

中国养老金改革的逻辑和福利效果: 基于人口“数量-质量”转换的视角^{*}

张 熠 张书博 汪润泉

内容提要: 本文通过世代交叠模型, 分析了我国人口结构沿着“数量-质量”前沿边界转换, 即少子化伴随人力资本进步时, 养老金体系最优的代际再分配水平。分析表明: (1) 如果仅面临少子化, 我国养老金体系的确应转向积累制, 但快速的人力资本进步完全可以补偿少子化的冲击, 使现收现付制仍是最优选择。如果考虑到代际公平性, 甚至政府需要加强现收现付制, 使老年世代分享经济发展的成果。(2) 基于“数量-质量”前沿边界转换的模型内生出我国较高的养老保险费率、退休前工资替代率和制度内赡养率, 较低的社会平均工资替代率和人口赡养率等许多令人困惑的典型事实。(3) 在此理论框架下, 无论不做实个人账户, 还是加强缴费受益关联、延迟退休以及建立“养老金-工资挂钩机制”都有利于增进社会福利。看似矛盾的中国养老金改革实践或许具有逻辑自洽性。

关键词: 人力资本 少子化 现收现付制 代际公平 养老金改革

一、引 言

现代社会中, 现收现付社会养老保险体系扮演了资源代际转移支付以维持代际公平的功能。但这种制度的可持续性依赖于下一代的赡养能力, 包括下一代劳动力的数量增长和质量提升。我国正在经历不断加速的少子化和老龄化趋势, 这必然削弱下一代向父母一代转移支付的能力。根据历次人口普查数据, 我国的人口出生率已从 1987 年的 23.33‰, 降至 2015 年的 12.07‰。城市 50 岁以上人口与 20—49 岁人口的比例则从 1990 年的 1:3 上升到了 2015 年的 1:1.7。然而, 从另一方面看, 少子化伴随着快速的人力资本进步, 也就是说, 尽管下一代人口数量减少, 但质量有显著的提高。根据《中国人力资本报告》的数据, 2015 年我国城镇人均人力资本为 1985 年的 4.7 倍, 年增长率达 5.3%。Whalley & Zhao (2013) 发现 1978—2008 年间, 中国人力资本年增长率高达 7.59%。因此, 我国的人口结构并不是简单的少子化、老龄化, 而是在沿着一条人口“数量-质量”前沿边界变动, 从一个高生育率、低人力资本积累的经济体, 变为了低生育率、高人力资本积累的经济体。在这一转换过程中, 或许下一代的总赡养能力并没有因生育率下降而降低。所谓“数量-质量”前沿边界是指家庭和社会所面临的在人口数量和质量之间转换的边界。这种转换既是生育政策“控制人口数量、提高人口素质”的目标, 也符合人口经济学中的“数量-质量权衡”理论。伴随着快速的人力资本进步, 经过一代人之后, 收入水平显著提高, 为了让老年人分享经济发展成果, 具有代际再分配功能的现收现付养老金体系或许不仅不该抛弃, 甚至应当得到加强。

^{*} 张熠、张书博, 上海财经大学公共经济与管理学院, 邮政编码: 200433, 电子信箱: zhang_yi@mail.shufe.edu.cn, zhangshuboshufe@163.com; 汪润泉, 南京财经大学公共管理学院, 邮政编码: 210023, 电子信箱: wrq8910@163.com。作者感谢国家自然科学基金项目(71503154)资助, 感谢匿名审稿专家的宝贵意见, 文责自负。

根据上述现实观察,本文提出三个层次递进的研究问题:其一,当经济体沿着“数量-质量”前沿边界发生人口数量质量转换时,从社会福利最大化的角度看,养老金体系最优的现收现付水平或代际再分配水平应当如何变动?人口结构质量的提高可以弥补数量的不足,从而使得现收现付养老金体系仍然是社会最优的选择吗?如果是这样,那么我国养老金改革没有完全遵循从现收现付制转向积累制的预定改革路径或许并不意味着改革失败,而是可理解的。其二,当经济体人口结构从数量转向质量时,最优养老金体系中各项参数,如缴费率、退休前工资替代率、平均工资替代率、人口赡养率、制度内赡养率等应当如何变动,这些参数的变动与我国现实情形一致吗?如果答案是肯定的,就意味着人口结构从数量到质量的转换很可能深刻影响着我国养老金体系的各项表现,有助于深化对于这一体系的理解。不能否认当前社会保障体系仍有不少不足之处,但许多看似糟糕的表现或许至少部分原因是来自人力资本进步等真实的冲击,而不是制度设计或执行中的错误导致的。其三,当经济体沿着“数量-质量”前沿边界转换时,最优养老金体系的各项制度参数应当如何动态调整?过去二十多年来我国养老金改革实践与这些调整一致吗?我国已经实施的和正在酝酿的各项养老金改革是福利增进的吗?如果这些问题的答案也是肯定的,那么中国养老金改革所选择的路径就是可理解的和具有内在逻辑自洽性的。

上述三个层面的问题都具有重要的现实意义。就第一个层面的问题而言,长期以来,现收现付制和积累制的选择始终是困扰中国社会保障学术界和改革者的难题,也是引起争论的热点问题。从20世纪90年代开始,如世界银行和一些学者(World Bank, 1994; Feldstein, 1999; Estelle, 2002)所倡议的,我国选择了从现收现付制转向积累制的养老金系统性改革(systematic reform)道路。系统性改革是反政府干预主张,一直被认为具有鲜明的新自由主义特征(Williamson & Deitelbaum, 2005)。按照当初设想的方式(国务院发展研究中心社会保障课题组, 2000; 孙祁祥, 2001),我国养老保险个人账户应当与社会统筹账户分离,并成为积累制“实账”运行。2005年《国务院关于完善企业职工基本养老保险制度的决定》(以下简称“38号文件”)也提出“做实个人账户”的目标。然而2006年以来的改革实践中,个人账户始终“空账”运行,也没能与社会统筹账户分离,并且在个人账户没有做实的情况下,政府反而将资金用于连续多年提高养老金待遇,实际上是在“向左走”,即加强了现收现付制和代际转移支付力度。如果从系统性改革理论出发,改革无疑是滞后或失败的。然而如果上述第一个问题的回答是肯定的,那么从快速人力资本进步的角度看,这种现实选择则是可理解的。

就第一个层面来说, Song et al. (2015) 与本文有着相似的研究出发点,都着眼于高增长经济体中的代际转移支付问题。Song et al. (2015) 通过数值模拟认为对于中国这样的高增长经济体来说,未来年轻世代较当前老年世代更富有,因此有较强的代际转移需求,所以机械地将成熟经济体系统改革经验照搬到发展经济体中是不合适的。然而,本文在研究思路、模型、传导机制、政策实验和研究范围等方面都与之截然不同。主要区别在:第一,本文带有鲜明的政治经济学特点,即代际人力资本进步这种生产力变化是如何影响社会保障等上层建筑的,着眼于使社会福利最大化的最优养老金体系分析。第二,模型上,本文建立在 Diamond (1965)、Rey et. al (2013) 等的稳态和平衡增长路径福利分析基础上,得出了许多定性的结论,有助于理解代际转移支付影响我国养老金体系和改革的机制,而不侧重于转移动态数值模拟。第三,传导机制和政策实验上的差异。本文更强调机制传导过程,且能分别刻画少子化、人力资本进步、代际公平性和道德风险各自的作用,并为众多养老金改革建立了一个广泛的政策实验框架。第四,研究范围上的差异。Song et al. (2015) 的研究仅涉及高增长经济体的代际转移支付问题,而没有涉及本文所说的第二和第三层面的问题。

就第二个层面来说,我国养老保险缴费率偏高,在人口赡养率相对较低,平均工资替代率水平不高时,却面临着严重的收支压力,许多省份陆续出现征缴收入不抵待遇支出的情况。另一方面,养老金退休前工资替代率和制度内赡养率又明显偏高,前者容易诱发提前退休等道德风险,后者直接影响基金收支平衡。这些典型特征是如何形成的? 本文从人力资本进步和代际公平性视角,构建了一个统一的解释性理论,有助于重新认识和思考我国养老金体系的表现。在这一点上,本文与 Fang & Zhang(2018) 的主题较为接近。Fang & Zhang(2018) 总结了我国养老金体系的若干典型事实,并且通过公平工资理论和人力资本进步为这些典型事实提供了统一的解释框架,首次指出我国社会保障体系中种种令人困惑的表现是一种“成长的烦恼”,而不是人口老龄化或制度设计缺陷导致的。本文与 Fang & Zhang(2018) 的研究主要有两大区别: 第一,核心机制不同。本文是从人口“数量-质量”前沿边界转换角度出发的,而 Fang & Zhang(2018) 的研究则是从企业公平工资理论出发的。第二, Fang & Zhang(2018) 没有探讨前文提到的中国养老金改革实践中矛盾和令人困惑的现象即第三个层面的问题,而这恰是本文的重点。

事实上,第三个层面的问题最值得深思,却在既往文献中未得到重视。我国养老金改革实践存在看似矛盾的现象,既包括前文提及的在改革方向上“向右看、向左走”,也表现在再分配工具选择方面“舍代内、取代际”,在道德风险管理方面“加油门、踩刹车”的特点上。从 2005 年起,在大幅提高退休者养老金待遇的同时,政府削弱了养老金体系的代内再分配力度。例如,38 号文件提高了养老金待遇和缴费之间的关联程度,更强调多缴多得,长缴多得,并且近年来开始酝酿延迟退休年龄改革和名义账户制改革。这种在再分配工具选择中“舍代内、取代际”的做法值得探讨。而且,加强缴费受益关联与提高养老金待遇也存在矛盾,前者意在“刹”住道德风险,产生对继续工作和缴费的激励性,后者则会助长道德风险,产生“引致退休效应”(Feldstein, 1974; Samwick, 1998)。因此,一个非常关键的问题是,看起来充满矛盾的重大改革实践的社会福利效果如何呢? 它们是否具有自洽的内在理论逻辑呢?

本文的主要贡献在于: 从人口结构沿“数量-质量”前沿边界转换的视角,说明我国不仅存在少子化和老龄化,还伴随着快速的人力资本进步,后者造成两代人力资本巨大的差距,此时以实现社会福利最大化,代际转移支付为主要特征的现收现付制应当加强而非削弱; 模型内生出高缴费率、低平均工资替代率、高退休前工资替代率、高制度内赡养率等许多我国社会保障体系的典型事实,这些看似糟糕的制度参数是否都是制度设计和执行失误所导致,抑或由可能是由人力资本进步等真实的生产力因素所引起; 在本文理论模型框架下,无论放弃系统性改革,还是提高养老金待遇、加强缴费受益关联以及延迟退休年龄,都是增进社会福利的。这促使我们重新思考我国养老金改革的路径,看似失败和矛盾的实践是否可能是基于现实的自然选择? 既往养老金改革研究中是否政府失灵因素被夸大,而忽略了对基础性的生产力因素的分析?

本文接下来的安排如下: 第二节建立世代交叠经济模型,模型中存在外生的法定退休年龄,在此框架下阐述了少子化与人力资本进步对养老金体系影响相互冲销的等价性原理。第三节拓展理论模型,先引入代际公平性因素,再引入内生的劳动者退休年龄选择,拟合我国养老金体系的主要特征。在第四节,通过实证数据校准理论模型,展开反事实分析,说明当经济体沿着“数量-质量”前沿边界转换时可以内生出我国养老金体系的典型事实,并分解少子化、人力资本进步、代际公平性以及道德风险等因素各自的作用。第五节检验当前重要的养老金改革的社会福利效果,并阐述改革实践背后的理论逻辑。第一项改革为转向积累制的系统性养老金改革,做实个人账户即属于此类改革; 第二项为养老金工资挂钩增长机制,包括 2005 年以来养老金随工资连续调整以及建立养老金正常调整机制都属于此类; 第三项为延迟法定退休年龄改革; 第四项为缴费受益关联改革,

例如“38 号文件”规定将养老金待遇和缴费年限、缴费工资挂钩以及名义账户制等都属于此类。最后,对全文进行总结。

二、基础模型

在人口数量减少但质量(人力资本)提高的情况下,从社会福利最大化的角度看,政府是否有必要改革现收现付制,转向积累制呢?相对于积累制,现收现付制最显著的特征是具有代际再分配的功能,即由当前年轻人的缴费负担退休者的养老金待遇。个人账户积累制不具有代际再分配功能,因此一般认为对宏观经济和社会福利的影响是中性的(Samuelson, 1975)。下面通过引入世代交叠模型来证明如下等价性原理:当人口数量减少而人力资本提升时,使社会福利最大化的最优养老金体系中,现收现付的成分或者说代际再分配力度可以保持不变。^①

(一) 模型框架

假设经济体在 t 期存在两代人:其一为 t 期出生的年轻世代,人数为 N_t ,其中第 i 个年轻人获得的人力资本禀赋为 $H_t^{i,i}$,年轻人无弹性地供给 1 单位劳动力;其二为 $t-1$ 期出生的世代,在 t 期成为老年人,人数为 $N_{t-1} = N_t / (1+n)$ 。这部分老年人如果工作则继续使用上一期形成的人力资本,第 j 个老年人的人力资本禀赋为 $H_t^{j,j} = \varphi H_{t-1}^{j,j} = \varphi H_{t-1}^{j,j} / (1+g)$ 。 $1+n$ 和 $1+g$ 分别为人口增长率和禀赋增长率, φ 为老年劳动者的经验溢价。假设采用线性生产技术(Shi & Zhang, 2009),全部产品为 $Y_t = \sum_{i=1}^{N_t} H_t^{i,i} + \sum_{j=1}^{N_{t-1}} \pi_t^j H_t^{j,j}$,其中 $\pi_t^j < 1$ 为 t 期第 j 个老年人的劳动参与率, $1+\pi_t^j$ 为行为人生工作的时长,而退休的时长 $1-\pi_t^j$,实际退休年龄为 $R^j = A + T \times (1+\pi_t^j)$,其中 A 表示参加工作年龄, T 为生命周期中每一期的长度。在基础模型以及部分拓展模型中,假设老年劳动参与率 π 以及退休年龄 R 都是由政府法定退休年龄政策外生决定的。

假设经济体中 t 期出生的第 i 个代表性行为人的终生效用函数为 $U_t^i = U^i(c_t^{y,i}, c_{t+1}^{o,i})$ 。 $c_t^{y,i}$ 和 $c_{t+1}^{o,i}$ 分别为年轻和年老时的消费。设效用函数是拟凹的 k 次齐次函数。众所周知,在无限期生存模型中,代表性个人的效用最大化和社会福利最大化是完全统一的,在带有增长的世代交叠模型中评价社会福利则面临一些困难。我们采用了 Rey et al. (2013) 提出的平衡增长路径方法,这种方法优势在于仅在宽松的序数效用假设下,即仅假设拟凹和齐次性,就可确保中央计划者社会福利函数和个人效用最大化的统一。社会福利水平取决于平衡增长路径上的消费率,是一个有限值。这种方法实际上是 Diamond(1965) 世代交叠模型中社会福利函数拓展到有增长情形时的版本。社会所面临的资源约束为:

$$\sum_{i=1}^{N_t} H_t^{i,i} + \sum_{j=1}^{N_{t-1}} \pi_t^j H_t^{j,j} = \sum_{i=1}^{N_t} c_t^{y,i} + \sum_{j=1}^{N_{t-1}} c_t^{o,j} \quad (1)$$

定义青年人消费率为 $\tilde{c}_t^{y,i} = c_t^{y,i} / H_t^{i,i}$,老年人消费率为 $\tilde{c}_t^{o,j} = c_t^{o,j} / H_t^{j,j}$ 。假设所有个体是同质的,即 $c_t^{y,i} = c_t^y, c_{t+1}^{o,i} = c_{t+1}^o, \tilde{c}_t^{y,i} = \tilde{c}_t^y, \tilde{c}_{t+1}^{o,i} = \tilde{c}_{t+1}^o, H_t^{i,i} = H_t^i$ 。在个体同质性假设下,对公式(1)两端除以全社会青年人的有效劳动 $H_t^i N_t$ 进行标准化,可整理为:

$$1 + \varphi \pi_t / [(1+g)(1+n)] = \tilde{c}_t^y + \tilde{c}_t^o / [(1+n)(1+g)] \quad (2)$$

上式左端为标准化后的社会总资源,右端则为年轻人和老年人可获得的消费率。在平衡增长路径上, $\tilde{c}_t^y$ 和 $\tilde{c}_t^o$ 均为常数,而人均消费 c_t^y 和 c_t^o 随着 H_t^i 按照 $1+g$ 的固定速度增长。

^① 设老年人退休前工资为 4000 元,如下两个经济体具有等价性:经济体 A 中 2 个年轻人赡养 1 个老年人,年轻人工资为 5000 元,缴费率为 20%,则每个年轻人缴费 1000 元,老年人获得养老金 2000 元,退休前工资替代率 $2000/4000 = 50\%$;经济体 B 中每个老年人只有 1 个年轻人赡养,年轻人工资 10000 元,缴费率仍为 20%,则年轻人缴费 2000 元,退休前工资替代率仍为 50%。经济体从 A 变为 B 时,缴费率、替代率等核心参数都无须调整。

(二) 最优资源配置

假设中央计划者面临的社会福利函数为:

$$\tilde{U}_t = U(\tilde{c}_t^y, \tilde{c}_{t+1}^o) = U(c_t^y/H_t^y, c_{t+1}^o/H_t^o) = (1/H_t^y)^k U(c_t^y, c_{t+1}^o) = (1/H_t^y)^k U_t \quad (3)$$

(3) 式中第三个等式使用了效用函数的齐次性假设。新的效用函数是通过原有效用函数正则变换得到的, 所以社会福利函数保留了个人效用函数同样的序数偏好。位于 $(\tilde{c}_t^y, \tilde{c}_{t+1}^o)$ 平面上的无差异曲线与 (c_t^y, c_{t+1}^o) 平面上的无差异曲线是一一对应的(Rey et al., 2013)。整个经济的最优资源配置定义如下:

定义1(最优资源配置): 最优资源配置是指, 在一条平衡增长路径上的消费率组合 $(\tilde{c}^y, \tilde{c}^o)$, 使得(3)式中稳态社会福利函数在(2)式列示的社会资源可行性约束下达到最大化, 其特征是满足公式(2)和公式(4)。

$$(\partial \tilde{U} / \partial \tilde{c}^y) / (\partial \tilde{U} / \partial \tilde{c}^o) = M(\tilde{c}_{t+1}^o / \tilde{c}_t^y) = (1+n)(1+g) \quad (4)$$

上述中央计划者的最优资源配置问题对应于宏观经济学著名的黄金律条件(golden rule)(Phelps, 1961)。 $M(\tilde{c}_{t+1}^o / \tilde{c}_t^y)$ 为两期消费边际替代率, 取决于两期消费率相对比值 $\tilde{c}_{t+1}^o / \tilde{c}_t^y$ 。由于边际替代率递减, 所以 $M' > 0$ 。Diamond(1965)提出的世代交叠模型黄金律条件 $(\partial \tilde{U} / \partial \tilde{c}^y) / (\partial \tilde{U} / \partial \tilde{c}^o) = 1+n$ 实际上是公式(4)在无增长($g=0$)时的特例。

(三) 竞争性经济的资源分配

下面讨论分散竞争性经济体如何通过现收现付养老金体系实现黄金律资源配置。经济体中代表性行为人生存两期, 年轻时期和年老时期的消费以及标准化的消费率分别为:

$$c_t^{y,i} = H_t^{y,i}(1-\tau) + W_t^{y,i} - S_t^i, \tilde{c}_t^{y,i} = (1-\tau) + \bar{w}_t^{y,i} - s_t^i \quad (5)$$

$$\begin{aligned} c_{t+1}^{o,i} &= \pi_{t+1}^i H_{t+1}^{o,i}(1-\tau) + (1-\pi_{t+1}^i) p_{t+1}^i + W_{t+1}^{o,i} + (1+r_t) S_t^i \\ \tilde{c}_{t+1}^{o,i} &= (1-\tau) \varphi \pi_{t+1}^i + (1-\pi_{t+1}^i) \delta_{t+1}^i + \bar{w}_{t+1}^{o,i} - (1+r_{t+1}) s_t^i \end{aligned} \quad (6)$$

S_t 为个人间的资金借贷, $S_t > 0$ 为贷出资金, $S_t < 0$ 为借入资金。上式中 τ 为现收现付养老保险缴费率, p_{t+1}^i 为退休后领取到的养老金, $\delta^i = p_{t+1}^i / H_{t+1}^i$ 代表退休前工资替代率。政府面临一个预算平衡关系:

$$\tau \sum_{i=1}^{N_{t+1}} H_{t+1}^{y,i} + \tau \sum_{i=1}^{N_t} H_{t+1}^{o,i} \pi_{t+1}^i = p_{t+1} \sum_{i=1}^{N_t} (1-\pi_{t+1}^i) \quad (7)$$

(7) 式左端为养老保险总缴费, 包括年轻人的缴费和老年人的缴费, 右端为老年人退休后的养老金, 老年人实际退休年龄为 $1+\pi$ 。根据效用函数和公式(5)——(7)求得一阶条件:

$$(\partial U / \partial c_t^y) / (\partial U / \partial c_{t+1}^o) = M(c_{t+1}^o / c_t^y) = 1+r_t \quad (8)$$

上式即跨期决策的欧拉方程, 函数 $M(c_{t+1}^o / c_t^y)$ 仍为边际替代率。齐次性假设保证了公式(8)中的边际替代率与公式(4)中是一一对应的。由于模型是一个同质性禀赋经济, 在市场均衡时, 同代人储蓄之和应当为零, 金融市场出清: $\sum_{i=1}^{N_t} S_t^i = 0$ 。那么最优的养老金体系应当是怎样的呢? 最优的现收现付体系应当实现社会福利的最大化, 其具体定义为:

定义2(最优养老金体系): 最优养老金体系是指, 存在一个现收现付养老保险费率、退休前工资替代率组合 (τ^*, δ^*) , 使得竞争分散性经济中的平衡增长路径 $(\tilde{c}^y, \tilde{c}^o)$, 满足公式(2)和公式(4)中最优资源配置条件, 则称社会保障体系是最优的。

瓦尔拉斯法则表明, 金融市场出清可以保证公式(2)中产品市场出清。对比公式(4)和公式(8)可以发现, 当满足 $1+r_t = (1+n)(1+g)$ 时, 分散竞争性经济可实现资源最优配置, 即利率等于人口增长率和工资增长率之和。这一条件与著名的艾隆条件(Aaron, 1966)存在紧密关联。在艾隆条件下, $1+r < (1+n)(1+g)$, 这意味着 $\tilde{c}^o / \tilde{c}^y$ 小于黄金律时的水平, 此时通过具有代际转移支付功能的现收现付制, 老年人消费提高, 年轻人消费下降, 利率回升, $(\tilde{c}^y, \tilde{c}^o)$ 重新回到黄金律水平

的平衡增长路径,带来社会福利提升。由于行为人是同质的,从公式(7)可得:

$$\delta_{t+1} = p_{t+1}/H'_{t+1} = \tau [(1+n)(1+g) + \varphi\pi_{t+1}]/[\varphi(1-\pi_{t+1})] \quad (9)$$

令 $\chi = (1+n)(1+g)$, 并定义为社会总代际赡养能力。需要注意的是,公式(5) — (9) 构成的分散经济框架下,人口增长率 $1+n$ 和人力资本进步 $1+g$ 始终同时出现,两者的效应是完全等价的。模型中最优资源分配状态和最优养老金制度都仅取决于总代际赡养能力 χ , 而与人口增长率或人力资本进步单个因素无关,即存在如下定理:

定理 1(数量-质量等价性): 在公式(5) — (9) 所描述的分散竞争性经济环境中,如果存在两个经济体: 第一个经济体生育率 $1+n$ 高; 第二个经济体代际人力资本进步速度快, 即 $1+g$ 高。如果两个经济体具有相同的总代际赡养系数 $\chi = (1+n)(1+g)$, 则它们具有相同的最优社会福利水平 $\tilde{U}$ 、最优消费组合 $\{\tilde{c}^y, \tilde{c}^o\}$ 、最优现收现付社会保障缴费率 τ^* , 以及最优退休前工资替代率 δ^* 。

定理 1 意味着当一个经济体从高生育率、低人力资本进步的状态转向低生育率、高人力资本进步时,原先的现收现付养老金体系仍然使稳态社会福利最大化,可以保证经济处于最优黄金律时的平衡增长路径,无须进行调整和改革。消费路径和社会福利水平也没有变化。此时,人口质量的提升,即人力资本进步,完全可以补偿少子化。在面临人口老龄化时,现收现付制的财务可持续性常被诟病,但如果社会投入巨大的资源维持现收现付体系的运转,总是可以保证其财务可持续性的。一个更重要的概念是社会可持续性 (social sustainability) (Grech, 2015), 即养老金体系持续达成社会目标的能力。定理 1 证明了现收现付制的社会可持续性。即便面临少子化,只要有足够的人力资本进步,现收现付制仍然可以是实现社会福利最大化的最优选择。

(四) 平均工资替代率

尽管养老金缴费率、退休前工资替代率、年轻人和老年人消费路径以及社会福利水平对经济沿着“数量-质量前沿边界”变动免疫,但是养老金平均工资替代率却会发生变动。养老金平均工资替代率是由退休者养老金待遇除以社会平均工资得到的。养老金平均工资替代率为:

$$\bar{\delta}^* = p_{t+1}/\bar{H}_{t+1} = \tau^* [(1+n) + \pi]/(1-\pi) \quad (10)$$

其中, $\bar{H}_{t+1} = [H'_{t+1}N_{t+1} + \pi_{t+1}H'_tN_t]/(N_{t+1} + \pi_{t+1}N_t)$, 代表社会平均工资。根据公式(10), $\partial\bar{\delta}^*/\partial(1+n)|_{\chi=\bar{\chi}} > 0$, 也就是说即便社会总赡养能力 $\chi = (1+n)(1+g)$ 不变,养老金平均工资替代率仍将随着少子化而下降。根据定义, $\bar{\delta}^* = p_{t+1}/H'_{t+1} \cdot H'_t/\bar{H}_{t+1} \approx \delta \cdot \varphi/(1+g)$ 。第三个等式近似成立是年轻劳动者人数远高于老年劳动力,故平均工资接近于年轻人工资。所以,平均工资替代率下降的原因是,老年人的养老金待遇虽然相对自身退休前工资保持不变(δ 不变),但人力资本进步越快,这一养老金待遇与年轻人工资相比就越来越少。

平均工资替代率下降是我国近年来养老金体系中存在的典型事实。根据历年《中国统计年鉴》数据,我国企业职工养老金平均替代率从 2000 年的 80% 左右下降到目前的约 50%。这与模型推导结果是吻合的。同时,平均工资替代率和退休前工资替代率背离的核心环节是年轻人相对老年人工资增长更快,或者说 $\varphi/(1+g)$ 下降。利用 CFPS 数据,我们也观察到这一事实。由于工作经验产生的溢价,老年人薪酬通常应当高于年轻人。Lagakos et al. (2018) 研究了 18 个国家的“经验-工资”分布,结果显示多数国家工资峰值出现在工作年限 25—35 年的劳动者群体。但从 CFPS2016 年的数据来看,中国年轻群体的薪酬约为临近退休老年人的 1.36 倍,说明存在促使养老金平均工资替代率与退休前工资替代率背离的中间机制。

在基础模型中,当经济体同时存在少子化和人力资本进步而社会总赡养能力不变时,尽管缴费

率、退休前工资替代率等参数保持不变。但随着人力资本进步, 年轻人工资增长, 如果老年人养老金没有增长, 平均工资替代率将持续下降。平均工资替代率是衡量老年人在社会中相对经济地位的重要指标, 漠视平均替代率的下降可能导致严重的代际公平性问题, 社会必须允许老年人分享经济发展的成果。在现实政策中, 我国社会养老保险体系也考虑了代际公平性问题: 第一, 从养老金计发公式看, 代表性退休者养老金与退休前一年的社会平均工资挂钩, 这使得老年人的养老金与当前经济发展和工资水平而非职业生涯平均工资水平相关联; 第二, 尽管我国养老金待遇调整还未建立一个正常的调节机制, 但 2005 年以来政府每年的养老金待遇调整都会考虑社会平均工资的增长, 使退休者分享经济发展成果。如图 1 所示, 2005 年以前, 平均养老金待遇增长率与平均工资增长率之间呈反向变动, 但 2005 年以来, 两者挂钩程度大幅提高。在下一节, 我们将考虑在基础模型中加入代际公平性后对最优养老金体系的影响。

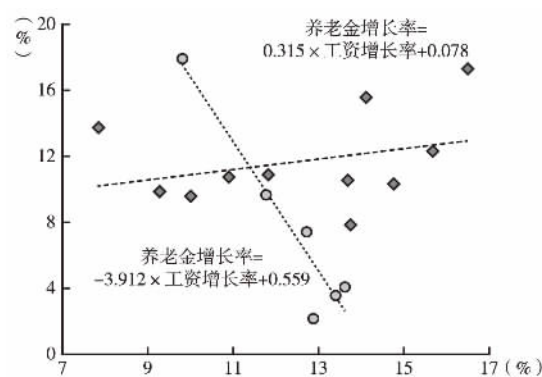


图 1 我国养老金与工资增长率挂钩情况

三、拓展模型

本节对基础模型做两方面的拓展: 首先通过内生代际公平性需求, 探讨当经济体沿着人口“数量-质量”前沿边界转换时, 人们对代际公平性的追求对社会保障体系的影响; 其次, 基础模型中假设退休年龄是由政府强制退休政策外生决定的, 现实中即便存在法定退休制度, 劳动者仍然会根据健康、人力资本等因素选择提前退休、在法定退休年龄退休以及退而不休。事实上我国提前退休的比例相当高, 男性在 60 岁之前退休的约占 54%, 女性在 50 岁之前退休的约占 30% (封进和胡岩, 2008), 退而不休的比例也很高 (程杰, 2014)。另外, 从政府角度看, 激励劳动者继续工作以及纠正道德风险问题也一直是我国养老保险制度设计中考虑的重要方面, 38 号文加强了待遇与缴费年限、退休年龄的挂钩, 更强调长缴长得, 将模型拓展到内生退休情形有助于理解这些道德风险现象、问题和政策改革动因。

(一) 内生代际公平性

通过行为经济学中的参照依赖理论, 本文在保持个人效用和社会福利对应关系的前提下内生地引入老年人对代际公平性的需求。假设行为人的效用函数和与之对应的社会福利函数为:

$$U_t^i = U^i(c_t^{y,i}, c_{t+1}^{o,i} - \lambda X_{t+1}), \tilde{U}_t = U(\tilde{c}_t^y, \hat{c}_{t+1}^o) = U(\tilde{c}_t^y, \tilde{c}_{t+1}^o - \lambda \tilde{X}_{t+1}) \quad (11)$$

代际公平性意味着老年人会以当前时点社会平均消费为参考, 由于年轻人占比较高, 不失一般性, 设 $X_{t+1} = c_{t+1}^y$, $\tilde{X}_{t+1} = X_{t+1}/H_t^y = \tilde{c}_{t+1}^y$, 这时当年轻人人力资本进步和工资增长, 老年人有分享经济发展成果的需求。 $0 < \lambda < 1$, 为参考偏好的强度, 在 $\lambda = 0$ 时, 公式 (11) 中的效用函数恢复为基础模型的形式。仍然假设效用函数具有齐次性。可以证明如下定理:

定理 2(代际公平性): 公式(5) — (7)、公式(11) 构成的分散竞争性经济要实现最优资源配置需满足的条件为: $M(\hat{c}^o/\hat{c}^y) = 1 + r = (1 + n)(1 + g) + \lambda(1 + g)$ 。设总赡养系数 $\chi = (1 + n)(1 + g)$ 不变时存在少子化和人力资本进步, 即 $d(1 + n) < 0, d(1 + g) > 0$, 且假设社会存在对代际公平性的需求 ($\lambda > 0, X_{t+1} = c_{t+1}^y$), 则 $d\tau^* > 0, d\delta^* > 0, d(\bar{\delta}^*/\delta^*) < 0$, 即最优现收现付养老保险费率上升, 养老金退休前工资替代率上升, 且与平均工资替代率存在背离。

根据《中国统计年鉴(2016)》, 2015 年我国企业职工养老金的社会平均工资替代率较低, 仅为 50% 左右。虽然面临快速老龄化问题, 但目前城市 60 岁以上人口赡养率仍仅为 17.7%, 几乎低于所有的 OECD 国家, 然而在法定缴费率高达 24% 的情况下, 养老保险基金依然面临着很大的失衡风险。而且, 尽管平均工资替代率较低, 退休前工资替代率却很高。在 OECD 的 *Pension at a Glance* 报告中, 中国养老金退休前工资替代率为 74%, 而与退休前实际收入相比的净替代率则为 81%。在一些微观调查数据库中, 有相当多的退休人员实际养老金待遇高于退休前实际工资。^①

定理 2 表明模型可以内生出上述典型特征。在引入老年人对代际公平性的要求后, 当经济体沿着“数量—质量”前沿边界转换时, 现收现付的力度应当加大, 而不是转向积累制。直观来看, 引入代际公平性, 意味着老年人有更迫切的愿望分享经济发展成果, 因此在高速人力资本进步过程中, 现收现付社会保障成分应当进一步提高。随着更强调代际利益分享, 老年人养老金待遇上升, 退休前工资替代率增加。但是, 由于养老金增长落后于年轻人工资的高速增长, 平均工资替代率反而会下降。在人力资本进步和代际公平性的考量下, 模型同时内生出了高缴费率、低平均工资替代率、高退休前工资替代率等典型事实。那么, 高缴费率和高退休前工资替代率不可避免会影响劳动者的退休行为, 带来道德风险问题, 即老年人工作与不工作时的收入差距缩小, 产生引致退休。因此, 有必要通过内生退休年龄模型来分析这种道德风险对最优养老金体系的影响。

(二) 退休行为内生与道德风险

在内生退休模型下, 假设代表性行为人的效用函数及对应的社会福利函数为:

$$\begin{aligned} U_t &= U(c_t^y, \bar{c}_{t+1}^o) = U(c_t^y, c_{t+1}^o - \lambda X_{t+1} - H_{t+1}' V(\pi)) \\ \tilde{U}_t &= \tilde{U}(\hat{c}_t^y, \hat{c}_{t+1}^o) = \tilde{U}(\hat{c}_t^y, \hat{c}_{t+1}^o - \lambda \bar{X}_{t+1} - \varphi V(\pi)) \end{aligned} \quad (12)$$

由于齐次性假设, 公式(12) 中的个人效用与社会福利函数仍然是对应的, 代表了同样的偏好和无差异曲线。在劳动负效用函数设定中, 使用了 Greenwood et al. (1988)、Diamond(1998)、Chetty & Saez(2010) 等物质化劳动负效用的方式, $H_{t+1}' V(\pi)$ 代表继续工作产生的负效用, 且 $V'(\pi) > 0, V''(\pi) < 0$ 。在内生退休年龄后, 养老金体系可能由于信息不对称而产生道德风险问题。如果政府完全不掌握个人劳动负效用信息, 就只能在道德风险存在的情况下选择次优政策, 反之则可以通过调节法定退休年龄或者加强继续工作的激励来影响个体退休行为, 最终恢复最优的资源分配状态。下面分别讨论次优政策以及政府如何实施干预恢复最优资源配置状态。

1. 现实中的次优政策

在次优政策中, 政府面临类似于“Baily-Chetty 失业保险模型”(Baily, 1978; Chetty, 2008) 中“流动性效应”和“道德风险”之间的权衡。政府一方面希望通过现收现付制实现代际公平, 另一方面又要防止道德风险。政府和个人之间存在一个序贯博弈: 政府选择养老保险费率和待遇水平; 然后个人根据政府政策参数, 选择最优的消费组合和劳动参与率。运用逆向求解法, 先求解个人最优的劳动参与率选择。当存在信息不对称时, 第 i 个行为人面临的决策问题为根据公式(12) 的效用函

^① 根据中国健康与养老追踪调查(CHARLS) 2011 年的计算, 净退休前替代率为 124.39%, 2013 年的替代率为 117.4%, 根据中国家庭追踪调查(CFPS) 2012 年的数据, 替代率为 93.0%, 都远高于企业职工平均工资替代率。

数,结合公式(5) — (6) 的预算约束,选择最优的消费和劳动参与率。其中,最优消费的一阶条件仍为边际替代率 $M(\hat{c}_{t+1}^o/\hat{c}_t^y) = 1 + r_t$,最优劳动参与率的一阶条件为:

$$\pi_{t+1}^i: 1 - \tau - \delta^i + (1 - \pi_{t+1}^i) \alpha = V'(\pi_{t+1}^i) \quad (13)$$

公式右端为劳动边际负效用,代表老年人继续工作的边际成本,公式左端是继续工作的边际收益。各国养老金制度设计中通常都会不同程度强调长缴多得,即退休年龄越晚,养老金替代率越高。根据我国“38 号文件”的规定,对于代表性行为人,基础养老金待遇是社会平均工资乘以缴费年限再乘以 1%,可以证明替代率与工作时长是线性关系,即替代率对退休时间的导数 δ'_π 是常数,即公式(13) 中的 α 。政府面临的社会福利最大化问题是在公式(2)、(13) 构成的约束集下通过选择缴费率 τ 以最大化公式(12) 中的社会福利函数。可证明如下一阶条件:

$$\begin{aligned} 1 + r &= M(\hat{c}^o/\hat{c}^y) = (\partial \tilde{U}/\partial \hat{c}^y) / (\partial \tilde{U}/\partial \hat{c}^o) \\ &= (1 + n)(1 + g) + \lambda(1 + g) + \varphi(\tau + \delta) \pi'_\tau \quad (14) \\ \pi'_\tau &= - (1 - \pi)(\tau + \delta) / \{ \tau [V'' + \alpha(1 - \pi) + \tau + \delta] \} < 0 \end{aligned}$$

对比公式(14) 和定理 2 可以发现,上式右端第三项为道德风险部分, π'_τ 反映了缴费率升高对劳动参与率的负面影响。同时,求解中央计划经济下的最优政策,对比两者可以证明:

定理 3(次优和最优政策): 当存在信息不对称时,养老保险费率越高,老年劳动参与率和退休年龄越低。政府次优的现收现付养老保险费率须低于最优费率。而在中央计划者最优费率下应当满足: $M(\hat{c}^o/\hat{c}^y) = 1 + r = (1 + n)(1 + g) + \lambda(1 + g)$, 且 $V'(\pi) = 1$ 。

定理 3 表明,道德风险束缚了政府通过现收现付制进行代际转移支付的空间。如前所述,我国养老金退休前工资替代率较高,在这种情况下,继续提升代际转移支付将导致严重的道德风险,因此次优的现收现付养老保险费率应低于最优费率。

2. 政府干预政策

次优决策模型中的道德风险从本质上说是一种外部性: 个人在决策时仅考虑继续工作的边际收益(等于 $1 - \tau - \delta$) 和边际成本,而忽略了对总体基金平衡的影响或者说继续工作的社会边际收益(等于 1)。针对这种外部性存在两种干预方式(Weitzman, 1974): 第一种是“数量工具”,即政府外生地法定退休年龄控制在最优水平上,根据定理 3,退休年龄应满足 $V'(\pi^*) = 1$; 另一种方式为价格工具,即加强缴费受益关联。在缴费受益关联 α 达到 $(\tau + \delta)/(1 - \pi)$ 时, $V'(\pi) = 1$,道德风险消失,社会资源分配重新回到最优状态。可以证明:

定理 4(政府干预政策): 在分散竞争性经济中,假设总代际赡养系数 $\chi = (1 + n)(1 + g)$ 不变的同时存在生育率下降和人力资本进步,即 $d(1 + n) < 0$, $d(1 + g) > 0$, 那么 $d\tau^* > 0$, $d\delta^* > 0$, $d(\delta^*/\delta^*) < 0$, $d\pi < 0$, 即最优现收现付养老保险费率上升,养老金退休前工资替代率上升,且与平均工资替代率存在背离,老年劳动参与率下降。为了恢复最优资源配置,政府需要加强缴费收益关联(价格工具), $d\alpha > 0$, 或者将退休年龄控制在满足 $V'(\pi^*) = 1$ 的水平(数量工具)。

从数据来看,我国城市老年劳动参与率偏低,导致制度内赡养率远高于人口赡养率。第六次人口普查数据(以下简称“六普”)显示,我国目前 55—64 岁的男性老年人劳动参与率为 39.8%,而女性为 11.9%。这一老年劳动参与率远低于 OECD 国家平均水平。从人口赡养率看,城市 60 岁以上人口与 20—59 岁人口之比为 19.2%,也几乎低于所有的 OECD 国家。但制度内赡养率却达到 32%,2017 年进一步上升到 37.7%。定理 4 表明本文模型内生出了这种低劳动参与率、低人口赡养率和高制度内赡养率的现象。

其次,从政策层面看,我国政府一直努力控制提前退休现象和加强缴费受益关联,以提高老年劳动参与率。“38 号文件”将基础养老金待遇与缴费年限挂钩,以及将个人账户养老金待遇与退休年龄挂钩。在历次基本养老金调整中,也更强调体现“长缴多得”的激励机制。缴费受益关联使得

老年人早退休的机会成本增大,作为“价格工具”使得退休变得更加“昂贵”。此外,政府也正在酝酿使用延迟法定退休年龄等“数量工具”控制道德风险。定理 4 同样内生出了与实践相似的政策建议。定理 4 表明:当年轻人人力资本快速进步时,政府一方面需要加大代际转移支付,另一方面需要削弱代内转移支付,加强缴费受益关联,体现社会保障制度的激励性。此时,连续提高养老金待遇与延迟退休、加强缴费受益关联等控制道德风险的政策不但不矛盾,实际上还是相互配合的“左右手”:前者使老年人分享快速经济发展的成果,后者可以控制伴随高代际转移支付而来的道德风险。

(三) 模型机制总结

从基础模型开始,定理 1—4 陆续引入少子化、人力资本进步、代际公平性、道德风险和缴费受益关联等因素,概括起来,这些因素对于最优的现收现付养老金体系的影响如图 2 所示。

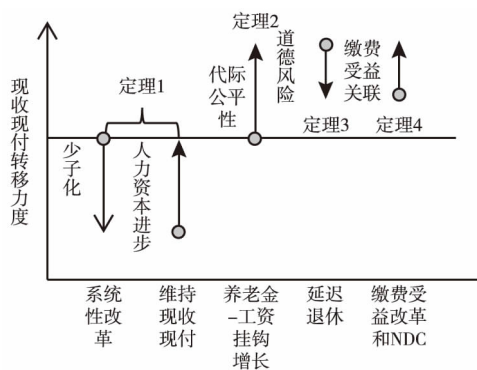


图 2 理论模型机制

图中纵轴为现收现付缴费率或养老金体系的代际转移支付力度。如图所示,定理 1—4 之间存在紧密的逻辑联系。定理 1 表明,如果仅有少子化的趋势,大幅缩小养老金体系现收现付成分,甚至实施系统性养老金改革转向积累制更有利于社会福利最大化。然而如果人力资本进步可以完全补偿少子化趋势,从而保持总体赡养能力不变时,则如定理 1 所示,最优现收现付制缴费率可以回到没有少子化时的水平。但是,平均工资替代率将出现下降,意味着老年人相对经济地位下降,损害代际公平。

考虑到社会对代际公平性的追求,定理 2 表明,在人力资本进步过程中,为了实现代际公平,满足老年人分享经济发展成果的需求,现收现付缴费率和退休前工资替代率都应当更高,超过没有少子化时的水平。在加入内生退休选择后,定理 3 表明,更高缴费率和养老金退休前工资替代率减弱了老年人继续工作的积极性,出现引致退休现象,道德风险限制了政府通过现收现付养老金体系进行代际转移支付的能力和空间,使得次优的养老保险费率下降。定理 4 表明,政府应当加强缴费受益关联或者延迟退休年龄,来控制道德风险的不利影响。道德风险的控制可以缓解政府在公平和效率之间面临的困境,能够更从容地实施更强的代际转移支付。此时,合意的现收现付缴费率水平也能有所恢复。当缴费受益关联足够强或延迟退休力度足够大时,代际转移支付力度可以完全恢复到最优的状态。

四、校准和模拟

(一) 校准

鉴于数据资料的完整性,本文中所有变量和参数均选取 2014 年或 2015 年的数据。假设效用函数为常见的齐次函数形式: $\ln U_t^i = \ln(c_t^y) + \beta \ln(c_{t+1}^o)$ 。劳动负效用为二次型,即 $V(\pi) = \phi \pi^2 / 2$, ϕ 为劳动负效用系数。模型中有 9 个外生参数: 人口增长率 $1+n$ 、代际人力资本进步 $1+g$ 、退休时期效用的权重 β 、年轻人和老年人的标准化非劳动收入 $\bar{w}^y$ 和 $\bar{w}^o$ 、缴费受益关联程度 α 、经验溢价 φ 、代际公平性考量 λ , 以及劳动负效用系数 ϕ 。模型有 3 个核心内生变量,即现收现付养老保险缴费率 τ 、老年劳动参与率 π 和退休前工资替代率 δ 。本文的校准策略是: 首先,前 6 个变量通过既往文献

或统计资料确定;其次,将3个核心内生变量作为校准模型最主要的目标变量(targeted endogenous variables),利用这些变量作为矩条件校准其余3个核心外生参数,从而识别出整个模型;最后,评估其余非目标内生变量的表现,来判断模型校准的合理性。

首先确定部分外生参数取值。本文将20—49岁设为第一期,50—76岁设定为第二期,两期或世代之间时间跨度为28年。这意味着代表性个体职业生涯长度或缴费年限为 $(1+\pi) \times 28$,退休年龄为 $(1+\pi) \times 28 + 20$ 。根据2015年1%人口抽样调查数据,20—49岁抽样人数为1037.38万人,50岁以上抽样人数为6251.09万人,两者之比为1.66,取整后本文将 $1+n$ 设定在1.7。根据《中国统计年鉴》,中国过去一个世代(1988—2015)产出年增长率为8.74%,Whalley & Zhao (2013)发现人力资本对中国经济增长的贡献率为38.12%,这意味着人力资本进步带来的年经济增长率约为3.33%,累计28年中由人力资本带动的收入增长为2.5倍。所以,本文将 $1+g$ 校准为2.5。根据2015年我国收入法统计的GDP,计算得劳动收入份额约为54%,这意味着一个世代中人力资本存量增加至5.45倍($2.5^{\frac{1}{0.54}}$),年增长率约为6.25%,这与既往研究基本一致。梁润等(2015)估算的结果为5.73%。《2017年中国人力资本报告》的数据显示,1987—2015年间,城镇居民人均人力资本存量年增长率5.88%,全国居民年增长率为7.02%。根据 $1+n$ 和 $1+g$ 的数值,将总赡养能力 $\chi = (1+n)(1+g)$ 设定在4.25。世代交叠模型中年度折现率常设定为3%,对应本文的 β 值为0.42。Yew & Zhang(2013)将 β 值设定为0.3。本文折中选择将 β 值设定在0.35。非劳动收入与我国收入法计算的GDP值相吻合,而根据CFPS2016年的数据,我国年轻人和老年人的非劳动收入的比值接近1:1,再考虑到人力资本进步,本文中年轻人标准化非劳动收入 $\overline{w}^y$ 和老年人非劳动收入 $\overline{w}^o$ 分别设定为0.6和1.5,此时劳动收入份额恰好为54%。

现实中次优养老金决策包括三个核心内生变量,即现收现付养老保险缴费率 τ 、老年劳动参与率 π 和退休前工资替代率 δ 。本文将 τ 设为0.17。根据《中国统计年鉴》2015年的人均养老金缴费(8778元)以及企业直报工资数据(53165元),计算出的实际缴费率十分接近校准值。根据2015年1%人口抽样统计结果中50岁以上老年劳动参与率,本文将 π 调校至0.3,对应的制度内赡养率为0.35。如前所述,OECD的*Pension at a Glance*将我国退休前工资替代率设定为0.74。我们也分析某省2017年全部新退休职工首月养老金的替代率,发现这一替代率为0.7089。故本文将 δ 设定为0.7。根据模型设定和“国发[2005]38号”文件规定,得到现实的缴费收益关联为0.664。^①

三个目标内生变量和公式(2)、(13)——(14)构成了三个矩条件,根据矩条件可以得到经验溢价 $\varphi = 1.646$ 、代际公平性考量 $\lambda = 0.183$ 以及劳动负效用系数 $\phi = 1.983$ 。这些校准得到的参数是否具有可信度呢?本文设计了两种常用检验方法,第一是直接将校准参数与现实可得的统计数据对比;第二是考察其他非目标内生变量的表现。在Lagakos et al. (2018)的研究中,发达国家的经验溢价约为1.816,而发展中国家为1.369,所以模型校准得到经验溢价1.646完全落在这一区间内。在 $\phi = 1.983$ 时,最优劳动参与率为0.504,对应退休年龄为62.1岁,这一数字与OECD国家的劳动参与率和退休年龄(男性64岁,女性63.1岁)是较为接近的。代际公平性考量 λ 控制了养老金与工资的挂钩增长情况。根据我国2005年以来职工养老金和工资增长率,回归得到增长率挂钩程度为0.315,而模型预测的挂钩程度为0.29,两者极为接近。利用2001—2015年省级面板数据,我们进行了OLS、双向固定和动态面板三项回归分析,两者的挂钩程度也在0.276—0.351之间。

表1中罗列了各种非目标内生变量,包括平均工资替代率、制度内赡养率、退休年龄、年轻人与

① 退休者基础养老金为: $\delta = (1 + \varphi\pi) / (1 + n + \pi) \times (1 + \pi) \times 28 \times 1\%$, $\alpha = d\delta/d\pi$,代入劳动参与率数值可得。

老年工资比、存贷款利率、净退休前工资替代率、年轻劳动力占劳动力市场比重、年轻人工资收入占全部工资收入比重等。如表 1 所示,模型中各非目标内生变量与实际变量比较接近。综合两种方法的检验结果,模型参数具有可信度。下面利用校准得到的参数进行反事实分析。

表 1 参数和变量

外生参数					
$1+n$	1.70	人口增长	β	0.35	主观折现率
$1+g$	2.50	人力资本进步	$\overline{w}^o$	1.50	老年人非劳动性收入
α	0.66	缴费受益关联水平	$\overline{w}^y$	0.60	年轻人非劳动性收入
目标内生变量			校准参数		
τ	0.17	缴费率	φ	1.65	经验溢价
π	0.30	老年劳动参与率	λ	0.18	参照依赖强度
δ	0.70	退休前工资替代率	ϕ	1.98	劳动负效用系数
非目标内生变量					
变量名	模型值	实际值	数据来源	数据年份	变量含义
20	0.49	0.53	国家统计局①	2015	平均工资替代率
D^I	0.35	0.35	国家统计局	2015	制度内赡养率
R^A	56.40	≈54	人社部②	2015	实际退休年龄
H^I/H^o	1.52	1.36	CFPS③	2015	劳动收入比
$1+r$	4.91	4.75—5.40	中国人民银行	2015.1	长期存贷款利率
δ^y	0.84	0.83	OECD	2016	净退休前工资替代率
ρ	0.85	0.87	CFPS	2015	年轻人占比
ρ^H	0.90	0.90	CFPS	2015	年轻人工资份额

注:①采用国家统计局企业直报数据。②来自李唐宁、李唯嘉(2016),《人社部:延迟退休对就业影响有限目前平均退休年龄仅 54 岁》,《经济参考报》2016 年 7 月 27 日。梁玉成(2007)表明 20 世纪 80 年代以来我国平均退休年龄 53—56 岁间震荡。③CFPS2016 显示年轻人(20—49 岁)和老年人(50—72 岁)平均工资收入比在 1.36 左右。

(二) 反事实分析:沿“数量—质量”前沿边界的转换

反事实分析是考察当一个经济体沿着“数量—质量”前沿边界移动时,生育率下降,但同时存在快速的代际人力资本进步,如果社会总赡养能力不变,那么其养老金体系将表现出如本文模型证明的特征,即高现收现付缴费率、低平均工资替代率、高退休前工资替代率、低人口赡养率、高制度内赡养率以及工资—年龄分布提前隆起等现象。为了验证这一点,假设 $1+n$ 从 1.7 的校准值恢复到更高水平,而 $1+g$ 则从目前 2.5 下降到更低水平,同时保持 $\chi = (1+n)(1+g) = 4.25$ 不变。图 3 三维空间 X 轴和 Y 轴分别为人口数量和质量,而 Z 轴为最优的现收现付缴费率。图中深色曲线刻画了最优缴费率与人口数量和质量之间的关系。曲线在“X—Y 平面”上的投影即代表了“数量—质量”前沿边界。

根据图 3(a),当经济体沿着“数量—质量”前沿返回到“高生育率、低人力资本”状态时,缴费率从现实的 17% 逐步下降到 9.9%,降幅约 42%,即根本不需要如此大的代际转移支付力度,最优养老金制度中完全可以有超过 7% 的积累制成分。图 3(b) 比较了沿着同样的“数量—质量前沿”移动时平均工资替代率和退休前工资替代率的变化。通常人们临近退休时的工资高于职业生涯平均工资,因此平均工资替代率应当高于退休前工资替代率。但当经济体沿着“数量—质量”前沿转换,随着年轻人人力资本进步和代际转移支付的加大,老年人养老金水平大幅提高,造成最优退休前工资替代率上升,而老年人工资水平相对年轻人工资下降,导致最优平均工资替代率下降,如图

3(b) 所示。图 3(c) 显示了内生退休年龄框架下, 高的缴费率和退休前工资替代率导致老年人继续工作的意愿更低, 产生引致退休现象, 造成低老年劳动参与率, 制度内赡养率也大幅上升, 与人口赡养率之间差距拉大。此外数据显示, 在 $1+g=1$ 时, 年轻人工资是老年人的 60.76%, 提前隆起现象也彻底消失。只有在 $1+g > \varphi = 1.65$ 时, 年轻人工资才会超过老年人。

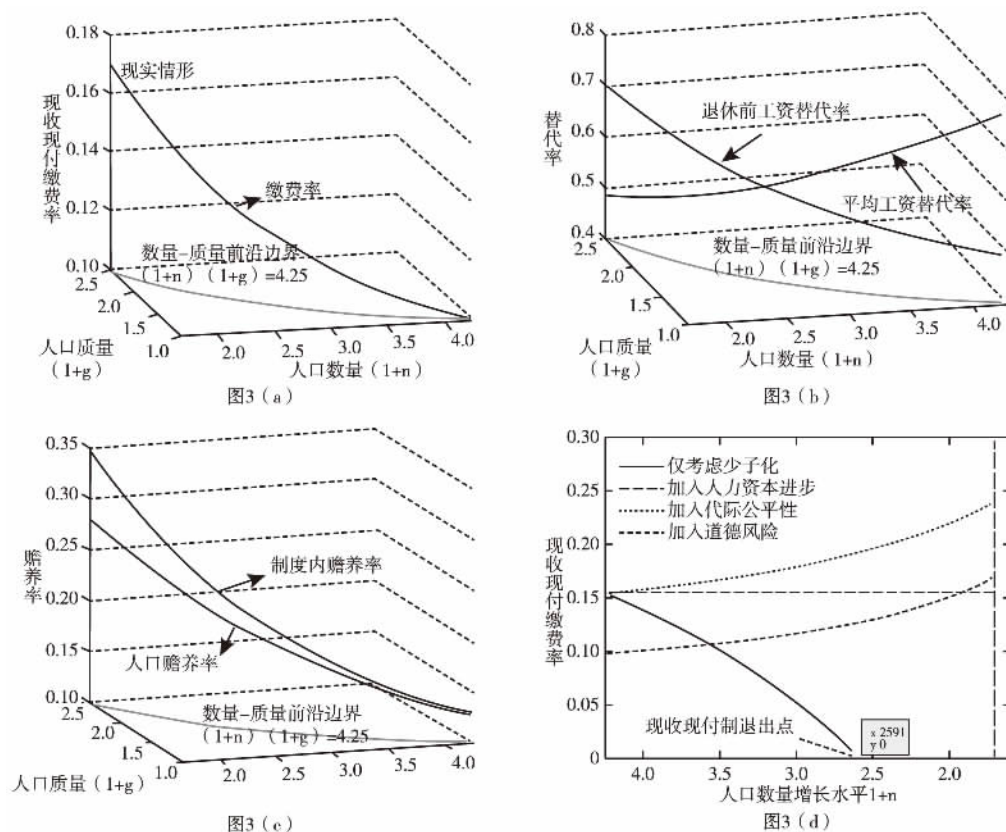


图3 反事实分析和影响因素分解

上述反事实分析验证了模型的确可以内生出我国养老金体系的各种典型特征, 这些特征是少子化和人力资本共同作用的综合效应, 图 3(d) 则验证了定理 1—4 的模型机制, 展示了少子化、人力资本进步、代际公平性以及内生退休时的道德风险因素对最优现收现付缴费率的影响。在假设 $1+n=4.25$ 而 $1+g=1$ 时, 现收现付制缴费率为 15%。引入少子化因素时, 如图中实线所示, 随着少子化程度的加剧, 社会人口赡养率上升, 最优缴费率下降。在 $1+n$ 降至 2.5 以下时, 最优缴费率降为 0, 此时现收现付制没有存在的必要性, 养老保险应当采用完全积累制。在此环境下, 系统性改革建议是恰当的。然而, 由于存在人力资本进步, 当经济体沿着“数量-质量”前沿转换时, 则如定理 1 中等价性原理以及图 3(d) 中短划线所示, 最优现收现付缴费率被恢复到 15% 的水平, 少子化带来的人口数量不足完全被质量进步所补偿。引入代际公平性后, 如定理 2 及图中点线所表明的, 最优现收现付缴费率更高, 且随着人力资本进步而上升至 24%。定理 3 中, 道德风险是抑制政府实施代际转移支付的重要原因。图 3(d) 显示, 如果考虑道德风险, 现收现付缴费率将重新大幅下降。现实中我国现收现付制具有一定的缴费受益关联激励机制, 这使得现收现付缴费率有所恢复, 回到图中点划线终端的与现实情景相吻合的 17% 的水平。事实上也可以对退休前工资替代率、平均工资替代率、劳动参与率、年轻人和老年人消费水平等各个参数进行类似分解, 追溯每个机制的作用。可以发现这些变量的变动与定理 1—4 中的分析结果保持一致, 不再赘述。

五、政策实验

政策实验目的是探讨当前重要的改革,如从现收现付转向积累制的系统性改革、养老金正常调整机制、延迟退休年龄和缴费受益关联等的社会福利效果如何。模拟结果如图 4 所示。

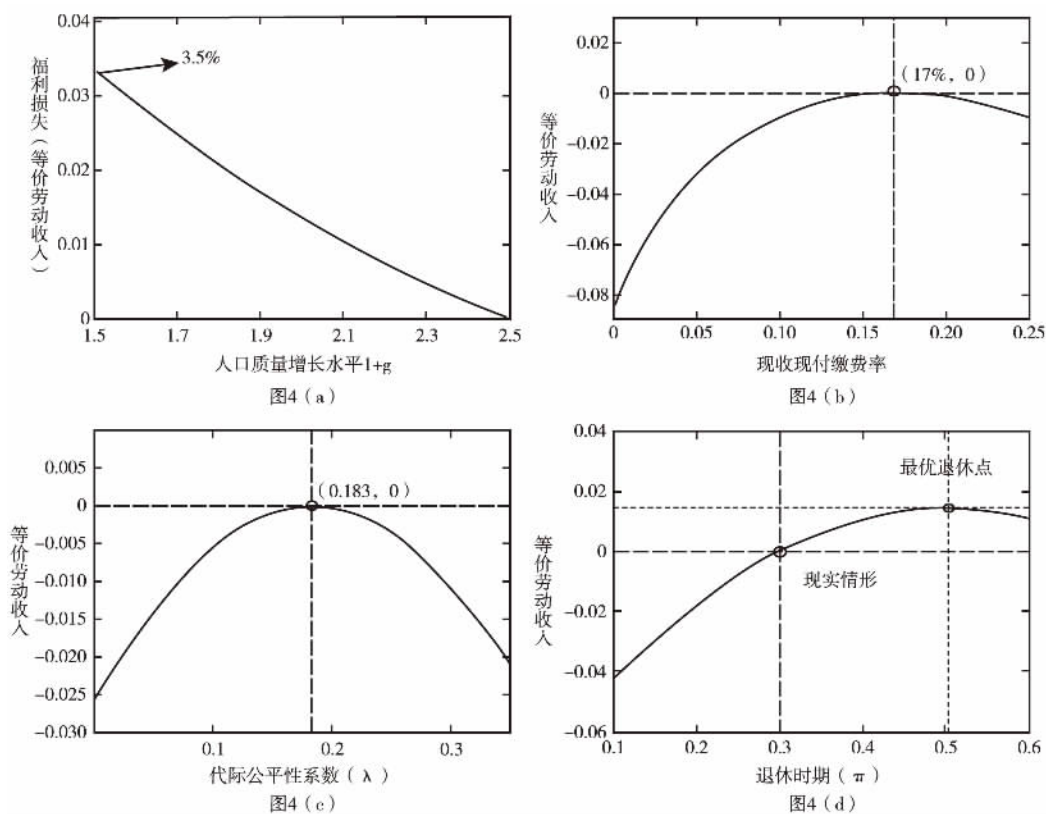


图 4 政策改革的社会福利效果

(一) 系统性养老金改革

从 20 世纪 90 年代开始,从现收现付部分转向积累制的系统性改革就一直是我国养老金改革的基本方向,并得到历史上多个政策文件的确认。然而,现实中,我国个人账户一直在空账运行,系统性改革已经处于停滞状态。为了说明系统性改革的福利效果,本文设计了两套模拟试验:第一,假设我国只面临少子化趋势,而没有快速的人力资本进步,在此假设基础上评估采用现收现付制度将造成多少福利损失;第二,从同时存在少子化和人力资本进步的现实情形出发,假设中国遵循系统性改革的建议,将付出多少福利损失。

图 4(a) 中横轴为人口质量进步速度,纵轴换算为等价劳动收入。如果人力资本进步较慢,而少子化程度较高,则维持当前现收现付制度将付出巨大的福利损失。而且,福利损失随人力资本进步减缓而急剧增加,在人力资本进步速度降至 1.5 时,福利损失达到终生劳动收入的 3.5%。这表明,如果没有足够的人力资本进步,在少子化过程中仍然维持如此高的现收现付缴费率,成本将十分高昂,此时部分转向积累制就是必要的。图 4(b) 中横轴为现收现付制缴费率,或者说代际转移支付比例。如图所示,在人力资本快速进步时,如果下调现收现付缴费率转向积累制,也将导致巨大的福利损失。如果完全采用积累制,那么福利损失将达到惊人的 8%。所以,在人口经济环境变化中选择恰当的改革策略至关重要。系统性养老金改革虽然考虑了中国严重的少子化和老龄化趋势,但没有考虑到我国同期快速的人口人力资本进步。中国

养老金改革“向右看、向左走”则可以避免这一高昂代价,在本文的框架下,恰恰是基于现实的合理选择。

(二) 养老金调整机制

早在 1995 年,国务院《关于深化企业职工养老保险制度改革的通知》中就明确提出“基本养老金可按当地职工上一年度平均工资增长率的一定比例进行调整”。自 2005 年起,我国连续 14 次调整退休者养老金待遇,养老金增长速度远快于 2005 年之前,体现了不同世代共享经济发展成果的理念。在本文模型中,参数 λ 体现了对代际公平性的追求,也是影响养老金-工资挂钩增长的关键因素。根据近年来实际养老金调整数据,参数 λ 在模型中被校准为 0.1826。如果降低 λ ,则养老金与工资的挂钩程度将下降。当 $\lambda = 0$ 时,养老金与工资增长率挂钩消失。

本文设计了如下反事实实验:假如行为人真实的公平性考量 λ 在 0.1826,而政府对个人公平性考虑判断为 $\hat{\lambda} \neq \lambda$,这一误判导致实际养老金和工资挂钩增长程度偏离最优。如图 4(c) 所示, $\hat{\lambda} = 0$,即在政府忽略老年人分享经济增长的需求时,福利损失等价于劳动收入的 2.5%。在本文理论框架下,我国政府近年来连续提高老年人养老金待遇,建立养老金增长与工资增长的挂钩机制,恰是对老年人分享经济增长成果的有效回应。

(三) 延迟退休年龄和缴费受益关联

在本文的模型中,当劳动参与率偏离社会最优水平时,有必要通过政府干预加以纠正。根据校准结果,现实中老年劳动参与率为 0.3,最优劳动参与率则是 $\pi^* = V'(1)^{(-1)} = 1/\varphi = 0.504$ 。政府可以通过数量干预和价格干预促使劳动参与率恢复到最优水平。

图 4(d) 中曲线是以现实为参照计算的不同老年劳动参与率水平对应的社会福利变化。图中曲线在 0.3 左侧为负值,代表福利损失;在右侧为正值,代表福利增进。如理论模型预测,曲线在劳动参与率为 0.504 时达到峰值,此时实际退休年龄大约为 62 岁。与其他纠正市场失灵的研究保持一致(Weitzman, 1974),劳动参与率逐渐恢复到最优值的过程中,福利增进递减。根据图 4(d),如果政府可以通过延迟退休年龄等外生干预政策,将老年劳动参与率提高到最优水平,带来的福利改进等价于代表性行为人全部终生收入的 1.45%,这一福利改进是比较显著的。因此,政府对老年劳动参与的干预措施是十分必要的。

缴费受益关联是“价格”干预措施,在个体内生退休年龄选择时,它使得个人过早退休面临的养老金损失增大,提高个人退休的边际成本,实质上是对退休施加的庇古税。缴费受益关联的干预效果与延迟退休年龄一一对应,因此图 4(d) 中任何一个老年劳动参与率都对应一个缴费受益关系。如果要使老年劳动参与率恢复到最优的水平,则缴费受益关联程度需要提高到 2 以上,即每个劳动者劳动参与率增加 1%,养老金替代率增加 2% 以上,远远超过了模型根据现实校准得到的 0.664 的水平。如果没有任何缴费受益关联,老年劳动参与率将下降到 0.1 左右,造成巨大的福利损失。因此,2005 年“38 号文件”提高了退休者基础养老金与退休年龄之间的挂钩程度,有助于提高社会福利水平,但仍与最优缴费受益关联存在很大差距。

综上所述,在本文框架下,无论延迟退休年龄还是加强缴费受益关联,都有助于提高我国老年劳动参与率,增进社会福利。而且,政府越希望通过代际转移支付来实现代际公平,越需要更有力地控制道德风险。如定理 4 以及实证模拟所表明的,这两个看似矛盾的改革策略或许能成为相互配合的“左右手”,中国社会保障改革的路径选择依然是可理解的。

六、结 论

本文通过世代交叠模型和基于平衡增长路径的社会福利分析框架,探讨了当经济体同时面临

少子化和代际人力资本进步时的最优现收现付养老金体系。本文的分析表明,从社会福利最大化角度看,如果仅仅存在少子化,确实有必要改革现收现付的养老金体系,然而,如果考虑到快速的人力资本进步,经济体沿着总赡养能力不变的“数量-质量”前沿边界转换,则有必要维持甚至加强代际转移支付,让老年人分享经济发展成果,此时现收现付养老保险缴费率应当上升。模型可以内生出当前我国养老金体系表现出的许多典型事实,如高缴费率、高退休前工资替代率、低平均工资替代率、高制度内赡养率等。

从中国的养老金改革实践看,模型同时内生出养老金改革中几个重要且令人困惑的表现。在养老金改革方向上,尽管我国选择从现收现付制转向部分积累制,但是随着快速的人力资本进步,政府需要加大代际转移支付以满足老年人分享经济发展成果的需求,所以表现出“向右看、向左走”的特点:保持个人账户“空账”运行,却将更多资金用于提高老年人养老金待遇。当代际转移支付变得更重要时,现收现付缴费率和退休前工资替代率提高,这对老年劳动参与率产生负面激励,政府有必要控制道德风险,在一定程度上牺牲代内转移支付,更强调“多缴多得、长缴多得”。2005年以来政府在再分配方面选择“舍代内、取代际”,加强了缴费受益关联。表面看来,无论缴费收益关联还是延迟退休年龄都是控制道德风险的重要手段,提高养老金待遇则会助长道德风险。这些“加油门、踩刹车”的行为看似矛盾,但在本文框架中则恰是相互配合的“左右手”。

综上所述,本文模型从社会福利最大化中内生出各项制度参数和改革路径,发现这些参数及其变动特征与我国现实是吻合的。因此,一个有趣的发现是,这些看似令人费解、相互矛盾的改革实践在本文理论框架中实际上都是福利增进的,中国养老金改革的路径依然是可理解的。这一发现为探索中国养老金改革理论逻辑问题提供了一个新的视角:各项“糟糕”的制度参数以及看似矛盾的改革策略是否都是制度设计和执行失误导致的?是否既往研究中有可能政府失灵被夸大,反而忽略了对基础性生产力因素的考量呢?本文模型采用了从中央计划者社会福利最大化中内生出各项制度参数的方法。就像真实经济周期理论是为了防止人们过度夸大市场失灵和政府干预的必要性,这种范式下的研究是为了防止过度夸大政府失灵和市场化的必要性。当然,现实中政策改革不可能完全按照这种“理性主义”方式运行,也不能否认我国社会保障体系仍存在不足之处,而且即便改革实践中体现出这种理性,也很可能不是改革主体自觉的,而是被改革现实需求所驱动的、非自觉的。但重要的是,在此范式下,我们表明当人口结构沿着“数量-质量前沿边界”转换时,基于这种理性主义的模型内生出的众多制度参数特征、变动方向以及最优改革策略与我国现实是基本一致的,故而应防止夸大现实政策中不完善成分的作用,而忽略了对基础生产力因素的分析 and 考量。

此外,人口老龄化是我国经济社会发展面临的重大挑战,它是由预期寿命延长和少子化两种因素所驱动的。预期寿命变动不但会引起与少子化完全等效的赡养率变化,还可能导致人们在生命周期决策中提高老年期效用的权重。尽管本文只分析了少子化问题,但如果忽略个体生命周期变动对两期效用相对权重的影响,本文结论完全适用于预期寿命延长造成的老龄化。最后,本文还与正在实施的“二胎”等计划生育政策调整有关。本文瞄准少子化、老龄化背景下,人力资本进步如何影响最优社会保障体系选择的问题。如果“二胎”政策是有效的,那么它将拓展我们面临的“数量-质量前沿边界”,使得不需要太高的人力资本进步就能保证“质量能够弥补数量的不足”。也就是说,本文的结论将更加稳健,结论所需要满足的参数前提条件也将更加的宽松。

参考文献

程杰,2014《退而不休的劳动者:转型中国的一个典型现象》,《劳动经济研究》第5期。

- 封进、胡岩, 2008 《中国城镇劳动力提前退休行为的研究》, 《中国人口科学》第4期。
- 国务院发展研究中心社会保障课题组, 2000 《分离体制转轨成本, 建立可持续发展制度——世纪之交的中国养老保障制度改革研究》, 《管理世界》第6期。
- 梁润、余静文、冯时, 2015 《人力资本对中国经济增长的贡献测算》, 《南方经济》第7期。
- 梁玉成, 2007 《市场转型过程中的国家与市场——一项基于劳动力退休年龄的考察》, 《中国社会科学》第5期。
- 孙祁祥, 2001 《“空账”与转轨成本——中国养老保险体制改革的效应分析》, 《经济研究》第5期。
- Aaron, Henry., 1966, “The Social Insurance Paradox”, *Canadian Journal of Economics & Political Science*, 32, 371—374.
- Baily, M. N., 1978, “Some Aspects of Optimal Unemployment Insurance”, *Journal of Public Economics*, 10(3), 379—402.
- Chetty, R., 2008, “Moral Hazard versus Liquidity and Optimal Unemployment Insurance”, *Journal of Political Economy*, 116 (2), 173—234.
- Chetty, Raj, and Emmanuel Saez, 2010, “Optimal Taxation and Social Insurance with Endogenous Private Insurance”, *American Economic Journal: Economic Policy*, 2(2), 85—116.
- Diamond, P. A., 1965, “National Debt in a Neoclassical Growth Model”, *American Economic Review*, 1965. 55(5), 1126—1150.
- Diamond, P. A., 1998, “Optimal Income Taxation: an Example with a U-shaped Pattern of Optimal Marginal Tax Rates”, *American Economic Review*, 88 (1), 83—95.
- Estelle, J., 2002, “How Can China Solve Its Old-age Security Problem? The Interaction between Pension, State Enterprise and Financial Market Reform”, *Journal of Pension Economics & Finance*, 1(1), 53—75.
- Fang, H., and Y. Zhang, 2018, “Growing Pains in Chinese Social Security System”, The Inaugural NBER-SAIF Conference “Retirement Security: Public and Private Pensions”, June 29—30, Shanghai, China
- Feldstein, M., 1974, “Social Security, Induced Retirement, and Aggregate Capital Accumulation”, *Journal of Political Economy*, 82 (5), 905—926.
- Feldstein, M., 1999, “Social Security Pension Reform in China”, *China Economic Review*, 10 (2), 99—107.
- Grech, A. G., 2015, “Convergence or Divergence? How the Financial Crisis Affected European Pensioners”, *International Social Security Review*, 68 (2), 43—62.
- Greenwood, J., Hercowitz, Z., and Huffman, G. W., 1988, “Investment, Capacity Utilization, and the Real Business Cycle”, *American Economic Review*, 78 (3), 402—417.
- Lagakos, D., Moll, B., Porzio, T., Qian, N., and Schoellman, T., 2018, “Life-cycle Wage Growth across Countries”, *Journal of Political Economy*, 126 (2), 797—849.
- Phelps, E., 1961, “The Golden Rule of Accumulation: a Fable for Growth Men”, *American Economic Review*, 51 (4), 638—643.
- Rey, E. D., and Lopez-Garcia, M. A., 2013, “Optimal Education and Pensions in an Endogenous Growth Model”, *Journal of Economic Theory*, 148 (4), 1737—1750.
- Samuelson, P. A., 1975, “Optimum Social Security in a Life-cycle Growth Model”, *International Economic Review*, 16 (3), 539—544.
- Samwick, A. A., 1998, “New Evidence on Pensions, Social Security, and the Timing of Retirement”, *Journal of Public Economics*, 70 (2), 207—236.
- Shi, Y., and Zhang J., 2009, “On High Fertility Rates in Developing Countries: Birth Limits, Birth Taxes, or Education Subsidies? ”, *Journal of Population Economics*, 22(3), 603—640.
- Song, Z., K. Storesletten, Y. Wang, and Zilibotti, F., 2015, “Sharing High Growth across Generations: Pensions and Demographic Transition in China”, *American Economic Journal: Macroeconomics*, 7 (2), 1—39.
- Weitzman, M. L., 1974, “Prices vs. Quantities”, *Review of Economic Studies*, 41 (4), 477—491.
- Whalley J., and Zhao, X., 2013, “The Contribution of Human Capital to China’s Economic Growth”, *China Economic Policy Review*, 2(1), 1—22
- Williamson, J. B., and Deitelbaum, C., 2005, “Social Security Reform: Does Partial Privatization Make Sense for China? ”, *Journal of Aging Studies*, 19(2), 257—271.
- Yew, S. L., and Zhang, J., 2013, “Socially Optimal Social Security and Education Subsidization in a Dynastic Model with Human Capital Externalities, Fertility and Endogenous Growth”, *Journal of Economic Dynamics & Control*, 37 (1), 154—175.

The Logic and Social Welfare Effects of China's Pension Reform: An Analysis from the Perspective of Population Quantity-Quality Transformation

ZHANG Yi^a, ZHANG Shubo^a and WANG Runquan^b

(a: School of Public Economics and Administration, Shanghai University of Finance and Economics;

b: School of Public Administration, Nanjing University of Finance and Economics)

Summary: To deal with its aging population, China started systemic pension reform through transferring from PAYGO to the partial fund program in the 1990s. However, in practice, the reform seems to be “looking right but going left.” The “funded individual accounts” emphasized in policy are operated as “non-financial accounts” in reality. The pension fund is meant to continuously improve the pension benefits of the elderly. In terms of redistribution, on the one hand, the government has strengthened the intergenerational transfer payment through the PAYGO system; on the other hand, it has weakened the intragenerational transfer payment and enhanced the connection between pension benefits and contributions. Strengthening the connection between benefits and contributions is intended to curb moral hazard. Conversely, improving the overall level of benefits might lead to early retirement and serious moral hazard problems. What are the social welfare outcomes of these seemingly contradictory reforms? Logically speaking, can these reforms be self-consistent?

It is worth noting that China's low fertility and aging population are accompanied by rapid intergenerational human capital progress. Therefore, whether the continuous improvement in population quality can make up for the shortage in population quantity, and whether the PAYGO system remains the optimal policy are questions to be answered. We use an overlapping generation model to assume that the economy moves along a frontier of a population quantity-quality tradeoff, which assumes that the fertility rate decreases and human capital increases. The research results show that when the rapid progress of human capital can fully compensate for the impact of the low fertility rate, the PAYGO system is the best choice. This finding can explain some of the characteristics of “looking right but going left” in China's pension reform.

The model, in incorporating intergenerational fairness, shows that in the rapid progress of human capital, the government should increase the degree of intergenerational transfer so that the elderly can share the achievements of economic development. This concern about intergenerational fairness might explain why the government uses funds to improve the overall pension benefit level rather than to finance personal accounts. However, a high contribution rate and high replacement rate might lead to moral hazard problems such as early retirement, restricting the term of the government's intergenerational transfer payment. To control moral hazard and expand the term of intergenerational transfer payment, it is necessary for the government to delay the statutory retirement age and strengthen the relationship between pension benefits and contributions. This way, the Chinese government might enhance intergenerational redistribution and weaken intragenerational redistribution.

By the calibration and numerical simulation, we find that the model based on the population quantity-quality frontier movement can endogenously generate many of the stylized facts given by Fang & Zhang (2018). Moreover, in this theoretical model, the government's various policies—including adopting non-financial accounts, strengthening the connection between benefits and contributions, delaying the statutory retirement age, and linking the growth rate between pension and wage—improve social welfare. Thus, the seemingly contradictory practices of China's pension reform have logical self-consistency given the progress of human capital and concerns about intergenerational fairness.

Following Song et al. (2015) and Fang & Zhang (2018), this study further explains the overall logic and social welfare outcomes of China's pension reform, and provides a new theoretical perspective through which to rethink its rationale. Previous studies may exaggerate the government's failure and neglect to examine the productivity factors that affect China's pension system. We find that a model based on the movement along the population quantity-quality frontier can better explain the stylized facts and the reform's path selection.

An aging population might be caused by lower fertility rates or longer life expectancy. Although this study focuses on lower fertility rate, the conclusions are applicable to an aging population that has emerged due to prolonged life expectancy.

Keywords: Human Capital; Low Fertility; Pay-As-You-Go; Intergenerational Fairness; Pension Reform

JEL Classification: D63, H55, J24, J14

(责任编辑:王利娜)(校对:南山)