



LNU

辽宁大学经济学部  
Faculty of Economics, Liaoning University

讨论稿系列  
Working Paper Series

No. C2022007

2022-12-05

## 参保人口缴费行为选择、养老金替代率可能性边界 与收支平衡<sup>①</sup>

陈曦 李佳雨 范璐璐

**摘要：**构建保障适度、可持续的养老保险体系，需要破解提高养老金保障水平与促进养老保险制度可持续发展的政策选择悖论。参保者缴费行为选择是影响养老给付水平和收支平衡的交叉性要素，本文以参保者缴费行为选择为切入点，分析提高养老金保障水平与改善收支平衡状况的帕累托改进可能性。研究发现：（1）城镇职工养老金社会平均工资替代率可能性边界下限为 12.76%–13.46%，上限为 51.43%–51.75%；（2）参保行为选择对养老金社会平均工资替代率具有直接影响，在参保缴费基数情景下，养老金社会平均工资替代率区间为 34.88%–35.60%，人口老龄化导致养老金给付平均水平下降约 0.8 个百分点；（3）延长缴费年限和降低参保起始年龄能够提高养老金替代率，但是会增加养老保险基金收支缺口。提高缴费基数可以实现提高保障水平与缩减养老保险收支缺口的帕累托改进；（4）城镇职工养老保险政策缴费激励导向长缴多得，参保者现实选择更加倾向于延长缴费年限，参保者缴费选择与政策激励导向相一致，缴费激励机制需要优化；（5）利用机器学习随机森林模型对参保者缴费档次选择的影响因素进行识别，参保者对政策信任程度等指标显著影响缴费档次选择。

**关键词：**参保缴费行为选择；养老金社会平均工资替代率；帕累托改进；随机森林模型

### 一、问题提出

党的十九大报告提出“全面建成覆盖全民、城乡统筹、权责清晰、保障适度、可持续的多层次社会保障体系”。人口老龄化持续提高过程中，面临合理提高养老保障水平与促进制度可持续发展的选择性悖论。通常情况下，养老金调整指数低于社会平均工资替代率，人口预期寿命延长导致参保者生命周期内养老金保障相对水平下降（穆怀中等，2021），提高养老金保障水平会降低养老保险制度可持续性。是否存在提高保障水平与促进制度可持续的帕累托改进可能性？是积极应对人口老龄化和构建保障适度、可持续养老保障体系面临的关键

<sup>①</sup> 基金项目：国家社科基金后期资助项目“城乡基础养老保险缴费率一元化适度水平及其实现路径（19FRKB001）阶段性研究成果。

作者简介：陈曦，辽宁大学公共管理学院，副教授，博士生导师；李佳雨，辽宁大学公共管理学院，博士研究生；范璐璐，辽宁大学经济学院，副教授。

问题。

养老金社会平均工资替代率是判断养老保障水平和影响制度收支平衡的交叉性指标,从交叉性元素切入,分析同时实现提高养老金保障水平和制度可持续发展的可能性路径是较为合理的研究视角。养老金社会平均工资替代率是人均养老金与社会平均工资之间的比例关系,反映全社会平均养老金对收入的替代和消费满足程度。在现行制度框架下,缴费标准直接决定参保个体养老金替代率,从而影响养老金社会平均工资替代率水平,缴费下限标准与缴费上限标准分别对应养老金社会平均工资替代率可能性边界下限与上限。养老金社会平均工资替代率可能性边界是检验养老保险收支平衡测算替代率指标参数设定合理性的有效工具,也是测度全社会平均养老保障水平范畴的合理指标。

在养老金社会平均工资替代率可能性边界区间内,探索同时实现合理提高养老保障水平与促进养老保险制度收支平衡,需要立足于养老保险给付与缴费对应的微观机理,即参保者缴费行为选择决定养老金给付水平,也影响养老保险基金征缴收入。参保者缴费行为选择是参保者在政策框架下依据个人经济、社会特征进行的参保决策,同时也是养老保险缴费激励导向的最终落脚点。判断提高保障水平与养老制度可持续性提升的帕累托改进可能性,需要分析哪种参保者缴费行为选择具有提高给付水平与改进收支平衡的效果。针对宏观政策问题,寻找微观破解逻辑,提供养老保险制度优化的新视角。

随着人口老龄化程度逐渐加深,养老保险制度可持续性越发受到学者关注。学者们通过构建养老保险基金收支平衡模型,测算人口老龄化过程中养老保险收支平衡发展趋势,实证检验缴费年限变化、征收体制改革、延迟退休等政策内部要素对养老保险制度可持续性的影响(杨一心、何文炯,2016;曾益等,2019;邓大松、仙蜜花,2015),以及零工经济、疫情冲击等政策外部要素对养老保险制度可持续性的冲击(张国英等,2022;赵亮、李灯强,2020)。特别是在养老保险降费率政策下,如何促进养老保险制度可持续发展成为学者们重点关注的问题(陈曦,2017;郑秉文,2016)。部分学者从微观参保决策视角出发,分析养老保险收支平衡对策。曾益等(2022)研究降低缴费率对缴费遵从度的影响,进而分析养老保险收支平衡变化。张川川、朱涵宇(2021)研究发现同村居民参与新农保的比例每增加10个百分点,个体参保概率显著增加4.24个百分点,有针对性地提高特定群体的参保率是提高社会保障项目整体参与率的有效手段。在积极应对人口老龄化过程中,养老金保障水平与制度可持续之间的矛盾开始显现。齐红倩、杨燕(2020)研究认为平均替代率对养老保险基金累计结余率有负向非线性影响。如何在保持养老金给付水平不下降的情况下,促进养老保险收支平衡可持续成为学者们关注的焦点,学者们利用代际交叠模型等方法从国有资本划转等视角分析养老给付水平约束下的收支平衡改进可能性(景鹏等,2020)。上述研究多是从宏观视角研究养老金保障水平与制度可持续之间的关系,缺少从微观参保主体选择视角破解两者矛盾的研究思路。

本文利用精算模型测算城镇职工养老金社会平均工资替代率的可能性边界,分析参保行

为选择对养老金社会平均工资替代率的影响,以及提高养老金保障水平与改善养老保险收支平衡状况的帕累托改进。构建养老保险缴费激励系数模型,判断现行缴费激励机制的多缴多得与长缴多得倾向性,分析参保者现实选择与缴费激励导向的匹配性,进一步利用机器学习随机森林模型识别参保者缴费行为选择影响因素,提出实现帕累托改进的对策。

## 二、养老金替代率可能性边界模型及测算

### (一) 养老金替代率可能性边界模型构建

养老金社会平均工资替代率是退休初始替代率、退休期养老金调整指数与工资增长率比值和退休人口结构共同作用的平均养老给付水平。本文构建养老金社会平均工资替代率的数理模型:

$$\bar{S}_t = \sum A_{i,j,t} \times S_{i,j,t} = \sum A_{i,j,t} \times S_0 \times \left( \frac{1+h}{1+g} \right)^{t-m} \quad (1)$$

其中,  $\bar{S}_t$  表示养老金社会平均工资替代率,  $A$  表示退休人口中不同年龄的人口占比,  $S_{i,j,t}$  表示不同年龄、性别退休人口的社会平均工资替代率,  $S_0$  表示退休初始社会平均工资替代率,  $h$  表示养老金调整指数,  $g$  表示工资增长率,  $m$  表示退休时点。

根据城镇职工养老保险给付模式政策规定,本文进一步设定退休初始社会平均工资替代率模型,其中模型右侧第一部分为基础养老金社会平均工资替代率,第二部分为个人账户养老金社会平均工资替代率。

$$S_0 = \frac{n\% \times (1+\beta)}{2 \times (1+g)} + \frac{\theta \times 12 \times (1+r)^{n+1} \times \left[ 1 - \left( \frac{1+g}{1+r} \right)^n \right]}{\delta \times (r-g) \times (1+g)} \quad (2)$$

其中,  $n$  表示缴费年限,  $\beta$  表示缴费基数占社会平均工资的比重,  $\theta$  表示个人账户缴费率,  $r$  表示个人账户基金投资收益率,  $\delta$  表示个人账户计发系数。将公式(2)代入公式(1)得到养老金社会平均工资替代率的最终模型:

$$\bar{S}_t = \sum A_{i,j,t} \times \left\{ \frac{n\% \times (1+\beta)}{2 \times (1+g)} + \frac{\theta \times 12 \times (1+r)^{n+1} \times \left[ 1 - \left( \frac{1+g}{1+r} \right)^n \right]}{\delta \times (r-g) \times (1+g)} \right\} \times \left( \frac{1+h}{1+g} \right)^{t-m} \quad (3)$$

### (二) 指标参数设定

本文对养老金社会平均工资替代率模型相关指标进行参数设定,确定养老金替代率可能性边界。

1. 缴费年限与缴费基数。根据养老保险政策规定,最低缴费年限为15年,缴费基数下

限为社会平均工资 60%，本文选择缴费年限 15 年、缴费基数为社会平均工资 60%作为替代率可能性边界下限指标参数。考虑到劳动初始年龄和退休年龄，将男性参保职工缴费年限上限选择 40 年，女性参保职工缴费年限上限选择 35 年，社会平均工资是不同收入群体的平均收入水平，平均缴费基数以不超过社会平均工资为上限。

2. 工资增长率与投资收益率。借鉴陈曦（2017）研究成果，工资增长率设定为 5%，与 2020 年实际工资增长率接近。考虑现阶段社会保障基金平均投资收益率，个人账户投资收益率设定为 3%。

3. 养老金调整指数。通过比较近年来城镇职工养老金调整指数与社会平均工资增长率之间的关系发现，养老金调整指数占社会平均工资增长率的比重约为 80%。本文设定养老金调整指数下限为 4%，养老金调整指数以不超过社会平均工资增长率为上限。

4. 退休年龄及退休人口中不同年龄人口占比。本文设定男性退休年龄为 60 周岁，女性退休年龄为 55 周岁。假定参保退休人口年龄构成与城镇人口年龄构成相同，利用人口普查数据对城镇年龄别人口进行预测，得到退休人口中不同年龄人口的占比。

### （三）养老金替代率可能性边界测算

根据测算发现，2021 年城镇职工养老金社会平均工资替代率可能性边界为 13.46%-51.43%，2050 年替代率可能性边界为 12.76%-51.75%。在全劳动期缴费、养老金调整指数与社会平均工资增长率相同等极限条件下，养老金社会平均工资替代率保持在约 51%，在测度城镇职工养老保险收支平衡以及养老金给付水平过程中，需要将社会平均工资替代率限定在可能性边界上限之内。同时，养老金调整指数低于社会平均工资增长率，将会导致退休期内参保个体的社会平均工资替代率下降，在人口老龄化的影响下，养老金替代率可能性边界下限将会下降，需要采用缴费激励等措施，提高养老金给付水平，确保养老金能够发挥保障基本生活需求的功能。

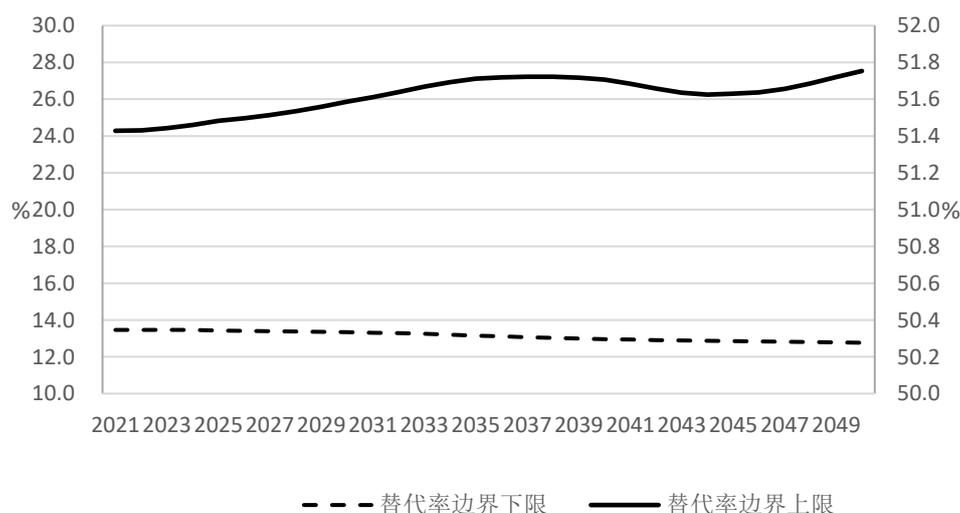


图 1 城镇职工养老金社会平均工资替代率边界

### 三、参保行为选择、养老金社会平均工资替代率与收支平衡变化

#### (一) 参保行为选择对养老金社会平均工资替代率的影响

根据养老金社会平均工资替代率可能性边界模型,缴费年限、缴费基数占社会平均工资比重等指标是影响替代率水平的关键因素。缴费年限与缴费基数是微观个体参保行为选择的指标体现,在最低缴费年限标准之上,参保者可以根据缴费能力、政策预期等因素主观选择缴费年限与缴费档次,特别是个体工商户和灵活就业人员拥有更多的参保自主权。参保行为选择不同将会导致养老金给付水平发生显著变化。

本文设定养老保险参保基准情景,在此基础上分析参保行为变化导致养老金社会平均工资替代率变动。<sup>①</sup>考虑参保人口在劳动期的就业转换、就业起始年龄等因素,设定基准情景下男性参保者平均缴费年限为30年、女性参保者平均缴费年限为25年。在现实生活中,参保者多数选择接近退休年龄时进行缴费,因此本文设定基准情景下参保者选择劳动期后期缴费。由于灵活就业人员最低缴费档次为社会平均工资60%,综合考虑灵活就业人员参保缴费档次选择等因素,设定基准情景下缴费基数占社会平均工资比重为80%。

在基准情景基础上,本文分析参保缴费行为选择变化与养老金社会平均工资替代率之间的关系。情景I:在基准情景下,养老金社会平均工资替代率区间为34.88%-35.60%,人口老龄化导致养老金给付平均水平下降约0.8个百分点。情景II:在参保缴费年限提高10年的情况下,养老金社会平均工资替代率提高至40.64%,可能性边界为39.78%-40.64%。情景III:在保持缴费年限与参保起始年龄不变情况下,平均缴费基数由社会平均工资80%提高至社会平均工资100%,养老金社会平均工资替代率提高至41.04%,可能性边界为40.21%-41.04%。情景IV:在保持缴费年限与缴费基数不变情况下,参保者由后期缴费变为前期缴费,即就业起始年龄开始参保缴费。降低参保起始年龄会提高养老金给付水平,养老金社会平均工资替代率由35.60%提高至36.17%,可能性区间为35.43%-36.17%。参保行为选择变化会对养老金给付水平产生显著影响,参保缴费年限越长、缴费基数越高、参保起始年龄越低,养老金社会平均工资替代率越高,可以抵消人口老龄化对养老金给付水平的影响。

表1 参保行为选择对养老金社会平均工资替代率的影响

年份	基准情景替代率 (%)	延长缴费年限替代率 (%)	提高缴费基数替代率 (%)	降低参保起始年龄替代率 (%)
2021	35.60	40.64	41.04	36.17
2025	35.61	40.65	41.05	36.18
2030	35.55	40.56	40.98	36.11
2035	35.39	40.37	40.80	35.96
2040	35.12	40.07	40.49	35.68

<sup>①</sup> 基准情景是对参保指标参数平均水平的综合判断,也是分析参保行为选择与养老金社会平均工资替代率变动之间关系的比较标准,基准情景水平变动不会改变比较结果的性质。

2045	34.91	39.83	40.24	35.46
2050	34.88	39.78	40.21	35.43

注：基准情景，男性缴费30年，女性缴费25年，后期缴费，缴费基数=社会平均工资80%，养老金调整指数=社会平均工资增长率80%；延长缴费年限情景，男性缴费40年，女性缴费35年，后期缴费，缴费基数=社会平均工资80%，养老金调整指数=社会平均工资增长率80%；提高缴费基数情景，男性缴费30年，女性缴费25年，后期缴费，缴费基数=社会平均工资，养老金调整指数=社会平均工资增长率80%；降低参保起始年龄情景，男性缴费30年，女性缴费25年，前期缴费，缴费基数=社会平均工资80%，养老金调整指数=社会平均工资增长率80%。

## （二）参保行为选择对养老保险基金收支平衡的影响

激励参保者延长缴费年限、提高缴费基数和降低初始参保年龄会提高养老金社会平均工资替代率，相应增加养老保险基金支出。同时，参保行为选择变化也会对养老保险基金收入相关参数产生影响，参保行为选择变化之后养老保险基金收入增幅大于基金支出增幅，则参保行为选择变化既能够提高养老金给付水平，也能够促进养老保险收支平衡。本文进一步检验参保行为变化是否具有同时实现养老金给付水平提升和制度可持续性优化的帕累托改进可能性。

参保者选择延长缴费年限，缴费年限占劳动期的比重提高，相应的养老保险遵缴率提升，养老保险基金收入增加，缴费年限提高对应养老金社会平均工资替代率提高，养老保险基金支出增加。以缴费年限延长10年为例，通过比较基金收入与支出增加幅度发现，延长缴费年限导致养老保险基金支出增加幅度会高于基金收入增加幅度，使得养老保险基金收支缺口增大，即延长缴费年限虽然能够提高保障水平，但会减弱养老保险制度可持续性。

参保者选择提高缴费基数会直接增加当期养老保险基金征缴收入，根据养老保险多缴多得原则，养老金社会平均工资替代率水平也会提高，养老保险基金支出增加。在缴费基数由社会平均工资80%提高至社会平均工资100%的情况下，提高缴费基数显著增加养老保险基金收入，在抵消基金支出增加的情况下仍有剩余，养老保险收支缺口缩小。提高缴费基数不仅可以提升养老保障水平，也可以促进养老保险制度可持续发展。

降低参保起始年龄是参保者在就业初始年龄选择参保，个人账户缴费增值期限延长，个人账户养老金给付水平有所提高，养老金社会平均工资替代率增加，对应的养老保险基金支出增加。在缴费年限的固定不变情况下，参保者选择前期缴费，在劳动期后期处于参保不缴费的状态，养老保险遵缴率会发生相应变化<sup>①</sup>。在降低参保初始年龄的情况下，虽然能够提高养老金社会平均工资替代率水平，但遵缴率会有一定的下降，养老保险基金收支缺口增加。

通过比较参保者行为选择变化发现，延长缴费年限和降低参保起始年龄能够提高养老金保障水平，但是不利于改善养老保险制度收支平衡状况。提高缴费基数既可以提高保障水平，又可以促进养老保险制度可持续发展，激励参保者提高缴费基数可以实现提高保障水平与缩

<sup>①</sup> 由于参保者遵缴程度数据难以有效获取，本文假定在基准情景下劳动者选择后期缴费，当期养老保险实际缴费人口为30-59周岁男性参保者和30-54周岁女性参保者，养老保险遵缴率=实际缴费人口/参保总人口。在降低参保初始年龄的情况下，劳动者选择前期缴费，当期养老保险实际缴费人口为20-49周岁男性参保者和20-44周岁女性参保者，随着实际缴费人口变化，养老保险遵缴率也发生相应变化，对基金收入具有直接影响。

减养老保险收支缺口的帕累托改进。

表 2 参保行为选择对养老保险收支缺口的影响

年份	基准情景收支平衡（亿元）	延长缴费年限收支平衡（亿元）	提高缴费基数收支平衡（亿元）	降低参保起始年龄收支平衡（亿元）
2021	5105	3893	9811	2537
2025	-3343	-5825	1026	-7084
2030	-19820	-24399	-16481	-24392
2035	-45573	-50585	-44267	-48350
2040	-80210	-89195	-81219	-86327
2045	-125274	-140999	-129187	-135761
2050	-183965	-209085	-191670	-192467

注：（1）本文以 2020 年数据为基础，假定养老保险劳动人口覆盖率和老年人口覆盖率逐步提高，在 2050 年达到 100%；（2）社会平均工资增长率为 5%，养老保险总缴费率为 24%，养老保险替代率为不同情景下养老金社会平均工资替代率，缴费基数与遵缴率按照不同情景缴费模式进行设定和测算。

#### 四、参保行为现实选择、政策激励导向与优化机制识别

##### （一）政策激励导向与参保行为现实选择的一致性

激励参保者提高缴费基数是合理提高养老保障水平与促进养老保险制度可持续发展的重要途径。本文首先测算养老保险政策对参保行为的激励导向，判断在现有政策框架下多缴多得与长缴多得哪种缴费选择更为有利。

本文比较在相同缴费预算条件下，参保者选择多缴多得与长缴多得对应的养老金给付水平及缴费激励系数，分析参保者经济理性假设下的参保行为选择倾向。为了方便进行对比分析，本文假定在有限预算约束下参保者面临两种模式选择：选择最低缴费年限 15 年，提高养老保险缴费基数；尽可能延长缴费年限，选择最低缴费档次。在选择最低缴费年限 15 年条件下，首先确定参保者缴费档次，进而明确养老金给付水平和缴费激励系数。缴费档次模型为：

$$w = \frac{PVFC \times (r - g)}{\theta \times (1 + r) \times \left[ 1 - \left( \frac{1 + g}{1 + r} \right)^n \right]} \quad (4)$$

其中， $w$  为缴费档次， $PVFC$  为参保者缴费现值， $r$  为贴现率， $g$  为社会平均工资增长率， $n$  为缴费年限， $\theta$  为养老保险总缴费率。

在选择最低缴费档次情况下，首先确定缴费预算对应的缴费年限，进而明确养老金给付水平和缴费激励系数。缴费年限模型为：

$$n = \log_{\frac{1+g}{1+r}} \left\{ 1 - \frac{PVFC \times (r-g)}{\theta \times w \times (1+r)} \right\} \quad (5)$$

本文对相同缴费预算条件下多缴多得与长缴多得对应的给付水平及缴费激励系数测算相关系数进行参数设定。(1) 社会平均工资设定为 1, 工资增长率为 5%, 贴现率为 3%; (2) 相比企业职工参保者, 灵活就业人员参保选择自主权更大, 本文重点分析灵活就业人员参保行为选择, 因此养老保险总缴费率设定为 20%; (3) 在上述参数条件下缴费预算, 即缴费现值大于 3, 才能够符合养老保险最低缴费年限 15 年等政策规定, 本文分析缴费现值分别为 4 和 5 的情况下, 养老保险缴费激励系数。

在缴费现值为 4 的情况下, 参保者选择最低缴费年限为 15 年, 缴费基数可以提高至社会平均工资 1.16 倍, 养老金给付现值为 5, 缴费激励系数为 1。<sup>①</sup>参保者选择最低缴费档次, 缴费年限可以延长至 26 年, 养老金给付现值为 9.54, 缴费激励系数为 5.54。在缴费现值为 5 的情况下, 参保者选择最低缴费年限为 15 年, 缴费基数可以提高至社会平均工资 1.45 倍, 养老金给付现值为 5.67, 缴费激励系数为 0.67。参保者选择最低缴费档次, 缴费年限可以延长至 31 年, 养老金给付现值为 12.26, 缴费激励系数为 7.26。

通过测算发现, 在相同缴费预算条件下, 参保者选择长缴、低缴的参保模式更加有利, 可以获得更高的养老金给付水平, 缴费激励政策对延长缴费年限的激励导向更强。

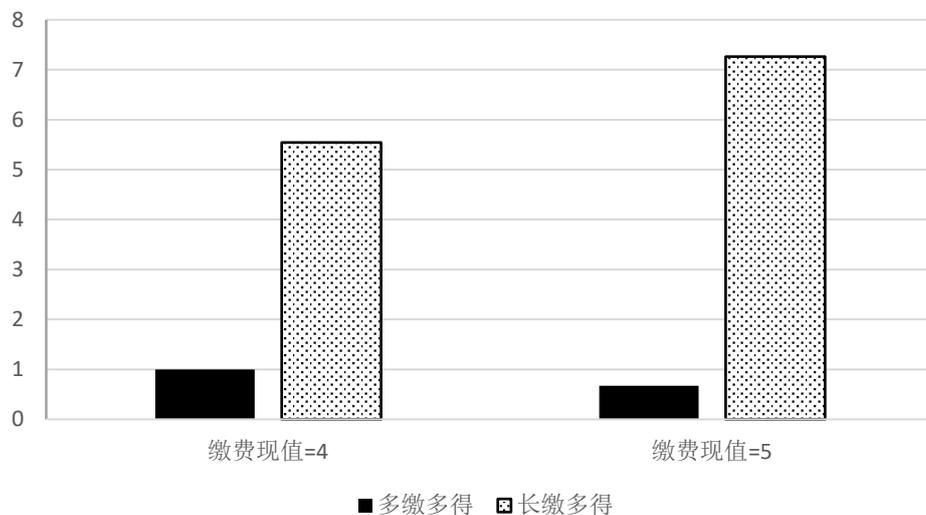


图 2 多缴多得与长缴多得的缴费激励效应比较

养老保险参保行为选择是参保者在收入预算既定情况下进行的长缴多得与多缴多得比较决策, 参保者在长缴多得与多缴多得之间进行选择。在预算约束下, 或者选择长缴费年限、低缴费基数的参保模式, 或者选择短缴费年限、高缴费基数的参保模式。<sup>②</sup>本文利用养老保

<sup>①</sup> 缴费激励系数=给付现值-缴费现值, 反映参保者养老保险净收益, 可以有效体现缴费激励程度。缴费激励系数是参保者经济理性情况下进行参保决策的核心依据。

<sup>②</sup> 参保行为选择还存在两种情形: 低收入参保者, 选择最低缴费年限与最低缴费基数, 政策激励难以改变预算约束对参保行为选择的影响; 高收入参保者, 既选择长缴费年限, 又选择高缴费基数, 参保决策不受预算制约, 政策激励效果不明显。本文重点研究政策激励对参保者缴费决策的影响, 因此主要比较“长缴、低缴”与“短缴、高缴”之间的差别。

险参保行为调查数据统计分析发现<sup>①</sup>，灵活就业人员参保人口中选择最低缴费档次（缴费基数为社会平均工资 60%）的比例达到 25%，缴费基数超过社会平均工资 100%的比例仅为 12%。在灵活就业人员参保样本中，倾向于在最低缴费年限标准基础上继续缴费的比例达到 78%，而缴费满 15 年之后不再继续缴费的比例约为 22%。通过数据可以发现，灵活就业人员参保缴费优先选择长缴多得，尽可能选择低缴费基数而延长缴费年限。

养老保险参保者的缴费行为现实选择与政策缴费激励导向具有一致性，政策激励更加向长缴多得倾斜，而参保者也更倾向于优先选择延长缴费年限。但通过参保行为选择、养老金给付水平与基金收支平衡的联动测算，参保缴费年限延长虽然能够提高养老金给付水平，但不利于实现养老金收支平衡，而提高缴费基数可以实现养老金给付水平提升与基金收支缺口缩减相联动的帕累托改进，需要重点激励参保者不断提高缴费基数，增强参保缴费积极性。

## （二）养老保险参保缴费档次决策影响因素识别

### 1. 随机森林模型

确定养老保险参保缴费档次决策的影响因素是建立合理的养老保险缴费档次激励机制的关键基础。参保者缴费档次选择不仅受经济约束、净收益激励等因素影响，也是参保者个体特征、职业特征等因素共同交织影响的结果。考虑到参保缴费档次与相关因素可能存在非线性关联，本文采用随机森林模型对灵活就业人员养老保险缴费档次决策的影响因素进行识别检验。随机森林模型具有较好的抗噪声能力，可以更好解决指标之间自相关等问题，处理高维度数据。

随机森林模型是将决策树作为基学习器，由若干基学习器组成的集成学习器，主要用于分类和回归分析。随机森林模型以诸多决策树分类结果为基础，按照多数投票原则确定出现最多的分类结果作为最终分类。作为基学习器的决策树模型可以通过信息熵和信息增益的过程进行理解。在多元分类数据集中，信息熵的确定模型为：

$$Ent(B) = -\sum_{k=1}^n x_k \log_2 x_k \quad (6)$$

其中， $Ent(B)$ 为信息熵， $x_k$ 为第 $k$ 类样本占全部数据集的比重，当 $k$ 越大时，信息熵越大。信息熵反映了信息量的大小，如在封闭箱子放入红球和白球各 1 个，随机抽取 1 个，抽取到红球和白球的概率分别为 1/2，确定的信息熵数值为 1，如果箱子中放入两个红色的球，随机抽取 1 个，由于抽取到红球的概率为 1，属于确定事件，信息量就为 0，确定的信息熵数值为 0。在确定数据集的信息熵之后，可以利用信息增益方式划分数据集。利用数据集的不同特征进行类别划分，在划分类别之后，测算信息熵在划分类别前后的差值，差值越大（信息增益）说明分类效果越好。信息增益的模型可以表达为：

<sup>①</sup> 养老保险参保行为调查数据是国家自然科学基金重点项目“新三支柱‘橄榄型’可持续发展养老制度设计与优化”课题组选择全国 21 个典型地区进行养老保险参保状况调研而得。

$$Gain(B, \alpha) = Ent(B) - \sum_{j=1}^{\lambda} \frac{|B^j|}{|B|} Ent(B^j) \quad (7)$$

其中， $Gain(B, \alpha)$ 为信息增益程度，特征 $\alpha$ 具有 $\lambda$ 个取值，利用特征 $\alpha$ 进行数据集分类，会产生 $\lambda$ 个分支点，在第 $j$ 个分支点包含了数据集 $B$ 中所有特征 $\alpha$ 取值为 $\alpha^j$ ，对应的样本为 $B^j$ ，对 $\lambda$ 个分支点赋予权重为 $|B^j|/|B|$ ，信息增益程度越大，说明分类确定性提高越多。

随机森林是包含了多个决策树的集成学习工具，随机森林对数据样本采用自主采样法（bootstrap sampling），即对数据集中的样本采用“放回式”的抽样方法，假设对数据集进行 $N$ 次放回式抽样，得到数量为 $N$ 的新样本，在抽样中，样本 $x$ 被抽到的概率为 $1/N$ ，未被抽到的概率为 $(1-1/N)$ ，在经过 $D$ 轮抽样之后，样本 $x$ 仍未被抽取到的概率为：

$$\lim_{D \rightarrow \infty} (1 - \frac{1}{N})^D \approx \frac{1}{e} \approx 0.368 \quad (8)$$

也就是说在 $D$ 轮抽样之后，未被抽取到的样本大约占36.8%。随机森林模型在每次迭代中利用自主抽样法生成的训练集对模型进行训练，在 $T$ 轮迭代之后会形成 $T$ 个基学习器，利用36.2%未被抽取到的样本对基学习器进行泛化误差的“外包估计”，最终通过 $T$ 个基学习器的分类结果进行简单投票形成数据集分类结果。

## 2. 参保行为决策影响因素识别

本文利用养老保险参保行为调查数据，选取以灵活就业人员身份参加养老保险的对象为样本，剔除数据缺失样本，样本总量为564个。本文从参保者个体特征、经济特征、参保倾向、社会特征等角度选择影响参保档次选择的相关指标。在个体特征层面，选择年龄、婚姻、职业等指标；在经济特征层面，选择家庭收入、消费等指标；在参保倾向层面选择参保年限、政策认知等指标；在社会特征层面，选择家庭赡养负担等指标。

表3 参保缴费档次影响因素指标体系

标签/特征序号	指标名称	指标内涵
标签	缴费档次	缴费基数 $\geq$ 社会平均工资 100%设定为 1 缴费基数 $<$ 社会平均工资 100%设定为 0
特征 1	参保者年龄	参保者年龄对缴费年限、领取养老金等待期等因素有关，会影响参保决策
特征 2	婚姻状况	家庭基本情况有可能会影响参保选择
特征 3	参保者职业	不同职业具有就业稳定性等方面的差异，会影响参保选择
特征 4	家庭赡养社会分担	父母参加社会养老保险情况
特征 5	家庭赡养市场分担	父母参加商业养老保险情况
特征 6	缴费连续性	是否出现过缴费中断

特征 7	参保年限	可以判断是否随着缴费年限增加而出现缴费偏好变化
特征 8	主动缴费程度	参保缴费的自愿性
特征 9	个人储蓄养老	是否有个人储蓄养老计划
特征 10	养老安排倾向性	在个人储蓄、子女养老、商业养老等选择中，国家养老倾向性的排序
特征 11	参保者性别	性别对应不同的退休年龄，有可能会影响缴费年限选择
特征 12	参保者学历情况	学历对经济理性等方面可能具有影响
特征 13	参保者家庭收入	收入水平决定了参保的经济约束条件
特征 14	参保者家庭债务	是否存在家庭欠债
特征 15	政策信任程度	信任基本养老保险政策的程度
特征 16	政策认知程度	对缴费政策了解程度
特征 17	政策信心	是否担忧未来 60 岁退休时政策变化而无法领取养老金
特征 18	就业稳定性	是否有可能去外地工作
特征 19	可支配收入	工资是否够日常花销
特征 20	收入变化预期	未来收入是否会有大幅提升
特征 21	消费变化预期	未来消费是否会有大幅提升
特征 22	自身健康评估	自身健康状况评价
特征 23	家人健康评估	家人健康状况评价

本文将数据随机划分为训练集和测试集，利用训练集进行模型训练，模型准确度达到约 86%，且在决策树数量为 60 左右准确度最高。因此，本文确定最终包含 60 个决策树的随机森林模型。

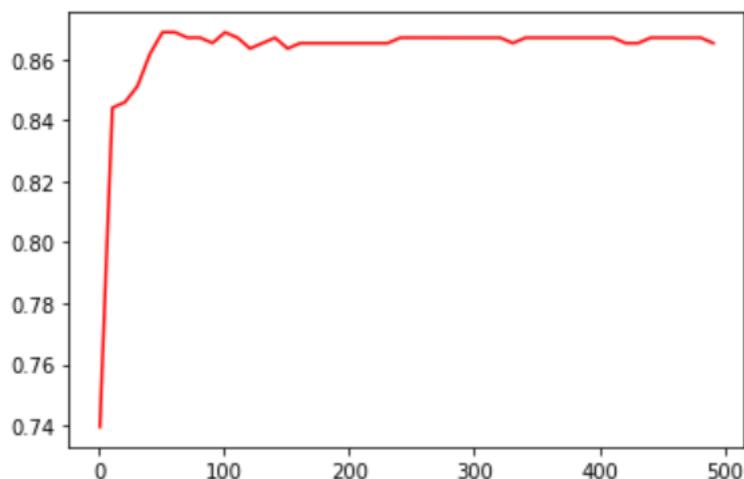


图 3 随机森林模型准确度

利用随机森林模型可以得到不同特征值的重要性，确定影响参保档次选择的重要指标。通过影响因素识别发现：（1）参保者缴费连续性是影响缴费档次选择的关键因素，参保者缴费保持连贯，会更加倾向于提高缴费档次；（2）参保者对政策信任程度和消费变化预期是影响缴费档次选择的重要指标，参保者对养老保险政策信任程度越高，越倾向于提高缴费档次，参保者预期未来消费平稳，则会提高缴费档次；（3）在参保者个体特征中，学历和职业是影响参保选择的显著因素，参保者学历越高，收入和政策认知程度越高，越有可能选择多缴多得；（4）参保者父母是否缴纳社会保险也会影响参保者行为选择，参保者父母已经参加社会保险会降低参保者赡养压力，提高参保者缴费预算，从而激励其提高缴费档次。（见图 4）

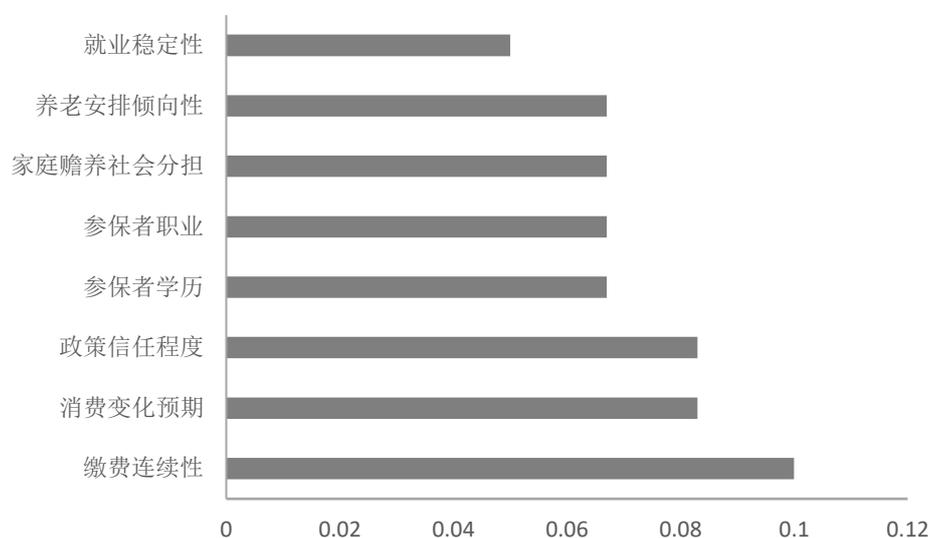


图 4 参保缴费档次影响因素重要性

## 五、结论及对策

本文构建城镇职工养老金社会平均工资替代率可能性边界模型，分析参保行为选择对养老金社会平均工资替代率的影响，以及同时实现提高养老金保障水平与改善养老保险收支平衡状况的可能性。构建养老保险缴费激励系数模型，判断现行缴费激励机制的多缴多得与长缴多得倾向性，分析参保者现实选择与缴费激励导向的匹配性及存在的问题，进一步利用机器学习随机森林模型识别参保者缴费行为选择影响因素。研究发现：（1）养老保险缴费最低标准对应养老金社会平均工资替代率边界下限，养老保险缴费上限标准对应养老金社会平均工资替代率上限，2021 年城镇职工养老金社会平均工资替代率可能性边界为 13.46%-51.43%，2050 年替代率可能性边界为 12.76%-51.75%；（2）参保行为选择对养老金社会平均工资替代率具有直接影响，在参保缴费基准情景下，养老金社会平均工资替代率区间为 34.88%-35.60%，人口老龄化导致养老金给付平均水平下降约 0.8 个百分点；（3）延长缴费

年限和降低参保起始年龄能够提高养老金替代率,但是会增加养老保险基金收支缺口。提高缴费基数既可以提高保障水平,又可以促进养老保险制度可持续发展,激励参保者提高缴费基数可以实现提高保障水平与缩减养老保险收支缺口的帕累托改进;(4)在相同缴费预算情况下,多缴多得缴费激励系数为1,长缴多得缴费激励系数为5.54,政策激励导向长缴多得。参保者现实选择更加倾向于延长缴费年限,参保者缴费选择与政策激励导向相一致;(5)多缴多得更加有利于提高保障水平与促进收支平衡,利用随机森林模型对参保者缴费档次选择的影响因素进行识别,参保者对政策信任程度等指标显著影响缴费档次选择。

根据上述研究结论,本文提出如下对策建议:(1)加强养老保险政策宣传,积极传递养老保险战略储备基金、全国统筹等政策信号,增强参保者对养老保险政策的信任程度,激励参保者提高缴费基数,实现多缴多得,促进养老保险收支平衡;(2)增强多缴多得缴费激励机制,在现行城镇职工养老金调整指数政策中,养老金调整与缴费年限挂钩,缺少与缴费档次挂钩。因此,在养老金调整指数设定过程中,增加养老金调整与缴费基数挂钩机制。

(3)促进参保者缴费连续稳定,健全养老金异地结转机制,逐步实现养老保险全国统筹,减少参保者就业变动对参保决策的抑制性影响。

#### 参考文献:

1. 穆怀中,范璐璐,陈曦.人口预期寿命延长,养老金保障风险与政策回应[J].人口研究,2021(1).
2. 杨一心,何文炯.养老保险缴费年限增加能够有效改善基金状况吗?——基于现行制度的代际赡养和同代自养之精算分析[J].人口研究,2016,40(3).
3. 曾益,魏晨雪,李晓琳,杨思琦.征收体制改革,延迟退休年龄与养老保险基金可持续性——基于“减税降费”背景的实证研究[J].公共管理学报,2019,16(4).
4. 邓大松,仙蜜花.延长退休年龄对基本养老保险统筹基金收支平衡的影响研究[J].江西财经大学学报,2015(5).
5. 赵亮,李灯强.我国城镇职工基本养老保险基金可持续性研究——基于新冠肺炎疫情冲击的影响[J].财经科学,2020(12).
6. 张国英,林伟垌,孙中伟.零工经济对城镇职工基本养老保险的冲击——基于对基金收支平衡的模拟计算[J].安徽师范大学学报:人文社会科学版,2022,50(1).
7. 郑秉文.供给侧:降费对社会保险结构性改革的意义[J].中国人口科学,2016(3).
8. 曾益,姚金.降低养老保险缴费率:“一举两得”抑或“得不偿失”?——基于缴费遵从度与基金可持续的视角[J].南方经济,2022(1).
9. 张川川,朱涵宇.新型农村社会养老保险参与决策中的同群效应[J].金融研究,2021(9).
10. 齐红倩,杨燕.人口老龄化,养老保障水平与我国养老保险基金结余[J].南京社会科学,2020(8).
11. 景鹏,王媛媛,胡秋明.国有资本划转养老保险基金能否破解降费率“不可能三角”[J].财政研究,2020(2).

# Choice of Contribution Behavior, Pension Replacement Rate Possibility Boundaries, and Balance of Payments

Chen Xi Li Jiayu Fan Lulu

**Abstract:** The construction of a moderately protected and sustainable pension insurance system needs to solve the paradox of policy choice between increasing the level of pension protection and promoting the sustainable development of the pension insurance system. The choice of participants' contribution behavior is a cross-cutting element that affects the level of pension benefits and the balance between income and expenditure. This paper takes the participants' choice of contribution behavior as the entry point to analyze the possibility of Pareto improvement between increasing the level of pension protection and improving the balance of payments. It is found that (1) the lower bound of the possibility of social average wage replacement rate for pension for urban workers is 12.76%-13.46% and the upper bound is 51.43%-51.75%. (2) The choice of participation behavior has a direct impact on social average wage replacement rate for pension. The social average wage replacement rate for pension ranges from 34.88% to 35.60% under the baseline scenario of insurance contributions, and the aging population leads to a decrease of about 0.8 percentage points in the average level of pension benefits. (3) Extending the number of years of contribution and lowering the starting age of participation can increase the pension replacement rate, but it will increase the shortfall of pension fund. Increasing the contribution base can achieve the Pareto improvement of increasing the protection level and reducing the pension fund gap. (4) The incentive of urban workers' pension insurance policy is to pay more for longer contributions, and the participants' realistic choice is more inclined to extend the contribution period. The participants' contribution choice is consistent with the policy incentive orientation, and the contribution incentive mechanism needs to be optimized. (5) Using the machine learning random forest model to identify the factors influencing the participants' choice of contribution bracket, the participants' trust in the policy and other indicators significantly affect the choice of contribution bracket.

**Keywords:** Choice of Participant Contribution Behavior; Social Average Wage Replacement Rate for Pension; Pareto Improvement; Random Forest Model.