



C2026004

2026-05-11

基于 CFPS 的社会经济研究：从现实数据到学术创新和 最优政策探索

易君健

北京大学中国经济研究中心

北京大学国家发展研究院

摘要：在当代社会科学研究中，数据的获取与运用正以史无前例的速度发生变革。从传统的问卷调查到如今的大数据，研究者的工具箱不断扩展，带来了前所未有的分析机会。然而，这种变革也引发了一系列重要问题：在以各种大数据为代表的独特数据日益受到青睐的背景下，传统调查数据是否仍具备竞争力？它在现代社会科学研究中的角色和意义又是什么？本文以 CFPS 为核心，探讨调查数据在社会科学研究中的独特价值。本文将首先回顾经济学研究与数据应用的广泛联系，并重点介绍中国的社会科学调查数据，特别是家庭调查的发展。随后，本文将以我与合作者的研究为例，分析如何基于 CFPS 进行社会经济研究。特别地，本文第三部分将围绕我与合作者发表在 *Journal of Political Economy* 上的论文“Multidimensional Premarital Investments with Imperfect Commitment”，探讨如何从现实数据中提炼学术观点，进行学术创新，并为政策制定提供科学依据。接着，本文将讨论调查数据在大数据时代的优势及其面临的挑战，尤其是与大数据的结合潜力。通过对这些问题的探讨，本文不仅旨在厘清调查数据在现代社会科学中的角色，还希望为社会科学研究者提供借鉴，以更好地利用数据服务于学术创新和政策建议。

一、经济学研究与数据应用

经济学研究可以按照两种不同的方式进行分类：理论分析（Theoretical Analysis）与经验分析（Empirical Analysis），或实证分析（Positive Analysis）与规范分析（Normative Analysis）。理论分析旨在构建解释经济现象的逻辑框架，而经验分析则通过数据验证这些理论的有效性。实证分析主要回答“世界是什么样的”和“为什么是这样”的问题；而规范分析则关注“世界应该是什么样的”，通常涉及福利

分析和政策分析。规范分析需要定义社会效用函数或政府目标函数，以提出政策改进建议。值得注意的是，这两种分类方式的研究往往交叉存在。比如，研究者往往需要通过实证分析深入地了解现状，然后在此基础上利用规范分析提出政策调整建议，而实证分析与规范分析都会涉及理论分析和经验分析的过程。

在经济学的经验分析中，数据尤为重要。数据不仅为描述和量化经济现象提供了工具，还为政策制定提供了科学依据。数据能够反映国民经济和社会经济活动的基本状态，同时为社会科学研究提供重要的现实基础。

数据在国民经济和社会经济活动中扮演着至关重要的角色，为政策制定、经济管理提供了底层依据。数据是了解国民经济运行状态的基础工具，它能够准确反映经济增长、就业、收入分配、消费支出等关键指标的动态变化。例如，GDP、失业率和通货膨胀等宏观经济数据为政府制定财政和货币政策提供了核心依据。同时，数据还能够揭示区域之间和行业之间的经济差异，从而帮助政府优化资源配置，促进区域协调发展。在社会经济活动中，数据也为企业决策提供了重要参考。通过分析消费者行为数据，企业可以更精准地制定营销策略、优化供应链并提升生产效率。在当前数字经济快速发展的背景下，数据还成为一种新的生产要素，其价值已渗透到金融、医疗、教育、环保、能源等多个领域。

数据对于社会科学研究至关重要，它是分析社会现象、验证理论假设和提出政策建议的重要工具。数据可以量化复杂的社会经济行为，帮助研究者揭示隐藏的规律和因果关系。同时，数据为理论模型的检验提供了必要支持，使得理论不仅停留在逻辑推理层面，还能在实际情境中验证其适用性。此外，高质量的数据能够支持跨时间、跨区域的比较研究，有助于识别社会结构和制度安排对行为的长期影响。

中国的社会科学数据收集起步较晚，特别是在家庭调查领域。家庭调查可以捕捉微观层面的家庭和个人行为特征，成为分析消费、投资、劳动供给等问题的关键。在 2010 年之前，中国的家庭调查主要有 CHNS、CHIP、中国城镇住户调查（UHS）、中国综合社会调查（CGSS），以及人口普查等。

CHNS 是美国北卡罗来纳大学教堂山分校人口中心与中国疾病预防控制中心营养与健康所合作开展的国际调查项目。研究团队自 1989 年至 2015 年共开展了 10 次调查，内容涵盖营养学、公共卫生、经济学、社会学和人口学等领域，并收集了食品市场、医疗设施、社会服务等社区数据。调查采用多阶段随机抽样方法，从 15 个省市中抽取约 7200 个家庭、超过 30000 名个体。

CHIP 由北京师范大学中国收入分配研究院分别于 1988 年、1995 年、2002 年、2007 年、2013 年和 2018 年开展，共进行了六轮。调查收集了家庭和个人的收支信息，涵盖城镇和农村住户。为反映农村向城镇迁移的日益重要性，CHIP 2002 年开始增加流动人口调查，形成城镇、农村和流动人口三个子样本。

UHS 由国家统计局城市社会经济调查司组织开展，主要内容包括家庭人口构成、现金收支、商品消费、劳动就业、居住条件及耐用消费品拥有量等。调查关注城镇住户，采用分层随机抽样方法。

CGSS 始于 2003 年，当前主要由中国人民大学中国调查与数据中心主持。截至 2022 年，CGSS 共进行了 15 次年度调查，较为系统、全面地收集了行为模式、思维模式、生活方式以及社会变迁方面的基本信息。

人口普查是由国家统计局组织实施的全国性调查，至今已开展了七次全国人口普查，分别在 1953 年、1964 年、1982 年、1990 年、2000 年、2010 年和 2020 年。此外，截至 2025 年 5 月，中国已于 1987 年、1995 年、2005 年和 2015 年开展了四次全国 1% 人口抽样调查，俗称“小普查”（Mini Census）。

这些数据为研究者提供了重要支持，但多为横截面数据，难以揭示动态变化趋势。自 2010 年起，中国相继开展了几项全国性面板数据调查，其中有代表性的包括 CFPS、CHARLS 和中国家庭金融调查（CHFS）等。这些数据各有侧重。

CFPS 由北京大学中国社会科学调查中心实施，是一项全国性、大规模、多学科的社会跟踪调查项目。该项目最初于 2008 年、2009 年开展测试调查，自 2010 年起开展正式访问，并根据 2010 年基线调查界定 CFPS 基线成员。此后 CFPS 每两年进行一次追访，跟踪收集个人、家庭、社区三个层次的数据，涵盖经济活动、教育成果、家庭关系与家庭动态、人口迁移、健康等诸多研究主题。CFPS 基线调查的样本覆盖 25 个省（自治区、直辖市），目标样本规模为 16000 户，调查对象包含样本家中的全部家庭成员。

CFPS 最大的优势在于它记录了完整的家庭结构。CFPS 的家庭结构数据分为 T1、T2 和 T3 三张表格，分别包含不同范围的家庭成员信息。T1 表记录的是居住在家庭内部、与家庭成员有血缘或婚姻关系的成员，例如父母子女、配偶等。T2 表则扩展到了虽然不居住在家庭内部，但与家庭成员存在血缘或婚姻关系的个体，例如居住在其他地方的父母、子女、兄弟姐妹或配偶等。T3 表的记录对象是物理上住在家庭内部，但与家庭成员没有血缘或婚姻关系的成员，例如保姆、租住者或其他寄宿人员。这种分类结构为研究家庭结构提供了基础，尤其是 T1 表和 T2 表，可以帮助研究者全面了解家庭成员间的动态关系，包括核心家庭和扩展家庭的互动模式。这种细致的分类对于分析家庭经济活动、代际支持和社会网络等问题具有极大的研究价值。

CHARLS 是由北京大学发起并支持的大型跨学科调查项目，2024 年起由北京大学和武汉大学两校共建，旨在收集全国 45 岁及以上中老年人及其配偶的高质量微观数据。该项目于 2011 年正式启动基线调查，覆盖全国 28 个省（自治区、直辖市）的 150 个县区、450 个社区（村），涉及约 1.7 万名受访者，每两到三年进行一次追踪

调查。调查内容包括个人基本人口结构、健康状况、医疗服务利用状况、工作与退休状况、收入与消费、资产以及社区基本情况等多个方面。

CHFS 是由西南财经大学中国家庭金融调查与研究 centers 于 2011 年设计实施的一项全国性抽样调查，旨在全面收集中国家庭在金融领域的微观数据。调查内容涵盖人口特征与就业、资产与负债、收入与消费、社会保障与保险及相关主观态度等多个方面。2025 年 7 月，CHFS 正式启动第八轮调查，样本覆盖全国 29 个省（自治区、直辖市）的 234 个县（区、县级市）、937 个村（居）委会、23000 余户家庭，具有全国、省级和副省级城市的代表性。CHFS 的数据为学术研究和政策制定提供了重要支持，填补了中国家庭金融领域的研究空白。

二、使用 CFPS 进行的研究

在第一部分中，我们讨论了经济学研究与数据应用的广泛联系，并强调了数据在推动实证研究中的重要作用。作为一项全国性家庭调查，CFPS 为探索中国社会经济规律提供了宝贵的数据支持。为了更直观地展示 CFPS 的潜力和价值，下文将回顾我与合作者使用 CFPS 完成的几项研究。这些研究涵盖了 we 已发表的学术文章和未发表的工作论文，充分体现了这一数据源在揭示社会现象、解释行为机制以及指导政策制定等方面的独特作用。

我与 Bhaskar 等人在研究中探讨了在缺乏完全承诺的情况下，婚前多维投资（包括人力资本和住房）的决定机制（Bhaskar et al., 2023）。我们首先利用了 2010 年 CFPS 基线数据中的住房和教育投资信息，发现在性别比例高度失衡的县，拥有男孩的家庭的住房投资显著多于拥有女孩的家庭，而教育投资则呈现相反的趋势。这种地区差异反映了性别失衡和婚姻市场竞争对投资行为的影响，为我们的理论研究提供了经验事实基础。随后，我们建立了一个理论模型，考虑不同投资形式（人力资本和住房）提供的不同承诺可能性。理论分析表明，在男性对收入拥有更高的议价能力，而女性在婚姻市场中更有竞争优势（婚姻市场男多女少）的情况下，男孩父母更倾向于增加住房投资而减少教育投资。我将在第三部分更加详细地介绍这项研究。

我与 Huang 等人探讨了 2011 年中国最高人民法院对《中华人民共和国婚姻法》的解释变化如何改变夫妻间的议价能力，并进一步影响家庭的资源分配和子女投资（Huang et al., 2023）。文章利用双重差分法，关注购房时间与法律变更的交互作用，以研究这一变化对家庭行为的影响。研究发现，法律的改变削弱了妻子在家庭内部的议价能力。具体来说，在 2011 年之前结婚并购房的家庭中，房产仅登记在丈夫名下的情况提高了女性的就业参与率，但同时也减少了子女教育的投入，并增加了丈夫对传统男性消费品（如酒精）的支出。研究使用的数据主要来自 CFPS，包括 2010 年[《婚姻法司法解释（三）》出台前]及 2012 年和 2014 年[《婚姻法司法解释（三）》

出台前]三轮数据。CFPS 提供了详尽的家庭房产信息，包括房产登记人和购房年份，使研究能够准确识别法律变更对不同家庭的影响。数据还覆盖了就业、消费及子女教育支出等多个关键变量，为评估政策影响提供了有力的支持。

在研究童年时期经历的困苦如何影响个体成年后的风险偏好和创业行为时（Yi et al., 2022），我们利用了中国 1959-1961 年的三年困难时期作为准自然实验，通过了解自然灾害导致不同地区粮食短缺的程度，考察个体早年的困苦对风险容忍度及创业参与的影响。我们发现，早年经历更多困苦的人更可能成为企业家，这种影响部分源于因果效应，而非单纯是自然灾害后的选择性存活。异质性分析显示，虽然早年的困苦使男性和女性在风险容忍度上均有所提高，但创业行为的增加主要体现在男性身上，这一性别差异可能反映了传统的社会规范。CFPS 数据提供了个人风险偏好的度量指标，为分析风险偏好作为行为机制提供了支持性证据。

我与 Fan 等人也探讨了中国自 1979 年经济体制改革以来的代际收入流动性（IGE）变化趋势及其背后的原因（Fan et al., 2021），CFPS 数据包含了父母和子女的收入，以及详细的社会经济特征。我们采用多种统计方法（如回归分析、排名相关性和过渡和矩阵）估计代际收入流动性，并校正生命周期偏差、选择偏差和衰减偏差，确保结果的稳健性，从而更精确地衡量代际收入流动性。研究发现，中国的代际收入流动性在 1970-1980 年出生的群体中为 0.390，而在 1981-1988 年出生的群体中上升到 0.442，收入代际传递性显著增强。这种趋势在城市和沿海地区居民中尤为明显。

在一篇工作论文中，我们研究了计划生育政策在废除后是否依然对生育率产生持续影响，以及这种影响的机制是什么（Guo et al., 2026）。我们建立了一个理论模型，提出了如下假说：通过限制母亲的生育数量，计划生育政策形成了新的生育规范，这种规范持续影响下一代女性的生育决策。为验证模型预测，我们利用了计划生育政策这一准自然实验，通过分析 20 世纪 70 年代出生的女性及其后代的生育行为，探讨政策的跨代效应。基于中国 2020 年人口普查、CFPS 以及其他数据，我们采用事件研究法发现：母亲受计划生育政策影响降低了女儿的生育率，母亲少生育一个孩子，下一代生育数量平均减少 0.67。CFPS 数据的应用体现在对个体的生育意愿和生育行为的分析上，有助于揭示生育规范的代际传播路径和效应。

我和 Yu 等人探讨了中国的计划生育政策如何通过城乡间的生育差异对代际收入流动性产生影响（Yu et al., 2026）。由于计划生育政策在城乡之间实施力度不均，我们利用这一政策特征提供的准自然实验，分析生育差异如何影响代际收入流动性。这项研究基于 CFPS 和 CHARLS，构建了全国范围内的父子对样本，估计代际收入流动性。结果显示，生育差异每增加一个单位，代际收入流动性（以收入排名相关性来衡量）下降约 53.2%。对此，我们给出的解释是，生育决策存在质量与数量的权衡取舍，受政策约束较大的城市家庭通常选择少生育，但对子女进行更多的人力资本

投资，受政策约束较小的农村家庭则反之。这种机制通过教育流动性也得到了验证：生育差异的扩大显著降低了教育层面的代际流动性。

我和 Qin 等人探讨了出生季度对长期结果的影响，特别是这一影响的性别差异，并发现，相较于其他季度出生的人，出生于第四季度的人拥有更好的教育结果和劳动市场表现，并且这种季度出生效应对女性的影响更大（Qin et al., 2025）。我们对此提出的假说是，农业生产的季节性与传统的性别偏好相结合，导致新生儿时期资源分配的性别差异，从而影响个体长期结果。在研究中我们采用 CFPS 数据中度量性别偏好的变量来验证我们的识别假设，并且使用 CFPS 的家庭消费记录构建家庭消费的季节性波动，以验证我们提出的机制。此外，我们还发现改革开放带来的经济发展缓解了季节性资源匮乏造成的影响。

2025 年，我和 Bhaskar 等人（Bhaskar et al., 2025）在研究中对我们于 2023 年提出的理论（参见 Bhaskar et al., 2023）进行了更加严格的经验分析。我们同样采用 CFPS 数据，揭示了地区性别比如何影响男孩与女孩父母的投资决策。CFPS 数据的家庭层面和个人层面信息，包括家庭迁移情况、住房状况和儿童教育支出等，为研究提供了翔实的基础。

三、实例：如何利用现实数据进行学术创新和最优政策探索

在前一节中，我们通过具体的研究实例，简要介绍了基于 CFPS 数据完成的几项学术研究。接下来，我们将进一步探讨如何从现实数据出发，通过严谨的分析方法实现学术创新，并提供政策建议。以 Bhaskar 等（2023）的研究为例，我们不仅将展示数据如何揭示经济现象的潜在机制，还将探索理论模型在解释复杂社会行为和优化政策选择中的应用，而 CFPS 数据在支持理论建模和经验验证中将起到关键作用，为未来的研究和政策实践提供思路。通过这些案例，我们希望展示学术研究如何有效联结数据、理论与政策，推动中国的知识创新与社会进步。

我们研究的起点是一个现实观察：在中国的婚姻市场中，无论是农村还是城市，在结婚时普遍需要购置房产。相比之下，其他类型的投资，特别是人力资本投资，在婚姻筹备中的优先级明显偏低。但从投资回报的角度考虑，投资房产的经济回报未必高于投资人力资本，尤其是在农村地区。这种现象引发了我们的疑问：为何房产投资在婚姻中占据如此重要的地位？其经济学理论基础究竟何在？这些经济学理论对国家的社会经济政策有何科学指导意义？

为了验证这一观察是否具有普遍性，我们借助 CFPS 数据系统地总结经验事实。CFPS 数据揭示了两个显著的趋势：第一，性别比例越失衡（男多女少）、婚姻市场竞争越激烈的地区，男性家庭购置房产的比例越高；第二，在这些地区，男性父母外出打工的比例也更高，而打工的首要目的是筹资买房，其次才是支持子女教育。这些趋势表明，房产投资在婚姻市场中的作用值得进一步探讨。

为了揭示这些现象背后的经济学原理，我们提出了一个理论：房产在婚姻中作为一种承诺机制（Commitment Device）发挥着重要作用。婚姻是一种不完全契约，夫妻双方难以转移人力资本产权，也难以就人力资本产出的分配原则在事前做出有效的承诺。而房产作为物质资产，其产权清晰且可转移，而且房产作为公共品能同时为婚姻中的双方带来效用，使其能够在婚姻契约中建立一种承诺。

从这一点出发，我们进一步构建了一个包含两个维度投资的婚姻匹配模型。在模型的婚前投资中，既包含可转移效用（Transferable Utility）的房产投资，也包含不可转移效用（Nontransferable Utility）的人力资本投资。这是婚姻匹配理论的一个创新发现：以往的婚姻市场模型要么单独考虑可转移效用投资，要么单独考虑不可转移效用投资，我们首次在同一框架下纳入这两种投资形式，并探讨其对婚姻市场均衡的影响。模型分析显示，当婚姻市场处于对称状态，即性别比例平衡，且家庭内部夫妻双方议价能力相同时，市场的竞争性均衡可以达到帕累托最优。然而，一旦性别比例失衡或家庭议价能力不对称，市场均衡将偏离最优状态，导致资源过多集中于房产投资，削弱了对人力资本的投入。

这一理论研究成果发表在 *Journal of Political Economy* 上，而我们的另一篇工作论文（Bhaskar et al., 2025）为理论预测提供了经验支持。这篇工作论文采用了 CFPS2010 年全国基线数据，利用家庭外出务工情况、住房状况和儿童教育支出等信息，得到的经验证据和理论预测一致：当性别比例失衡程度提高时，男孩的父母更倾向于通过外出务工提高收入，用以支持更大的住房投资，而教育投资则显著减少。同时，我们构建了一个新的工具变量解决了性别比例的内生性问题。我们将医疗从业人员因违规披露胎儿人性别而受到处罚的力度作为工具变量，这一变量通过减少性别选择服务的供给，降低了地方性别比例失衡的程度，同时与父母性别偏好的相关性较弱。我们发现性别比例每增加一个标准差，男孩家庭的住房建筑面积增加约 4.8%，而教育支出减少约 11.8%。此外，

性别比例越失衡，父母的劳动供给特别是迁移工作就越多。通过这种方式，我们的研究实现了一个从数据观察到学术理论创新，再从理论预测回到数据支持的循环。

然而针对同一种现象，人们总会做出不同的理论解释，不同的理论解释又对应不同的政策含义。一种替代性解释是，房产在婚姻市场中具有炫耀品（Status Good）的功能。相比承诺机制的解释，炫耀品假说指出，家庭可能通过购置房产展示社会地位。不同的机制对政策的影响截然不同：若房产是承诺机制的体现，政策应关注婚姻财产分配规则，若房产更多作为炫耀品，政策则应着眼于改变社会文化习俗。

为了提出可靠的政策建议，我们需要回归学术探索，借助更一般的理论和更丰富的数据识别不同的影响机制。我们正在尝试建立一个同时包含两种机制的婚姻匹配模型，即房产既有承诺机制的功能，也有炫耀品的功能。若能够获取更详细的信

息，就有可能量化不同机制的相对作用。因此，我们正在与 CFPS 紧密合作，重点收集更丰富的婚姻财产转移信息。在此基础上，我们将进行理论建模，展开结构计量估计，精确量化各机制的影响程度，并且探索最优的政策，为制定有针对性的政策提供科学依据。

四、调查数据在大数据时代的优势与挑战

大数据时代，随着技术的快速发展和数据获取手段的多样化，越来越多的研究开始依赖行政数据和其他独特数据源。这些数据因其大规模、实时性和高覆盖率的特点，在描述社会经济现象的宏观趋势时展现出强大的分析潜力。例如，税务记录、医疗信息以及社交媒体数据等被广泛用于研究个体行为和政策效果。然而，这种趋势也引发了一个重要问题：传统的调查数据是否因此失去了其在社会科学研究中的独特优势？事实上，调查数据以其设计的科学性和对复杂社会现象的细致捕捉而著称，能够提供许多大数据和行政数据难以替代的信息。我们将从数据的全面性和细致性、信息采集的灵活性、数据质量的可控性以及样本代表性四个方面，探讨调查数据在大数据时代依然不可或缺的价值。同时，我们也将指出调查数据所面临的挑战，并进一步分析其与大数据结合的潜力。

调查数据以其全面性和细致性在社会科学研究中具有重要地位。调查数据能够涵盖从个体到家庭的多维度信息，包括教育、收入、健康、消费以及社会网络等方面。这种细致性使研究者能够深入分析微观层面的社会经济行为及其背后的机制。以 CFPS 为例，其通过 T1、T2 和 T3 三张表细致记录了家庭成员的信息，分别涵盖了核心家庭、扩展家庭和住在家庭内部的非亲属成员。家庭结构表为分析家庭行为提供了独特的支持。例如，通过结合 T1 表和 T2 表的数据，研究者可以全面捕捉核心家庭与扩展家庭之间的互动模式，这种分类体系尤其适用于研究代际支持和家庭资源分配等议题。这样的数据结构不仅为涉及家庭内部资源配置的学术研究提供了更多的可能性，也为制定基于家庭的公共政策提供了更加全面、准确的事实依据。

调查数据的另一个重要优势在于其信息采集的灵活性。研究者可以根据具体的研究需求设计问卷，直接询问复杂且主观的问题，例如幸福感、信任度、政策偏好以及社会价值观等。这些主观信息对于研究个体行为背后的动机机制至关重要，而这些信息在大数据中往往难以捕捉或准确提取。大数据虽然具备规模大、实时性强的特点，但其主要依赖行为记录和自然语言处理，难以全面反映个体的内在感受与态度。仍以 CFPS 为例，通过面对面访谈的方式，访员能够收集丰富的主观和情感数据，例如受访者的社会信任度、对家庭生活的满意度等。

与大数据相比，调查数据在数据质量上具有更高的可控性，这体现在数据收集和处理的严格设计与执行上。调查数据常通过精心规划的抽样方法和规范化的操作流程来确保数据的可靠性，从而最大限度地降低测量误差和系统偏差。例如，CFPS

采用多阶段分层随机抽样设计，不仅在样本选择上实现了科学性，还在数据收集过程中对问卷执行进行了严格监控。每一轮调查都由受过专业训练的访员完成，辅以详细的调查手册和标准化流程，确保每个问题的含义被准确传达并正确记录。在数据清洗过程中，CFPS 在确认中断样本、清理各库编码、核查问卷逻辑、清理重点变量等环节也都有严格的标准控制。这种精细化的质量控制流程使得调查数据成为社会科学研究中可靠性较高的资料来源，为深入分析和得出可信结论提供了坚实的基础。

样本的代表性是调查数据的一大优势，这使其能够在全国范围内推断普遍性规律。CFPS 基线调查覆盖了中国 25 个省（自治区、直辖市），样本代表了 95% 以上的人口特征，包括城乡居民，不同性别、年龄和受教育程度的个体。相比之下，部分大数据偏向特定群体，例如社交媒体用户或某些消费平台的客户群体，这样的样本难以反映整体人口的真实特征，也限制了其在全局性政策分析中的适用性。调查数据在确保样本代表性方面的另一项重要措施是采用抽样权重。通过多阶段分层随机抽样方法，调查数据对样本中不同群体的选择概率进行了精准控制，确保不同人群的特征在样本中得到合理体现。随后，通过赋予样本权重，研究者能够在分析中调整因抽样过程导致的偏差，使推断结果更接近全国总体特征。

尽管调查数据具有上述优势，但也面临显著的挑战。第一，成本是制约调查数据收集的重要因素。设计、实施和维护大规模调查需要投入大量的人力和物力资源。第二，在人工智能快速发展的时代，社会经济现象的变化速度加快，传统调查数据难以及时更新，从而限制了其在动态分析中的应用。这些挑战需要通过技术创新和数据共享机制加以应对。

因此，大数据时代，我们需要将调查数据与大数据有机结合，为社会科学研究探索新的可能性。大数据具有实时性和规模优势，而调查数据则能够深入揭示微观机制，因此两者的结合可以形成优势互补。例如，将 CFPS 的结构化数据与互联网数据相结合，能够更全面地分析消费行为与心理因素之间的联系。再如，结合地理位置数据和调查数据，可以更精准地评估区域经济政策的效果。这种跨领域的数据融合有助于拓展研究的深度和广度。

五、结语

本文从经济学研究与数据应用的基本概念出发，介绍了中国主要的家庭调查数据，并通过使用 CFPS 的具体研究实例，展示了如何基于数据开展严谨的学术研究。在梳理已有研究成果的基础上，我们以婚姻市场中的投资行为研究为例，展示了如何从现实观察出发，通过数据分析发现规律，进而建立理论模型解释机制，最后回到数据验证理论并探讨政策含义。我们还探讨了大数据时代调查数据的独特优势和面临的挑战，指出了传统调查与新兴数据相结合的发展方向。

作为社会科学研究者，从现实数据上升到学术创新，再走向最优政策探索，我们应当秉持三重态度：对现实的尊重和谦逊、对科学的追求与严谨、对政策的负责与审慎。首先，社会科学研究的起点永远是现实。研究者需要以开放和谦逊的态度观察社会现象，倾听不同群体的声音，理解行为背后的深层动机。正如我们在婚姻市场研究中看到的，复杂的社会现象往往超出既有理论的解释范围，需要研究者放下成见，虚心求证。其次，面对复杂的社会现象，研究者必须秉持科学精神，通过严谨的方法寻求真知。这不仅体现在数据收集与分析的规范性上，更体现在理论推导的逻辑性和实证检验的可靠性上。最后，社会科学研究的最终目标是服务于人民的福祉。在提出政策建议时，研究者必须充分认识到人的行为和社会制度的复杂性以及政策效果的不确定性。我们需要全面考虑不同的理论解释对应的政策结果，审慎评估政策的潜在影响，避免因过度自信而导致失误。

这三重态度相互支撑、缺一不可。对现实的尊重和谦逊让我们能够发现真正值得研究的问题，对科学的并严谨使我们能够获得可靠的研究结论，而对政策的负责与审慎则确保研究成果能够真正造福社会。在复杂多变的现代社会中，只有始终保持这样的研究态度，社会科学才能在认识世界和改造世界的过程中发挥应有的作用。

参考文献

- Bhaskar, V., W. Li, and J. Yi (2023), “Multidimensional premarital investments with imperfect commitment”, *Journal of Political Economy*, 131 (10), 2893-2919.
- Bhaskar, V., W. Li, and J. Yi (2025), “Strategic parental investments in a competitive marriage market”, *European Economic Review*, 173 (C), 104958.
- Fan, Y., J. Yi, and J. Zhang (2021), “Rising intergenerational income persistence in China”, *American Economic Journal: Economic Policy*, 13 (1), 202-230.
- Guo, R., L. Lin, J. Yi, H. Zhu. (2026), “Intergenerational impact of birth-control policies: The role of norm”, *Economic Journal*, forthcoming.
- Huang, Y., J. Pantano, H. Ye, et al. (2023), “Property division upon divorce and household decisions”, *Journal of Human Resources*, 58 (2), 532-560.
- Qin, X., J. Yi, and H. Zhang (2025), “Quarter of birth, gender inequality, and economic development”, *Journal of Labor Economics*, Forthcoming.
- Yi, J., J. Chu, and I. P. Png (2022), “Early-life exposure to hardship increased risk tolerance and entrepreneurship in adulthood with gender differences”, *Proceedings of the National Academy of Sciences*, 119 (15), e2104033119.
- Yu, Y., Y. Fan, and J. Yi (2026), “China’s One-Child Policy Amplifies Intergenerational Inequality via Enlarged Differential Fertility”, *Proceedings of the National Academy of Sciences*, forthcoming.