

政策实施也要“高质量发展”，评估至关重要

——与部分中青年学者话“开局”⑤

本报见习记者 朱玲珑
本报记者 郭泉真

与记者交谈中纪园园表示,我国政策体系的建设正经历从注重“有没有”到追求“好不好”的转变,反映出国家治理体系和治理能力现代化向更高阶段迈进的新要求。

去年10月,党的二十届四中全会通过《中共中央关于制定国民经济和社会发展的第十五个五年规划的建议》(以下简称《建议》)提出:“增强宏观政策取向一致性,强化政策实施效果评价”“强化规划实施监测评估和监督,健全政策协调和工作协同机制,确保党中央决策部署落到实处”。纪园园认为:“夯实基础、全面发力”的“十五五”时期有一系列关键部署,最终会落地为政策实效。在此背景下,未来的政策评估更需注重由“量”转“质”,进一步促进公众的政策理解,在构建具有中国特色的政策评估体系同时,将评估体系嵌入治理过程,推动政策实施迈向“高质量发展”,助力实现中国式现代化。

政策评价为政策实施提供科学依据 未来政策需趋向合力评估

解放日报:您的研究领域主要是因果推断相关理论方法及其在现实问题中的应用研究。如何向大众解释政策评估研究的现实意义?它对于中国发展的价值在哪?

纪园园:虽然因果推断、政策评价听起来很高深,但可以用一句话简单概括:国家出台了一项政策,到底有没有发挥作用?

比如说,我们看到房价变化,企业创新产出增加,或者大学教育回报率在浮动,怎样判断这些变化归因于某项政策,还是自然发生。而且政策发挥了多大的作用,需要很精确、准确,仅凭直觉远远不够,甚至可能得出错误结论。

因果推断的意义就在于,将原本模糊的“经验”“感觉”转化为可验证、量化的证据。

正如医学上处理药剂检验的临床对照实验:一组病人吃药,一组病人不吃药,看药剂是否有效。评估政策时虽不完全一样,但两者的道理基本相同,公共政策同样需要科学的评估方法来识别其真实影响。用具有可比性的城市,一个受政策影响,一个不受政策影响,来看最终的政策效应。但不同于短期药效,政策发挥效果需要时间,政策评价基本上评估几年前的政策,有时候体现出滞后性,但这种滞后性在科学范围内。它的评估也属于平均效应,讨论的是平均意义下的效果,而不是个体效应。

近年来,我关注科技创新领域的政策,不仅会做政策评价,还进行内部调研和案例分析。现在我们做政策

评价,不仅要得出结果,还尝试根据结果提出落地的政策建议。中国正处在高质量发展的关键阶段,每一项政策背后都涉及庞大的人口与资源配置,因此政策评估至关重要。

解放日报:当前中国经济的复杂性与政策制定的综合性,给精准的政策评估带来了新挑战。近年来,我国在政策评价上是否面临一些问题和困难?要如何应对?

纪园园:我刚刚也提到,中国的政策工具越来越多元,政策目标越来越综合。我认为至少有三类评估挑战。

第一,政策实施往往是多目标和多手段,使得评估难以孤立观察。过去政策大多就是一个明确的目标,但在当前治理体系中,政策工具更多以组合形式推出,并同时承载增长、就业、创新、绿色发展等多重目标。比如产业政策,数字经济政策,涉及财政、金融、土地、科技等多部门协同推进。此类综合政策,往往叠加实施、联动发挥,效应彼此交织,难以直接判断和识别具体政策产生的独立效应。这意味着,政策评估不再适合采用单点式、线性化分析,而必须在更为系统、整体的框架下识别机制、厘清路径,以前某些方法就不太适用。

第二个,我们的经济结构也在加速变化,传统的评估指标难以反映真实情况。我国经济正经历由要素驱动向创新驱动的深刻转型,企业组织形式、就业结构、商业模式都发生了变化。在此背景下,传统统计指标的刻画维度过于粗放,不适用于反映新兴领域政策成效。比如,企业创新能力就不能只看专利数量,我如今在研究中会重点以新产品销售收入作为指标,即创新成果的市场价值。

第三类挑战就是数据。我们做研究时发现,可用于政策评估的高频度、跨部门且具可比性的数据仍然不充足。如果国内不同地方对于某项政策的执行力度不同,各省份关于宏观数据的统计口径就可能存在差异,从而可能对同一政策的评估结论造成系统性偏差。不同的地区之间的数据虽然有互相沟通,但仍存在壁垒。其实,我们的评估评价无法达到百分百精确,尽可能通过科学的手段接近精确,对因果识别和机制辨析提出了更高要求。

接下来,我觉得要多关注部门之间的合作。现在政策的制定和实施,很多时候都需要多部门联合作战。许多重大改革均由发改委、科技、财政等主要部门以联合评估的方式形成合力。以“双碳”政策评估为例,我们过去只看能源领域的数据,但实际上评估者也要看补贴政策、绿色金融、技术改造、地方绩效等,将“碳减排”置于完整产业链和政策体系中加以分析,更有科学性与解释力。

最后是加快使用大数据和更现代的方法。现在的宏观经济数

一项政策出台后就“一劳永逸”了吗?
如何检验“纸上”的政策构想真正走在了惠及民生的“路上”?
公众对政策效果的实际感受为何会与宏观数据存在“温差”?
尤其是未来五年,“关键时期”,“能效”“质效”已成热词——
怎么检验、怎么确保?

上海社会科学院经济研究所副研究员纪园园,多年来深耕中国政策评价领域,专注运用因果推断等前沿方法,以学者身份一直“盯着”产业政策的实施成效与优化路径。而这,正是“关键时期”确保基本实现社会主义现代化取得“决定性进展”的迫切需要。

中国政策评估强调系统化思维 呈现“先行先试”的问题导向逻辑

解放日报:我们国家的政策评价和国外有没有不一样的地方?有没有贡献研究成果以及形成一套自己的中国经验?

纪园园:政策评估不是单纯技术问题,它深受国家治理结构、发展阶段、经济体量和社会需求的影响。这个研究方向在国外很早就兴起了,我感知国内大概在2010年以后流行起来。因为国内的数据越来越开放,政策评估拥有足够数据作支撑。我们会使用国际上公认成熟的理论推导模型。但因为各国实际情况不同,评估方法在具体落实上有所不同。中国的政策评价与国外相比,我认为有两大显著不同:

一是中国政策评估更强调系统化思维。我觉得,有时候评估西方的政策,比评估中国政策反而更简单。因为西方的政策经常是单点实施,部分受限于国家的执行力度。我们国家的政策则是一套“组合拳”,政策体量,目标多,会实时调整。一项政策可能包含对产业、医疗等多领域的影响,所以有时候我们不可能完全借鉴西方的方法,还会进行理论创新,突出中国特色。这种政策外溢效应和系统性特征在国际比较中也非常具有特色。

二是中国快速试点一评估一推广的机制,在国际上独具特色。欧美国家的政策往往以长期研究为前置,随后再在小范围内开展试点验证;相比之下,中国更强调在实践中检验和优化政策路径,即通过先行试点、过程性评估,再依据数据和证据推动改革在更大范围扩展,也会有边试点边推广的情况。以医疗领域的药品集中带量采购为例,政策从最初的十几个试点城

市推广至全国,即在持续监测与评估基础上完成的。这种评估更具有实时性,政策调整也更加灵敏,有助于在复杂情境中及时识别问题、校正偏差。

我认为,中国的先行先试呈现出“问题导向”的评估逻辑。国际上其实很多评估一般以模型和实验设计为中心,政策评估常常是结束后的总结;国内则先问问题再建模型,重在解决什么问题。再加评估模式上具有灵活性、适应性,以及“从试点中学习”的特点,对许多处于快速转型阶段的发展中国家具有重要启示,能够被视为具有可复制性和可推广性的治理资源,为全球公共治理提供新的知识供给。

此外,中国经济体量大、区域差异显著,天然具备开展大规模自然实验和识别政策效应的条件,擅长用大数据讲好政策故事,将评估体系嵌入治理过程。不少学者推动大数据与因果推断的深度融合,为政策评估方法论的国际发展提供了新的思路与路径。

要用大众能听懂、看得懂的方式沟通 从平均效应到差异化研究

解放日报:有时候不管是政策评估的结论,还是统计数据结果,会与公众基于切身感受的认知产生距离。但站在学术的角度,这种政策评估模型或是数据结论遵循着科学依据,如何弥合这种信任与认知的落差?

纪园园:我觉得应该是双方面的。一方面你肯定得向公众普及更多,另一方面我觉得学术界也适当改变晦涩难懂的语言,不用人家听不懂的表述。现在提倡把论文写在祖国大地上,我们要用人家能听懂、看得懂的方式沟通。

首先,在面对公众沟通时,把复杂的模型翻译成简单的故事。我觉得学术圈有时候容易讲得太难、太专业化。比如“异方差稳健标准误”或“断点回归”识别策略,我觉得大众并不想知道这些。我们要回答的是,政策到底解决了什么问题?为什么有效?你的证据是什么?毕竟是学术界的论文,你前面可以有专业表达,但不要全篇都是晦涩

难懂的文字。理解产业政策,不一定非得弄个公式,你还可以去描述同类企业中哪些企业通过落实政策之后见效更快,让大众更容易理解。

其次,学术圈自己得先改变,主动解释方法的局限性,而不是一味强调结论。公众质疑科学的原因往往并非不信任数据,而是担心研究“掩盖不确定性”,显得过于绝对。因此,在沟通中应明确研究结论的适用范围,除了说明模型能够解释什么,更重要的是说明不能解释什么,大众也会和你有共鸣。例如,GDP作为关键宏观指标具有重要价值,但它无法反映收入分配状况。

此外,要让数据和公众联系起来,不是隔空对话。在数据结论中,我们经常讲平均值,很少讲分布、讲差异,导致公众难以将抽象统计结果与自身体验相联系。现在期刊上也会讲一些案例分析,不能全讲平均效应。我们近年来加入更多微观证据、典型样本或结构性差异的展示,使公众能够“看到自己在统计中的位置”。比如低收入群体、高收入群体、中等收入群体,考虑平均之后的差异化展示。

解放日报:现在网上流行一种“梗”,网友看到一些数据会自嘲:感觉自己“被平均了”,这是不是与您说的平均效应有关?一方面,官方学界给到大众的结果要准确真实,另一方面也需要让大众理解数据的科学依据。

纪园园:是的。我们现在大部分还是研究“平均”效应,但像细化、尾部的差异值,也是因果推断方法的一个发展趋势。我感觉最近几年学术界越来越接地气,已经在这方面努力。毕竟我们国家的人民群众可以分为不同群体,中低收入群体占其中一部分,有时需要分情况讨论。

评估标准须由“量”转向“质” 上海科创应持续发力基础研究

解放日报:去年,党的二十届四中全会通过的“十五五”规划《建议》提出“增强宏观政策取向一致性,强化政策实施效果评价”。从政策指引中,您看到了哪些值得关注的动向?

纪园园:从我的研究来看,主要关注三方面。

第一,《建议》提到“增强宏观政策取向一致性,强化政策实施效果评价”,也进一步强调“高质量发展”,意味着评估标准须由“量”转向“质”。以前我们看GDP增速、投资规模等数量性指标,但未来更强调创新驱动、绿色转型、区域协调、市场效率与就业质量等体现发展质量的内涵。从政策评估视角看,这也意味着评估体系要全面升级。

第二,关于科技自立自强。我觉得科技自立自强被置于前所未有的战略高度,研发政策和创新政策进一步要求系统化评估。我比较关注两个问题,一是科技投入如何影响异质性创新产出。不同企业和地区的技术效果不相

同,需要精准识别。某一项政策对小微企业、大型企业、中型企业,或者对劳动密集型、资本密集型企业的不同影响,要多做异质性分析。二是如何评估基础研究的长期效应。现在很多研究大部分关注短期效应,长期效应的评估难度很大。基础研究有太多不确定性场景,必须要发展新的研究方法,我们也在思考探索。

第三,根据区域发展进行政策的差异化评估。国家在因地制宜发展新质生产力,政策评价上不能一刀切,也得因地制宜。可能同一政策在西部地区和东部地区产生的效应不同,我们根据不同结果来调整不同政策倾向,或者是扶持手段。

解放日报:今年1月,上海“十五五”规划《建议》明确提出:“强化规划实施监测评估和监督,健全政策协调和工作协同机制”,以及“聚焦高风险、高价值基础研究,优化投入和评价机制,产出更多标志性原创成果”。在您参与有关上海的科创政策评价中,有哪些政策通过评估发现了不足,应该如何完善?

纪园园:我觉得有一点不足是成果转化。现在大家都提倡“产学研”,但实际上国内产学研的发展速度还不够理想,高校和企业这部分的转化能力还不足,转化渠道还没有完全打通。此外,要建立专门的转化服务部门,解决政策在落地过程中的困难。

我觉得上海国际科创中心建设还是得持续发力基础研究。比如人工智能的应用表现不错,但人工智能的基础研究较为薄弱。上海在一些基础研究的攻关项目上给了较大力度的支持,如何实现突破性创新仍然是难点。

解放日报:上海地区在政策评估和数字化转型上有哪些机遇和发展空间?

纪园园:上海的数字化转型,肯定不是简单的技术升级。我先说机遇,上海作为超大城市,它的治理场景特别丰富。政策来了,如果城市特别小,不方便进行一些试点。所以开展政策评估,很多时候都会让上海先行先试。它人口规模大、经济形态丰富、产业多元,有一定代表性,是很好的样本,如同天然试验田。

完善的数字基础设施也是一个机遇,方便及时评估、智能评估。上海有全国领先的“一网通办”,容易获取数据。随着智能治理体系的系统化构建,上海有望成为全球观察和研究智慧城市治理的重要窗口,在国际城市治理创新版图中发挥标杆作用。

上海还有一个特点,科研水平高。不少高校和科研机构对于政策评估都有研究,形成重要支撑。我觉得上海可以建立跨学科的评估平台,借助和国外的紧密联系,吸引国外专家参与,形成政府、高校、智库、产业的联合体系。

综合这三方面来看,上海在政策评价上具备制度创新的能力和优势。一些政策可以在上海率先探索,再推广复制到全国,上海很适合形成混合机制的评估模板。

文明餐饮 杜绝浪费

将光盘行动 进行到底