

我们的模型形式简洁方法科学，在经过合理评估后，我们认为我们的模型具有如下几个方面的优势：

1. 模型形式简洁，通过基本的计量手段揭示了抽象的社会科学变量之间存在的统计建模关系，并通过量化方法对问题提出合理的建议。

2. 模型策略合理，充分排除了无关变量干扰。对于面板数据，时间上的滞后是一种消除内生性的经典策略。这里经过多重对比，使用的一阶面板滞后分位数回归法相比于普通分位数回归不仅拟合效果上得到了提升，更多变量产生了显著效应并且有效消除了内生性因素。

3. 模型具有良好的可解释性。从模型的统计结果与显著性可以清晰看出变量之间的作用关系以及作用程度，能够比较直观地揭示其作用机理并映证当下社会热点问题，能够给出一些参考意见。

4. 模型形式直观，无论是 SEM 路径图还是热力图等都能直观反映研究结论。

6.2 模型的缺点

但就客观而言，我们认为我们的模型还存在以下不足：

1. 数据面板不够，样本单一，这一问题主要是受限于国家统计局按年度的数据且最多 20 年。并且本文暂时还没有分区域进行实证，没有考虑到地缘因素与经济发展水平差异对问题的影响。

2. 模型策略简单，内生性问题未能充分消除。这一问题仍然来自于面板数据过少且 2021 年才提出双减政策，缺乏面板数据与断点，难以使用断点回归等经典内生性消解策略。

6.3 模型的改进与扩展

我们认为这些问题都能通过数据的方式得到解决，在充分调研各不同地区的相关变量与发展水平以及获取更长时间的面板以后，仍然可以通过上述手段对问题进行更为合理的建模。

参考文献

- [1] 程开明. 结构方程模型的特点及应用[J]. 统计与决策, (10), 22-25, 2006.
- [2] 荆建华. 试论教育对人口的影响[J]. 河南教育学院学报(哲学社会科学版), (03), 1-7, 1994.
- [3] 邓利方,李铭杰.人口老龄化对不同层次城市医疗消费的影响分析——基于面板分位数回归模型[J].广东行政学院学报, 28(05), 82-90, 2016.
- [4] 周荣. 人口老龄化与财政医疗卫生支出, 影响机理与效应测度[D]. 安徽财经大学, 2021.
- [5]倪云松. 房价、房产与生育行为 ——基于 CHFS 数据的分析[J]. 经济问题, (11), 121-129, 2021.
- [6] 贺丰年. 因子分析法应用于人力资源配置的分析研究[J]. 企业经济, 32(06), 63-6, 2013.
- [7] 庄西真. 发展职业教育, 对冲人口结构变化的影响[J]. 职教论坛, 37(5), 1, 2021.
- [8] 王一帆, 罗淳. 促进还是抑制? 受教育水平对生育意愿的影响及内在机制分析[J]. 人口与发展, 27(5), 72-82, 2021.
- [9] 吕春娟, 李静. "三孩政策"下个税专项附加扣除制度优化设计[J]. 行政管理改革, 9(9), 35-43, 2021.

附录

Leslie 模型 Matlab 代码

```
clc,clear
p=0.48764 %2021 年女性所占总人口比例
a=xlsread('data3.xlsx');
N=a(:,1);
N0=N(~isnan(N)); %第 0 年（2021）年女性各年龄段人口数（万人）
N00=(N0/1000)' %第 0 年（2021）年女性各年龄段人口数（千万人）
A=eye(90);
b=a(:,8);
b0=b(~isnan(b));
b1=((b0([1:90],1))/100)'
for i=1:90
    A(i,:)=A(i,:)*b1(1,i);
end
c=a(:,9);
c0=c(~isnan(c));
c1=((c0([1:90],1))/1000)'; %修正后的生育率
M=sum(c1'); %总合生育率
d=zeros(91,1);
B=[c1;A];
L=[B,d]; %构造的 leslie 矩阵
G=[];
for i=0:15
    X=L^i*(N00)'; %第 i 年后女性各个年龄段的人口数（千万）
    G=[G,X];
    Z=X./p; %第 i 年在各个年龄段的人口总数预测
    K(i+1,1)=sum(Z);
    S1=sum(Z([1:15],:)); %第 i 年 0-14 岁的总人数
    D(1,i+1)=S1;
    S2=sum(Z([16:65],:)); %第 i 年 15-64 岁的总人数
    E(1,i+1)=S2;
    S3=sum(Z([66:91],:)); %第 i 年 65-95 岁的总人数
    F(1,i+1)=S3;
end
K %2020-2035 人口总数
D %年龄在 0-14 岁总人数（包括男女）
E %年龄在 15-64 岁总人数（包括男女）
F %年龄在 65 及 65 岁以上总人数（包括男女）
G %每个年龄段的女性人数
x=2020:2035;
y=K;
figure(1)
```

```

plot(x,y,'*')
title('我国全国总人口变化趋势');
xlabel('时间（单位：年）');
ylabel('人口数量（单位：千万）');
x=2020:2035;
y1=D';
y2=E';
y3=F';
figure(2)
plot(x,y1,'-or',x,y2,'-ob',x,y3,'-og')
title('我国全国各年龄段人口变化趋势');
xlabel('时间(单位：年)');
ylabel('人口数量（单位：千万）');
hold on;
legend('年龄在 0-14 岁总人数','年龄在 15-64 岁总人数','年龄在 65 及 65 岁以上总人数');
hold off;

```

分位数回归 Python 代码

```

# -*- coding: utf-8 -*-
"""
Created on Fri May 20 10:28:44 2022

@author: Mashituo
"""

import numpy as np
import pandas as pd
import statsmodels.api as sm
import statsmodels.formula.api as smf
import matplotlib.pyplot as plt
plt.rcParams['font.sans-serif'] = ['SimHei'] #显示中文
plt.rcParams['axes.unicode_minus']=False #用来正常显示负号
plt.style.use("ggplot")

data = sm.datasets.engel.load_pandas().data
data.head()
data=pd.read_excel("问题三数据.xlsx")
data=data.iloc[1:]
old_cols=data.columns
data.columns=['x%d'%i for i in range(21)]
fig, ax = plt.subplots(figsize=(16, 16))
for per in range(11,20):
    data['per']=data['x%d'%per]
    quantiles = np.arange(.1, .9, .1)
    mod = smf.quantreg('x20 ~ per', data)

```

```

def fit_model(q):
    res = mod.fit(q=q)
    return [q, res.params['Intercept'], res.params['per']] + res.conf_int().loc['per'].tolist()

models = [fit_model(x) for x in quantiles]
models = pd.DataFrame(models, columns=['q', 'a', 'b', 'lb', 'ub'])

ols = smf.ols('x20 ~ per', data).fit()
ols_ci = ols.conf_int().loc['per'].tolist()
ols = dict(a = ols.params['Intercept'],
           b = ols.params['per'],
           lb = ols_ci[0],
           ub = ols_ci[1])

print(models)
print(ols)

x = np.arange(data.per.min(), data.per.max(), 1)
get_y = lambda a, b: a + b * x
#fig, ax = plt.subplots(figsize=(8, 6))
plt.subplot(3,3,per-10)
for i in range(models.shape[0]):
    y = get_y(models.a[i], models.b[i])
    plt.plot(x, y, linestyle='dotted', color='grey')

y = get_y(ols['a'], ols['b'])

plt.plot(x, y, color='red', label='OLS')
plt.scatter(data.per, data.x20, alpha=.2)
#plt.ylim((0,0.2))
legend = plt.legend()
plt.xlabel(old_cols[per], fontsize=16)
plt.ylabel('出生率', fontsize=16);

```