



北京大学中国经济研究中心  
China Center for Economic Research

讨论稿系列  
Working Paper Series

No. C2023001

2023-03-27

## 国家治理指数和发展中国家现代化道路： 基于历史沿革、理论逻辑和构建方法的考察

北京大学国家发展研究院  
“国家治理指数构建”课题组<sup>1</sup>

### 摘要

国家治理体系和治理能力的现代化是发展中国家现代化的必要途径。国家治理能力的提升依赖于有效的国家治理指标评价体系。在实践中，发展中国家受制于西方主流国家治理理论，难以对自身经验和挑战进行准确的总结。本文探讨了国家治理指数的历史沿革、理论概念和建构方法，剖析了国家治理指数的政治经济学逻辑。通过对于世界治理指数、民主多样性、法治指数等代表性指数的研究，指出西方主流国家治理指数服务其战略利益的本质，揭示了西方主流国家治理指数应用于评价发展中国家治理体系和治理绩效的不足。在此基础上，本文探讨了构建国家治理指数中国方案的前景，提出国家治理指数应当以发展中国家为主体、以发展中国家人民为中心、以发展和治理的实绩为标尺、服务于南南合作与发展实践的建议。

<sup>1</sup> 课题组成员：席天扬、雷镇环、刘诗尧、李怡爽、刘珍岑、赵家琪。

## 一、 引论

在发展中国家现代化进程中，国家治理能力的提升是实现经济腾飞和长治久安的必要途径。美国政治学家亨廷顿在 1968 年出版的《变化世界中的政治秩序》开篇提出，“各国之间最重要的政治分野，不在于它们政府的形式，而在于它们政府的有效程度。...美国、英国和苏联... 每个国家自成一个政治共同体，人民对其政治制度的合法性有举国一致的共识。每个国家的公民及其领导人对社会公益和他们的政治共同体赖以立足的那些传统和原则，观点是相同的。三个国家都具备强大、能适应的、有内聚力的政治体制。”在针对二战以后第三世界社会动荡的分析中，亨廷顿进一步提出：第三世界国家的社会崩溃并非是因为经济停滞，而是政府治理能力不足以解决工业化造成的结构性变迁，不均衡发展累积的治理赤字最终反噬发展中国家的现代化蓝图(Huntington, 1968)。

如何判断国家治理体系和治理能力的有效性？这既是一个理论问题，也是一个科学问题。发展中国家面临的困境是：对于治理能力和绩效的评价深受西方主流理论话语影响，难以对于自身发展经验和挑战进行独立表述。在国际发展和治理实践中，对于发展中国家治理能力的评估严重依赖于西方的概念、理论和指标评价体系。发展中国家现代化所面临的挑战，在更深层次上也表现为理论话语的挑战，即能否建立一套反映发展中国家现实需求、准确度量发展中国家治理能力和治理绩效的指标体系。

冷战结束之后，西方国家在全球治理体系中占据了优势，“治理”超越“发展”成为西方国家国际秩序中对发展中国家输出的核心概念。与冷战时期“资本主义”和“社会主义”两大阵营的分野不同，西方国家在后冷战时代奉行民主叙事和新自由主义叙事：政治体制层面强调多党竞争、分权制衡；经济体制层面强调私有产权、价格竞争和市场开放。在上述理论体系上构建的国家治理指数，既是西方价值观的载体，也是服务于西方战略利益的工具。

无论是在世界银行所主导的世界治理指数排名，抑或《经济学人》智库主导的“民主指数”中，中国在全世界所有国家中均处在中偏后位置，与中国的发展现实和国民体验不符。在公共卫生领域，2019 由霍普金斯大学健康安全中心、核威胁倡议和经济学人智库制作的全球国家卫生安全能力的评估报告（Global Health

Security) 把中国排在所有国家中的第 51 名, 把美国排在第 1 名。新冠疫情爆发后, 美国彭博社发布的新冠疫情抗击能力指数 (Covid Resilience Ranking) 一度把美国排在世界首位, 把中国的表现排在靠后位置。这种排名将会影响发展中国家的公共卫生体系运行效果的认知, 并影响其在公共卫生决策中是借鉴中国经验还是照搬西方模式。

本文首先对于流行的国家治理指数进行介绍, 阐述治理指数历史严格以及现实影响。在此基础上, 结合国家治理概念分析主流治理指数体系在理论、方法和价值层面的缺陷。本文的分析以世界治理指数、民主多样性和法治指数等三个具有广泛影响的数据为例, 指出这些指数的共同问题: 首先, 发展中国家的主体性缺失, 造成国家治理指数对于发展中国家制度和治理结构的多样性反映不足; 其次, 发展中国家人民视角的缺失, 造成国家治理指数过于倚重专家和精英意见, 忽视普通民众感受; 最后, 实践理性的缺失, 造成国家治理指数过于重视基于文本和规则的正式制度, 轻视治理实绩。在此基础上, 本文将探讨如何构建体现发展中国家需求、以发展中国家人民为中心、以发展和治理的实绩为标尺的国家治理指数中国方案。

## 二、 国家治理指数的历史沿革

国际组织和研究机构投入大量资源编制各领域国家治理指数, 并广泛地借助指数进行决策、推动政策和治理议题。根据编制机构和政策特征差异, 可以把国家治理指数分为四类<sup>2</sup>。

第一类是由大型国际组织所编制的综合治理指数。这些指数对于各国的制度和综合治理能力进行多维度评价, 指标体系往往反映组织的目标和议程。比如, 经济合作与发展组织从 2009 年起发布“社会制度与性别指数 (Social Institutions and Gender Index)”年度报告, 对世界各国在法律和社会政策中的性别平等进行评估。世界卫生组织每年发布“全球健康支出 (Global Health Expenditure Database)”包括健康指数、健康支出和宏观指标三大类共 300 多个指标, “全球健康观察”

<sup>2</sup> 张萌萌和杨雪冬(2022)把跨国政治评估指数按照议题区分为: 政治制度类型评估指数、政治规范评估指数、制度绩效评估指数;按照构建主体分为: 政府间机构、跨国非政府组织和学术研究机构指数。

(Global Health Observatory) 包含了 1500 多个与国民健康相关的国别指标。世界银行编制和发布的世界发展指数 (World Development Indicator), 人力资本指数 (Human Capital Index), 营商环境指数 (Doing Business Database), 企业家调查 (Enterprise Surveys) 以及全球金融包容性指数 (Global Financial Inclusion (Findex) Database) 在各国经济发展实践和研究中均有广泛影响。

第二类治理指数是由高校和学术机构所编制的治理指数。这些指数往往由学者牵头编制, 服务于特定学术研究目的。代表性指数包括瑞典乌普萨拉大学的和平与冲突研究系 (Department of Peace and Conflict Research) 创建的乌普萨拉冲突数据项目 (Uppsala Conflict Data Program (UCDP)), 美国圣母大学 (University of Notre Dame) 环境变化倡议 (Environmental Change Initiative, ND-ECI) 发布的全球适应性倡议国家指数 (Global Adaptation Initiative Country Index ND-GAIN) 等。民主多样性数据 (Varieties of Democracy, 简称 V-DEM) 由哥德堡大学政治学者斯塔凡·林德伯格 (Staffan Lindberg) 于 2014 年创立, 把民主概念分解为 5 个核心维度指标, 87 个中层指标和 473 个底层指标<sup>3</sup>, 成为西方政界和学界评价各国政治体制的重要依据。

第三类是由西方国家智库和非政府组织编制的专门指数, 往往集中在具体的治理概念。代表性指数包括透明国际 (Transparency International) 的腐败感知度 (Corruption Perception Index), 和平基金会 (Fund for Peace) 的脆弱国家指数 (Fragile States Index), 加图研究所 (Cato Institute) 的人类自由指数 (Human Freedom Index), 菲沙研究所 (Fraser Institute) 的世界经济自由报告 (Economic Freedom of the World), 系统和平研究所 (Center for Systemic Peace) 的国家脆弱性指数全球趋势 (Global Trends, State Fragility Index)。

第四类是由政府机构所编制的政策类指数。例如美国国家科学基金会从 1985 年起每年发布科学与工程指数 (Science and Engineering Indicators), 美国国际发展署每年发布人口与健康调查 (Demographic and Health Surveys (DHS))、市民社会组织可持续发展指数 (Civil Society Organization (CSO) Sustainability Index)、正义与安全 (Justice and Security Index) 等报告。

国家治理指数构建不是单纯的科学求知, 其规则反映了主导国家的战略利益

详见 V-Dem Methodology V12 和 V-Dem Codebook V12。

和核心价值。国际政治经济学实证研究表明，国际组织的决策受制于出资国主流政治观念和意识形态（McLean, 2012; Bayram & Graham, 2015）。诺贝尔经济学奖得主、世界银行前首席经济学家斯蒂格利茨因为公开发表不同于美国财政部的观点而辞职（Wade 2002）。编制治理指数的研究机构往往具有官方或半官方背景。例如，美国约翰斯霍普金斯大学发布的全球健康安全指数年度报告由总部位于美国华盛顿的智库核威胁倡议（the Nuclear Threat Initiative）、隶属于约翰斯霍普金斯大学博隆伯格公共卫生学院的健康安全研究中心，以及隶属于经济学人杂志的“经济学人影响力”（Economist Impact）项目合作发布。核威胁倡议的创始人是美国前联邦参议员、参议员军事委员会主席山姆努恩（Sam Nunn），具有国家安全和情报工作色彩。西方学术界影响深远的政体数据库（Polity data series）的创始人为美国著名国际关系学者 Ted Robert Gurr，项目经费来自于美国中央情报局拨款的政治不稳定性任务组（Political Instability Task Force）。

在实践中，国际组织发布的治理指数往往成为评价发展中国家政府治理绩效的标尺。美国政府“千禧年挑战对外援助计划”（the Millennium Challenge Account）每年向低收入国家提供 50 亿美元的援助资金，要求赢得援助资金的国家必须在世界治理指数的至少四个子指数上有较好表现，对于腐败控制的子指数得分需要在低收入国家中排名在前 50%才有资格申请 MCA 项目援助（Kaufmann and Kraay, 2002）。荷兰政府也利用世界治理指数来考核被援助国家的资金效益。一些商业评级公司也用世界治理指数考察和评价各国的投资环境（World Bank, 2006）。美国国会要求国务卿根据该指数的汇编报告对世界各国政府的廉洁和民主状况提交年度评估报告（Haber and Kononykhina, 2018）<sup>4</sup>。

从上述讨论可以看出，治理指数是西方主导的全球治理体系的一个环节。美国国际发展署根据受援助的发展中国家在世界治理指数、透明度指数、民主指数、法治指数等维度的“进步”来评估其发展援助政策在地缘政治意义上成功与否。是否“进步”的评估标准来自于民主叙事和新自由叙事主导的话语体系。与其说是援助和政策导致了治理的成功，不如通过援助和评估体系建立了和西方国家联系

<sup>4</sup>“38 个数据库”统计自 WGI2022。WGI 使用的数据库数量在不同年度的版本之间会发生变化，比如 2022 年版本为了实现 WGI 底层数据的完全公开，WGI 选择放弃一个保密数据库（Country Policy and Institutional Assessment data）作为数据源，同时 WGI 会基于新的数据组成对往年指标进行回溯更新。

更加紧密、治理模式更相近的盟友。美国国际开发署在其官方主页上的自身使命中就把“促进美国的国家安全和经济繁荣”置于“促进接受援助的国家走向能够自力更生和保持活力的道路。<sup>5</sup>”之前作为首要的目标。

基于西方外交利益和国家安全决定对外发展援助的做法可以追溯到“冷战”时期。发展经济学家罗斯托（Walt Rostow）1960年出版《经济发展阶段》，提出著名的经济现代化五阶段论，认为从经济落后走向现代化的过程可以概括为“传统社会—具备经济起飞的前提条件—经济起飞—经济成熟—高大众消费社会”的转型过程，其中第二阶段和第三阶段最需要资本积累，也是发展中国家最难以逾越的阶段。罗斯托提出，美国在冷战对抗中立足于不败之地的最好策略是帮助第三世界中受到共产主义思潮影响第三世界国家贫困人口脱贫，通过援助建立和巩固亲西方政权，帮助这些国家实现资本积累，完成经济起飞。1961年，肯尼迪总统通过行政命令宣布建立美国国际开发署，罗斯托也因对于美国外交和国家安全事务的影响力在1966到1969年期间担任约翰逊总统的国家安全顾问（Easterly, 2002; Gilman, 2007）。

另一位塑造了美国冷战时期对外援助政策的政治学家李普塞（Martin Lipset）先后担任了美国和平研究所、民主基金会等智库的顾问。李普塞1959年的论文“民主的社会条件：经济发展与政治合法性”认为，在跨国比较中，经济发展水平和西方式民主之间具有统计相关性，这一表述此后被西方学术理论界概述为“现代化理论”（Modernization Theory）：即经济发展有助于促进第三世界国家在政治体制和治理方式上变得更加“现代”，更接近于西方发达国家的模式。现代化理论影响了肯尼迪和约翰逊政府，以“先发展后民主”的思路，在对外援助上优先考虑对亲西方政权，在短期内以受援国的经济增长而不是政治体制和治理的改革为目标（Gilman, 2007）。

随着苏联解体和新自由主义思潮的兴起，国际发展的主流范式从“先发展后民主”转变为更体现西方价值观的“先民主后发展”和“先治理后发展”。全球发展经济学领域的主导范式也从1960年代的结构主义转变为1990年代的新自由主义（林毅夫，2013）。亲西方非民选政权遭受更多批评，发展援助的重点目标也从1960-70年代的资本积累转变为到强调个体自由、市场经济和建立更符合西方

详见：<https://www.usaid.gov/2022-at-USAID>，2023年1月15日。

价值观的治理体系。世界银行在 1996 年首次推出世界治理指数，正是在后冷战时代背景下国际发展援助范式转变的反映。

### 三、 何谓国家治理？

英语世界治理(governance)一词的含义较为宽泛：小到公司、村社，大到市场、政府、国家、乃至全球体系，都可以成为治理的主体。根据 1990 年代以来由新自由主义思潮所主导的发展政策思维，国家治理的核心首先是管理政府的规则——“一国赖以行使其权威的传统和制度”，而这些制度又可以具体地分解为产权保护、法治、对于政府的监督、廉洁、政府对于民众的回应性和问责性、政治稳定性以及监管负担等指标(Kaufman and Krray, 2008)。上述清单包含了新自由主义思潮关于“什么是好的制度”和“西方为什么在经济和军事上领先”的主流理论叙事，同时也是支撑西方国家主导的国际组织和发展援助议程的概念体系。

新自由主义经济发展理论认为，以西方选举制度和分权体制为特征的民主制度有利于促进经济发展。这种观点背后的理论来自于诺斯、奥尔森等学者关于产权促进投资激励的政治经济学模型（姚洋，2022）。根据这些模型，刚性产权保护能够提升投资激励、促进创新和企业家精神，有利于经济发展。这种论述认为行政部门是对于私人产权最大的威胁，而西方民主则通过选举加强了民众对于政府的监督和限制。这个论述忽略了无论选举政治，还是当代西方的直接民主形式，都提高了收入较低阶层的政策影响力，增强了政策从富人向穷人的收入再分配倾向。因此，詹姆斯·麦迪逊和马克思都把西方民主制度看作一种对于私人产权制度的破坏性力量（Madison, 1787; Knack and Keefer, 1997）。即使根据主流新古典经济学理论，西方民主制度对于经济发展的影响也不确定。从经验来看，基于 1990 年代之前历史的研究很少发现西方民主制度有利于经济发展，而 1990 年代之后声称西方民主有利于经济发展的文献则大幅增加(Przeworski and Limongi, 1993)。

对于民主理论和新自由主义叙事的依赖弱化了第三世界国家的治理独立性。世界治理指数对于政府有效性的评分以政府更少干预市场为标准，忽略政府引导经济发展的积极作用（Kurtz and Schrank, 2007）。这反映了西方主流学术界对治理的理解至今仍在延续奥尔森关于掠夺性国家的传统模型，而很少讨论国家能力

对经济发展的重要性 (Fukuyama 2013)。与之对照,“发展型国家”(developmental state)理论认为,后发经济体成功离不开政府对于企业和支持,东亚“四小龙”等国家和地区的快速增长得益于合宜的产业政策以及政府与企业的良好互动 (Johnson 1982; Wade 2003)。已有的国家治理指数很少关注政府在产业协同和市场增进中的积极作用。

鉴于西式民主对于发展和治理存在复杂和不确定的影响,没有理由认为基于西方国家标准模型的制度设计能够取代发展中国家自身的治理经验,成为制定政策的主要依据。正是意识到主流发展和治理理论存在的问题,林毅夫(2014)强调不存在一个适用所有国家的发展政策,发展中国家在政策制定中首先要甄别和发展经济比较优势吗,决定发展和治理的政策。

在理论层面,大多数治理指数底层逻辑可以追溯到西方民主概念。影响广泛的世界治理指数包含六个治理维度:(1)表达和问责;(2)政治稳定与无暴力;(3)政府有效性;(4)监管质量;(5)法治;(6)腐败遏制。其中第一个维度“发声与问责”的主要标准是国家领导人是否由民众直接投票选举产生。其隐含的理论假设是:良治(good governance)的前提是政治民主,而政治民主又被先验定义为选举政治。然而以选举政治为基准来评判治理水平缺乏理论依据。根据政治学家赫希曼(Albert Hirschman)1970年的经典著作《退出,呼吁和忠诚》(Exit, Voice and Loyalty)对于发声(voice)的阐述,发声包括以各种形式向政府表达诉求以获得解决问题和改变现状的目的(Thomas, 2010),选举官员既不是公民能否表达诉求的充分条件,也不是必要条件。在实践中,民众的建议建言、人民来信、信访、市长信箱、媒体平台是中国民众表达诉求的重要渠道。研究显示中国市长所领导下的市政府比美国地方政府官员更积极回应普通民众的来信,提供他们需要的信息或解决他们的困难 (Koehler et al., 2021; Distelhorst and Hou, 2017)。这些机制也是中国全过程民主的重要特征。如果先入为主的把选举认定为民众发声的唯一渠道,就会低估中国的治理体系在确保民众表达意见、回应民众诉求上的效能。

## 四、 方法论问题

### (一) 发展中国家治理主体性和指数建构有效性

“幸福的家庭都是相似的，不幸的家庭各有各的不幸。”套用《安娜·卡列尼娜》开头的表述，治理指数的构建应当能够充分反映发展中国家的治理体系和治理能力的差异。在社会科学中，建构有效性(construct validity)是衡量评价指标的科学性和有效性的的重要依据(Adcock& Collier, 2012)。根据这一理论，如果某个治理维度在实践层面与一系列可观察的变量高度相关，那么对于该制度维度的测量指标也应该跟这些变量高度相关。

在实践中，变量方差分布是反映建构有效性的重要特征。团队内部个体成员之间在能力、努力和实绩上存在较大差异，但大多数人考绩相同，该指标就不满足建构有效性。一个好的指标变量需要具有足够的方差，同时样本方差在理论上能够反映个体之间的真实表现的差异。

按照这个逻辑，把 2019 年世界治理指数中所有国家分为发达与发展中国家分为两组（根据世界银行的标准）：发达国家(62 个)与发展中国家(123 个)的组间方差显著大于发展中国家的组内方差<sup>6</sup>。这意味着在世界治理指数体系里，过去几十年间世界各个国家和地区之间的治理指标的变化主要来自于各个国家之间的横向差距，而不是每个国家自身治理水平的变化。而国家之间的横向差异又在很大程度上能够被收入水平这单一指标所解释。

从表面上看世界治理指数方差分布合乎逻辑：发展中国家与发达国家经济基础的不同，制度上层建筑和治理水平存在更大的整体差异。但实践层面，更重要的不是治理指数能否有效地区分发达和发展中国家，而是治理指数能否有效地识别出发展中国家中绩效更好、治理能力更强，治理改善速度更快的国家。2019 年全世界三分之二国家属于发展中国家（地区），全球近八十亿人口中的八分之七居住于发展中国家，每一个发展中国家所面临的发展和治理问题都是具体、多样的。想要解释好发展中国家之间治理体系的差异，需要建立一套充分反映发展中国家主体性和治理差异性的指数，以发达国家为标尺的指数难堪大任。

以埃及和阿根廷两个中等国家为案例。两国在 WGI 法治维度下得分相近、得分互相落入对方的置信区间（见左图），这说明在 WGI 的指标体系下埃及和阿

<sup>6</sup> 平均组内方差反映的是一个发展中国家（或发达国家）与一个“平均”的发展中国家（或发达国家）相比在某一个分指标上，其分布的离散程度。组间方差反映的是一个“平均”的发达国家与一个“平均”的发展中国家相比在某一个分指标上相比，其分布的离散程度。组间方差大于平均组内方差表明在某一个分指标上，一个“平均”的发达国家和一个“平均”的发展中国家之间的差异要大于发展中国家（或发达国家）内部的差异。

根廷被认为有着相近的法治能力。而在现实中，自 2011 年以来，埃及社会局势始终较为动荡，而同一时段内阿根廷社会和政治保持相对稳定的状态，WGI 无法反映两国在法治维度的差异性。

图 1 世界治理指数组内、组间差异对比

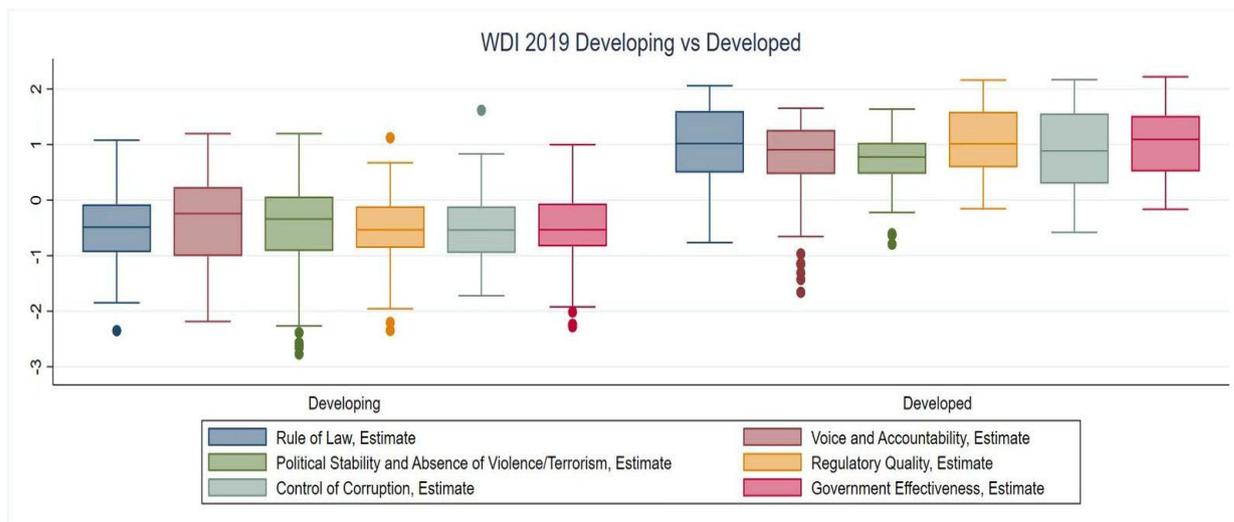
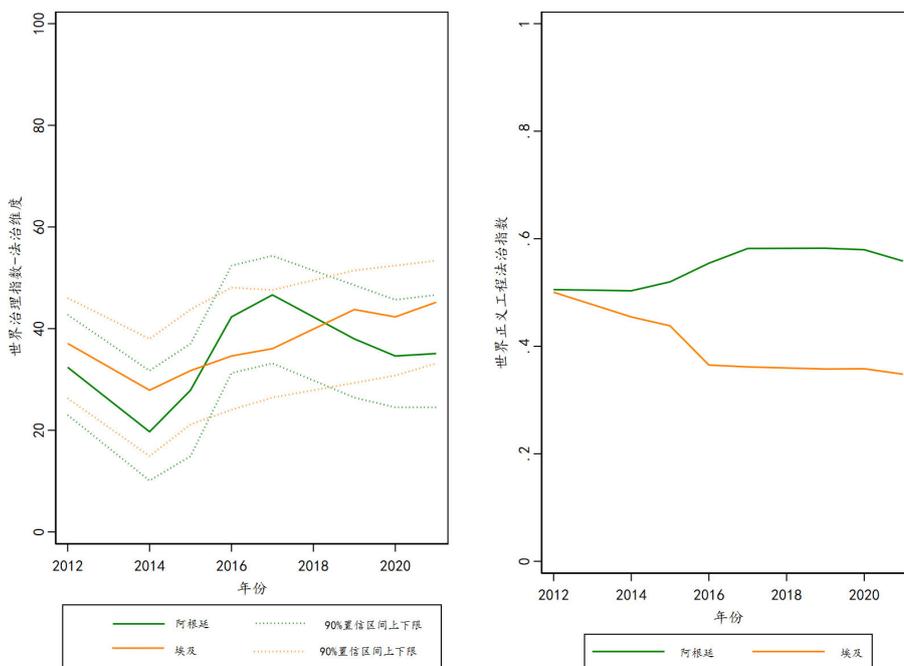


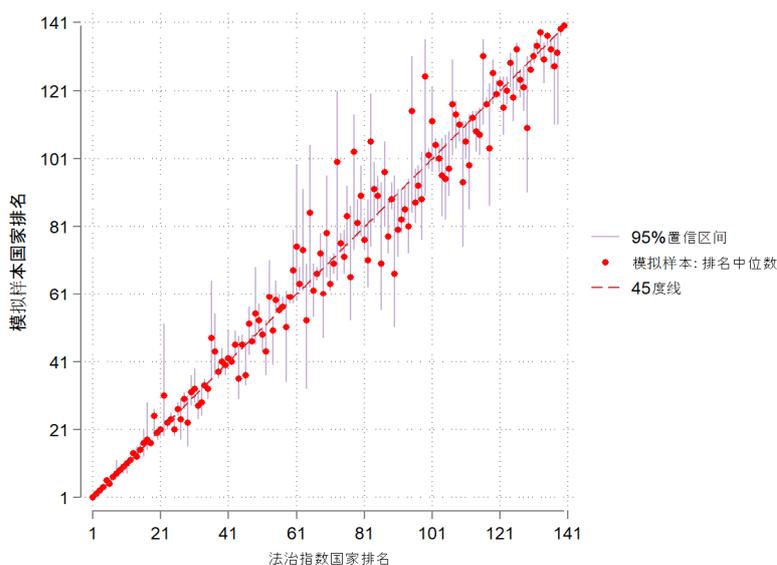
图 2：世界治理指数——埃及和阿根廷对比



注：左图是 WGI 法治维度下阿根廷和埃及的对比，虚线为 WGI 计算的 90%置信区间上下限。右图是 WJP 法治指数下阿根廷和埃及对比。

治理指数建构有效性上的缺失同样表现为不同的治理维度之间缺少有机的理论联系，造成了指标体系的合成误差。以法治指数为例，有些国家在对政府的约束方面得分较高，另一些国家在维护社会秩序上得分较高，如何对于不同的维度进行加总？世界正义工程的做法是直接赋予了八个子指数完全相同的权重。这种加权方法假设一个在廉洁程度上有所欠缺的国家，可以通过同等程度地提高其法律的执行力弥补。通过蒙特卡洛模拟(Monte Carlo Simulation)随机生成了 1000 组权重，重新计算了每个国家的法治指数和排名，与实际发布的国家排名进行对比。图 3 显示，不同维度权重的随机变化可能导致国家排名超过 20 位的上下浮动<sup>7</sup>。这种偏离在排名中等和较差的国家往往更明显，而这些国家也一般是发展中国家。

图 3 随机权重对指数国家排名的影响



注：图中反映了随机生成子指数权重对国家排名的影响。x 轴为 2022 年实际发布的法治指数国家排名，y 轴为我们随机生成 1000 组权重并重新计算的法治指数国家排名（每个国家取 1000 组样本中的中位数）。

## （二）人民主体性：精英和民众评价的差异

<sup>7</sup> 我们计算了 1000 组样本中各国排名的中位数，大部分也与实际发布的国家排名有差距（偏离 45 度线）。

习近平总书记在党的二十大报告中提出“坚持以人民为中心的发展思想”是党和国家前进道路上必需把握的重大原则之一。就国家治理指数而言，对于治理绩效的评价是否以人民的感受、认知、获得感为基础，将是国家治理指数有效性的重要决定因素。缺乏民众的理解和支持，许多初衷良好的政策将成为空中楼阁。

从表 1 可以看出，西方主流国家治理指数更多依赖于专家评分而非民众的直接反馈。有些（比如透明国际腐败感知指数）完全依赖专家打分来构建。针对非专家群体（例如企业主、居民等）的问卷调查数据则很少被使用。世界治理指数的子维度“发声和问责”只利用了非洲晴雨表的三个居民问卷问题，对于其他地区却没有使用任何居民问卷调查数据。相反，该子指数却使用了 7 种来自于自由之家（Freedom House）的专家评分。

表 1 主要治理指数的数据来源结构

指数名称	客观数据	专家评分	问卷数据	其他指数
Transparency Index	1.00	0.00	0.00	0.00
Rule of Law Index	0.08	0.66	0.26	0.00
Bertelsmann Transformation Index	0.00	1.00	0.00	0.00
Comparative Constitutions Project	0.00	0.00	0.00	1.00
Worldwide Governance Indicators	0.00	0.79	0.21	0.00
Corruption Perception Index	0.00	1.00	0.00	0.00
Democracy Index	0.00	0.71	0.29	0.00
V-Dem	0.00	0.77	0.00	0.23
E-Government Survey	0.67	0.00	0.33	0.00
Environmental Performance Index	1.00	0.00	0.00	0.00
Freedom in the World	0.00	1.00	0.00	0.00
Global Competitiveness Report	0.65	0.34	0.00	0.00
Global Integrity Report	0.00	1.00	0.00	0.00
Global Peace Index	0.69	0.27	0.04	0.00
Institutional Profiles Database	0.00	1.00	0.00	0.00
Legatum Prosperity Index	0.65	0.22	0.13	0.00
Migrant Integration Policy Index	0.00	1.00	0.00	0.00
Open Budget Survey	0.00	1.00	0.00	0.00
Social Progress Index	0.70	0.22	0.08	0.00
Sustainable Governance Indicators	0.28	0.72	0.00	0.00
World Competitiveness Yearbook	0.67	0.33	0.00	0.00
平均值	0.30	0.57	0.06	0.06

注：客观数据指统计数据、法定制度等事实性一手数据，专家评分指邀请专家对特定事项打分的数据，问卷数据指民意调查、社会调查等问卷数据，现有指数定义为其他治理指数等二手数据源。该表格中比重计算规则如下：（1）先计算子指数的数据构成。优先使用指数技术文档披露的数据源比重；如果上述权重不可得，则计算底层指标中使用各类数据的比

---

例。(2) 根据子指数在总指数的权重,将子指数的数据构成加权平均到总指数层面。如果指数数据源披露不完善、或子指数权重不确定,则无法计算数据构成。按此方法计算,最终得到 21 个指数的数据来源比重。其中,WGI 的全部数据均来自外部数据库,该指数专家评分和问卷数据的所占比重是通过外部数据库的数据源计算得到。

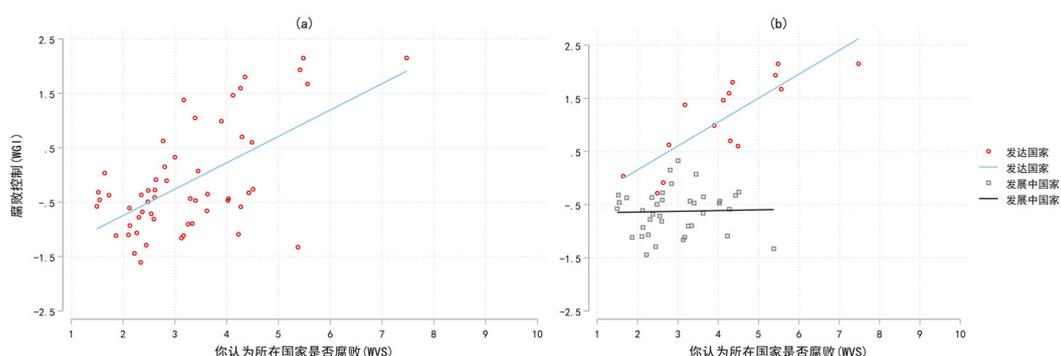
西方专家评分为基础的治理指数难以反映发展中国家的真实治理水平。在实证研究中,专家打分和从业者汇报自身真实经历是测量政府的廉洁程度的两种常见手段。由个人汇报自身经历虽然可能受到撒谎或者瞒报影响,但这种基于经历而的测量方式显然比专家评分更可能接近真实的政府腐败状况。对比以经历为基础的政府清廉度和专家评分的清廉指数,大多数研究都发现二者之间存在明显的差距。Razafindrakoto and Roubaud (2010) 通过对比八个非洲国家民众和专家问卷,发现专家往往会高估当地的腐败程度。Treisman (2007) 对比了世界银行 WGI 腐败控制指数(2005 年版本)和同年的一个跨国居民问卷,该问卷询问了受访者是否在过去一年为了获得政府服务而额外交钱。虽然 WGI 腐败控制指数和居民的回复之间有明显的相关性,但是该相关性主要集中在 WGI 腐败控制指数得分较高的国家中间。

专家评分不准确的一个原因是专家们使用收入水平和国家制度来代替对治理表现的评价。Donchev and Ujhelyi (2014) 发现在控制了人均国民生产总值和政权类型之后,WGI 腐败控制指数与民众的腐败经历变得不相关。与之相反,居民问卷中民众对政府清廉程度的主观评价在控制了民众和所在国的许多协变量后,依然与民众的腐败经历有非常稳健的正相关。这说明专家在评价政府清廉程度时,更关注该国的收入水平和选举制度,而忽视了真正发生的腐败现象;而民众在判断政府廉洁时,他们过去一年是否看到过腐败现象是一个主要因素。相比民众的直接感知,专家判断更容易受到经济发展水平和政治制度的影响来判断一个国家的政府是否廉洁。这造成专家严重高估了非洲国家的腐败现象(Donchev and Ujhelyi,2014)。

已有研究揭示民众对公共部门腐败的判断比专家意见更加可靠。Mungiu-Pippidi et al. (2018) 通过政府采购中使用非公开招标方式的比例,来测量该国政府的廉洁程度。这背后的逻辑是廉洁的政府会更少地指定特定供应商,而更多地通过公开招标来采购商品和服务。作者利用欧洲国家的采购数据和民众问卷,发现民众认为更腐败的政府,确实会更多地使用非公开招标方式。

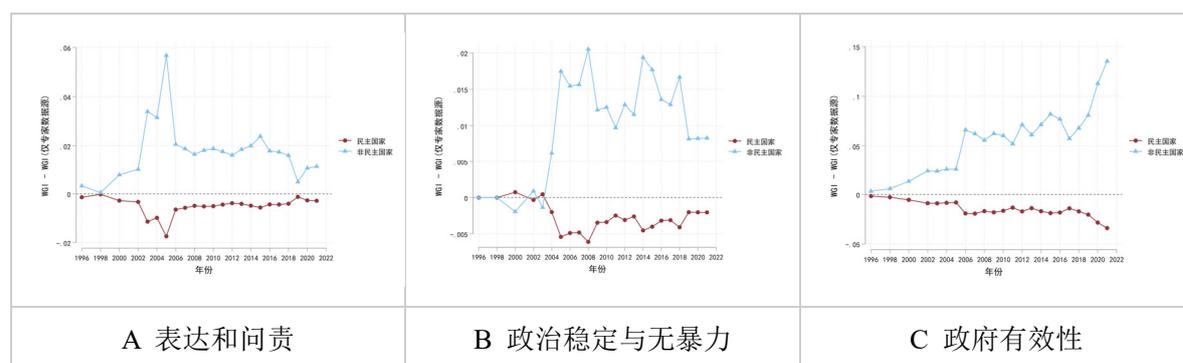
我们把世界价值观调查（World Value Survey，以下简称 WVS）中民众对于本国腐败的感知和世界治理指数中的专家评分进行了对比。图 4 显示：关于腐败程度的专家评分和民众感知在发达国家样本中高度一致，而在发展中国家中几乎没有相关性。这显示专家的认知会因为研究对象国是否为发达国家而产生系统偏差。政治体制也可能造成类似的专家偏差。如图 5 所示，如果在世界治理指数的数据源中剔除来自于民众的数据，把产生的新指数和原有指数对比会发现：对于采取西方选取体制的国家，剔除民众数据会使得该国治理水平得分变高；而对于非民主国家，剔除民众数据会使得该国治理水平得分下降。这说明西方专家对于西方民主国家的治理水平的评价要系统性地优于这些国家的民众对其政府治理的评价，而专家对于非西方民主国家的评价要比这些国家国民的评价更低。一些学者认为世界治理指数不能反映中国真实的治理表现（Fukuyama 2013）。

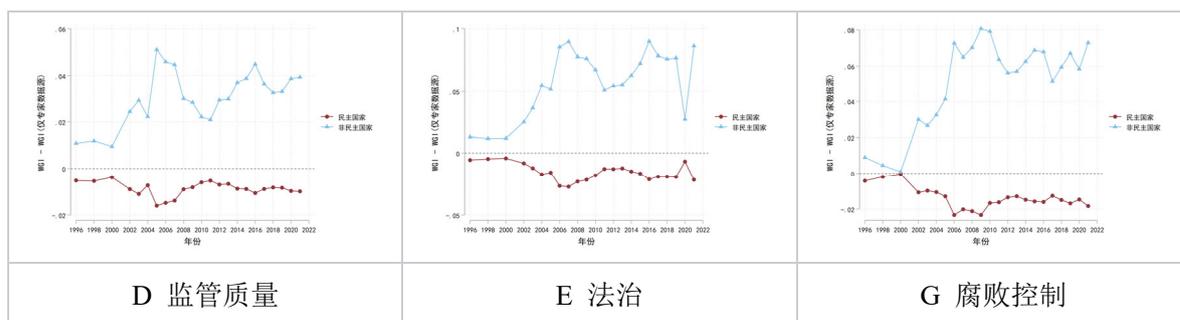
图 4 WGI 子指数和 WVS 代表指标的相关性：腐败控制



注：图中展示了 WGI 子指数和 WVS 指标国家均值的相关性，其中斜线为拟合线。横轴为 WVS 指标的国家均值，手动调整指标取值，使 1 代表最低，10 代表最高。纵轴为 WGI 指数，保留指数原本的取值范围。

图 5 WGI 框架下专家数据和民众调查数据的比较





### (三) 是否以实绩为准绳

专家评分外，治理指数信息来源的 30%是客观的发展或治理数据，比如国家的土地面积、人口、自然资源数量，也包括国内生产总值这样的虽然人为建构但是建立在可测量指标基础上的数据。从指数构建的实践来看，专家评分占据了更大权重，这也导致治理指数在时间序列上变化缓慢，这使得研究者几乎无法通过比较对于一个国家的治理能力的内生变化做出有效的评估。

治理指数脱离一国经济发展与社会治理的实绩(performance)，导致在实践上几乎不可能对于同一个国家的治理水平进行时序上的比较，对于发展援助和政策评估能力也产生了负面影响。Heywood and Rose (2014) 在讨论透明国际的政府廉洁指数 CPI 和世界银行腐败遏制指数时发现，2011 年数据只有 10--15%的变化不能被 2000 年的数据解释：两个政府廉洁指数无法反映 11 年间大多数国家的反腐成效。

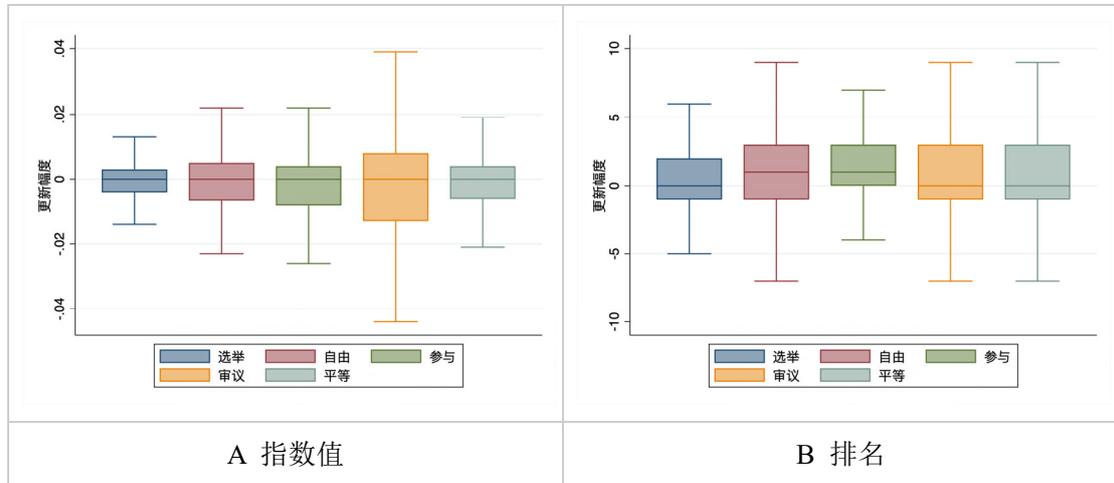
依赖专家打分的民主多样性指数 (V-Dem) 也存在类似问题。单个国家指标更新频率间隔较大，在最上层的五个维度指数中，除选举民主指数基本做到每年更新外，其他四个指数的独特部分更新频率基本在 2-3 年左右 (附录图 A1)；绝大部分国家的指标更新幅度极小。

如图 6.A 所示，五个组分的箱型图箱体都在 -0.05 到 0.05 之间，说明变化幅度在 25-75 百分位数的这一半国家，指数基本没有发生更新。即使看 1.5 四分位间距以外的边界，变化最大的审议民主指数幅度也不超过 0.1，远低于指数可信区间宽度<sup>8</sup>。因此，我们很难判断指数变动是一国政治情景的变化导致，还是专家打分出现的随机扰动。图 6.B 用排名代替指数绘制箱型图也有相似发现，中间 50%

<sup>8</sup> V-Dem 指数由 IRT 模型估计得到，其可信区间宽度普遍大于 0.1。

的国家排名更新幅度都在 3 以内，对比 V-Dem 的国家数目 179，这仍然是非常小的幅度。

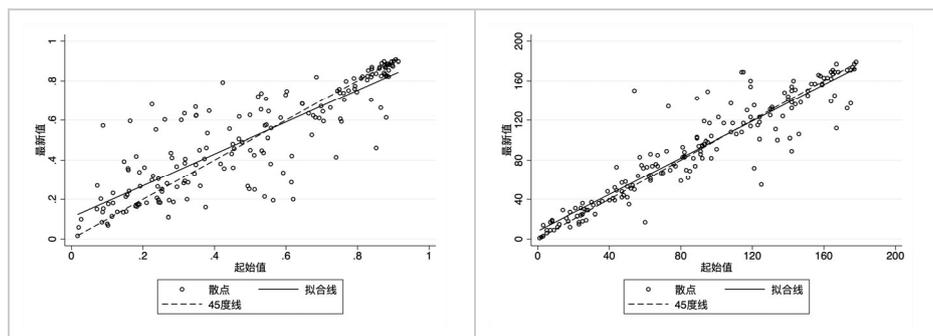
图 6 V-Dem 核心指标的更新幅度



注：图中统计了 2000 年后 V-Dem 各国数据的更新频次和幅度。为可视化目的，箱型图没有展示 1.5 四分位间距以外的离群值。

把民主多样性指数在 2000 年后的选举民主指数与其最新得分相比，绘制成散点图，如果散点普遍分布于 45 度线附近，说明在 V-Dem 中大部分国家 20 年内没有发生变化。图 6A 表明，指数拟合线斜率仅仅略低于 45 度线，而大部分散点均位于 45 度线附近，只有部分离群值推动拟合线发生偏离。对比国家民主名次（图 6B）上述趋势更加明显。图 5 中大部分国家更新幅度接近 0，以及图 6 早期指数值和近期指数值极为接近的结论说明，驱动 V-Dem 整体指数变化的，主要是某几个国家的离群值，而非各国整体的均匀变动。V-Dem 指数更容易捕捉一国政治环境的突变，而非稳定、正常状态下治理绩效的长期变化。

图 7 V-Dem 选举民主指数 2000 年至今的变化幅度



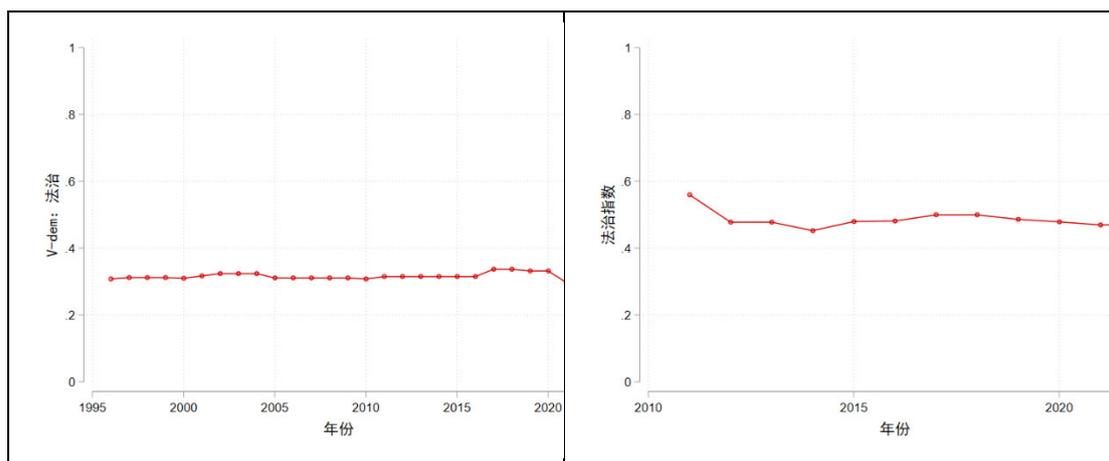
A 指数值	B 名次
-------	------

注：图中展示了各国 V-Dem 选举民主指数在 2000 年后最早一次取值和最晚一次取值的相关性，实线为拟合线，虚线为 45 度线。

治理指数脱离实绩的后果是，当国家内部的治理结构或机制发生重大变革，治理指数仍维持原状或变化不足以反映改革深度。以中国近年司法改革为例，最高人民法院分别于 1999 年、2005 年、2009 年、2015 年和 2019 年发布了五个《人民法院五年改革纲要》，作为指导人民法院深化司法体制综合配套改革工作的重要纲领。具体措施涉及法院组织体系、法官制度、诉讼程序、审判方式、执行制度、司法管理等多个司法领域。2007 年起的省高院院长异地调任制度成为加强司法监督和遏制司法腐败的重要保障 (Li and Wang, 2022)。2015 年推行省以下法院人财物统一管理改革以减少地方政府对司法的干预为核心，削弱地方保护主义，提高了基层法官审判的独立性 (Lei and Li, 2022)。法官额制在减少法院内部行政权力影响的同时，提高了司法队伍的专业化程度 (Sun and Fu, 2022)。优化案件分配机制，提高司法审判效率，实现“繁案精审”、“简案快审”；建立审判流程、庭审活动、裁判文书、执行信息四大信息平台，加强司法公开的广度和深度，有效提高司法公信力；面对灵活多变的市场需求，金融法院、互联网法院、知识产权法院等专门法院也随之建立。这一系列改革措施遵循司法规律且立足中国国情，从影响司法公正、制约司法能力的深层次问题改起，从解决人民最关心、最直接、最现实的问题入手，并以人民的获得感为最终目标。然而，现有的治理指数却并没有充分反映出这些改革的效果。

图 7.A 展示了 V-dem 对中国司法透明度、司法公平与独立性、执行力等维度加总后的综合评分，可以看出自 1995 年至今变化幅度不超过 0.05。由于专家打分的时间间隔较大，在 2006-2009、2011-2015 年之间指数几乎没有变化。世界正义工程建立的法治指数 (图 7.B) 近年来对中国的法治评估也没有体现中国司法几轮改革的影响。由于这些指数过度依赖专家评分、信息更新频率较低，无法体现中国在司法建设的治理实绩和各项创新性措施的成效。

### 图 8 V-dem 和法治指数对中国的评估



注：图中分别展示了 V-dem（1995 年起）和法治指数（2011 年起）对中国法治水平的评估。

## 五 结论：构建治理指数的中国方案

本文对于西方世界发起和主导的国家治理指数的历史沿革、发展现状、理论逻辑和构建方法进行了梳理，指出西方主流国家治理指数并非对于世界各国治理体系和治理能力的全面客观反映，而是在后冷战时代的大背景下，以西方主流价值观为中心，以民主叙事和新自由主义叙事为经纬，服务于西方世界主导的国际政治经济秩序的话语体系的组成部分。厘清西方世界主导的国家治理指数的本质，有助于客观评价发展中国家在发展和治理中取得的实绩，正确对待发展中国家和发达国家在不同治理维度上的差异和差距，在政策实践中不以西方主流国家治理指数为唯一准绳，不以追求在跨国组织的治理指数排名的爬升为政绩，提升道路自信、理论自信、制度自信、文化自信。

党的二十大提出中国式现代化的宏伟蓝图，指出“中国式现代化为人类实现现代化提供了新的选择，中国共产党和中国人民为解决人类面临的共同问题提供更多更好的中国智慧、中国方案、中国力量”。习总书记进一步指出，马克思主义中国化和新时代中国特色社会主义思想的发展包含着胸怀天下的国际主义维度，“要拓展世界眼光，深刻洞察人类发展进步潮流，积极回应各国人民普遍关切，为解决人类面临的共同问题作出贡献，以海纳百川的宽阔胸襟借鉴吸收人类一切优秀文明成果，推动建设更加美好的世界。”中国政府和社会各界近年来在南南国家合作和发展中所做的工作，势必要超越短期利益和地缘政治的狭隘维度，

从全球南方立场出发，以促进人类命运共同体福祉的情怀和担当来研究发展中国家的治理模式和经验，共享发展和治理的成果。

从这个意义来说，构建与中国自身发展经验和治理模式自洽的国家治理指数，其重要性不仅在于“争夺”关于国家治理体系评价的话语权，而是南南国家合作与发展实践的迫切现实需求。比如，当中国学者向发展中国家介绍中国经验、解释中国发展和治理模式优越性时，应该依据什么标准？如何理解中国在发展、治理绩效上的成功及其与西方主流治理叙事的差异？中国发展、治理成功的经验是独一无二的特例还是具有普遍意义？如果中国经验具有普遍意义，中国的研究者又应该如何从中国经验出发，去构建适合全球南方的一般理论概念和指标体系，评价其他发展中国家的治理绩效？在实践层面，当中国向发展中国家提供援助之后，如何系统性地评价这些项目在促进经济发展和社会治理上面的实施效果？相应的评价标准和理论基础又是什么？对于上述问题的回答都呼唤构建国家治理指数的中国方案。近年来，国内的一些高校研究机构在国家治理指数构建上进行了有益的尝试，也产生了一定影响的成果<sup>9</sup>。但总体而言，由国内智库和研究机构构建的国家治理指数，在指数的数量、规模、涉及政策维度范围、国际学术和政策影响力方面，与西方主流国家治理指数相比还存在差距。基于本文分析，我们提出构建国家治理指数中国方案的几点建议。

- (1) 国家治理指数构建体现发展中国家主体性，充分认识和尊重发展中国家在发展阶段、治理风格、社会结构、文化传统上的差异性，识别发展中国家治理体系核心问题，避免对于国家治理能力进行单一维度的加总排名，构建包含多种不同治理维度、体现不同治理模式的概念和度量体系
- (2) 国家治理指数构建体现人民主体性，摆脱治理指数对于西方精英和专家打分的过度依赖，把本国民众对于发展和治理的真实感受作为国家治理指数的主要信息来源。
- (3) 国家治理指数构建体现治理绩效本位，更加重视各国经济发展和社会治理的实际绩效，以各国实际的绩效在不同时点前后变化为衡量治理绩效改善的为标尺，对于国家治理指数的指标评价体系进行校准。
- (4) 国家治理指数构建坚持实践导向性，把治理体系、治理能力和治理绩效

<sup>9</sup> 有代表性的综合类国家治理指数包括华东政法大学政治学研究院发布的《国家治理指数报告》（2025-2022）。

的评估纳入南南国家合作与发展工作议程，建立分国别、地区、部门和治理议题的国家治理指数，以治理指数为抓手，推进对于国际发展援助和南南合作项目实施效果的系统评估。

## 参考文献

### 中文

- 林毅夫.(2014). 新结构经济学:反思经济发展与政策的理论框架(增订版). 北京大学出版社.
- 姚洋.(2018). 理解中国共产党体制.《中国新叙事:中国特色政治、经济体制的运行机制分析》
- 姚洋、席天扬主编.上海:格致出版社、上海人民出版社.
- 姚洋(2022).《制度与经济增长》,文汇出版社。
- 张萌萌、杨雪冬(2022). 测量抑或规训:对政治评估指数的评估,《社会科学》。

### 英文

- [1] Andersson, B., & Xin, T. (2018). Large sample confidence intervals for item response theory reliability coefficients. *Educational and Psychological Measurement*, 78(1), 32-45.
- [2] Arrington, N., Bass, L., Glynn, A., Staton, J. K., Delgado, B., & Lindberg, S. I. (2021). Constitutional Reform and the Gender Diversification of Peak Courts. *American Political Science Review*, 115(3), 851-868.
- [3] Bizzarro, F., Gerring, J., Knutsen, C. H., Hicken, A., Bernhard, M., Skaaning, S. E., ... & Lindberg, S. I. (2018). Party strength and economic growth. *World Politics*, 70(2), 275-320.
- [4] Botero, J. C., & Ponce, A. (2011). Measuring the Rule of Law. *The World Justice Project - Working Paper Series*. Available at SSRN: <https://ssrn.com/abstract=1966257>.
- [5] Dahlum, S., Knutsen, C. H., & Wig, T. (2019). Who revolts? Empirically revisiting the social origins of democracy. *The Journal of Politics*, 81(4), 1494-1499.
- [6] Djuve, V. L., Knutsen, C. H., & Wig, T. (2020). Patterns of regime breakdown since the French revolution. *Comparative Political Studies*, 53(6), 923-958.
- [7] Donno, D., Morrison, K., & Savun, B. (2022). Not All Elections Are Created Equal: Election Quality and Civil Conflict. *The Journal of Politics*, 84(1), 134-147.
- [8] Easterly, W., (2002). *The Elusive Quest for Growth: Economists' Adventures and Misadventures in the Tropics*. The MIT Press. Chapter 2.
- [9] Gerring, J., Gjerløw, H., & Knutsen, C. H. (2022). Regimes and industrialization. *World Development*, 152, 105791.
- [10] Haggard, S., MacIntyre, A., & Tiede, L. (2008). The Rule of Law and Economic Development. *Annual Review of Political Science*, 11, 205-34.

- [11] Heywood, P. M., & Rose, J. (2014). "Close but no Cigar": the measurement of corruption. *Journal of Public Policy*, 34(3), 507-529.
- [12] Hummel, C., Gerring, J., & Burt, T. (2021). Do political finance reforms reduce corruption?. *British Journal of Political Science*, 51(2), 869-889.
- [13] Knack, S., & Keefer, P., (1997) Does Inequality Harm Growth Only in Democracies? A Replication and Extension. *American Journal of Political Science*, 41(1), 323-332.
- [14] Lachapelle, J., Levitsky, S., Way, L. A., & Casey, A. E. (2020). Social revolution and authoritarian durability. *World Politics*, 72(4), 557-600.
- [15] Lührmann, A., Marquardt, K. L., & Mechkova, V. (2020). Constraining governments: New indices of vertical, horizontal, and diagonal accountability. *American Political Science Review*, 114(3), 811-820.
- [16] Marquardt, K. L., & Pemstein, D. (2018). IRT models for expert-coded panel data. *Political Analysis*, 26(4), 431-456.
- [17] Nye, C. D., Joo, S. H., Zhang, B., & Stark, S. (2020). Advancing and evaluating IRT model data fit indices in organizational research. *Organizational Research Methods*, 23(3), 457-486.
- [18] Ping, S. N., Wang, Y. T., & Chang, W. Y. (2022). The Effects of China's Development Projects on Political Accountability. *British Journal of Political Science*, 52(1), 65-84.
- [19] Sebhatu, A., Wennberg, K., Arora-Jonsson, S., & Lindberg, S. I. (2020). Explaining the homogeneous diffusion of COVID-19 nonpharmaceutical interventions across heterogeneous countries. *Proceedings of the National Academy of Sciences*, 117(35), 21201-21208.
- [20] Sundström, A., Paxton, P., Wang, Y. T., & Lindberg, S. I. (2017). Women's political empowerment: A new global index, 1900–2012. *World Development*, 94, 321-335.
- [21] Wilson, S. L., & Wiysonge, C. (2020). Social media and vaccine hesitancy. *BMJ global health*, 5(10), e004206.
- [22] Andrews, Matt. 2010. "Good government means different things in different countries." *Governance* 23(1): 7-35.
- [23] Distelhorst, Greg, and Yue Hou. 2017. "Constituency Service under Nondemocratic Rule: Evidence from China." *Journal of Politics*, 79(3): 1024-1040.
- [24] Donchev, Dilyan, and Gergely Ujhelyi. 2014. "What Do Corruption Indices Measure?" *Economics & Politics*, 26(2): 309--331.
- [25] Gilman, N., (2007). *Mandarins of the Future Modernization Theory in Cold War America*. Johns Hopkins University Press.
- [26] Haber, Matthias, and Olga Kononykhina. 2018. "A Comparative Classification and Assessment of Governance Indices." In *Governance Indicators: Approaches, Progress, Promise*, edited by Helmut K. Anheier. Oxford, UK: Oxford University Press.
- [27] Heywood, Paul M., and Jonathan Rose. 2014. "'Close but no Cigar": the measurement of corruption." *Journal of Public Policy*, 34(3): 507-529.
- [28] Hirschman, Albert O. 1970. *Exit, Voice, and Loyalty: Responses to Decline in*

- Firms, Organizations, and States. Harvard University Press.
- [29] Holmberg, Sören, Bo Rothstein, and Naghmeh Nasiritousi. 2009. "Quality of government: What you get." *Annual review of political science*, (12): 135-161.
- [30] Hollyer, James R. 2018. "Measuring Governance: Objective and Perceptions-based Governance Indicators." In *Governance Indicators: Approaches, Progress, Promise*, edited by Helmut K. Anheier. Oxford, UK: Oxford University Press.
- [31] Høyland, Bjørn, Karl Moene, and Fredrik Willumsen. 2012. "The tyranny of international index rankings." *Journal of Development economics* 97(1): 1-14.
- [32] Johnson, Chalmers A. 1982. *MITI and the Japanese Miracle: The Growth of Industrial Policy, 1925-1975*. Stanford University Press.
- [33] Kaufmann, Daniel, and Aart Kraay. 2002. "Governance Indicators, Aid Allocation, and the Millennium Challenge Account." Available at <https://dx.doi.org/10.2139/ssrn.534063>.
- [34] Kelley, Judith G., and Beth A. Simmons. 2019. "Introduction: The power of global performance indicators." *International Organization*, 73(3): 491-510.
- [35] Koehler, Ekkehard A. and Matsusaka, John G. and Wu, Yanhui. 2021. "Street Level Responsiveness of City Governments in China, Germany, and the United States." Working Paper, available at [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3907862](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3907862).
- [36] Kurtz, N Marcus J., and Andrew Schrank. 2007. "Growth and Governance: Models, Measures, and Mechanisms." *The Journal of Politics*, 69(2): 538--554.
- [37] Lei, Z., & Li, Y., (2022). Making Local Courts Work: The Judicial Recentralization Reform and Local Protectionism in China. Available at SSRN 4052769.
- [38] Li, Z., and Wang, Z., (2022). Judicial Recentralization as Political Control: A Difference-in-Differences Analysis of Judicial Leader Rotation in China. Available at SSRN 3884714.
- [39] Razafindrakoto, M., & Roubaud, F., 2010. "Are International Databases on Corruption Reliable? A Comparison of Expert Opinion Surveys and Household Surveys in Sub-Saharan Africa." *World Development*, 38(8): 1057--1069.
- [40] Rothstein, Bo O., and Jan AN Teorell. (2008) "What is quality of government? A theory of impartial government institutions." *Governance*, 21(2): 165-190.
- [41] Skaaning, S. E. (2010). Measuring the Rule of Law. *Political Research Quarterly*, 63(2), 449-460.
- [42] Sun, Y, & Fu. H., (2022). Of Judge Quota and Judicial Autonomy: An Enduring Professionalization Project in China. *The China Quarterly* 251: 866-887.
- [43] Treisman, Daniel. 2007. "What Have We Learned About the Causes of Corruption from Ten Years of Cross-National Empirical Research?" *Annual Review of Political Science*, 10: 211-244.
- [44] Thomas, Melissa A. 2010. "What do the worldwide governance indicators measure?" *The European Journal of Development Research*, 22(1): 31-54.
- [45] Wade, Robert Hunter. 2002. "US hegemony and the World Bank: the fight over people and ideas." *Review of International Political Economy*, 9(2): 215-243.
- [46] Wade, Robert. 2003. *Governing the Market: Economic Theory and the Role of*

Government in East Asian Industrialization. Princeton University Press.

[47] World Bank. (2006). World Bank releases largest available governance data source. Washington DC: World Bank.

[48] World Justice Project. (2022). The World Justice Project Rule of Law Index 2022. Washington D.C.: World Justice Project.

[49] World Justice Project. (2022). Variables Used to Construct the WJP Rule of Law Index 2022. Washington D.C.: World Justice Project.

## 附录（不计入正文）

### 一、正文中涉及关键指数简介

#### 1. 世界治理指数（Worldwide Governance Indicator）

WGI 指数具有全面性、延续性和公开性等优点，这使得它成为最具有影响力的综合指标之一。在全面性上，WGI 基于不可观测变量模型（Unobserved Components Model，以下简称 UCM）对不同数据源的指标进行处理，为数据源之间的可比搭建了桥梁，可以进行跨国比较，弥补了其他测量指标覆盖面狭窄的不足，与其他指标形成了较好的互补。在延续性上，WGI 每年追踪底层数据库的更新，从而实现 WGI 的逐年更新。指数体系的价值之一在于可以及时全面地反映各国在年度之间的相对变化，WGI 的延续性进一步提高了其应用价值。在公开性上，世界银行为 WGI 搭建了网站并披露了详细的技术文档和原始数据供使用者进行参考使用。同时面对关于 WGI 的争议和批评，WGI 的提出者 Daniel Kaufmann 及其合作者多年来对各方质疑逐一进行公开回应，学术讨论提高了 WGI 的严谨度和可信度，也进一步提高了 WGI 的关注度。WGI 自提出以来不断对所用方法和数据进行修正和升级，但是在定义、使用数据和加总方法上仍然存在一些问题，这些问题使得 WGI 的评价结果可能有偏。

#### 2. 民主多样性数据库（V-Dem）

民主多样性数据（Varieties of Democracy，简称 V-DEM）也是由高校学者创建，在学术研究中被广泛使用的数据库。该指数由哥德堡大学政治学者斯塔凡·林德伯格（Staffan Lindberg）于 2014 年创立。V-Dem 有两个突出特征，第一是向前追溯编制了历史数据，最早的观测点可追溯到 1789 年。第二是对民主概念进行了细致的分解。2022 年的第 12 版的数据共囊括 473 个底层指标，87 个中层指标，5 个核心维度指标。其中，指数最重要的五个指标为选举民主指数（Electoral Democracy Index），自由民主指数（Liberal Democracy Index），参与民主指数

(Participatory Democracy Index), 审议民主指数 (Deliberate Democracy Index) 和平等民主指数 (Egalitarian Democracy Index)。由于其对民主治理多维度、细颗粒的刻画, V-Dem 成为西方主导的政界和学界量化各国良治程度的重要工具。

由于多维度、细颗粒带来的繁重工作量, V-Dem 采用汇编事实数据和专家调查数据结合的编制方法, 并未加入大众调查的成分。V-Dem 的变量共有五类。A 类数据为特征事实, 由 V-dem 项目团队从现有数据源填写 (如 Polity IV 等现有治理指数)。B 类数据为特征事实, 但由目标国家的在读或毕业研究生填写。C 类数据为评价类指标, 由项目专家填写。D 类数据为 A-C 类数据的组合。E 类数据为未经加工直接使用的外部数据。本质而言, 这些数据要么依赖现有数据 (如 A 类数据), 要么依赖个人主观打分 (如 B 类和 C 类数据)。V-Dem 的专家分布具有两个特征: (1) 以学界人士为主, 专家库的 3700 余人中 85% 是学者, 超过 2/3 是具有博士、硕士学位或者隶属于研究机构。(2) 每个国家有 5 名左右的专家负责打分, 但有 40% 左右的专家没有居住在自己评价的国家。

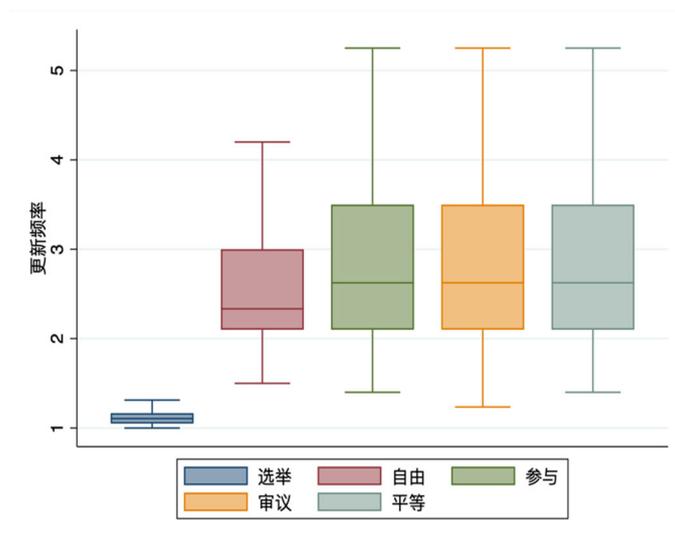
### 3. 法治指数 (Rule of Law Index)

法治指数的数据来源分三类: 民意调查 (General Population Poll)、专业受访者问卷 (Qualified Respondents' Questionnaires)、以及第三方统计数据或现有指数, 其中前两类为主要原始数据来源。民意调查每 2-3 年开展一次, 每个国家抽样 1000 人。专业受访者问卷涵盖与民商法、宪法、劳工法、公共卫生相关的共四份问卷, 每年开展一次。受访对象为各国的律师、大学及研究机构学者、NGO 工作人员、以及经世界正义工程推荐的司法工作者。2022 年共有 3600 位本地专家参与了调查, 平均每个国家 26 位 (World Justice Project, 2022)。

法治指数的计算方法相对透明, 由上述五份问卷中部分问题得分加权得到二级指数评分, 再由二级指数评分以简单平均的方式加总得到八个子指数评分和一个综合评分。为确保指数衡量的有效性, 世界正义工程还将每年的各项指数评分与其他机构发布的相关指数进行交叉比对 (Botero and Ponce, 2010), 并由第三方机构 (如欧盟委员会联合研究中心) 对最终发布的指数进行统计检验。

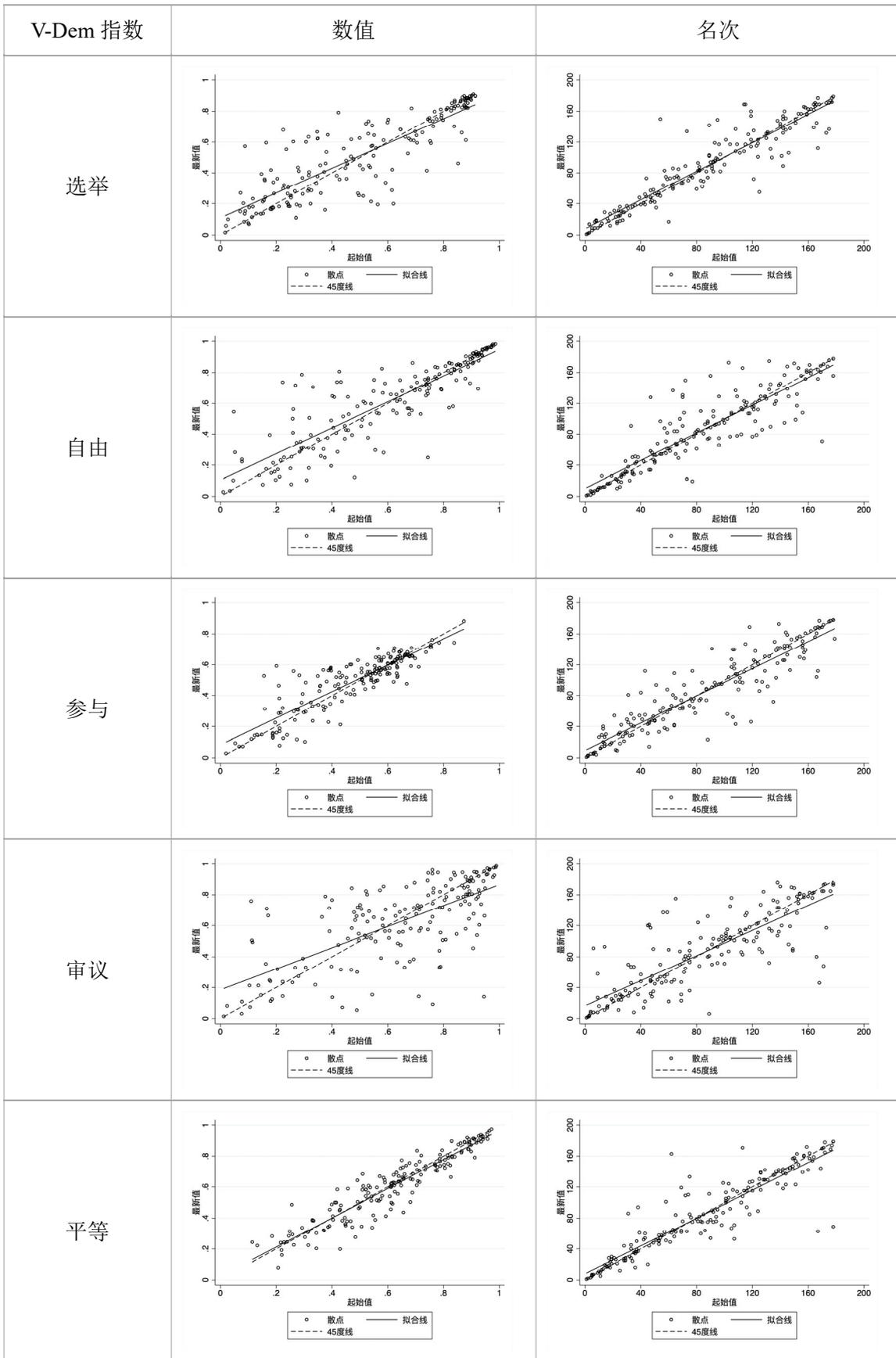
## 二、图片

图 A1 V-Dem 指数更新频率



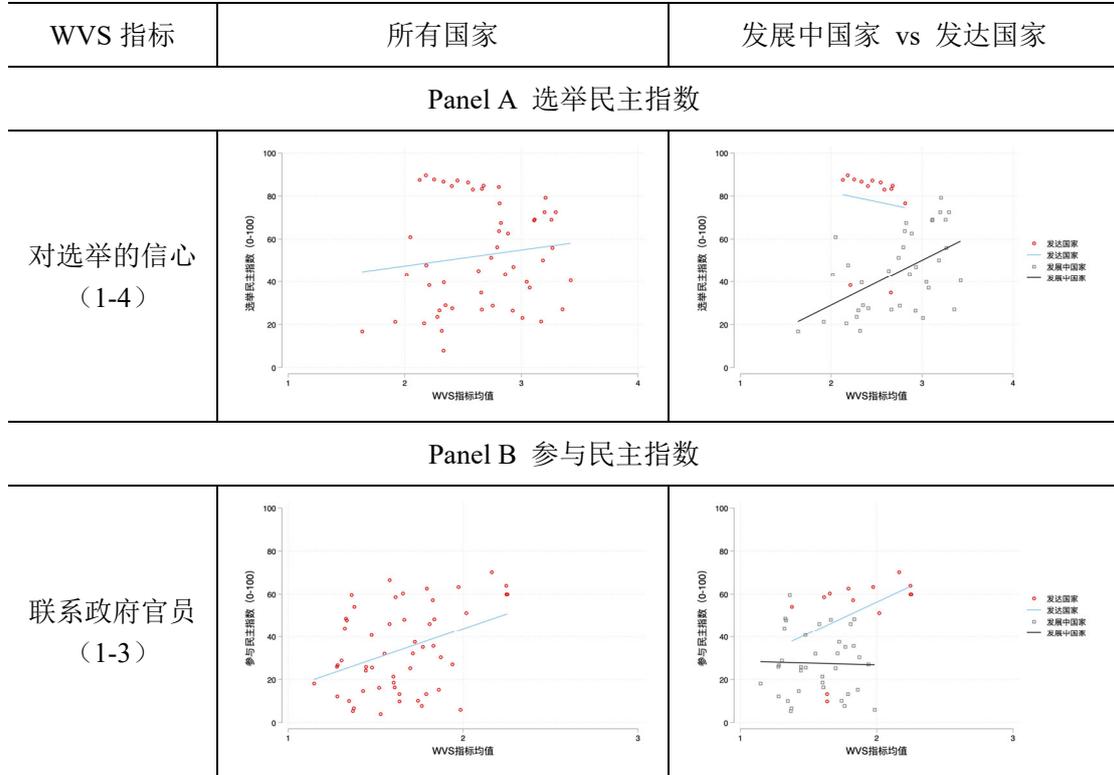
注：图中统计了 2000 年后 V-Dem 各国数据的更新频次。为可视化目的，箱型图没有展示 1.5 四分位间距以外的离群值。

图 A2 V-Dem 核心指标 2000 年至今的变化幅度



注：图中为 V-Dem 指数在 2000 年后期初和期末值相关性，实线为拟合线，虚线为 45 度线。

图 A3 V-Dem 子指数和 WVS 代表性指标的相关性



注：图中展示了 V-Dem 子指数和 WVS 指标国家均值的相关性，其中斜线为拟合线。横轴为 WVS 指标的国家均值，我们手动调整指标取值，使 1 代表最低，3 (1-3) 或 4 (1-4) 代表最高。纵轴为 V-Dem 的选举民主指数，为可视化方便，我们将其放缩到 0-100。



### 三、表格

表 A1 WGI 数据来源

	表达和问责	政治稳定与 无暴力	政府有效性	监管质量	法治	腐败控制
专家评价	0.66	0.96	0.67	0.92	0.84	0.7
民众调查	0.34	0.04	0.33	0.08	0.16	0.3
民众调查按照受访者分类:						
普通民众	0.34	0	0.14	0	0.06	0.14
企业/商人	0	0.04	0.19	0.08	0.1	0.16

表 A2 选举民主指数和 WVS 指标的相关性（所有指标）

变量名	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	对选举的信心 (1-4)			计票公平 (1-4)			选民遭受贿赂 (1-4)			选举工作人员公平公正 (1-4)		
国家组别	全部	发达	发展中	全部	发达	发展中	全部	发达	发展中	全部	发达	发展中
选举民主	0.332	-0.200	0.908**	0.184	-0.087	0.693*	-0.213	0.500	-0.743**	0.281	-0.166	0.796***
指数	(0.263)	(0.308)	(0.373)	(0.267)	(0.202)	(0.375)	(0.278)	(0.565)	(0.364)	(0.220)	(0.303)	(0.285)
观测数	74,007	21,073	49,946	69,616	20,517	49,099	68,148	19,770	48,378	68,710	20,151	48,559
拟合优度	0.040	0.075	0.078	0.101	0.103	0.032	0.125	0.091	0.072	0.088	0.117	0.036

注：括号内为国家层面聚类的标准误。回归样本为 WVS 第七轮的受访者截面。所有回归均控制受访者年龄、性别和受教育程度，国家对数人均 GDP 和受访年份固定效应。回归以一国受访者数目进行概率加权。我们手动调整 WVS 指标取值，使 1 代表最低，3 (1-3) 或 4 (1-4) 代表最高。\*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ 。

表 A3 参与民主指数和 WVS 指标的相关性（所有指标）

变量名	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
国家组别	签署请愿（1-3）			参加抵制（1-3）			联系政府官员（1-3）			鼓励他人开展政治行动（1-3）		
	全部	发达	发展中	全部	发达	发展中	全部	发达	发展中	全部	发达	发展中
参与民主指数	0.989***	1.751***	0.670**	0.400	1.007**	-0.083	0.432***	0.911***	-0.154	0.685***	1.152***	0.087
数	(0.265)	(0.321)	(0.263)	(0.239)	(0.347)	(0.199)	(0.161)	(0.161)	(0.200)	(0.205)	(0.269)	(0.244)
观测数	74,918	21,264	50,668	74,140	20,890	50,270	74,858	21,032	50,835	70,866	18,415	49,463
拟合优度	0.307	0.296	0.085	0.166	0.169	0.059	0.145	0.188	0.043	0.170	0.207	0.042

注：括号内为国家层面聚类的标准误。回归样本为 WVS 第七轮的受访者截面。所有回归均控制受访者年龄、性别和受教育程度，国家对数人均 GDP 和受访年份固定效应。回归以一国受访者数目进行概率加权。我们手动调整 WVS 指标取值，使 1 代表最低，3（1-3）或 4（1-4）代表最高。\*  $p < 0.1$ ，\*\*  $p < 0.05$ ，\*\*\*  $p < 0.01$ 。

表 A4 法治指数子指数数据来源

	对政府约束力	廉洁程度	政府透明度	基本权利	社会秩序与安全	法律执行力	民事司法公平	刑事司法公平
专家评分	0.667	0.676	0.618	0.756	0.148	0.767	0.936	0.736
公众评分	0.313	0.324	0.257	0.226	0.385	0.233	0.064	0.264
统计数据	0.021	0.000	0.125	0.018	0.467	0.000	0.000	0.000

注：此表格根据 2022 法治指数技术文件提供的各指数（二级指数）涵盖的问题数量统计，并按照官方赋予的权重进行计算。

**表 A5 法治指数子指数评估框架**

一级指数	主要衡量标准
对政府约束力	政府及其官员的权力通过宪法或其他基本法加以界定和限制；通过政府机构之间的制衡与非政府力量的制约受到限制；政府官员、军队、警察严格依法办事，承担相应的法律责任。
廉洁程度	议会、政府官员、警察、军队等不得滥用职权谋取私利、超越职权范围和侵犯他人基本权利。
政府透明度	法律应具备的明确、公开、稳定、公正等要素。
基本权利	维护法治发展所必须保护的基本权利：个人迁徙自由、意志自由，保护隐私权、言论自由、集会自由、结社自由和劳资谈判之权利等。
社会秩序与安全	人身安全、财产安全、社会稳定等。
法律执行力	法律颁布、实施和执行程序的开放性、公平性和高效性。
民事司法公平	建立公正、负责的司法队伍（法官、检察官和其他司法工作者）；替代性纠纷解决机制（传统、社区、宗教）在提供独立、公正的司法救济方面的作用。
刑事司法公平	
非正式司法制度	国家间多样性，难以衡量约束力和公正与否，因此不计入总分。

表 A6 民主化与法治指数（含子指数）的相关性

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	综合	政府约束	廉洁程度	透明度	基本权利	社会秩序	法律执行	民事公平	刑事公平
民主化	0.003*** (0.001)	0.007*** (0.001)	-0.001 (0.001)	0.004*** (0.001)	0.006*** (0.001)	-0.001 (0.001)	0.003*** (0.001)	0.002** (0.001)	0.002** (0.001)
观测数	1,020	1,020	1,020	1,020	1,020	1,020	1,020	1,020	1,020
拟合优度	0.983	0.960	0.977	0.934	0.971	0.934	0.967	0.960	0.965

注：民主化的衡量标准来自 Center for Systematic Peace 发布的 Polity V 指数。括号内为国家层面聚类的标准误。所有回归均控制国家和年份固定效应。

表 A7 WGI 子指数和 WVS 代表指标的相关性：腐败控制

	(1)	(2)	(3)
变量名	你认为所在国家是否腐败 (1-10)		
国家组别	全部	发达	发展中
腐败控制	0.581*	1.589***	0.331
	(1.974)	(3.298)	(0.925)
法治			
表达和问责			
观测数	77,866	23,191	54,675
拟合优度	0.127	0.198	0.033

注：括号内为国家层面聚类的标准误。回归样本为 WVS 第七轮的受访者截面。所有回归均控制受访者年龄、性别和受教育程度，国家对数人均 GDP 和受访年份固定效应。回归以一国受访者数目进行概率加权。我们手动调整 WVS 指标取值，使 1 代表最低，10 代表最高。\*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ 。

表 A8 世界治理指数组内、组间差距

指标	平均组内方差	组间方差
法治水平 (Rule of Law)	0.62	1.06
表达和问责水平 (Voice and Accountability)	0.80	0.76
政治稳定和无暴力水平 (Political Stability and Absence of Violence)	0.78	0.79
监管质量 (Regulatory Quality)	0.59	1.09
腐败控制水平 (Control of Corruption)	0.68	1.03
政府有效性 (Government Effectiveness)	0.63	1.09