

社会保障与经济增长:基于拟自然实验的分析

贾俊雪, 李紫霄, 秦 聪

[摘要] 20 世纪八九十年代大量发展中国家陆续建立了社会保障制度,以此为契机,本文对社会保障制度的经济影响进行拟自然实验分析。特别地,本文利用 114 个发展中国家 1981—2005 年的面板数据和倾向得分匹配—双差分法构造反事实,较好地校正了选择偏差,识别出社会保障制度对经济增长的因果处置效应。研究表明,社会保障制度对经济增长具有显著的负影响,导致实际人均 GDP 增长率下降了 2.5957 个百分点。这一不利影响倾向于主要通过物质资本积累机制发挥作用,导致物质资本投资率增长率下降了 173.94%,对人力资本积累的影响则较弱。制度安排尤其政策设计对社会保障制度的经济增长效应具有突出影响:较低的待遇水平和社保缴税(费)率会加剧社会保障制度的不利影响。较低的发展水平和较低的储蓄倾向具有类似影响,人口老龄化和较低的生育率则可缓解社会保障制度的不利影响。这对于中国长期可持续社会保障制度的建设具有良好的启示。

[关键词] 社会保障制度; 可持续性; 经济增长; 拟自然实验分析; 倾向得分匹配—双差分法

[中图分类号]F120 **[文献标识码]**A **[文章编号]**1006-480X(2018)11-0042-19

一、引言

改革开放以来,历经 40 年的不懈探索,中国逐步建立起世界上最大的社会保障体系,为增进居民福祉乃至世界社会保障发展做出了重要贡献。^①但不容忽视的是,相较发达国家而言,中国的社会保障制度建设起步较晚,当前面临的困难和矛盾依然较为突出。尤其是进入 21 世纪以来,伴随着人口老龄化进程的日益加快^②,转型期社会风险日渐凸显,社会基本矛盾亦发生了深刻变化,对中国社会保障制度的健全与完善提出了更高要求。无疑,一个完备、普惠的社会保障体系不仅可为中国经

[收稿日期] 2018-03-30

[基金项目] 国家自然科学基金面上项目“基于中国实践的财政分权理论”(批准号 71673279);国家社会科学基金重大项目“推动中国经济中高速可持续发展的突破性改革:地方政府治理体系改革”(批准号 17ZDA048)。

[作者简介] 贾俊雪,中国人民大学财政金融学院、中国财政金融政策研究中心教授,博士生导师,经济学博士;李紫霄,中国人民大学财政金融学院博士研究生;秦聪,中国人民大学国家发展与战略研究院讲师,经济学博士。通讯作者:秦聪,电子邮箱:qincong@ruc.edu.cn。感谢专家和编辑部的宝贵意见以及应世为的助研工作,当然文责自负。

^① 2016 年 11 月,国际社会保障协会在第 32 届全球大会上授予了中国政府“社会保障杰出成就奖”。

^② 按照国际惯例,60 岁以上人口占总人口 10%以上或者 65 岁以上人口占总人口 7%以上,即为老龄化社会。根据国家统计局数据,2001 年,中国 65 岁以上人口占总人口的比重首次超过 7%,开始步入老龄化社会。

济社会的转型升级提供强大助力,亦是提高居民福利、实现社会公平的有效措施。然而,2008年以来,欧洲一些国家爆发的主权债务危机在很大程度上与这些国家高福利的社会保障制度有关,暴露出慷慨的福利体系在抑制经济活力、加剧政府债务风险等方面存在的突出问题(郑秉文,2011)。与此同时,关于中国社会保障制度能否可持续、是否会“崩溃”的担忧近年来也甚嚣尘上,颇有愈演愈烈之势。因此,如何有效规避对经济增长的潜在不利影响、确保民生发展与经济增长的协调共进就成为中国政府在构建长期可持续的社会保障制度中亟需解决的一个重大问题。^①这在当前中国经济增长新旧动能接续转换、地方政府债务规模持续高企的现实背景下尤为关键。^②

显然,要想较好地解决这一问题就需深入厘清社会保障制度对经济增长的影响及其作用机理,以及制约影响社会保障制度经济增长效应的重要因素。长期以来,经济学家们一直积极致力于这方面的研究探索。大量理论文献深入剖析了社会保障制度对经济增长的影响,识别出两个极为重要的作用渠道——通过影响物质资本积累和人力资本积累进而影响经济增长的机制(Feldstein,1974;Barro,1974;Zhang,1995;Ehrlich and Kim,2007;Bruce and Turnovsky,2013;Gustman and Steinmeier,2015;郭庆旺等,2007;郭凯明和龚六堂,2012),为深刻认识和理解社会保障制度的经济影响提供了良好的理论基础和分析视角。但已有研究关于社会保障制度通过这两个渠道对经济增长究竟会产生怎样的影响以及哪些因素在其中发挥着重要作用还莫衷一是,尚未给出一致结论。这促使经济学家们更多转向经验研究。一类文献主要探究了某一国家(或地区)社会保障制度的经济影响(Feldstein,1996;Cigno et al.,2003;赵怡,2007;何立新等,2008;白重恩等,2012);另一些研究则聚焦于跨国分析,以期深入揭示社会保障制度经济影响的一般经验规律以及决定社会保障制度经济增长效应的潜在重要因素(Ehrlich and Zhong,1998;Zhang and Zhang,2004;Ehrlich and Kim,2007;贾俊雪等,2011)。

然而,值得注意的是:^①目前,此类文献主要是利用社会保障税费率(或社会保障支出占GDP的比率)等单一指标来考察社会保障制度的经济增长效应,无法全面捕捉其整体经济影响——社会保障涉及收支两个层面,而很多国家(包括中国)的社会保障存在收支缺口,收支变化往往并不同步(Romer and Romer,2016)。^②已有文献虽然较深入地探讨了经济发展水平等因素对社会保障制度经济增长效应的影响,但忽略了如下一些重要问题——人口老龄化是否会加剧社会保障制度对经济的潜在不利影响,形成负向冲击的交互叠加?生育水平和储蓄倾向又具有怎样的作用?这些问题的深入厘清无疑具有很好的现实意义。^③如何有效矫正选择偏差问题也是此类研究面临的一个较大挑战。显然,一国社会保障制度的建立实施以及具体的制度政策安排并非是完全外生的、随机的,而通常是一个由一系列经济和社会等因素决定的自选择过程,这些因素亦会同时影响经济增长,故可能存在突出的内生选择问题。目前,已有研究主要利用工具变量法试图解决这一问题。然而,构造一个良好的工具变量往往具有较大困难;更为重要的是,一些研究指出,即便是工具变量法也可能无法有效解决选择偏差问题(Black and Smith,2004;Caliendo and Kopeinig,2008)。

本文旨在较好地解决上述问题,拓展和深化关于社会保障制度经济影响的认识和理解,为中国可持续的社会保障制度的健全与完善提供一些有益借鉴。20世纪八九十年代,大量发展中国家陆

① 近年来,如何确保社会保障制度的可持续性成为世界各国政府和学术界关注的重点,很多改革措施被提出,包括降低保障水平、提高缴费率和延迟退休等(Kitao,2014)。

② 根据中国全国人大常委会规定,地方政府债务率不得超过100%。但是,2013年国家审计署第24号审计结果显示,36个样本地方政府中,9个省会城市的政府债务率超标。截至2016年1月,贵州、辽宁、云南和内蒙古四省份的债务率已越红线,其中辽宁的债务率高达197.5%。

续建立了社会保障制度。以这些重要的制度变迁为契机,本文对社会保障制度的经济影响进行拟自然实验分析。具体而言,本文以 114 个发展中国家 1981—2005 年的面板数据为基础,利用倾向得分匹配—双差分法(Propensity Score Matching with Difference-in-Difference)构造反事实,在拟自然实验环境下识别出社会保障制度对经济增长、物质资本积累和人力资本积累的因果处置效应。^① Caliendo and Kopeinig(2008)等指出,与工具变量等方法相比,倾向得分匹配—双差分策略可更好地控制观测和非观测因素(共同冲击和个体特定因素)的影响(这对于本文进行的跨国分析而言尤为重要,因为不同国家的传统文化信念、地理环境和资源禀赋等存在明显差异),故可有效校正选择偏差问题。因此,采用这一方法有助于更加准确地识别出社会保障制度对经济增长的因果性影响及其主要机理——反事实分析即平行趋势检验(Parallel-trend Test)证实了这一计量策略的有效性。而且,处置分析框架的应用也有利于较好地克服使用社会保障支出比率(或社会保障税费率)等单一指标的弊端,更为全面地揭示出社会保障制度的整体经济影响。最后,本文亦利用分样本分析考察了制度安排(基金积累制,特别是部分积累制)和政策设计(社会保障的待遇水平和税费负担)以及经济发展水平、储蓄倾向、人口老龄化和生育水平对社会保障制度经济增长效应的影响。这有利于得到更加丰富可靠的结论,对中国可持续的社会保障制度建设也更具借鉴意义。

本文余下部分的结构安排如下:第二节对相关文献进行简要综述;第三节给出本文的计量策略;第四节介绍了数据及其来源,给出主要变量的描述性统计分析;第五节给出基准回归分析和稳健性检验结果;第六节考察了制约影响社会保障制度经济增长效应的重要因素;最后为本文的结论和政策建议部分。

二、文献综述

现代经济社会中,社会保障制度或许是最重要的一项制度,亦可能是最庞大的一个公共项目。正因如此,其对经济的影响及作用机理一直是学术界关注的焦点。伴随着经济社会实践和现代经济理论的发展,人们对于这一问题的认识和理解亦经历了一个不断丰富、深化的过程。

早在 20 世纪 30 年代美国“大萧条”时期,凯恩斯经济学家就主张将社会保障作为刺激总需求、治理衰退的一个重要手段(Harris, 1941),这被认为是美国社会保障制度建立的一个重要理论基础。此后,大量研究秉承新古典思想,将分析重点转向社会保障制度对居民储蓄进而物质资本积累的影响。Feldstein(1974)较早考察了这一问题,指出社会保障制度会通过资产替代效应抑制居民储蓄,亦会通过引致退休效应促进居民储蓄,故对居民储蓄进而物质资本积累的影响不明确。不过,他的分析缺乏微观基础,而在新古典范式内为社会保障制度分析建立起正式及较完备的微观基础则主要归功于 Samuelson(1958)^②和 Barro(1974)。Barro(1974)首次利用一个代际交叠模型(Overlapping Generations Model, OLG)指出,当居民具有较强的遗产动机时,社会保障制度不会改变理性居民的跨期预算约束,故对居民储蓄和物质资本积累的影响是中性的。后续研究则大多表明社会保障制度会抑制居民储蓄和物质资本积累。这些文献往往考虑了更为复杂的经济环境。例如, Bruce and Turnovsky(2013)引入了更加符合现实的人口结构, Gustman and Steinmeier(2015)考虑了居民时间

① 处置分析框架主要用于微观政策(如劳动管制政策)的评估分析,近年来也开始较广泛地用于宏观制度和政策变化的影响分析。例如, Mendonça and Souza(2012)利用这一分析框架考察了通胀目标制是否有助于物价稳定。

② Samuelson(1958)最早利用 OLG 模型考察了社会保障制度的经济影响,但他没有考虑居民储蓄和物质资本积累。

偏好的异质性(储蓄倾向等差异)。

此外,受 Lucas(1988)内生增长理论的影响,经济学家们还尤为关注社会保障制度通过影响人力资本积累对经济增长的作用机理,对居民子女教育投资行为的影响成为其中一个重要的分析视角。Pogue and Sgontz(1977)很早就认识到,社会保障制度可能会激励人们增加子女教育投入。Zhang(1995)利用一个 OLG 模型对此进行了较为严谨的理论分析。他沿循 Becker and Barro(1988)的思路,考虑了父母的利他主义并将居民生育决策内生化,发现现收现付制会促使居民减少生育数量而更关注子女质量,故会增加子女教育投资,促进人力资本积累和经济增长,基金积累制则具有相反作用;但他忽略了父母的利己主义(家庭保障)动机和子女对父母的赡养行为。此后的一些理论研究表明,考虑了这些因素后,现收现付制更倾向于削弱家庭保障机制、抑制人力资本积累(Ehrlich and Lui, 1998; Ehrlich and Kim, 2007; 郭庆旺等, 2007)。^①社会保障制度对公共教育的影响亦是文献关注的重点,分析的焦点在于其如何影响民众对公共教育的政治支持。Soares(2006)指出,现收现付制会削弱民众对公共教育的政治支持,抑制人力资本积累。Kaganovich and Zilcha(2012)则认为公共教育投入增加可提升社会保障收益,故社会保障制度(尤其基金积累制)会强化民众对公共教育的政治支持。

上述分析表明,已有研究关于社会保障制度通过物质资本积累和人力资本积累机制对经济增长究竟会产生怎样的影响以及哪些因素在其中扮演着重要角色还远未形成一致看法。现实经济中,影响社会保障制度通过这两种机制发挥作用的因素众多。已有理论研究出于模型可行性的考虑和关注问题的不同均忽略了一些重要因素,故需要经验研究予以厘清(Auerbach and Kotlikoff, 1987)。这也正是本文研究的一个主要目的,而此无疑具有很好的现实意义。特别地,在人口老龄化的背景下,人们之所以会对社会保障制度的可持续性感到担忧,主要是源于赡养比增加会加重居民保障负担。如果社会保障制度确实会抑制居民储蓄,则将会进一步加剧这一问题,这就需要相应的制度变革(如削减保障水平)来化解潜在的保障危机。但是,如果社会保障制度可以促进人力资本积累,提升劳动生产率,那么收入水平的提高就可部分(甚至全部)抵消赡养比增加带来的不利冲击,社会保障制度则无需较大调整(Kaganovich and Zilcha, 2012)。

为此,学术界进行了大量经验研究。Feldstein(1974, 1996)利用美国时序数据发现,社会保障税增加显著抑制了居民储蓄和物质资本积累。赵怡(2007)针对中国的研究也得到了类似结论,Cigno et al.(2003)基于德国的经验分析则给出相反结论。还有一些文献聚焦于跨国分析。Ehrlich and Zhong(1998)基于 49 个国家的分析表明,社会保障税(以社会保障收益占 GDP 的比值度量)提高不利于物质资本积累和经济增长。Ehrlich and Kim(2007)利用 57 个国家的面板数据和类似指标发现,社会保障对人力资本积累具有显著的抑制作用。Zhang and Zhang(2004)的跨国分析则表明,社会保障支出比率对物质资本积累的影响是中性的,对人力资本积累和经济增长则具有促进作用。

不过,这些文献大多没有很好地解决社会保障制度可能存在的自选择问题。一些研究尝试利用工具变量法来校正这一问题。例如,Zhang and Zhang(2004)以及 Ehrlich and Kim(2007)利用社会保障制度的建立年限作为工具变量,但这并不能很好地矫正制度本身存在的选择偏差问题。为此,Feldstein and Liebman(2002)主张利用拟自然实验分析来有效解决这一问题。白重恩等(2012)利用倾向得分匹配—双差分法构造反事实,在拟自然实验环境下有效校正了选择偏差,估计出中国新农

① 郭凯明和龚六堂(2012)的理论分析表明,社会保障制度会削弱家庭保障机制,促使居民减少生育数量,故有更多资源用于子女教育投资,因而有利于人力资本积累和经济增长。张川川和陈斌开(2014)利用中国新型农村社会养老保险的数据发现,社会保障制度对家庭保障机制具有一定的替代作用。

合制度对居民预防性储蓄的因果处置效应。王天宇和彭晓博(2015)利用同样的方法分析了新农合制度对居民生育意愿的影响。基于20世纪八九十年代大量发展中国家的制度变迁(陆续建立了社会保障制度),本文采取相似的计量策略。本文的另一个主要特色在于:比较全面地揭示出制约影响社会保障制度经济增长效应的重要因素——已有理论文献强调很多因素都可能对社会保障制度的增长效应产生重要影响,但经验研究还主要停留在经济发展水平的作用上。^①不过,限于数据的可获取性,本文也并非面面俱到,而是紧密结合理论与实践(尤其结合中国经济社会发展实践),着重考察了社会保障的制度安排、政策设计以及发展水平、储蓄倾向、人口老龄化和生育水平的制约影响。

三、计量策略

20世纪八九十年代,大量发展中国家陆续进行改革,建立实施了社会保障制度。以这些重要的制度变迁为契机,本文对社会保障制度的经济影响进行拟自然实验分析。特别地,本文旨在回答的问题是:在其他因素保持不变的情况下,若一国建立了社会保障制度,那么其经济增长将会发生怎样的变化?倘若能同时观测到每个国家处于改革和非改革两种状态下的经济增长状况,那么社会保障制度的因果处置效应就是它们之间的差异。然而,现实经济中只能观测到每个国家处于改革或是非改革某一实际状态下的增长状况,这就需要构造反事实(即非观测状态的)结果来识别社会保障制度对经济增长的因果处置效应。显然,任何一个国家的社会保障制度的建立及其制度政策安排并非是完全外生的、随机的,而通常是一个由一系列经济和社会等因素决定的自选择过程,而这些因素也会同时影响经济增长,故可能存在突出的自选择问题——很难区分究竟是制度建立造成了增长差异还是增长差异决定了制度的建立与否。因此,简单地将非改革国家的结果作为改革国家的反事实结果会带来选择偏差。

本文利用倾向得分匹配法来解决这一问题(Rosenbaum and Rubin, 1983)。首先估计出每个国家的倾向得分 $P(X)$,即在其他特征 X 给定的情况下,一国建立社会保障制度的预测概率;然后据此将改革国家(处置组)与非改革国家(对照组)样本进行匹配,以匹配成功后(即综合特征最相近)的对照组国家的结果作为处置组国家的反事实结果。在匹配样本满足条件独立分布(从而确保构造一个拟自然实验环境)和共同支撑假设下^②,社会保障制度的因果效应即平均处置效应(The Average Treatment Effect on the Treated, ATT)就是匹配两组在共同支撑域 S_p 上结果变量之差的均值,即:

$$\tau_{ATT}^{PSM} = \frac{1}{n_1} \sum_{i \in I_1 \cap S_p} \left\{ Y_{1i} - \sum_{j \in I_0 \cap S_p} W(i, j) Y_{0j} \right\} \quad (1)$$

其中, I_1 和 I_0 分别代表处置组和对照组国家, n_1 为落入 $I_1 \cap S_p$ 区域内的样本数量, Y_{1i} 和 Y_{0j} 分别为处置组和对照组国家的结果变量即经济增长率(本文还考虑了物质资本投资和人力资本投资), $W(i, j)$ 为对照组国家 j 的权重。目前,基于权重 $W(i, j)$ 的构造方法,匹配方法可分为内核匹配法(Kernel Matching)和最邻近匹配法(Nearest-neighbor Matching)。

倾向得分匹配法在校正选择偏差方面具有明显的优势,但其只依赖于观测变量、无法消除非观

① 例如,Ehrlich and Zhong(1998)发现,经济发展水平越高,社会保障的不利影响越突出。

② 条件独立分布假设是指在给定倾向得分的条件下,一国是否建立社会保障制度与其经济增长状况之间是独立的。满足这一条件意味着制度建立是条件外生的,因此可以通过比较处置组和对照组的经济增长状况来识别制度建立的因果效应。共同支撑假设是指所有的处置组和对照组国家建立社会保障制度的预测概率严格处在0—1之间,从而排除了极端情况,在此情况下寻求这两组预测概率的最大交集即共同支撑域。共同支撑假设排除了预测得分分布在两端的样本,从而提高了匹配质量,当然也会导致样本量减小。

测因素可能带来的估计偏差(Dehejia, 2005)。这一不足对于本文的跨国分析而言尤为突出,因为不同国家的传统文化信念、地理环境和资源禀赋等因素往往存在明显差异,而世界金融危机等亦会对各国经济产生共同冲击影响。为此,Heckman et al.(1998)以及 Caliendo and Kopeinig(2008)主张将倾向得分匹配和双差分法结合起来使用,以充分发挥双差分法的优势——利用面板数据和双差分有效控制非观测因素包括共同冲击和不随时间变化(或持续性很强)的个体特定因素的影响。Glazerman et al.(2003)也指出:倾向得分匹配法是一种能显著降低偏差的非参数估计方法,尤其与双差分等方法相结合会更加有效。而且,这一组合利用非参数估计构建双差分,亦可弥补双差分法在估算结果变量条件期望时要求函数必须为线性的不足(Smith and Todd, 2005)。

因此,本文以跨国面板数据为基础,利用倾向得分匹配—双差分法控制观测和非观测因素的影响,有效校正选择偏差,识别出社会保障制度对经济增长的平均处置效应,即:

$$\tau_{ATT}^{PSM-DID} = \frac{1}{n_1} \sum_{i \in I_1 \cap S_p} \left\{ (Y_{li}^{t_1} - Y_{li}^{t_0}) - \sum_{j \in I_0 \cap S_p} W(i, j) (Y_{0j}^{t_1} - Y_{0j}^{t_0}) \right\} \quad (2)$$

其中, t_0 和 t_1 分别为改革前时点和改革后时点(即事前和事后时点)。(2)式表明,倾向得分匹配—双差分法通过匹配后的处置组和对照组自身的一次差分即 $(Y_{li}^{t_1} - Y_{li}^{t_0})$ 和 $(Y_{0j}^{t_1} - Y_{0j}^{t_0})$ 消除了两组样本自身的变化趋势,这两项的再次差分给出社会保障制度的因果影响。

四、数据与描述性统计

本文使用的是 114 个发展中国家 1981—2005 年的面板数据。^①之所以选择以发展中国家作为研究对象,主要是因为:①20 世纪八九十年代以来,大量发展中国家陆续建立了社会保障制度,这为本文运用拟自然实验分析考察社会保障制度的经济影响提供了一个良好契机;②发达国家建立社会保障制度的时间普遍较早(例如,美国在 1935 年就已建立了社会保障制度),数据收集存在很大困难。

本文数据来源如下:样本国家建立社会保障制度的时间和制度类型等信息来自国际社会保障

① 发展中国家是按照国际货币基金组织(IMF)2005 年发布的《世界经济展望》中的标准划分的,包括非洲:阿尔及利亚、安哥拉、贝宁、博茨瓦纳、布基纳法索、布隆迪、佛得角、中非共和国、乍得、科摩罗、刚果民主共和国、刚果共和国、哥斯达黎加、科特迪瓦、吉布提、赤道几内亚、厄立特里亚、埃塞俄比亚、加蓬、冈比亚、加纳、几内亚、几内亚比绍、利比里亚、利比亚、马拉维、马里、毛里求斯、摩洛哥、莫桑比克、纳米比亚、尼日利亚、卢旺达、圣多美和普林西比共和国、塞内加尔、塞舌尔、塞拉利昂、索马里、苏丹、坦桑尼亚联合共和国、多哥、突尼斯、乌干达、赞比亚、津巴布韦;亚洲和大洋洲:阿富汗、巴林、孟加拉、柬埔寨、斐济、伊朗、伊拉克、科威特、老挝、马来西亚、马尔代夫、蒙古、缅甸、阿曼、巴基斯坦、卡塔尔、萨摩亚、沙特阿拉伯、所罗门群岛、斯里兰卡、泰国、汤加、图瓦卢、瓦努阿图、越南、也门;美洲:巴哈马群岛、伯利兹、哥伦比亚、古巴、多米尼克、多米尼加、厄瓜多尔、格林纳达、圭亚那、洪都拉斯、牙买加、巴拿马、巴拉圭、秘鲁、波多黎各、委内瑞拉;欧洲:亚美尼亚、阿塞拜疆、白俄罗斯、波斯尼亚和黑塞哥维那、保加利亚、克罗地亚、捷克共和国、爱沙尼亚、格鲁吉亚、匈牙利、哈萨克斯坦、吉尔吉斯共和国、拉脱维亚、立陶宛、马其顿、马耳他、摩尔多瓦、罗马尼亚、塞尔维亚、斯洛伐克、斯洛文尼亚、塔吉克斯坦、土耳其、土库曼斯坦、乌兹别克斯坦(国际社会保障协会网站把亚美尼亚、阿塞拜疆、格鲁吉亚、乌兹别克斯坦、哈萨克斯坦、吉尔吉斯共和国、塔吉克斯坦、土库曼斯坦都归入欧洲地区)。由于印度、巴西和南非建立社会保障制度的时间较早(分别为 1952 年、1923 年和 1928 年),数据难以获取,故本文没有考虑这些国家。由于数据可比性的问题,本文也没有考虑中国。

网站)^①，相应国家的社会保障缴费(税)率和社会保障支出数据来自国际货币基金组织(IMF)的 GFS (Government Finance Statistics)数据库和国际劳工组织(International Labor Office,ILO)出版的 The Cost of Social Security,实际人均 GDP(2005 年为基期,购买力平价调整后)和物质资本投资数据主要来自 Heston et al.(2009)构造的 PWT(Penn World Table) 6.3。为了尽可能较为全面地考虑影响社会保障制度建立的因素,本文从联合国网站数据库(<http://data.un.org>)和世界银行网站数据库(<http://data.worldbank.org.cn>)补充了其他数据。

表 1 给出样本国家的地区分布以及处置组国家建立社会保障制度的时间分布和制度类型。从中可知,114 个样本国家中非洲国家最多(47 个),美洲国家最少(16 个),其中有 73 个为对照组国家(样本期内一直没有建立社会保障制度的国家)。处置组国家(此间建立实施了社会保障制度的国家)有 41 个(占全部样本国家的 35.9649%),其中绝大部分国家采取的是基金积累制(包括部分基金制和完全基金制,共 40 个,占全部处置组国家的 97.5610%)。这些处置组国家建立社会保障制度的时间分布在 1986—2001 年,尤以 1986—1991 年(12 个,占全部处置组国家的 29.2683%)和 1997—2001 年(22 个,占全部处置组国家的 53.6585%)这两个时期较为集中。本文以 1981—1985 年作为事前时点,以确保事前两组样本(此间的所有样本国家)均未建立社会保障制度,这样处理也可减少改革预期效应造成的估计偏差;将 2002—2005 年作为事后时点,以尽可能捕捉社会保障制度较长期的影响,同时也是为了确保有足够多的对照组样本以保障匹配质量——2005 年以后,几乎所有样本国家都建立了社会保障制度。

表 1 样本国家的地区分布与处置组国家社会保障制度建立的时间分布和制度类型

地区	样本 国家数	处置组国家社会保障制度类型			处置组国家建立社会保障制度的时间分布				
		现收 现付制	基金 积累制	合计	1981— 1985	1986— 1991	1992— 1996	1997— 2001	2002— 2005
非洲	47	0	13	13	0	2	3	8	0
亚洲和大洋洲	25	0	9	9	0	6	2	1	0
美洲	16	0	8	8	0	3	2	3	0
欧洲	26	1	10	11	0	1	0	10	0
合计	114	1	40	41	0	12	7	22	0
		(2.4390)	(97.5610)			(29.2683)	(17.0732)	(53.6585)	

注:基金积累制包括部分和完全基金积累制,小括号中的数字为该类型处置组国家占全部处置组国家的比重(%)。

对于结果变量,本文利用实际人均 GDP 增长率刻画一国的经济增长状况。为了检验影响机理,本文也考虑了物质资本投资和人力资本投资;就物质资本投资而言,本文采用的是物质资本投资率

① 国际社会保障网站(<http://www.issa.int>)是国际社会保障协会(International Social Security Association)的官方网站。根据这一协会的定义,社会保障是指通过立法或其他强制性安排建立的为个人提供的各种保障计划,主要包括社会养老保障、医疗保险、失业保险、工伤保险和最低生活保障等内容。因此,本文研究的社会保障是指这样的一揽子保障计划。本文的样本国家均为发展中国家,这些国家在建立社会保障制度时普遍是以发达国家的社会保障制度为蓝本,借鉴了其中的很多做法。例如,同时建立了涵盖养老、医疗和失业等一揽子保障计划,规定了较为严格的享受养老保障的资格和条件(普遍需要达到规定的退休年龄和养老保险的缴纳期限等)。但受自身经济社会发展等因素的影响,采取的制度类型(例如,部分基金制还是完全基金积累制)以及社会保障水平等方面存在较为明显的差异。

增长率;由于发展中国家普遍缺乏受教育年限的数据,本文利用基础教育净入学率增长率刻画人力资本投资^①。表 2 给出事前和事后时点整个样本、处置组和对照组国家结果变量的均值和标准差,以及处置组和对照组结果变量的均值差异。

表 2 显示,样本期内,整个样本国家的实际人均 GDP 增长率从事前时点(1981—1985 年)的-0.8167%增加到事后时点(2002—2005 年)的 3.3101%,提升了 4.1268 个百分点,这一变化具有良好的统计显著性。处置组和对照组的状况没有明显差异;处置组国家的实际人均 GDP 增长率由事前时点的 0.0491%增加到事后时点 3.3471%,提升了 3.2980 个百分点;对照组国家的状况类似,实际人均 GDP 增长率从事前时点的-1.3261%增加到了事后时点的 3.2884%,提升了 4.6144 个百分点。两组国家的实际人均 GDP 增长率变化的差异为-1.3164 个百分点,但不具有统计显著性(见表 2 第(4)列 Δ_t 行)。物质资本投资也呈现出类似的变化特征:样本期内,处置组的物质资本投资率增长率增加了 4.9539 个百分点,对照组增加了 6.4236 个百分点,两组间的差异为-1.4697 个百分点,但不具有统计显著性。样本期内,处置组(对照组)国家的基础教育净入学率增长率略有下降(上

表 2 结果变量的统计描述

结果变量	年	均值(标准差)			差异: (2)-(3)= (4)
		全样本 (1)	处置组 (2)	对照组 (3)	
实际人均 GDP 增长率(%)	1981—1985	-0.8167 (4.1173)	0.0491 (3.6074)	-1.3261 (4.3344)	1.3752* (0.8134)
	2002—2005	3.3101 (3.3141)	3.3471 (3.2266)	3.2884 (3.3881)	0.0587 (0.6635)
	Δ_t	4.1268*** (0.5320)	3.2980*** (0.8372)	4.6144*** (0.6842)	-1.3164 (1.0994)
	样本量	108	40	68	
物质资本投资率增长率(%)	1981—1985	-3.2312 (7.0402)	-3.0004 (5.1802)	-3.3696 (7.9865)	0.3693 (1.4325)
	2002—2005	2.6413 (8.1385)	1.9535 (6.9115)	3.0540 (8.8189)	-1.1005 (1.6529)
	Δ_t	5.8725*** (1.0961)	4.9539*** (1.4164)	6.4236*** (1.5383)	-1.4697 (2.2705)
	样本量	104	39	65	
基础教育净入学率增长率(%)	1981—1985	0.5908 (1.7838)	0.4801 (1.8973)	0.6576 (1.7252)	-0.1775 (0.3834)
	2002—2005	0.5955 (2.0604)	0.1493 (1.2214)	0.8648 (2.4013)	-0.7156 (0.4370)
	Δ_t	0.0047 (0.2699)	-0.3308 (0.3319)	0.2072 (0.3833)	-0.5380 (0.5573)
	样本量	93	35	58	

注:由于部分国家在事前时点(1981—1985 年)或事后时点(2002—2005 年)数据缺失,为确保结果的可比性,本文在计量分析中仅保留了两个时点均有完整数据记录的样本。 Δ_t 行的数值是其上一行与上两行的差,第(4)列数值是第(2)列与第(3)列数值的差, Δ_t 行和第(4)列括号里的数字为 t 检验的标准误,*、** 和 *** 分别表示在 10%、5% 和 1% 的置信水平上显著。

① 本文也尝试利用识字率(样本数量明显较少)来刻画人力资本投资,结果保持了较好的一致性。

升),但两组间的差异也不显著。不过,由于可能存在选择偏差,很难简单认为社会保障制度的建立对经济增长、物质资本投资和人力资本投资没有影响。

匹配变量的选取(即倾向得分估计模型的设定)对于处置组和对照组样本匹配质量(进而选择偏差的矫正)至关重要。为此,结合已有文献(Zhang and Zhang, 2004; Ehrlich and Kim, 2007)和各国社会保障制度建立实践,本文选取了如下 8 个匹配变量:实际人均 GDP(取自然对数)、人口老龄化程度(用 65 岁及以上人口占总人口的比重度量)、出生率(每千人中当年新生并存活下来的人口数)、死亡率(每千人中当年死亡的人口数)、人口规模(总人口数,取自然对数)、农村人口比率(农村人口占总人口比重)、开放度和消费率。由下文的平衡性检验看,这组匹配变量确保了较高的样本匹配质量。

这些变量采用的都是 1981—1985 年(即事前时点)的均值。这样的处理不仅可以确保它们不受社会保障制度建立及其预期效应的影响,也可以避免短期波动的影响。对于这些变量,本文剔除了小于 1%分位数和大于 99%分位数的观测值以避免异常值的影响。由表 3 可知:匹配前,相较于对照组而言,处置组在事前时点具有较低的死亡率和消费率,表明社会保障制度的建立并非随机的,需要校正选择偏差问题。

表 3 匹配变量的描述性统计和倾向得分估计结果

事前时点	均值(标准差)			差异: (2)-(3)= (4)	倾向得分估计: probit 回归 (5)
	全样本 (1)	处置组 (2)	对照组 (3)		
实际人均 GDP(取自然对数)	8.2991 (0.9335)	8.4447 (0.6790)	8.2049 (1.0623)	0.2398 (0.2082)	-0.2397 (0.3269)
人口老龄化程度(%)	4.6764 (2.5933)	4.8237 (2.5650)	4.5811 (2.6325)	0.2425 (0.5823)	0.2059 (0.1507)
出生率(‰)	35.1670 (10.8056)	33.9197 (9.8631)	35.9741 (11.3961)	-2.0544 (2.4181)	0.0573 (0.0409)
死亡率(‰)	10.7659 (4.1919)	9.5637 (3.4892)	11.5438 (4.4505)	-1.9801** (0.9165)	-0.1736** (0.0725)
人口规模(取自然对数)	15.0560 (1.7701)	15.0939 (1.8174)	15.0315 (1.7567)	0.0624 (0.3978)	0.0346 (0.0956)
农村人口比率(%)	57.6373 (19.4022)	55.9661 (18.1746)	58.7188 (20.2599)	-2.7527 (4.3503)	0.0016 (0.0123)
开放度(%)	83.2589 (53.7210)	84.1276 (44.0590)	82.6969 (59.5607)	1.4307 (12.0736)	-0.0022 (0.0031)
消费率(%)	66.1596 (19.3953)	61.1519 (15.9324)	69.3999 (20.8504)	-8.2480* (4.2632)	-0.0172* (0.0095)
R ²					0.0958
样本数	84	33	51		84

注:为确保描述性统计和倾向得分估计的样本量是一致的,从而具有良好的可比性,本文在描述性统计中剔除了匹配变量缺失的样本。第(4)和(5)列小括号里的数字为标准误。倾向得分估计模型的回归结果略去了截距项的回归系数。*、** 和 *** 分别表示在 10%、5%和 1%的置信水平上显著。

五、社会保障制度的经济增长效应

本节首先对处置组和对照组样本进行匹配,在确保匹配样本较好地满足平衡性和共同支撑条件后,识别出社会保障制度对经济增长的因果处置效应及其机理,最后进行稳健性分析。

1. 样本匹配

本文采用如下社会保障制度建立实施概率的 probit 模型,估计出每个样本国家建立社会保障制度的预测概率,即倾向得分:^①

$$\text{probit}(\text{treat}_i=1)=\alpha+\beta X_i+\varepsilon_i \quad (3)$$

其中, treat_i 为社会保障制度建立哑变量:样本国家若在 1981—2005 年建立了社会保障制度,则赋值为 1,否则为零。 X_i 为匹配变量,如前所述,包括实际人均 GDP(取自然对数)、人口老龄化程度、出生率、死亡率、人口规模(取自然对数)、农村人口比率、开放度和消费率。

表 3 第(5)列给出模型(3)的具体估计结果。从中可知:一国的死亡率和消费率越低,其越有可能建立实施社会保障制度,这与前文简单的数据统计描述结果一致,也进一步表明社会保障制度的建立并非随机的,存在样本选择问题。依据倾向得分,本文对处置组和对照组样本进行匹配。借鉴已有文献的做法(万海远和李实,2013),本文采取了内核匹配(Epanechnikov 内核,带宽为 0.10)和 5 对 1 最邻近匹配法以确保结果的稳健性。^②

2. 平衡性与共同支撑检验

本节首先对匹配后样本进行平衡性检验,即检验匹配后的样本是否满足条件独立分布假设。这一识别条件要求匹配后样本不存在系统性差异,这得到了检验结果的良好证明(见表 4)。特别地,t 检验表明,匹配后样本的所有匹配变量都不存在显著的组间均值差异,而且各变量的标准化差异较匹配前均有较大幅度的下降(除人口规模以外)。此外,基于匹配后样本估计的 probit 模型的 R^2 很小,表明此时匹配变量对于一国是否建立社会保障制度的解释力很弱,即社会保障制度的建立与否对于匹配后样本而言是条件随机的。^③

此外,正如前文指出的,倾向得分匹配法还需要满足共同支撑条件。这一条件可以确保匹配后样本具有良好的可比性,因此可以显著提高样本匹配质量、增加倾向得分匹配法估计的有效性(Heckman et al., 1998)。但需要注意的是,共同支撑域是整个匹配样本的一个子集,因此,倘若共同支撑域内的匹配样本数量较少,那么倾向得分匹配法识别出的只是一个子集效应(Lechner, 2001)。为此,本文也检验了这一条件。图 1 显示:匹配前,两组样本倾向得分的概率分布存在明显差异且重叠区域(共同支撑域)较小;匹配后,它们的分布具有较好的一致性,且重合区域足够大。因此,上述问题并不突出(只有少数样本落在共同支撑域外,估计处置效应时剔除了这些样本),可以确保估计的平均处置效应是准确可靠的。

① 本文也尝试采用 logit 模型来估计倾向得分,结果没有明显变化。

② 在使用内核匹配法时,本文也尝试了 Gaussian 内核,结果没有明显差异。

③ 处置组和对照组匹配变量的标准化差异为: $(\bar{X}_1 - \bar{X}_0) / \sqrt{0.5(S_1^2 + S_0^2)}$, $\bar{X}_1$ 和 S_1^2 分别为处置组匹配变量的均值和方差, $\bar{X}_0$ 和 S_0^2 分别为对照组匹配变量的均值和方差。该值越小意味组间差异越小。目前,实践中通常使用 10%、15%,甚至 20%作为认定平衡性的标准。就整体平衡性检验而言,匹配后匹配变量对制度建立概率的解释力越差(即基于匹配后样本估算的 probit 倾向得分模型的 R^2 越小),匹配后两组的分布差异越小,匹配质量越高(Caliendo and Kopeinig, 2008)。

表 4 平衡性检验结果						
变量	样本	均值差异检验			标准化差异检验	
		处置组均值	对照组均值	t 检验 (p 值)	标准化差异	降幅 (%)
实际人均 GDP(取自然对数)	匹配前	8.4447	8.2049	1.1500 (0.2530)	26.9000	96.4000
	匹配后	8.4416	8.4502	-0.0400 (0.9680)	-1.0000	
人口老龄化程度 (%)	匹配前	4.8237	4.5811	0.4200 (0.6780)	9.3000	36.1000
	匹配后	4.8391	4.9942	-0.2300 (0.8170)	-6.0000	
出生率 (‰)	匹配前	33.9200	35.9740	-0.8500 (0.3980)	-19.3000	82.7000
	匹配后	33.7390	33.3840	0.1300 (0.8940)	3.3000	
死亡率 (‰)	匹配前	9.5637	11.5440	-2.1600 (0.0340)	-49.5000	99.2000
	匹配后	9.6746	9.6586	0.0200 (0.9850)	0.4000	
人口规模(取自然对数)	匹配前	15.0940	15.0310	0.1600 (0.8760)	3.5000	-155.6000
	匹配后	15.1920	15.0320	0.3600 (0.7170)	8.9000	
农村人口比率 (%)	匹配前	55.9660	58.7190	-0.6300 (0.5290)	-14.3000	59.9000
	匹配后	56.1150	55.0120	0.2300 (0.8190)	5.7000	
开放度 (%)	匹配前	84.1280	82.6970	0.1200 (0.9060)	2.7000	41.7000
	匹配后	83.9940	84.8280	-0.0600 (0.9540)	-1.6000	
消费率 (%)	匹配前	61.1520	69.4000	-1.9300 (0.0560)	-44.5000	97.6000
	匹配后	60.9850	61.1860	-0.0500 (0.9640)	-1.1000	
Pseudo R ²	匹配前				0.0960	
	匹配后				0.0050	

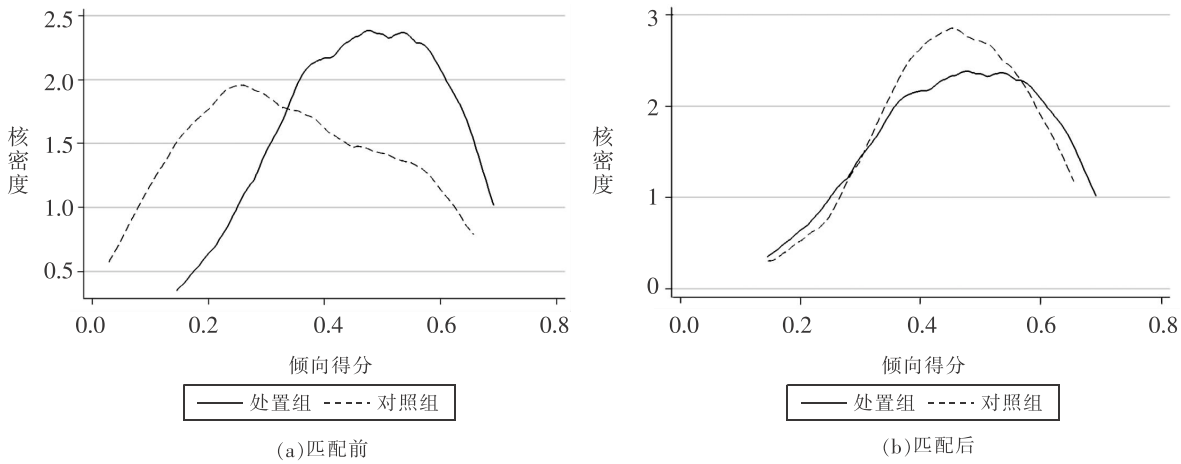


图 1 样本匹配前后倾向得分的核密度

3. 因果处置效应

在确保匹配样本较好地满足了条件独立分布和共同支撑条件后,本文依据公式(2)估计出社会保障制度对实际人均 GDP 增长率、物质资本投资率增长率和基础教育净入学率增长率的平均处置效应,具体结果见表 5。总体而言,内核匹配法和 5 对 1 最邻近匹配法估计得到的平均处置效应具有较好的一致性。

具体而言,社会保障制度对实际人均 GDP 增长率的平均处置效应为负,且具有较好的统计显著性。这一负效应在内核匹配法和 5 对 1 最邻近匹配法下分别为-2.6269 和-2.5645,且都在 5%的置信水平上显著,表明样本期内社会保障制度导致处置组国家的实际人均 GDP 增长率平均下降了 2.5957 个百分点。这一不利影响倾向于主要是通过物质资本积累渠道发挥作用的:社会保障制度对物质资本投资率增长率的平均处置效应为-4.7651 和-5.6726,且至少在 10%的置信水平上显著,导致处置组国家的物质资本投资率增长率平均减少了 5.2189 个百分点,较 1981—1985 年下降了 173.94%。上述结果与大多数已有研究的结论相一致(Feldstein, 1996)。与之不同的是:社会保障制度对人力资本积累的影响较弱,表现在对基础教育净入学率增长率的处置效应为负,但不具有统计显著性。

表 5 社会保障制度的平均处置效应

	内核匹配 (1)	5 对 1 最近邻匹配 (2)
实际人均 GDP 增长率		
平均处置效应	-2.6269** (1.0953)	-2.5645** (1.1351)
处置组样本	32	32
对照组样本	51	51
总样本	83	83
物质资本投资率增长率		
平均处置效应	-4.7651* (2.6758)	-5.6726** (2.7144)
处置组样本	30	30
对照组样本	50	50
总样本	80	80
基础教育净入学率增长率		
平均处置效应	-0.1132 (0.7747)	-0.0590 (0.7267)
处置组样本	27	27
对照组样本	41	41
总样本	68	68

注:小括号中的数字为标准误。***、** 和 * 分别表示在 1%、5%和 10%水平上显著。下表同。

4. 稳健性分析

接下来,本文进行稳健性分析以检验基准结论的可靠性,具体检验结果见表 6。^①考虑到倾向得分匹配的共同支撑条件可能受到倾向得分分布的尾部的影响(Black and Smith, 2004),本文利用修

① 这一节和下节的分析只给出 5 对 1 最邻近匹配法的估计结果,内核匹配法的结果存在一些差异,但总体结论保持了较好的一致性。遵循一些文献的做法(Abadie et al., 2010; Billmeier and Nannicini, 2013; 刘甲炎和范子英, 2013; 刘乃全和吴友, 2017), 本文也尝试利用合成控制法(Synthetic Control Method)考察社会保障制度的经济影响,但合成效果较差。原因可能在于:本文的样本期跨度较长(1981—2005 年)而且样本国家均为发展中国家,数据搜集存在一定困难,导致数据缺失较多,很大程度上影响了合成控制法的效果。而且,合成控制法只能就单个分析单元的评估变量进行合成,若对多个分析单元进行整体评估时,则需要将多个分析单元合并为一个分析单元(Abadie et al., 2010; 刘甲炎和范子英, 2013; 刘乃全和吴友, 2017)。本文为跨国分析,处置组国家存在较大的异质性,加总为一个单元的做法很可能会带来较大问题,这也在较大程度上限制了合成控制法在本文的应用。实际上,本文曾尝试了多种方案,例如,以各大洲或较小的特定地区(如东欧地区)为单位将处置组国家加总,但效果都较差。今后,在更好地获取数据的基础上,利用合成控制法对社会保障制度的经济影响进行更加细致准确的评估无疑是一个非常重要的主题,笔者也将持续追踪这一问题。十分感谢匿名评审专家就此提出的宝贵建议。

剪策略(Trimming Strategy)考察基准结果对处置组倾向得分分布尾部部分极端样本的稳健性。若修剪之后的结果仍与基准结果一致，则表明基准结果并不依赖于倾向得分分布的尾部，是较为可靠的。具体而言，本文分别尝试了 2%、5% 和 10% 三种修剪水平，即剔除处置组倾向得分分布的尾部 2%、5% 和 10% 的样本，结果详见表 6 的第(1)—(3)列。^① 可以看出，基于不同修剪水平修剪后样本的结果与基准结果具有较好的一致性，表明本文基准结果的可靠性较高。

表 6		稳健性检验结果		
5 对 1 最邻近匹配	修剪水平			平行趋势检验(安慰剂检验): 样本期限定在 1981—1985 年 (4)
	2% (1)	5% (2)	10% (3)	
实际人均 GDP 增长率				
平均处置效应	-2.5704** (1.1383)	-2.5645** (1.1351)	-2.3616** (1.1590)	-0.6756 (1.9519)
处置组样本	33	32	30	30
对照组样本	51	51	51	47
总样本	84	83	81	77
物质资本投资率增长率				
平均处置效应	-5.5194** (2.6961)	-5.6726** (2.7144)	-5.9493** (2.7708)	3.1301 (7.3254)
处置组样本	31	30	28	30
对照组样本	50	50	50	50
总样本	81	80	78	80
基础教育净入学率增长率				
平均处置效应	0.1341 (0.7397)	0.0935 (0.7378)	-0.0590 (0.7267)	-0.4426 (0.6331)
处置组样本	29	28	27	27
对照组样本	41	41	41	40
总样本	70	69	68	67

本文将倾向得分和双差分法结合起来使用主要是希望充分发挥双差分法的优势，以消除非观测因素带来的估计偏差。但能否有效做到这一点还取决于平行趋势假设是否成立，即处置组和对照组的结果变量在事前时点是否具有相同的变化趋势。若这一识别假设不成立，即两组国家的结果变量在制度建立前就已经存在着趋势差异，则估计结果将会有偏。为检验该条件是否成立，本文构造了一个安慰剂检验(Placebo Test)：将样本期限定为事前时点(即 1981—1985 年)，并假定处置组国家此间就已建立了社会保障制度，利用倾向得分匹配—双差分法估计出平均处置效应。这是一个反事实分析：处置组国家在 1981—1985 年并没有建立社会保障制度，因此如果得到的处置效应是显著的，说明平行趋势假设不成立；反之，该假设成立。表 6 第(4)列给出的检验结果显示：此时，社会保障制度对实际人均 GDP 增长率、物质资本投资率增长率和基础教育净入学率增长率的处置效应都不具有统计显著性，故平行趋势假设成立。这较好地证实了本文的计量策略是有效的，估计结果是准确可靠的。

① 需要说明的是，为满足共同支撑条件，本文在估算基准结果时剔除了倾向得分高于和低于对照组得分最大值和最小值的处置组样本；而在使用修剪策略时，本文剔除的是处置组倾向得分分布尾部的一定比例(修剪水平)的样本。由于这两种方法剔除的目标样本不同，因此在修剪策略下处置组样本的数量可能多于或少于基准分析的数量。

六、社会保障制度经济增长效应的影响因素

本节进一步通过分样本分析来考察社会保障制度对经济增长的异质性影响,识别制约影响社会保障制度经济增长效应的重要因素。

1. 制度安排和政策设计

前文分析较好地揭示出社会保障制度的经济影响,但忽略了一个重要问题,即制度安排的影响。正如表 1 所示,处置组国家采取的社会保障制度类型存在较为明显的差异,而已有研究普遍认为,现收现付制、部分基金制和完全基金制对微观主体行为进而对经济增长的影响不同(Feldstein and Liebman, 2002)。但遗憾的是,在本文的样本中,仅有一个国家采取的是现收现付制,因此,本文无法剖析现收现付制与基金积累制的影响差异。本文首先尝试将这一样本国家剔除掉,结果没有变化。^①进而,本文尝试将处置组国家分为两组,即部分基金制组和完全基金制组,分别估计这两组的平均处置效应。但由于采取完全基金制的样本国家数量也比较少,致使本文无法准确识别出其平均处置效应,故表 7 第(2)列仅给出部分基金制的平均处置效应。从中可以看出,结果均不显著。因此,单纯从经济增长和物质资本积累的角度看,部分基金制似乎更为可取。当然,实践中是否应采取部分基金制还需要综合考虑基金管理成本、金融投资风险和收入分配等问题(Feldstein and Liebman, 2002)。

政策设计无疑也是决定社会保障制度经济影响的一个重要维度。目前,已有研究主要利用社会保障支出比率(或社会保障税费率)等单一指标来捕捉社会保障的影响。但社会保障涉及收支两个层面,而现实经济中各国的社会保障普遍存在收支缺口,收支变化往往并不同步(Cigno et al., 2003; Romer and Romer, 2016)。为此,本文分别从待遇和税费水平两个方面来考察政策设计的影响:①将处置组国家分为:高支出组,即样本期内社会保障支出比率均值大于处置组中位数(3.6068%)的国家;低支出组,即样本期内社会保障支出比率均值小于中位数的国家,分别估计了它们的处置效应。结果表明:对于高社会保障支出组而言,处置效应均不显著;与之不同,低社会保障支出组的结果与基准结果较为一致,对经济增长和物质资本积累均具有显著的负影响(见表 7 第(3)和(4)列)。②将处置组国家分为:高缴费组,即样本期内社保缴费率均值大于处置组中位数(4.4317%)的国家;低缴费组,即样本期内社保缴费率均值小于中位数的国家。表 7 第(5)和(6)列显示:与按社保支出比率分组的结果类似,低缴费组的社会保障制度的不利影响更加突出。究其原因,可能在于本文的样本国家均为发展中国家,这些国家社会保障的待遇水平普遍偏低,故而不利于经济增长。因此,一个待遇水平适度(不能过低亦不能过高)的社保体系可能更为可取。

2. 发展水平和储蓄倾向

Ehrlich and Zhong(1998)发现,相比于发展水平较低的国家而言,社会保障对高收入国家经济增长的不利影响更加突出。本文也尝试将处置组国家划分为两组:较高发展组,即事前时点的实际人均 GDP 大于中位数(4637.68 美元)的国家;较低发展组,即事前时点的实际人均 GDP 小于中位数的国家,来考察经济发展水平的影响。^②表 8 第(1)和(2)列显示,虽然同属于发展中国家,期初收入水平不同的国家的社会保障制度的影响存在明显差异:对于较高发展组而言,社会保障制度的影

① 在估算基准结果时,由于现收现付制国家样本落在共同支撑域外,已经被剔除,故基金积累制样本的结果与基准结果相同。

② 之所以以事前时点而非整个样本期内的实际人均 GDP 作为分组依据,是为了避免社会保障制度的影响带来的分组困扰。

表 7 制度安排和政策设计对社会保障制度增长效应的影响

5 对 1 最邻近匹配	基金积累制	部分基金制	高社会保障 支出组	低社会保障 支出组	高缴费率组	低缴费率组
	(1)	(2)	(3)	(4)	(5)	(6)
实际人均 GDP 增长率						
平均处置效应	-2.5645** (1.1351)	-0.6738 (1.1328)	-2.3761 (1.8432)	-2.5739** (1.1615)	-0.5613 (1.6339)	-3.1667** (1.3652)
处置组样本	32	27	14	13	9	10
对照组样本	51	51	51	51	51	51
总样本	83	78	65	64	60	61
物质资本投资率增长率						
平均处置效应	-5.6726** (2.7144)	-2.9576 (2.7303)	-0.9239 (3.1352)	-6.0563* (3.3168)	-2.0982 (3.2922)	-6.1452* (3.4206)
处置组样本	30	24	13	14	7	10
对照组样本	50	50	50	50	50	50
总样本	80	74	63	64	57	60
基础教育净入学率增长率						
平均处置效应	-0.0590 (0.7267)	0.0965 (0.8497)	-0.1632 (0.9112)	-0.2927 (0.9089)	0.3631 (1.0147)	0.0898 (1.1440)
处置组样本	27	24	12	13	8	10
对照组样本	41	41	41	41	41	41
总样本	68	65	53	54	49	51

注：基金积累制包括部分基金制和完全基金制。

响不显著；而对较低发展组的经济增长和物质资本积累则具有显著的抑制作用，可能是由于这些较落后的国家存在着更加突出的制度和市场缺陷所致。这一结果与 Ehrlich and Zhong(1998)的发现有所不同，可能的原因在于：Ehrlich and Zhong(1998)没有很好地解决选择偏差问题，而且他们的样本国家数量较少且以发达国家为主。

直观上，若一国居民普遍具有良好的节俭习惯(较低的时间偏好率)进而较高的储蓄倾向，那么社会保障制度对居民储蓄进而对物质资本积累的不利影响可能较弱 (Gustman and Steinmeier, 2015)。本文依据处置组国家在事前时点的储蓄率将它们分为两组：高储蓄组即储蓄率大于中位数(21.5421%)的国家和低储蓄组即储蓄率小于中位数的国家，借此来考察储蓄倾向的影响。这一做法并不完美，但能在较大程度上捕捉节俭习惯和储蓄倾向的影响。影响一国储蓄率的因素很多，而居民的节俭习惯和储蓄倾向无疑是最重要、影响也最持久的一个因素。表 8 第(3)和(4)列显示：对于那些期初具有较低储蓄率(较低储蓄倾向)的国家而言，社会保障制度的不利影响更加突出，而对高储蓄组的影响不显著，这较好地证实了上述推断。

3. 人口老龄化和生育水平

无疑，人口老龄化是近年来关于社会保障制度可持续性及其改革争论的一个重要导火索。一般而言，人口老龄化会减少劳动供给，加重整个社会的保障负担，进而可能会强化社会保障制度对经济增长的不利影响，形成负向效应的交互叠加。但也有研究指出：预期寿命的延长会促使人们在年轻时增加储蓄(Bloom et al., 2003)，且较低的生育率也可降低生活负担，促使居民增加储蓄、加大对子女的教育投资(Zhang, 1995)，这些都可一定程度上缓解社会保障制度的不利影响。为了厘清这一问题，本文将处置组国家分为两组：高老龄化组，即事前时点的人口老龄化程度高于中位数

表 8 发展水平、储蓄倾向、人口老龄化和生育水平对社会保障制度增长效应的影响

5 对 1 最邻近匹配	较高 发展组	较低 发展组	高 储蓄组	低 储蓄组	高 老龄化组	低 老龄化组	高 生育率组	低 生育率组
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
实际人均 GDP 增长率								
平均处置效应	-1.4674 (1.4225)	-2.2559** (1.1413)	-1.6411 (1.6548)	-2.9503** (1.3860)	-0.2511 (1.2900)	-4.4184*** (1.6046)	-2.8879* (1.5930)	-0.6372 (1.3840)
处置组样本	15	15	15	16	13	16	16	13
对照组样本	51	51	51	51	51	51	51	51
总样本	66	66	66	67	64	67	67	64
物质资本投资率增长率								
平均处置效应	-3.2500 (2.7375)	-9.2211*** (3.5887)	-4.4782 (3.0863)	-5.8792* (3.3919)	-5.5726* (3.1056)	-3.5400 (3.6970)	-4.5962 (3.8486)	-2.5112 (2.9517)
处置组样本	15	12	14	16	12	16	16	12
对照组样本	50	50	50	50	50	50	50	50
总样本	65	62	64	66	62	66	66	62
基础教育净入学率增长率								
平均处置效应	0.9610 (0.6989)	-2.3386** (1.0724)	0.5339 (0.7872)	0.3317 (0.9326)	0.3733 (0.7678)	-1.8720* (1.0361)	-2.4034** (1.0674)	0.8077 (0.7383)
处置组样本	15	12	11	12	13	12	8	13
对照组样本	41	41	41	41	41	41	41	41
总样本	56	53	52	53	54	53	49	54

(3.9120%) 的国家和低老龄化组即事前时点的人口老龄化程度低于中位数的国家,分别估计了这两组的平均处置效应。表 8 第(5)和(6)列显示:后一种机制倾向于更强,体现在社会保障制度对高老龄化组的经济增长的影响不显著,而显著降低了低老龄化组的实际人均 GDP 增长率和基础教育净入学率增长率。这与按生育水平分组的结果保持了较好的逻辑一致性:社会保障制度对那些期初生育水平较低的国家(低生育组即生育率低于中位数 4.4066% 的国家)的经济影响不显著,对高生育组的实际人均 GDP 增长率和基础教育净入学率增长率则具有显著的负影响(见表 8 第(7)和(8)列)。就此来看,人口老龄化和较低的生育水平并未加剧反而有利于缓解社会保障制度的不利影响。这是一个比较有意思的结果,意味着人口老龄化和低生育率这样的人口结构性变化固然会带来这样那样的问题,但亦有其积极有利的一面。

七、结论与政策建议

深入厘清社会保障制度的经济影响及其主要作用机理对于构建长期可持续的社会保障体系至关重要。20 世纪八九十年代,大量发展中国家陆续建立实施了社会保障制度。基于这些重要的制度变迁,本文对社会保障制度的经济影响进行拟自然实验分析。具体而言,本文利用 114 个发展中国家 1981—2005 年的面板数据和倾向得分匹配—双差分法构造反事实,有效校正了选择偏差问题,识别出社会保障制度对经济增长的因果处置效应。

研究表明,社会保障制度对经济增长具有显著的负影响,导致处置组国家实际人均 GDP 增长率平均下降了 2.5957 个百分点。这一不利影响倾向于主要通过物质资本积累机制发挥作用,导致物质资本投资率增长率平均减少了 5.2189 个百分点,较 1981—1985 年下降了 173.94%;对人力资

本积累的影响则较弱,体现在对基础教育净入学率增长率的处置效应不具有统计显著性。稳健性检验和平行趋势检验较好地证实了上述结果的可靠性。异质性效应分析还表明,较低的社保支出和税费水平会加剧社会保障制度的不利影响,较低的发展水平和储蓄倾向具有类似影响,人口老龄化和较低的生育率则有助于缓解社会保障制度的不利影响。

无疑,社会保障制度在增进居民福利、实现社会公平中扮演着极为重要的角色,但亦有可能对经济产生不利冲击、加剧财政风险。因此,如何有效促进民生发展和经济增长的协调共进始终是各国政府在健全完善社会保障体系时面临的一个巨大挑战。就本文的研究看,不同国家的做法应有所不同,应充分结合自身经济社会发展实践采取更具针对性的措施。本文的研究主要针对发展中国家,因此虽并未直接涉及中国,但其中的很多结论无疑对于中国长期可持续的社会保障制度建设具有良好的启示意义。特别地,改革开放以来,中国业已实现了40年的高速增长,收入水平得到了极大提升,预期寿命明显增加,而且居民始终保持着良好的节俭习惯和较高的储蓄倾向;与此同时,计划生育政策的实施也使得中国的生育水平大幅下降。依据本文的研究结论看,这些都意味着中国的社会保障制度对经济增长的不利影响很可能较弱。这为中国进一步提升社会保障水平以更好地增进居民福祉和维护社会公平、有效化解社会基本矛盾提供了较大的政策空间。不过,也应清醒地看到:目前,中国的社会保障综合缴费率已较高,社保资金缺口问题也较为突出。因此,在新时代,中国政府应:①坚持以人民为中心的发展思想,从顺应人民对美好生活需要的高度出发健全完善社会保障制度,建立与经济发展阶段相适应的中国特色社会保障体系;②避免社会保障支出的过快增长,逐步探索构建社会保障收益水平的动态调整机制,根据经济社会的发展变化和财政承受能力,适时适度地进行优化调整,既要尽力而为又要量力而行;③适度降低社会保障综合缴费率,切实减轻个人和企业等微观主体的负担,有效激发经济活力;④优化制度安排,逐步做实个人账户,积极探索有利于人力资本积累的体制机制设计。

〔参考文献〕

- [1]白重恩,李宏彬,吴斌珍. 医疗保险与消费:来自新型农村合作医疗的证据[J]. 经济研究, 2012,(2): 41-53.
- [2]郭凯明,龚六堂. 社会保障、家庭养老与经济增长[J]. 金融研究, 2012,(1):78-90.
- [3]郭庆旺,贾俊雪,赵志耘. 中国传统文化信念人力资本积累与家庭养老保障机制[J]. 经济研究, 2007,(8):58-72.
- [4]何立新,封进,佐藤宏. 养老保险改革对家庭储蓄率的影响:中国的经验证据[J]. 经济研究, 2008,(10):117-130.
- [5]贾俊雪,郭庆旺,宁静. 传统文化信念、社会保障与经济增长[J]. 世界经济, 2011,(9):3-18.
- [6]刘甲炎,范子英. 中国房产税试点的效果评估:基于合成控制法的研究[J]. 世界经济, 2013,(11):117-135.
- [7]刘乃全,吴友. 长三角扩容能促进区域经济共同增长吗[J]. 中国工业经济, 2017,(6):79-97.
- [8]万海远,李实. 户籍歧视对城乡收入差距的影响[J]. 经济研究, 2013,(9):43-55.
- [9]王天宇,彭晓博. 社会保障对生育意愿的影响:来自新型农村合作医疗的证据[J]. 经济研究, 2015,(2):103-117.
- [10]张川川,陈斌开. “社会养老”能否替代“家庭养老”? ——来自中国新型农村社会养老保险的证据[J]. 经济研究, 2014,(11):102-115.
- [11]赵怡. 我国社会保障与经济增长关系研究[J]. 管理世界, 2007,(12):152-153.
- [12]郑秉文. 欧债危机下的养老金制度改革——从福利国家到高债国家的教训[J]. 中国人口科学, 2011,(5):2-15.
- [13]Abadie, A., A. Diamond, and J. Hainmueller. Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program [J]. Journal of the American Statistical Association, 2010,105(490):493-505.
- [14]Auerbach, A. J., and L. J. Kotlikoff. Dynamic Fiscal Policy[M]. Cambridge: Cambridge University Press, 1987.
- [15]Barro, R. J. Are Government Bonds Net Wealth[J]. Journal of Political Economy, 1974,82(6):1095-1117.

- [16]Becker, G. S., and R. J. Barro. A Reformulation of the Economic Theory of Fertility [J]. *Quarterly Journal of Economics*, 1988,103(1):1–25.
- [17]Billmeier, A., and T. Nannicini. Assessing Economic Liberalization Episodes: A Synthetic Control Approach [J]. *The Review of Economics and Statistics*, 2013,95(3):983–1001.
- [18]Black, D. A., and J. A. Smith. How Robust Is the Evidence on the Effects of College Quality? Evidence from Matching[J]. *Journal of Econometrics*, 2004,121(1):99–124.
- [19]Bloom, D., D. Canning, and B. Graham. Longevity and Life Cycle Savings [J]. *Scandinavian Journal of Economics*, 2003,105(3):319–338.
- [20]Bruce, N., and S. J. Turnovsky. Social Security, Growth, and Welfare in Overlapping Generations Economies with or without Annuities[J]. *Journal of Public Economics*, 2013,101(1):12–24.
- [21]Caliendo, M., and S. Kopeinig. Some Practical Guidance for the Implementation of Propensity Score Matching [J]. *Journal of Economic Surveys*, 2008,22(1):31–72.
- [22]Cigno, A., L. Casolaro, and F. C. Rosati. The Impact of Social Security on Saving and Fertility in Germany[J]. *FinanzArchiv: Public Finance Analysis*, 2003,59(2):189–211.
- [23]Dehejia, R. Practical Propensity Score Matching: A Reply to Smith and Todd [J]. *Journal of Econometrics*, 2005,125(1):355–364.
- [24]Ehrlich, I., and F. T. Lui. Social Security, the Family, and Economic Growth [J]. *Economic Inquiry*, 1998, 36(3):390–409.
- [25]Ehrlich, I., and J. Kim. Social Security and Demographic Trends: Theory and Evidence from the International Experience[J]. *Review of Economic Dynamics*, 2007,10(1):55–77.
- [26]Ehrlich, I., and J. G. Zhong. Social Security and the Real Economy: An Inquiry into Some Neglected Issues[J]. *American Economic Review*, 1998,88(2):151–157.
- [27]Feldstein, M. Social Security, Induced Retirement, and Aggregate Capital Accumulation [J]. *Journal of Political Economy*, 1974,82(5):905–926.
- [28]Feldstein, M. Social Security and Saving: New Time Series Evidence [J]. *National Tax Journal*, 1996,49(2): 151–169.
- [29]Feldstein, M., and J. B. Liebman. Social Security[J]. *Handbook of Public Economics*, 2002,4(8):2245–2324.
- [30]Glazerman, S., D. M. Levy, and D. Myers. Nonexperimental versus Experimental Estimates of Earnings Impacts[J]. *Annals of the American Academy of Political and Social Science*, 2003,589(1):63–93.
- [31]Gustman, A. L., and T. L. Steinmeier. Effects of Social Security Policies on Benefit Claiming, Retirement and Saving[J]. *Journal of Public Economics*, 2015,(129):51–62.
- [32]Harris, S. E. *Economics of Social Security*[M]. New York:McGraw–Hill, 1941.
- [33]Heckman, J., H. Ichimura, and P. Todd. *Characterizing Selection Bias Using Experimental Data* [R]. NBER Working Paper, 1998.
- [34]Heston, A., R. Summers, and B. Aten. *Penn World Table Version 6.3* [R]. Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania, 2009.
- [35]Kaganovich, M., and I. Zilcha. Pay-as-you-go or Funded Social Security? A General Equilibrium Comparison[J]. *Journal of Economic Dynamics and Control*, 2012,36(4):455–467.
- [36]Kitao, S. Sustainable Social Security: Four Options[J]. *Review of Economic Dynamics*, 2014,17(4):756–779.
- [37]Lechner, M. Identification and Estimation of Causal Effects of Multiple Treatments under the Conditional Independence Assumption[J]. *Econometric evaluation of Labour Market Policies*, 2001,13(3):43–58.
- [38]Lucas, R. E. On the Mechanics of Economic Development [J]. *Journal of Monetary Economics*, 1988,22(1): 3–42.

- [39]Mendonça, H. F. D., and G. J. D. G. Souza. Is Inflation Targeting a Good Remedy to Control Inflation [J]. *Journal of Development Economics*, 2012,98(2):178–191.
- [40]Pogue, T. F., and, L. G. Sgontz. Social Security and Investment in Human Capital [J]. *National Tax Journal*, 1977,30(2):157–169.
- [41]Romer, C. D., and D. H. Romer. Transfer Payments and the Macroeconomy: The Effects of Social Security Benefit Increases,1952—1991[J]. *American Economic Journal: Macroeconomics*, 2016,8(4):1–42.
- [42]Rosenbaum, P. R., and D. B. Rubin. The Central Role of The Propensity Score in Observational Studies for Causal Effects[J]. *Biometrika*, 1983,70(1):41–55.
- [43]Samuelson, P. A. An Exact Consumption–loan Model of Interest with or without the Social Contrivance of Money[J]. *Journal of Political Economy*, 1958,66(6):467–482.
- [44]Smith, J. A., and P. E. Todd. Does Matching Overcome LaLonde’s Critique of Nonexperimental Estimators[J]. *Journal of Econometrics*, 2005,125(1):305–353.
- [45]Soares, R. R. The Effect of Longevity on Schooling and Fertility: Evidence from the Brazilian Demographic and Health Survey[J]. *Journal of Population Economics*, 2006,19(1):71–97.
- [46]Zhang, J. Social Security and Endogenous Growth[J]. *Journal of Public Economics*, 1995,58(2):185–213.
- [47]Zhang, J., and J. Zhang. How Does Social Security Affect Economic Growth? Evidence from Cross–country Data[J]. *Journal of Population Economics*, 2004,17(3):473–500.

Social Security and Economic Growth: An Analysis of Quasi–Natural Experiments

JIA Jun–xue¹, LI Zi–xiao¹, QIN Cong²

(1. School of Finance, Renmin University of China,

Beijing 100872, China;

2. National Academy of Development and Strategy, Renmin University of China, Beijing 100872, China)

Abstract: Beginning at the 1980s and 1990s, a large number of developing countries gradually adopted the social security system. Drawn on this quasi–experiment, the current paper analyzes the impact of the establishment of social security system on economic growth. In particular, using a panel dataset of 114 developing countries from 1981 to 2005, this paper combines the propensity score matching method with the difference–in–difference approach to construct counterfactuals, which allows us to address the potential selection bias and therefore identifies the treatment effect of social security system. The results show that the growth rate of real GDP per capita decreases 2.5957 percentage points, indicating that social security system has a significant negative effect on economic growth. Rather than the lack of human capital accumulation, the main engine of the negative effect attributes to the insufficient physical capital accumulation—the growth rate of the physical capital investment rate has decreased 173.94%. Moreover, institutional arrangements and policy designs also play an important role on the causal effects of the social security system. Specifically, a social security system with lower benefits and taxes (or fees) is more likely to exacerbate the negative effects of the system on economic growth. We also find a similar scenario for the level of development and the propensity to save, but an opposite effect for the aging of population and fertility rate. The above findings should add sound knowledge for China in constructing a sustainable social security system in the long–term.

Key Words: social security system; sustainability; economic growth; quasi–natural experiments; propensity score matching with difference–in–difference

JEL Classification: H55 O40 O50

[责任编辑:王燕梅]