

绩效信息公开与政府响应

——基于政府门户网站建设第三方评估数据的断点回归分析

李晓方 孟庆国 王友奎*

【摘要】发布绩效评估榜单是第三方评估过程中经常采取的一种日常操作。在实践中，作为一种选择方式，许多项目评估方会选择性地将绩效优良的评估对象的工作业绩，而对较差部门的评估结果不予以公开。那么，这种选择性公开的行为会对评估对象的响应行为造成怎样的影响？以政府门户网站建设第三方评估数据为基础，通过断点回归设计实证检验了第三方评估的绩效公开与政府响应之间的关系。研究发现：公开第三方评估的结果有利于增强评估对象之间的竞争；对于进入公开榜单的优秀单位而言，公开会增强其持续学习行为；同时，与未进入公开榜单的对象相比，公开榜单的行为会降低评估对象在既有评估指标上的努力。

【关键词】第三方评估 绩效公开 政府响应 门户网站评估

【中图分类号】 D63

【文献标识码】 A

【文章编号】 1674-2486 (2019) 05-0115-18

一、文献综述与问题提出

作为一种由相对独立的组织实施绩效评估方式，第三方评估在引导组织自身建设、加强监督管理、促进社会进步等方面发挥了巨大作用。近年第三方绩效评估也多次被列入国务院常务会议的重要内容。因此，关注探索政府绩效的第三方评估问题，有较强的必要性和现实意义。

* 李晓方，南京理工大学公共事务学院，讲师；孟庆国，清华大学公共管理学院，教授；王友奎，清华大学公共管理学院，博士研究生。感谢匿名评审人的意见。

基金项目：中央高校基本科研业务费专项资金资助项目“大数据视域下‘互联网+机构编制管理’应用方法与路径研究”（30918013117）、国家社科基金项目“基于数据驱动的公共政策问题感知与推演研究”（17BZZ079）、国家社科基金项目“基于大数据应用的地方政府治理模式创新研究”（16BGL155）。

在理论研究方面，围绕第三方评估的定义、模式、特征以及第三方评估应用中的可行性和局限性等问题学者已经进行了较为详尽的分析。徐双敏和李跃（2000）认为，我国目前存在专家学者第三方、专业公司第三方、社会代表第三方和普通民众第三方四种不同第三方评估形式。包国宪和张志栋（2008）的研究关注到第三方评估的自律机制方面问题，针对我国第三方评估起步较晚的现实，提出应该借鉴组织自律的相关知识强化第三方评估的自律性。尚虎平和王春婷（2016）则更加关注第三方评估的国际经验对中国的借鉴意义，在系统回顾美国、英国和澳大利亚等国家第三方评估规律的基础上，提出中国政府绩效的第三方评估应该基于“有效把握公共受托责任”“纳入第三方智慧”“利益相关者参与的第三方评估”等规律展开。

从实证分析角度看，学者的探讨主要围绕采纳第三方评估的动因、第三方评估与绩效之间的相互关系以及第三方评估有效性的合理评价三个方面问题展开。在面对多种形式的绩效评估方式时，组织为什么选择由第三方实施外部评估而非选择内部评估的形式？对这一问题的回答，有的学者从组织自身角度出发，关注到“知识、成本、灵活性、目标、可问责性以及组织自身乐于接受外部批评程度”等系列因素影响（Conley - Tyle, 2005: 5）。同时，也有学者提出地方评估方式的选择离不开总体的社会政治、经济等环境变化的影响（Pattyn & Brans, 2013）。在分析国内政府第三方评估的选择时，包国宪等（2010）曾提出政府绩效第三方评估甘肃模式的选择是非公有制经济发展、中国加入 WTO 的客观环境以及政府创新等多重影响的客观结果。

第三方评估与绩效之间的相互关系实际上涉及到第三方评估方案、组织、实施、结果运用的过程与评估效果之间的复杂关系。逻辑上讲，又可以进一步拆分成三个密切相关的子问题。即第三方评估有没有效果？哪些因素影响以及如何影响这一效果？如何进一步改进第三方评估的效果？以环境保护领域为例，亚伦·查特吉（Aaron. K. Chatterji）等（Chatterji et al., 2009）对第三方评估与绩效之间的关系进行了探讨，认为作为企业社会责任第三方评估的多米尼 400 社会指数（Domini 400 Social Index）较好地反映了企业过去的绩效，并且在第三方评估中排名较好的企业，下一期往往会出现环保违规行为增多的趋势。阎波（2018）以比较案例研究为基础，对第三方评估方案的要素构成及其效果之间的关系进行了分析，提出评估方案和机制设计的科学性直接影响到评估效果，设计不当的评估机制可能诱发机会主义行为。在影响机制方面，组织学习、绩效反馈等因素作为连接评估和绩效之间的桥梁因素，得到学者广泛的认同（Sanderson, 2001; Moynihan, 2005）。马亮（2018）在一篇回溯性的文章中对

这些要素进行了总结,将绩效反馈、绩效问责、组织学习和组织创新等视为中介变量,而评估目的、评估参与、评估支持和评估使用等视为左右第三方评估奏效程度的自变量和调节变量。

第三方评估有效性的评价则主要围绕评估自身而展开,是一种对评价再评价的“元评价”问题(Espeland & Sauder, 2007)。新公共管理运动开始以来,随着政府绩效第三方评价和排名日益增多,如何评估和看待这些评估本身的建构效度问题,成为方法论上不可避免的问题(Gormley, 1999)。奥利弗·詹姆斯(Oliver James)和卡洛琳·皮特森(Carolyn Petersen)(James & Petersen, 2017)以复制试验(Replication Experiment)为基本方法对联合国电子政务调查的影响进行了评价,发现相对于以国家为基础的绩效评估而言,国际组织的评价和排名有利增强公民对政府绩效的感知和信任感。在国内研究方面,马亮和于文轩(2013)以比较研究为基础对国内政府绩效的一些典型第三方评价进行再评价,发现同一事项的多个评价虽然具有较强的相关性,但也存在一些重要的差异,分析和总结这些差异对于改进第三方评估的质量意义重大。

综上,已有研究为更好的理解第三方评估、以及评估与绩效之间的相互关系提供了有益的借鉴。但值得注意的是,第三方评估是一个涉及多领域、包含多环节的实践活动,既有研究多以案例和理论研究为主,对变量间因果关系的考察则相对薄弱。因此,对评估活动和效果之间关系的探索还有待更多领域以及更深一步的实证分析。本研究主要关注第三方评价与评价效果之间的关系问题。具体来说,主要分析在第三方评估的结果信息公开对政府响应行为的影响。在研究数据的来源上,主要选择政府门户网站绩效的第三方评估数据进行验证分析。基于门户网站第三方评估的实证研究,既可以拓宽第三方评估理论在不同领域适用性的认识,同时也可以加深对第三方评估过程中的操作性差异影响的认识。

文章后续内容做如下安排。首先介绍政府门户网站建设绩效第三方评估的实施背景和开展情况,并结合已有研究提出研究假设。其次汇报实证研究的方法、数据来源和结果。最后,文章对研究的结论、局限和下一步的研究方向进行进一步的总结和讨论。

二、政策背景与理论假设

(一) 政府门户网站建设与第三方评估

门户网站是政府向公众和社会提供信息和政务服务的窗口。较早地,从20

世纪90年代中期开始,陆续有政府架设门户网站向公民提供各类服务。1999年由中国电信和国家经贸委经济信息中心联合全国人民代表大会以及四十多家部委召开“政府上网工程启动大会”,中国网上政府建设快速展开。2005年中央人民政府门户网站正式试运行。门户网站建设在全国范围内体系化展开。2006年以后,随《政府信息公开条例》的颁布和实施,门户网站作为政府信息公开的重要载体,在内容建设和政府服务方面日益发挥着重要的作用(孟庆国、李晓方,2017)

与政府门户网站发展几乎同步,从2002年开始,赛迪顾问股份有限公司(中国软件测评中心)主导的“中国门户网站建设绩效评估”持续展开。整个评估工作覆盖了国务院部委及相关单位的部门网站、省级政府网站以及200多个地级市政府网站和部分县级市政府网站。截至2017年,门户网站建设评估工作已经连续开展了十六届。在评估指标的设计上,软件测评中心采取了“总体稳定”和“动态调整”相结合的策略。每年在保持基础性指标稳定的基础上,结合国务院最新工作要求,评估指标都会动态调整。在评估工作的组织实施和结果发布方面,测评中心也结合评估经验进行了动态调整(表1)。

表1 中国软件测评中心2010—2015年地级市政府门户网站绩效评估指标

年份	网站评估指标(一级指标)	一级指标变化情况
2010	信息公开、互动交流、日常保障、用户调查、教育服务、社保服务、就业服务、医疗服务、住房服务、交通服务、证件办理、企业开办、资质认定	重点将在线服务指标细化为10大领域,以一级指标的方式展示
2011	信息公开、互动交流、办事服务、日常保障、教育服务、社保服务、就业服务、医疗服务、住房服务、交通服务、证件办理、企业开办、资质认定、婚育收养、公用事业、经营纳税	新增3个办事领域,并以一级指标的方式展示
2012	信息公开、互动交流、办事服务、运维保障、新技术应用、舆情引导	新增舆情引导、新技术服务能力指标。将重点办事领域合并至办事服务指标,并以一级指标的方式展示
2013	信息公开、政策引导、互动交流、在线办事、新技术应用能力、特色服务	新增特色服务指标

(续上表)

年份	网站评估指标 (一级指标)	一级指标变化情况
2014	信息公开、办事服务、互动交流、舆论引导、网站功能与管理、用户满意度、数据开放	整合新设功能与管理指标、数据开放指标
2015	健康状况、信息公开、办事服务、互动交流、回应关切、网站功能、优秀案例	增加优秀案例指标, 整合新设健康状况一级指标

资料来源: 中国软件测评中心政府门户网站绩效评估报告 (2010—2015)。

长期积累的评估数据和评估过程的适应性调整为分析第三方评估如何影响政府行为提供了丰富的素材。就本研究而言, 主要关注绩效评估结果发布方式的调整对政府行为的影响。在中国软件测评中心“政府门户网站建设”的评估实践中, 从2013年起就不再发布地市以及区县绩效排名在100名以外的榜单数据。这一变化意味着评估工作以全面公开转变为以“奖优”为主, 同时也为分析第三方评估的结果公开对政府行为的影响提供了鲜活的数据。实证研究部分主要使用2013年、2014年两年评估数据。

(二) 理论与假设

第三方评估对政府行为和绩效的影响受制于特定的条件和方式。首先, 评估者自身的权威性和公信力被认为是评估产生影响的重要因素 (James & Petersen, 2017)。独立客观并有着较高感知有用性和知识性的第三方评估评估结果往往容易引起人们的重视, 并发挥影响。其次, 评估的目的、评估方案自身设计的科学性和合理性的影响也不容忽视。趋同的评估指标设计、泛滥的评估模式与政府绩效不彰共存是目前绩效评估的一大困境 (尚虎平、赵盼盼, 2012)。再者, 过程性因素也被认为是影响评估效果的重要机制。利益相关者能否从评估中获益 (Fletcher & Dyson, 2013)、评估过程中的参与方式以及评估者与被评者之间的是否相互信任均能影响评估效果的发挥 (Helden et al., 2012)。最后, 评估结果的使用方式也被普遍认为是影响评估发挥作用的重要条件和因素。评估结果的可用性、易用性 (Saunders, 2012) 以及管理者是否象征性地使用评估结果都对评估的影响关系重大 (Weiss, 1998)。

在影响机制方面, 第三方评估发挥作用的方式可以依主体不同而有所差异。对于管理者而言, 重点在于加强监督管理; 对于被评价者而言, 则表现为引导支持; 从社会层面看, 则表现为透明机制引导下的社会选择 (崔月琴、龚晓蝶, 2017)。三类主体的三种作用方式实际上代表了三种不同的类型机制。其中, 监

督和管理与绩效问责和强制有关，而引导支持则与组织的学习过程密切相关，社会选择机制意味着竞争和资源的争夺（王名等，2001）。

就本研究而言，主要关注第三方评估绩效榜单的发布方式对政府行为的影响。评估榜单的发布作为第三方评估结果运用中的一个重要组成部分，对其是否发挥作用和发挥作用的方式的探讨自然也离不开问责、组织学习和竞争机制等重点要素。因此，有关绩效公开对政府回应行为的影响从如下三个维度进行展开。

1. 绩效公开与政府的执行改进

关于评估信息公开功能的认识，有着两种不同的观点。一种观点是将信息公开视为一种合法化机制。它主要强调组织结构及其行为应与制度环境的价值和规范要求保持一致（Thornton & Ocasio, 1999）。根据这一理论，绩效信息较差的单位会在下一期工作中更加努力，以增强自身合法性（Cho et al., 2006）。查特吉和迈克尔·托菲尔（Michael W. Toffel）（Chatterji & Toffel, 2010）的研究关注第三方评估中绩效较好的行动者在下一期的反映。在对多米尼 400 社会指数与企业的环境保护行为的关系进行分析后，他发现那些过去在多米尼社会指数上取得较好成绩的单位，在下一期其环境污染和环境违法的数量会有显著的轻微上升。本文关注的中国软件测评中心公开的绩效榜单是一种“优秀”榜单（评估得分居前 100 名）。因此，结合已有研究提出如下假设：

H1：与进入绩效信息公开榜单的政府相比较，因成绩相对较差而未被公开成绩的地方政府会更加努力改进其现有工作的不足。

2. 绩效公开与政府的持续创新

公开的绩效信息还可以视为对政府过往行为的一种绩效反馈。特别是当评估指标随评估经验的积累不断变动时，持续进行第三方评估还可以起到引领和指导作用。实际上，根据组织学习理论，信息反馈引致的组织学习有单循环学习和双循环学习之分。单循环学习是以改变组织行动错误或不足为目标的工具性学习，它以保持基本价值不变的方式改变行动的策略。而双循环学习除了发现和更正组织既有的不足外，更多的是对组织既有的工作范式、流程和目标等提出较大的修正。双循环学习是行动的策略和价值、目标同时发生变化的学习（Argyris & Schön, 1997）。假设 H1 强调的实际上是单循环学习。公开的“优秀”榜单会给地方政府在已有指标方面做得不错的印象，从而降低其在既有指标上的努力。而对于未进入公开榜单数据的政府而言，其私下则可以通过主动问询的方式了解自己的行为表现。这样，评估实践可以为地方政府部门提供了

一个系统的组织诊断，为其识别差距和找准方向提供条件。本研究的第二个假设主要关注的则是组织学习的双循环反馈效应。根据双循环的反馈学习机制和组织运行的合法性机制，可以预期那些试图保持“优秀”位置的地方政府，总会更加积极地关注作为新价值引入的新增评估指标上的工作。

H2：与未进入绩效信息公开榜单的政府相比较，进入绩效公开榜单的单位会更加重视评估新增指标上的持续学习和创新。即以“奖优”为主的绩效公开会促进地方政府的持续学习和创新。

3. 绩效公开与政府之间的竞争程度

信息公开还可以被视为是降低信息不对称性、激发竞争和社会选择的一种机制。在对第三方评估排名和企业回应的一项研究中，查特吉和托菲尔（Chatterji & Toffel, 2010）指出第三方的评估排名能够影响消费者和投资者等利益相关者选择，从而对被评价对象产生影响。同样，在政府治理的过程中，独立客观的第三方评估结果有利于增强公民对政府的信任，提高政府的合法性（James & Peterse, 2017）。第三方评估和排名的社会实践实际创造了一种独特的声誉激励和羞辱式治理（Jin & Leslie, 2009）。因此可以预期公开的绩效评估榜单可以诱发地方政府之间的更为激烈的竞争。

H3：与绩效不公开的情形相比，在绩效公开的情况下地方政府在下一年度绩效评估中竞争更加激烈。

三、模型、方法与数据

（一）模型与估计方法介绍

信息公开榜单的发布可以视为第三方评估机构对地方政府门户网站建设的一种干预。对于干预效应检验常用的方法包括倾向得分匹配、双重差分以及断点回归等。其中，断点回归方式适用于处理变量完全由某变量是否超过某断点来决定情形。某种意义上，断点回归可视为“局部随机试验”，它能较好地处理可忽略性假设带来的问题（安格里斯特、皮施克，2012）。因此本文用断点回归进行研究。

本研究关注的主要处理变量为绩效信息是否公开发布。在中国软件测评中心的评估实践中，从2013年起便不再发布排名100名以后的地级和县级地方政府“门户网站建设绩效测评”的数据，而采用由地方政府自主咨询评估成绩的方式。绩效信息是否发布实际上依据的是排名数据是否超过100。因此，文章采

用清晰断点回归设计的方式对因果效用进行检验。在具体的模型选择方面，主要选择局部线性回归模型进行估计，并在后文中进一步提供二阶、三阶局部多项式回归的结果对稳健性进行检验。局部线性回归模型如下：

$$Y_i = \alpha + \beta(rank_i - 100) + \delta D_i + \gamma D_i(rank_i - 100) + \theta S_i + \varepsilon_i (i = 1, 2, \dots, n)$$

其中， Y_i 为被解释变量，表示研究关注政府的回应行为，本研究中主要包括三类回应行为。 $rank_i$ 表示地级市政府的排名，是配置变量。当 $rank_i > 100$ 时， D_i 的取值为 1，其余则为 0。 D_i 为处理变量，表示该年度地方政府门户网站建设第三方绩效评估数据是否公开。 D_i 的取值为 1 表示未公开。 S_i 为协变量，表示系列与 Y_i 相关的影响因素。 δ 即处理效应，表示公开不公开绩效评估数据对 Y_i 的影响。

样本数据选择方面，文章选择中国软件测评中心 2013—2014 两年地级市政府门户网站的测评数据进行评估。之所以选择这两年数据是 2013 年是地级市的网站评估数据选择性公开的第一年，且从 2013 年到 2014 年绩效评估指标有部分调整（表 1）。选择这两年数据有助于考察分析绩效公开的对地方政府执行改进以及持续创新的影响。文章剔除部分缺失数据以后共获得 243 个地级市的观察数据。文章使用 Stata 的“rd”组件进行估计和检验（Nichols，2011），基于局部多项式稳健性检验则通过 R 软件的“rdrobust”系列组件进行估计。

（二）变量具体解释和测量

被解释变量。政府回应行为是本文关注的核心解释变量。根据地方政府面对绩效公开榜单这一干预因素时行为反应的不同，我们进一步区分了三种类型的回应。一种是执行改进，关注的是政府在面对评估信息反馈时对既有工作内容的加强和改进，用地方政府在 2014 年的信息公开指数得分与 2013 年公开指数的得分变化来表示。根据绩效评估理论，第三方评估提供的绩效反馈可以为评估对象提供组织诊断，识别工作不足从而推动其工作改进。这反映了干预在促进组织单循环学习方面的作用。第二种回应行为是地方政府持续创新，用地方政府在 2014 年中国软件测评中心新增指标网站功能指数方面的得分来表示。根据测评中心的测评实践，它每年会根据新技术发展和国务院重点工作安排对指标进行适当调整。其中，新增指标代表了一种新的技术应用或者是新的工作重点。用地方政府在该指标上的表现来表征其持续创新行为，实际上是对地方政府双循环学习的一种反映，考察其是否能够根据技术发展要求和中央精神及

时跟进和调整工作中心。最后一种是地方政府之间的竞争程度,用相比2013年地方政府在2014年排名变化的程度来测量。如果公开造成了地方政府之间竞争程度的增强,可以预期其排名变化幅度应该会有相应的增加。

处理变量与配置变量。处理变量表征的是一种干预。文章中表示是否公开地方政府的评估绩效。即地方政府是否进入了公开的绩效榜单。在中国软件测试中心的评估实践中,从2013年开始,是否公开地级市政府门户网站建设的评估绩效是依据其排名来进行的。如果排名在100名以内(含100名)则公开其绩效,否则不予公开,没被公开的地方政府可以自愿咨询评估绩效。这实际上是将排名作为配置变量,依据排名是否大于100,作为处理分组。因此,文章以地方政府在2013年的绩效评估排名作为配置变量,而将地方政府评估绩效是否公开作为处理变量。

控制变量(协变量)。由于断点回归可以视为局部随机试验,是否包含随机变量并不影响断点回归估计量的一致性。加入协变量的好处在于减少扰动方差,使估计更准确(陈强,2014)。本文选择人均地区生产总值、人均公共预算收入和互联网的普及率作为控制变量。其中,互联网普及率用国际互联网的宽带接入户数除以该地区的年末常住人口数表示。之所以选择这三个指标作为控制变量,一方面是因为这些要素与政府信息公开的动力和能力要素密切相关,另一方面在已有研究中学者已经注意到并验证过这些要素对信息公开的各种行为的影响(马亮,2012;阎波等,2013)。变量数据主要来自《中国城市统计年鉴》和地方政府统计公报。

表2 主要变量及其计算方式

变量名称	变量含义	计算方法
excu	执行改进	2014年信息公开指数得分除以2013年信息公开指数得分
continvo	持续创新	2014年网站功能指数得分
compet	竞争程度	2014年排名减去2013年排名的绝对值
rank	排名	2013年网站建设绩效排名
d	是否公开	排名大于100时不公开其信息,此时值为1,其余为0
lnpergdp	人均地区生产总值对数值	地区生产总值取对数
lnperfina	人均公共预算收入对数值	人均公共预算收入取对数
pernet	网络普及率	国际互联网宽带用户数除以年末常住人口数量

资料来源:作者自制。

四、实证结果

（一）描述性统计与局部处理效应

表3展示了公开组和非公开组在核心变量上的描述性统计结果。从数据可以看出公开组在执行改进、持续创新和竞争程度三个变量上的均值都高于非公开组。其中，竞争程度变量因为其反映的是排名变化，因计量单位原因方差较大。

表3 核心变量描述性统计

变量名	执行改进		持续创新		竞争程度	
分组	公开	不公开	公开	不公开	公开	不公开
样本数	99	145	99	145	99	145
均值	0.88	0.47	0.57	0.27	37	32
方差	0.01	0.06	0.01	0.03	1 094	892
最大值	1.1	1.01	0.79	0.63	126	122
最小值	0.48	0.19	0.34	0.11	1	0

资料来源：中国软件测评中心政府门户网站绩效评估报告（2013—2014）。

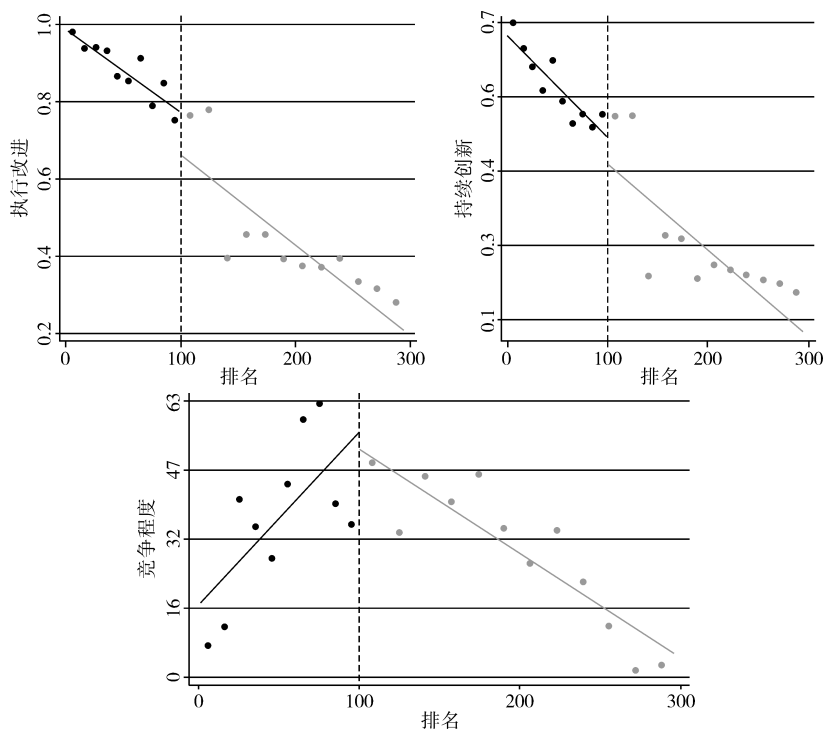


图1 榜单城市与非榜单城市断点附近因变量值比较

资料来源：中国软件测评中心政府门户网站绩效评估报告（2013—2014）。

更进一步，图 1 以图形化的方式展示以全样本数据进行断点回归的结果。图形化的方式有利于断点两侧的响应差异。从图形情况来看断点处均出现明显的跳跃。但考虑到断点的选择是依据其上一年度的总分排名，因此使用全样本分析很难判定这种差异是公开本身造成的影响还是其他因素（如自选择偏差）所造成的影响。下一节将进一步通过局部线性回归对结果进行验证分析。

（二）绩效公开与政府回应：局部线性回归的基本结果

以 2013 年的排名数据作为配置变量，使用三角核（Triangular Kernel）估计的方法对回归方程进行非参数估计。表 4 展示了估计结果。从表 4 中可以看出，在不添加协变量的情形下，是否进入公开榜单对政府回应行为影响在估计带宽为二分之一最优带宽的情形下有较为显著的影响。即公开绩效评估榜单，以“奖优”方式对地方政府进行引导的干预行为对政府绩效改进、持续创新和竞争程度等都有较为显著的影响。具体表现为，相对于进入榜单的地方政府而言，未进入公开榜单的地方政府在下一期会更加努力改变既有指标上的工作绩效。而进入榜单的地方政府在持续创新方面表现更加。“奖优”的公开行为促进地方政府的双循环学习和持续创新。在促进竞争方面，公开增了地方政府之间的竞争程度。在添加协变量以后，在带宽为二分之一最优带宽的情况下，公开对持续学习、竞争的影响表现出了一定的稳定性。公开对执行改进的影响则转为不显著。在最优带宽、两倍最优带宽情形下，结果稳定性并不理想。

（三）模型设定检验

1. 断点的连续性和内生性分组问题检验

在断点回归时，如果个体事先知道分组规则，并通过自身努力而完全控制分组变量，则可自行进入处理组和控制组，导致断点附近的是内生分组而非随机分组，引起断点失效（陈强，2014）。就本研究而言：一方面 2013 年是以排名 100 为断点决定绩效是否公开第一年，报告发布之前并未明确告知地方政府。另一方面，排名结果取决于地方政府在网站建设方面的相对努力程度，单一地方政府很难进行控制，因此理论上内生性分组的问题不存在。就数据检验方法而言，常用的检验方法是 McCrary 检验。其检验的原假设为配置变量的密度函数在断点处连续，即： $H_0: \theta = \lim_{x \downarrow c} f(x) - \lim_{x \uparrow c} f(x) = 0$

根据回归结果 $\hat{\theta} = 0.146$ ，而标准误差为 0.402（P 值为 0.72）。因此，不能拒绝原假设，说明模型内生性分组问题不存在。图 2 展示了断点附近的概率密度情况。

表 4 绩效公开与政府响应的关系检验

被解释变量	执行政进			持续创新			竞争程度		
	lwald	lwald50	lwald200	lwald	lwald50	lwald200	lwald	lwald50	lwald200
不包含协变量情形									
处理效应(δ)	.092 (1.34)	.195*** (12.35)	.008 (0.91)	-.214*** (-5.01)	-.112*** (-7.54)	-.090 (-1.33)	-27.4 (-2.19)	-56.201*** (-2.87)	.863 (0.04)
Bandwidth	4.034	2.017	8.07	3.313	1.657	6.626	12.94	6.47	25.89
含协变量情形									
处理效应(δ)	-.0049 (-0.04)	0	.0117 (0.15)	-.088*** (-2.09)	-.104	-.0447 (-1.18)	-29.697 (0.11)	-56.391*** (-2.74)	-6.399 (-0.36)
Bandwidth	4.034	2.017	8.069	3.313	1.657	6.626	12.944	6.472	25.89

注：lwald 表示局部 Wald 估计值。后面 50、100 的数字表示带宽与最优带宽的比例。没有数字表示带宽为依 Ibens & Kalyanaraman 计算方法计算的最优带宽。50 表示二分之一最优带宽，200 表示两倍最优带宽。***、**、* 分别表示系数在 0.05、0.10 的显著水平上显著。

资料来源：中国软件测评中心政府门户网站绩效评估报告(2013—2014)。

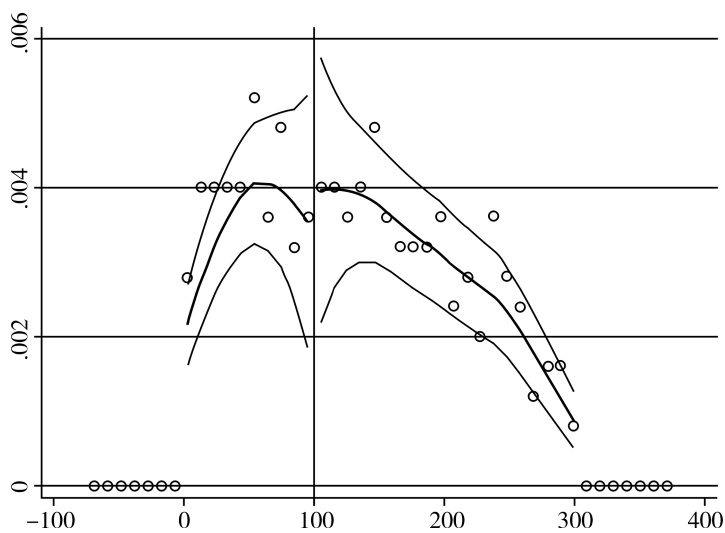


图2 断点附近分组变量的概率密度图

资料来源：中国软件测评中心政府门户网站绩效评估报告（2013—2014）。

2. 协变量在断点附近的连续性检验

在进行断点回归分析时，如果协变量在断点处也存在跳跃，此时不宜将处理变量 d 的系数全部归功于处理效应（陈强，2014）。因此，对处理效应的确认需要考察协变量在断点处是否存连续。表5展示了带宽为二分之一最优带宽的条件下连续性的检验结果。可以发现，在被解释变量为执行改进和持续创新时， $\ln\text{pergdp}$ 在断点附近连续性假设被拒绝，但结合没有协变量情形的模型检验结果，可以认为公开对政府回应的显著影响依然可以得到确认。

表5 协变量在断点附近的连续性检验

协变量 \ 被解释变量	执行改进	持续创新	竞争程度
$\ln\text{pergdp}$	-2.093976 ** (-0.56)	- .849144 ** (-2.70)	- .6277112 (-1.03)
$\ln\text{perfina}$	- .7336173 (-0.56)	- .7336173 (-5.6)	- .5803716 (-0.53)
pernet	- .0097745 (-0.29)	- .0097745 (-0.29)	.0209202 (0.39)

注：**表示系数在0.05的显著水平上显著。

资料来源：中国软件测评中心政府门户网站绩效评估报告（2013—2014）。

（四）稳健性检验

1. 基于安慰剂效应的稳健性检验

安慰剂效应的基本思想是如果断点选择任何一个位置处理效应依然存在，那么，回归的显著性可能只是一种机械的相关关系（刘行等，2017）。因此，基于安慰剂效应的稳健性检验往往都选择任意非断点的其他值作为假想断点对处理效应进行估计（李江一、李涵，2017）。表6展示了带宽为二分之一最优带宽，断点为62、102时的处理效应，结果显示在排名为62、102名附近回应行为并没有显著变化。

表6 安慰剂效应检验

被解释变量	执行改进		持续创新		竞争程度	
断点	62	102	62	102	62	102
处理效应	0	0	0	0	-30.35991 (-0.45)	18.86862 (0.632)
Bandwidth	1.902	2.10	1.47	1.393	7.896	9.25
协变量	否	否	否	否	否	否

资料来源：软件测评中心政府门户网站绩效评估报告（2013—2014）

2. 基于局部多项式回归的稳健性检验

局部线性回归是建立在断点两侧数据分布的线性趋势基础上的。当断点两端的函数形式是曲线时，线性估计的方式可能会夸大断点附近的处理效应。因此，近年有部分学者开始倾向于用局部多项式回归的方式进行处理效应检验（刘生龙等，2016）。但值得注意的是，局部多项式回归实质上要求一个中等或较大规模的样本（Calonico et al., 2018）。由于本文选择的以排名为配置变量的研究设计，所用数据为年度数据，断点两侧数据点较少，所以采用了局部线性回归的方式。采用局部多项式回归时，回归方程的拟合估计需要较大的带宽，这会进一步影响结果的可信性。作为对照，表7给出了以“msetwo”为最优带宽的估计方法时，二阶、三阶局部多项式回归的结果。从回归结果来看，当采用二阶、三阶多项式估计时，仅有极个别的因变量的处理效应较为显著。这也提示我们进一步关注到对局部线性回归结果未来需要大样本情形的进一步确认。

表 7 基于局部多项式的稳健性检验

被解释变量	执行改进	持续创新	竞争程度	执行改进	持续创新	竞争程度
处理效应 (δ)	-0.005	0.000	19.004	-0.201	-0.052 *	20.486
最优带宽 (左侧)	17.9	31.1	30.2	14.1	21.1	15.8
最优带宽 (右侧)	34.3	33.2	62.7	12.88	8.3	34.6
多项式阶数	2	2	2	3	3	3

注：*表示系数在 0.10 的显著水平上显著。
资料来源：中国软件测评中心政府门户网站绩效评估报告（2013—2014）。

五、结论与讨论

作为一种独立评估方式，第三方评估在引导和改进政府行为方面发挥了日益重要的作用。已有研究文献从理论和案例的角度出发做出了许多有益的研究发现。但总体来看对第三方评估如何以及在何种情况下影响政府绩效尚缺乏大样本数据的实证研究（马亮，2018）。本文以量化研究为基本方法对第三方评估过程中榜单发布这一行为影响进行验证分析。研究发现：公开第三方评估的结果有利于增强评估对象之间的竞争；对于进入公开榜单的优秀单位而言，公开会增强其持续学习行为；同时，与未进入公开榜单的对象相比，公开榜单的行为会降低评估对象在既有评估指标上的努力。研究加深了对绩效公开和政府回应行为之间关系的认识。值得注意的是，在第三方评估的过程中，评估方案的设计、评估前的沟通、评估过程中的参与等诸多要素都会影响评估效果的发挥。不同的因素与政府行为之间的关系有待进一步的实证分析和检验。在研究效度方面，需要说明的是，由于文章选择“门户网站建设绩效评估”的数据通过局部线性回归的方式进行断点分析，因此在研究结论的一般化推广受制于三方面因素的影响。

这三方面因素为：（1）“政府门户网站建设绩效评估”实施者自身的独特性。“政府门户网站建设绩效评估”是由中国软件测评中心具体负责实施的第三方评估。从隶属关系上讲，中国软件测评中心是赛迪顾问集团的组成单位，而赛迪又直属于国家工业和信息化产业部，可以说这样的第三方评估是有一定公信力的评估。评估机构本身的权威性和可信度可能是评估结果发挥影响的更为首要性要素。本研究关注的则是给定评估机构自身地位基础上，结果沟通方式变化对政府行为的影响。那么，对于社会独立第三方评估而言，沟通方式的调整会对政府的响应行为造成怎样影响呢？对这一问题的回答一方面依赖于被评

估者看待第三方评估的态度。只有行为者重视的第三方评估才可能引起反馈和学习效应。另一方面，社会独立第三方评估的社会影响力、评估的客观性和公正性对政府响应的影响可能需要进一步分析。(2)从评估设计和实施的目的来看，“门户网站”的绩效评估方案属于支持引导性的评估。评估结果并不与部门的奖惩和问责相挂钩。自然，公开透明则成为评估发挥作用的核心支撑机制。这也是考察结论的外部有效性时需要注意的一个问题。(3)从方法上看，断点回归方法虽然可以通过局部随机试验的方式对因果关系进行验证，但该推断仅局限于断点，并不一定能推广到其他样本，其外部有效性受限。这一点从实证结果对带宽的敏感性也可以看出。未来可能需要进一步的量化和质性研究来解释从结果公布到政府响应之间的内在机制。

参考文献

- 包国宪、张志栋(2008). 我国第三方政府绩效评价组织的自律实现问题探析. 中国行政管理, 23(1): 49-51.
- 包国宪、董静、郎玫、王浩权、周云飞、何文盛(2010). 第三方政府绩效评价的实践探索与理论研究——甘肃模式的解析. 行政论坛, 17(4): 7-11.
- 陈强(2014). 高级计量经济学及Stata应用. 北京: 高等教育出版社.
- 崔月琴、龚小碟(2017). 支持性评估与社会组织治理转型——基于第三方评估机构的实践分析. 国家行政学院学报, 4: 55-60.
- 李江一、李涵(2017). 新型农村社会养老保险对老年人劳动参与的影响——来自断点回归的经验证据. 经济学动态, 3: 62-73.
- 刘行、赵健宇、叶康涛(2017). 企业避税、债务融资与债务融资来源——基于所得税征管体制改革断点回归分析. 管理世界, 10: 113-129.
- 刘生龙、周绍杰、胡鞍钢(2016). 义务教育法与中国城镇教育回报率: 基于断点回归设计. 经济研究, 2: 154-167.
- 马亮(2012). 政府信息公开的影响因素: 中国地级市的实证研究. 情报杂志, 9: 142-146.
- 马亮、于文轩(2013). 第三方公共服务绩效评价的评价: 一项比较案例研究. 南京社会科学, 5: 55-63.
- 马亮(2018). 第三方评估提升政府绩效的理论框架与研究展望. 江苏师范大学学报(哲学社会科学版), 44(2): 68-78.
- 孟庆国、李晓方(2017). 全面推进政务公开: 内涵诠释、实践特色与发展理路. 河南师范大学学报(哲学社会科学版), 2: 19-25.
- 乔舒亚·安格里斯特、约恩·斯特芬·皮施克(2012). 基本无害的计量经济学: 实证研究者指南. 郎金煊、李井奎译. 上海: 上海人民出版社.
- 尚虎平、赵盼盼(2012). 绩效评估模式泛滥与绩效不彰困境——基于42个案例的分析. 中国

- 行政管理, 11: 18-24.
- 尚虎平、王春婷(2016). 政府绩效评估中“第三方评估”的适用范围与限度——以先行国家为标杆的探索. 理论探讨, 190(3): 12-18.
- 王名、刘国翰、何建宇(2001). 中国社团改革: 从政府选择到社会选择. 北京: 社会科学文献出版社.
- 徐双敏、李跃(2010). 基于利益相关者理论的政府绩效第三方评估主体分析. 中国行政管理学会2010年会暨“政府管理创新”研讨会论文集.
- 阎波、李泓波、吴佳顺、吴建南(2013). 政府信息公开的影响因素: 中国省级政府的实证研究. 当代经济科学, 35(6): 67-77.
- 阎波(2018). 公共组织第三方评估何以可为? 一项双案例比较研究. 经济社会体制比较, 3: 125-135.
- Argyris, C. & Schön, D. A. (1997). Organizational Learning: A Theory of Action Perspective. *Reis*, 78(77): 345-348.
- Chatterji, A. K. & Toffel, M. W. (2010). How Firms Respond to Being Rated. *Strategic Management Journal*, 31(9): 917-945.
- Chatterji, A. K., Levine, D. I. & Toffel, M. W. (2009). How Well Do Social Ratings Actually Measure Corporate Social Responsibility?. *Journal of Economics & Management Strategy*, 18(1): 125-169.
- Calonico, S., Cattaneo, M. D. & Farrell, M. H. (2018). On the Effect of Bias Estimation on Coverage Accuracy in Nonparametric Inference. *Journal of the American Statistical Association*, 113(52): 767-779.
- Cho, C. H. & Patten, D. M. (2006). The Role of Environmental Disclosures as Tools of Legitimacy: A Research Note Accounting. *Organizations and Society*, 32(78): 639-647.
- Conley-Tyler, M. (2005). A Fundamental Choice: Internal or External Evaluation?. *Evaluation Journal of Australasia*, 4(12): 3-11.
- Espeland, W. N. & Sauder, M. (2007). Rankings and Reactivity: How Public Measures Recreate Social Worlds. *American Journal of Sociology*, 113(1): 1-40.
- Fletcher, G. & Dyson, S. (2013). Evaluation as Work in Progress: Stories of Shared Learning and Development. *Evaluation*, 19(4): 419-430.
- Gormley, W. T. (1999). *Organizational Report Cards*. Cambridge: Harvard University Press.
- Helden, G. J. V., Johnsen, Å. & Vakkuri, J. (2012). The Life-Cycle Approach to Performance Management: Implications for Public Management and Evaluation. *Evaluation*, 18(2): 159-175.
- Jin, G. Z. & Leslie, P. (2009). Reputational Incentives for Restaurant Hygiene. *American Economic Journal: Microeconomics*, 1(1): 237-67.
- James, O. & Petersen, C. (2017). International Rankings of Government Performance and Source

- Credibility for Citizens: Experiments about E – Government Rankings in the UK and the Netherlands. *Public Management Review*, 20(4): 469 – 484.
- Moynihan, D. P. (2005). Goal Based Learning and The Future of Performance Management. *Public Administration Review*, 65(2): 203 – 216.
- Nichols, A. (2011). RD: Stata Module for Regression Discontinuity Estimation. Available at (Feb. 15, 2019): <http://ideas.repec.org/c/boc/bocode/s456888.html>.
- Pattyn, V. & Brans, M. (2013). Outsource Versus In – House? An Identification of Organizational Conditions Influencing the Choice For Internal or External Evaluators. *Canadian Journal of Program Evaluation*, 28(2): 25 – 76
- Sanderson, I. (2001). Performance Management, Evaluation and Learning in Modern Local Government. *Public Administration*, 79(2): 297 – 313.
- Saunders, M. (2012). The Use and Usability of Evaluation Outputs: A Social Practice Approach. *Evaluation*, 18(4): 421 – 436.
- Thornton, P. H. & Ocasio, W. (1999). Institutional Logics and the Historical Contingency of Power in Organizations: Executive Succession in the Higher Education Publishing Industry, 1958—1990. *American Journal of Sociology*, 105(3): 801 – 843.
- Weiss, C. H. (1998). Have We Learned Anything New about the Use of Evaluation?. *American Journal of Evaluation*, 19(1): 21 – 33.