

# 断点回归设计方法应用的研究综述

谢 谦<sup>1</sup> 薛仙玲<sup>2</sup> 付明卫<sup>1</sup>

(1. 中国社会科学院经济研究所,北京 100836; 2. 中国社会科学院研究生院,北京 102488)

**[摘 要]** 近年来,国内经济学界颇为关注断点回归设计(RDD)方法,运用 RDD 的文章日益增多。首先概述 Lee 和 Lemieux (2010)提出的运用 RDD 的规范,然后基于 2011-2017 年五大英文顶级经济学期刊发表的 RDD 应用研究论文,归纳出运用 RDD 的三个新动作:新的最优带宽确定方法、甜甜圈 RDD 和参数估计中最高只能使用二次多项式。以国外的 RDD 用法为参照,我们发现,国内熟悉 Lee 和 Lemieux (2010)提出的规范,也在吸收国外文献中出现的新动作,但运用 RDD 时存在如下几个突出问题:不做适用性检验、不重视描述统计图和不交代清楚非参数估计的关键细节。建议国内运用 RDD 时需要增加如下动作:非参数估计应该同时尝试 CV、IK 和 CCT 三种带宽确定方法;在配置变量存在堆积现象或被操纵的可能时,使用甜甜圈 RDD;尝试到四次多项式。

**[关键词]** 断点回归设计;参数估计;非参数估计;最优带宽;甜甜圈 RDD

**[DOI 编码]** 10.13962/j.cnki.37-1486/f.2019.02.006

**[中图分类号]**F224 **[文献标识码]**A **[文章编号]**2095-3410(2019)02-0069-11

## 一、引言

20 世纪 80 年代以来,应用微观计量研究经历了一场可信性革命,变得十分强调研究设计(research design)。这种转变导致工具变量法(Instrumental Variable, IV)、双重差分法(Difference-in-Difference, DID)和断点回归设计(Regression Discontinuity Design, RDD)成为应用微观计量研究中运用最广泛的方法<sup>[1]</sup>。与工具变量法和双重差分法相比, RDD 更接近于随机试验,因而从理论上讲是更好的因果识别方法<sup>[2]</sup>。RDD 虽然早在 1960 年就被 Thistlethwaite 和 Campbell (1960)<sup>[3]</sup>提出,但 20 世纪 90 年代末之后才被国外大量应用于经济学研究。尽管运用 RDD 要求在断点附近有较多观测值,对数据要求很高,但过去几年的相关研究并不少。2011-2017 年,五大英文顶级经济学期刊<sup>①</sup>发表的运用 RDD 做的经验研究论文(下文简称“五大 RDD 论文”)共 39 篇<sup>②</sup>,

**[基金项目]** 国家社会科学基金项目“我国医院行业市场机制有效性的实证研究”(17BGL170)

**[作者简介]** 谢谦(1982-),男,河北保定人,中国社会科学院经济研究所助理研究员。主要研究方向:国际经济学。

<sup>①</sup>分别为 American Economic Review(AER)、Econometrica(ECMA)、Journal of Political Economy(JPE)、Quarterly Journal of Economics(QJE)和 Review of Economic Studies(RES)。

<sup>②</sup>不包括五大上关于 RDD 方法的理论研究论文。我们在 Jstor 数据库中用全文任一地方出现“regression discontinuity”这个关键词检索,然后剔除不是运用 RDD 做的经验研究论文,譬如 RDD 的理论研究论文、仅论文献综述或脚注中出现“regression discontinuity”的论文等。

各年分别为6篇、3篇、3篇、11篇、5篇、7篇和4篇(如图1所示),表明RDD在国际上已经成为一种主流的经验研究方法。根据在中国知网的检索,中文期刊发表的头两篇运用RDD做的经验研究论文是“退休会影响健康吗?”<sup>[4]</sup>和“城市群落的崛起、经济绩效与区域收入差距”<sup>[5]</sup>;2011-2017年间,中文CSSCI期刊共发表运用RDD做的经验研究论文46篇,各年分别为1篇、1篇、0篇、5篇、8篇、8篇和23篇(如图1所示)。我们预计,大数据时代的来临会极大地打破运用RDD面临的数据限制<sup>①</sup>,运用RDD的相关研究会越来越多。

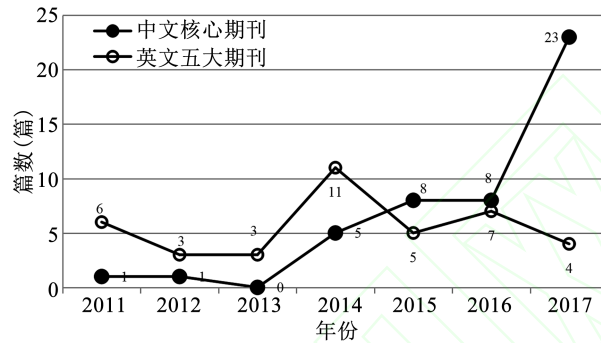


图1 中文核心期刊和英文五大期刊上发表的断点回归设计经验研究论文数<sup>②</sup>

Imbens 和 Lemieux (2008)<sup>[6]</sup>、van der Klaauw (2008)<sup>[7]</sup>、Lee 和 Lemieux (2010)<sup>[2]</sup>、Skovron 和 Titiunik (2017)<sup>[8]</sup>、Hausman 和 Repson (2018)<sup>[9]</sup>专门综述了RDD的理论和应用研究。余静文和王春超(2011)<sup>[10]</sup>简要介绍了断点回归设计的发展历史、理论、实施步骤和应用情况。但是,由于RDD这种方法的理论研究仍很活跃,运用RDD涉及参数估计和非参数估计的选择、参数估计中多项式次数的选择和非参数估计最优带宽的确定等众多技术细节,文献中在RDD的具体用法上存在很大差别,特别是中文文献运用RDD的规范程度明显滞后于外文文献,有时还存在误用。因此,十分有必要归纳RDD应用的新进展,为国内学者规范运用RDD提供借鉴。由于本文强调被五大RDD论文运用过的新进展,侧重于应用,我们不会涉及多配置变量RDD(RD designs with assignment variables)<sup>[11]</sup>、分位数RDD<sup>[12]</sup>、拐点回归设计(regression kink designs)<sup>[13]</sup>、多断点RDD(RD designs with multiple cutoffs)<sup>[14]</sup>、远离断点处的处理效应的识别方法(methods for extrapolation away from the cutoff)<sup>[15][16]</sup>、离散型配置变量RDD<sup>[17]</sup>等五大RDD论文中未运用的新进展。

本文综述 Lee 和 Lemieux (2010)<sup>[2]</sup>以来RDD的新进展。选择以 Lee 和 Lemieux

<sup>①</sup>这方面的一个表现是,大量行政管理数据(administrative data)被用于经济学经验研究。根据英国《经济学家》(The Economist)的报道,NBER(National Bureau of Economic Research)工作论文摘要里出现“administrative data”的论文数,2000年是0篇,之后快速增长,2017年达到28篇。报道链接为:<https://www.economist.com/international/2018/05/26/government-data-are-ever-more-important-to-economic-research>。

<sup>②</sup>中文核心期刊论文数的检索方法如下:在中国知网检索出主题包含“断点回归”、发表年度为2011年至2017年的中文社会科学引文索引(CSSCI)期刊论文,然后逐一检查、剔除非经验研究论文,最后剩下46篇。将检索限定在中文社会科学引文索引期刊,是为了保证检索到的论文的质量。英文五大期刊论文数的检索方法如下:在Jstor数据库中用全文任一地方出现“regression discontinuity”这个关键词、发表年度为2011至2017年、期刊为五大期刊检索到50篇,然后逐一检查、剔除非经验研究论文,最后剩下39篇。

(2010)<sup>[2]</sup>为基点的理由如下:首先,与 Imbens 和 Lemieux(2008)<sup>[6]</sup>、van der Klaauw (2008)<sup>[7]</sup>相比, Lee 和 Lemieux (2010)<sup>[2]</sup>的时间要晚。其次,与 Skovron 和 Titiunik (2017)<sup>[8]</sup>综述政治学中断点回归设计的应用情况、Hausman 和 Repon(2018)<sup>[9]</sup>综述时间作为配置变量(也叫驱动变量)的 RDD 应用情况相比, Lee 和 Lemieux (2010)<sup>[2]</sup>综述的是经济学中的应用情况,包括各种类型的配置变量,更综合、更全面。Lee 和 Lemieux (2010)<sup>[2]</sup>提出了运用 RDD 做经验研究的规范,算是一个共识。但是,通过归纳 2011-2017 年的五大 RDD 论文,我们发现:一方面很多论文都没有遵守此规范。另一方面涌现出了很多被广泛应用的新进展,譬如非参数估计中确定带宽的 IK 法<sup>①</sup><sup>[18]</sup>和 CCT 法<sup>[19]</sup>、参数估计中只应使用配置变量的低次项和甜甜圈 RDD (Donut hole RDD)等<sup>②</sup>。通过归纳 2011-2017 年 CSSCI 期刊发表的 RDD 经验研究论文,我们发现,国内总体上了解运用 RDD 的规范,但仍然存在不少问题,譬如不做适用性检验、不重视描述统计图、不交代清楚非参数估计的关键细节等。

## 二、运用断点回归设计的规范和新进展

### (一)运用断点回归设计做研究的既定规范

Lee 和 Lemieux (2010)<sup>[2]</sup>提出了运用 RDD 做经验研究的规范,概述如下:

1.利用配置变量检验 RDD 的适用性。RDD 的前提条件是个体不能精准操控 (precisely manipulate) 配置变量。此条件可利用配置变量来检验。首先,选定一定数量的箱体画出配置变量的历史直方图。如果频数 (frequencies) 在邻近断点处的两个箱体间存在跳跃式变化,则很可能意味着前提条件不成立。其次,做配置变量的 McCrary 检验<sup>[20]</sup>。此检验的 stata 命令 DCdensity 是外部命令,可在下述网址下载后安装: <https://eml.berkeley.edu/~jmccrary/DCdensity/>。

2.利用前定变量检验 RDD 的适用性。首先,选定一定数量的箱体,求前定变量在每个箱体内的均值,画出均值对箱体中间点的散点图;接着再画出使用多项式模型对前定变量做出的拟合曲线,其中多项式一般是 4 次多项式<sup>③</sup>。如果前定变量的均值在断点处存在跳跃,则意味着前提条件很可能不成立。其次,用前定变量对常数项、处理变量、配置变量的多项式、处理变量和配置变量多项式的交互项做回归。如果前定变量比较多,那么随机因素也可能导致某个前定变量存在显著的断点。因此,有必要把众多检验合并为一个检验所有前定变量都不存在断点的统计量。这时,应该用似不相关回归 (Seemingly Unrelated Regression, SUR)。

3.画结果变量的图。选定一定数量的箱体,求结果变量在每个箱体内的均值,画出均值对箱体中间点的散点图;接着再画出使用多项式模型对结果变量做出的拟合曲线,其中多项式一般是 4 次多项式。一定要求了局部平均值后再画图。如果直接画原始数据的散点图,那么噪音太大,看不出潜在函数的形状。如果结果变量的均值在断点处存在跳跃,则意味着处理变量有影响。用图来显示结果变量和配置变量之间的关系,已经成为运用 RDD 的标准作法。Lee

① Imbens 和 Kalyanaraman (2009) 是工作论文,后来于 2012 年发表在《Review of Economic Studies》上。

② 用 Stata 软件实现这些新进展的操作命令可以参考 Cattaneo et al. (2018), 更详细的操作命令细节可参考本文所引用文献的网上附件。这些网上附件可在发表各引用文献的期刊官网下载。

③ 作者文中用的都是 4 次多项式,没解释不用其他次数多项式的原因。

和 Lemieux (2010)<sup>[2]</sup>指出,画图在 RDD 经验文章中是如此重要,以致没有图的文章会让人自然而然地怀疑图形证据对文章不利。

4.全样本参数估计。使用全样本数据进行参数估计。以精确断点回归为例,估计模型如下:

$$y = \alpha + \rho \times \text{Treatment} + \sum_{k=1}^K \beta_k \times (x-c)^k + \sum_{k=1}^K \gamma_k \times \text{Treatment} \times (x-c)^k + W + \varepsilon \quad (1)$$

其中  $y$ 、 $x$ 、 $\text{treatment}$  和  $W$  分别为结果变量、配置变量、处理变量(当  $x \geq c$  时,  $\text{treatment} = 1$ , 否则  $\text{treatment} = 0$ )和前定变量。 $\alpha$ 、 $\rho$ 、 $\beta$  和  $\gamma$  为待估计的参数,  $\varepsilon$  为随机扰动项。 $\rho$  显著不为 0,表明处理对结果变量有影响。

参数估计的要点是确定模型(1)中的多项式次数  $K$ 。一般从一次尝试到八九次<sup>①</sup>,然后挑选赤池信息准则(AIC)取值最小的模型。另外一个比赤池信息准则更好的方法是,在多项式模型的自变量中先加入表示箱体的虚拟变量,然后不断加入配置变量的多次项,只到下述原假设成立:所有箱体虚拟变量的系数均等于零。

5.子样本非参数估计。非参数估计就是对最优带宽内的子样本做加权局部线性回归。非参数估计的要点是确定最优带宽。确定最优带宽的方法有两种<sup>②</sup>:拇指规则法(rule of thumb, ROT)和交叉验证法(cross validation procedure, CV)。加权可用矩形核密度函数、三角形核密度函数等。其中,使用矩形核密度函数加权的非参数估计,等于是基于子样本的参数估计。使用三角形核密度函数与使用矩形核密度函数的唯一区别是,前者给临近断点的观测值更大的权重。但是,要想给临近断点的观测值更大的权重,更透明的做法是在更小的带宽内使用矩形核密度函数。使用矩形核密度函数得到的结果也更易解释。值得庆幸的是,实践中不同核密度函数得到的结果一般是一致的。具体研究中,参数估计和非参数估计都要做,在两种方法下都稳定的结果更可信。

## (二)运用断点回归设计做研究的新进展

Lee 和 Lemieux (2010)<sup>[2]</sup>之后的7年里(2011-2017年),学者在运用 RDD 时出现了不少 Lee 和 Lemieux (2010)<sup>[2]</sup>未提及的新做法,譬如非参数估计中确定带宽的 IK 法和 CCT 法、甜甜圈 RDD(Donut hole RDD)、参数估计中只应使用配置变量的低次项和利用配置变量做适用性检验的 Frandsen 法<sup>[21]</sup>等。此处归纳比较重要的三种新做法如下:

1. 非参数估计中最优带宽的确定。Lee 和 Lemieux(2010)<sup>[2]</sup>介绍了 ROT 和 CV 两种确定最优带宽的方法。但是,2011-2017年间五大 RDD 论文使用得更多的是 IK 法和 CCT 法。stata 中用非参数法做断点回归估计时的命令 rd,就是用 IK 法确定最优带宽,而命令 rdrobust 提供 CV、IK、CCT 三种不同的最优带宽计算方法选项。Imbens 和 Kalyanaraman(2012)<sup>[18]</sup>指出,ROT 和 CV 两种方法对于在整个支撑集(support set)里估计回归方程而言是最优的,但 RDD 只关心断点处的回归方程估计,故 ROT 和 CV 对于 RDD 而言不是最优的,而 IK 法对于 RDD 而言

<sup>①</sup>作者说这是经验上的做法,没解释这样做的原因。

<sup>②</sup>Lee 和 Lemieux (2010)在脚注中提到了之后应用广泛的 IK 法(Imbens 和 Kalyanaraman, 2009),但是文中举例只使用拇指规则法和交叉验证法,没使用 IK 法。

是渐近最优的;另外,CV法需要研究人员自己设定一个调节参数(tuning parameter),而IK法是完全数据驱动的。Imbens和Kalyanaraman(2012)<sup>[18]</sup>进一步指出,尽管IK法具备这些优点,但研究人员不能只看这一个最优带宽下的估计结果,而应把这一带宽作为基准、检验结果对不同带宽的稳健性。Calonico、Cattaneo和Titiunik(2014)<sup>[19]</sup>认为,ROT、CV和IK这些方法得出的最优带宽过大,导致相应的置信区间有偏,会过度拒绝“没有处理效应”的原假设,结果把“没有”处理效应说成“有”。CCT法校正了过大带宽带来的偏误。

2. 甜甜圈RDD(Donut hole RDD)。甜甜圈RDD用来克服数据堆积(heaping)问题对估计结果的影响<sup>①</sup>。数据堆积指取配置变量某些值的观测值过多的现象。导致这一现象的原因包括受访者自报告某些信息时倾向于向某个数值近似,测量标尺的精读有限等。譬如,调查数据中受访者自报告的身高在160cm、170cm和180cm等取值上堆积,新生儿体重会在3000克、3500克和4000克等取值上堆积。数据操纵现象源自经济个体的逐利动机,只出现在断点处。然而,数据堆积不是源自经济个体的逐利动机,且可能出现在除断点之外的其他地方。如果结果变量受配置变量的堆积现象影响,那么RDD估计量可能是有偏的<sup>[22]</sup>。此时,可以去掉断点附近的某些观测值再做RDD估计。由于去掉断点附近的某些观测值后的数据就像一个“甜甜圈”,故称作“甜甜圈RDD”。至于去掉断点附近多少观测值为适,文献中尚未达成共识。

3. 全样本参数估计的缺点。根据Lee和Lemieux(2010)<sup>[2]</sup>,配置变量多项式要尝试到八九次。但是,Gelman和Imbens(2014)<sup>[23]</sup>认为,不能尝试配置变量多项式的高次项,只能使用局部一次或局部二次多项式,理由有三点:第一,断点回归设计的估计量可以写成处理组结果的加权平均值和控制组结果的加权平均值之差,其中加权权重是配置变量的函数。运用全域高阶多项式方法时,当配置变量取值较大时,使用不同次数多项式得到的权重差别巨大。然而,局部线性回归方法给临近断点处的观测值赋予要大的权重,理论上讲更合理。第二,处理效应的估计值对全域高阶多项式的次数十分敏感。使用不同次数多项式得到的估计值差别很大。第三,在本身不存在断点的情况下,全域多项式方法得到错误结果(认为有断点)的概率高于实际水平,而局部一次(或二次)多项式方法得到错误结果的概率和实际水平差不多。

### 三、断点回归设计的国外应用情况

2011-2017年,五大英文顶级经济学期刊共发表39篇运用RDD做的经验研究论文,涉及公共经济学、新政治经济学、劳动经济学和教育经济学等众多领域。39篇论文中有33篇使用行政管理数据<sup>②</sup>,由此可见开放行政管理数据对于RDD应用的重要性。第二部分提到的三个新进展的应用情况如下:首先,使用非参数估计的26篇论文中,用IK法、CCT法确定最优带宽的分别为13篇、6篇,表明这两种最优带宽确定方法已被学界接受;其次,在甜甜圈RDD被提出的2011年之后,33篇论文中有5篇用了此方法;最后,Dell(2015)<sup>[24]</sup>和Pinotti(2017)<sup>[25]</sup>引用了Gelman和Imbens(2014)<sup>[23]</sup>,但还使用超过二次的配置变量多项式做稳健性检验。

这些文献对RDD的运用与Lee和Lemieux(2010)<sup>[2]</sup>提出的规范大体一致,譬如20篇(占

<sup>①</sup>实际上,甜甜圈RDD也可以克服数据操纵问题对估计结果的影响。

<sup>②</sup>分别来自美国、巴西、挪威、意大利等17个国家。

51.28%) 论文引用了 Lee 和 Lemieux (2010)<sup>[2]</sup>, 33 篇(占 84.62%) 做了适用性检验。与 Lee 和 Lemieux (2010)<sup>[2]</sup> 提出的规范不一致的情况, 包括如下几点:

第一, Lee 和 Lemieux (2010)<sup>[2]</sup> 建议参数估计和非参数估计均要做<sup>①</sup>, 但这样做的只有 16 篇。只使用参数估计、只使用非参数估计的分别为 13 篇和 10 篇。在 16 篇同时使用了两种估计方法的文章中, 将两种方法视为同等重要的有 4 篇, 参数估计仅只是做稳健性检验、非参数仅只是做稳健性检验的分别为 4 篇和 8 篇。因此, 总体来看, 文献中偏重于参数估计。值得一提的是, ECMA 上的 4 篇文章都只用非参数估计, 而 QJE 上的 5 篇文章, 除 Malamud 和 Pop-Eleches (2011)<sup>[26]</sup> 同时使用了两种方法外, 其余 4 篇都只用参数估计。ECMA 由国际计量经济学会主办, QJE 由哈佛大学经济系主办。二者在 RDD 应用上的差异可能体现了两个主办方不同的学术风格。

第二, Lee 和 Lemieux (2010)<sup>[2]</sup> 建议用前定变量做适用性检验时, 应该做似不相关检验, 但没有一篇文献这样做。

第三, Lee 和 Lemieux (2010)<sup>[2]</sup> 建议画描述统计图中的拟合曲线时, 用配置变量的 4 次多项式, 但仅有 3 篇论文用了 4 次多项式。39 篇论文中, 仅有 Chetty et al (2014)<sup>[27]</sup> 这一篇没在描述统计图中画了对散点的拟合曲线。在画了的 38 篇中, 未说明用的什么方法、使用局部线性回归(local linear regression)、使用局部线性平滑(local linear smoother)、移动平均法(running-mean smoothing)和多项式方法的分别为 3 篇、3 篇、3 篇、1 篇和 28 篇, 可见多项式方法是主流方法。在画拟合曲线所用方法上的差异具有明显的期刊特色: 使用局部线性回归法的 3 篇文章有 1 篇来自 QJE, 2 篇来自 JPE, 使用局部线性平滑法的 3 篇文章均来自 ECMA, 使用移动平均法的一篇文章来自 RES, AER 上的文章都使用多项式方法。在使用多项式方法的 28 篇论文中, 使用 1 次、2 次、3 次、4 次、8 次及同时使用 1 次和 2 次多项式的分别为 8 篇、11 篇、4 篇、3 篇、1 篇和 1 篇, 使用 1 次多项式和 2 次多项式的明显居多。值得一提的是, Meng (2017)<sup>[28]</sup> 在同一张图中同时画了 1 次多项式和 2 次多项式的拟合曲线。

第四, Lee 和 Lemieux (2010)<sup>[2]</sup> 建议参数估计时的配置变量多项式要尝试到八九次, 然后挑选赤池信息准则(AIC)取值最小的模型。但是, 使用过参数估计的 29 篇论文中, 仅有 Clark 和 Martorell (2014)<sup>[29]</sup> 和 Deshpande (2016)<sup>[30]</sup> 根据 AIC 来挑选模型。在明确说明尝试过的多项式次数的论文中, 仅有 Auffhammer 和 Kellogg (2011)<sup>[31]</sup>、Pinotti (2017)<sup>[25]</sup> 尝试过超过 4 次的多项式, 绝大多数只尝试到 4 次多项式。

第五, Lee 和 Lemieux (2010)<sup>[2]</sup> 建议非参数估计时用矩形核密度函数。但是, 在使用非参数估计且说明了所用核密度函数的 16 篇文献中, 使用矩形核密度函数的为 7 篇, 而使用三角形核密度函数的有 8 篇, 还有一篇使用帐篷型核密度函数(tent-shaped edge kernel)。

综上所述, Lee 和 Lemieux (2010)<sup>[2]</sup> 之后, 运用 RDD 出现了 IK 和 CCT 两种最优带宽确定方法以及甜甜圈 RDD 两个新操作环节。Gelman 和 Imbens (2014)<sup>[23]</sup> 不该使用超过两次的多项式的建议, 尚未被学界普遍接受。同时, 学界在 Lee 和 Lemieux (2010)<sup>[2]</sup> 涉及的 RDD 运用

<sup>①</sup> 为方便讨论, 使用矩形核密度函数的非参数估计虽然等同于参数估计, 但此处只将其归为非参数估计。

环节上未达成共识。Lee 和 Lemieux (2010)<sup>[2]</sup>发表在权威经济学期刊《Journal of Economic Literature》上,其两位作者均是国际上知名的经济学者。他们提出的 RDD 应用规范未被学界完全接受,并不因为他们的观点不准确、不权威,而是因为 RDD 的理论研究仍很活跃,故应用上达成的共识不多;因为 RDD 应用涉及众多技术环节,故更难达成共识。作为一个对比,IV 和 DID 的理论研究相对成熟、应用涉及的技术环节单一<sup>①</sup>,学者运用这两种方法做研究时的动作比较一致。

#### 四、断点回归设计的国内应用情况

2011-2017 年间,CSSCI 期刊共发表 RDD 经验研究论文 46 篇,涉及公共经济学、环境经济学、财政学等众多领域,其中使用行政管理数据的仅有王骏和孙志军(2015)<sup>[32]</sup>一篇。总体而言,对于两篇权威的 RDD 综述文章——Imbens 和 Lemieux (2008)<sup>[6]</sup>、Lee 和 Lemieux (2010)<sup>[2]</sup>,至少引用过其中一篇的中文论文有 32 篇,占 69.57%,这意味着国内学者总体上了解运用 RDD 涉及的众多技术环节。RDD 应用的新进展也在被国内文献吸收。首先,确定最优带宽的 IK 法已被广泛应用,CCT 法也开始被使用。使用非参数估计且说明了带宽确定方法的 24 篇文章中,使用了 IK 法、CCT 法的分别有 10 篇、3 篇。其次,李宏彬等(2014)<sup>[33]</sup>、张川川等(2015)<sup>[34]</sup>和张明(2017)<sup>[35]</sup>都使用了甜甜圈 RDD。最后,马光荣等(2016)<sup>[36]</sup>、何文剑(2016)<sup>[37]</sup>、李江一和李涵(2017)<sup>[38]</sup>以及王家庭等(2017)<sup>[39]</sup>共 4 篇文章引用了 Gelman 和 Imbens(2014)<sup>[22]</sup>。

与国外文献一样,国内文献运用 RDD 也存在很多与 Lee 和 Lemieux (2010)<sup>[2]</sup>规范不一致的情况,具体如下:首先,同时使用参数估计和非参数估计的只有 18 篇(占 39.13%)。只使用参数估计、只使用非参数估计的分别为 17 篇和 11 篇。在 18 篇同时使用了两种估计方法的文章中,将两种方法视为同等重要的有 3 篇,参数估计仅只是做稳健性检验、非参数仅只是做稳健性检验的分别为 2 篇和 13 篇。因此,总体来看,国内文献与国外文献一样,偏重于参数估计。其次,用前定变量做适用性检验时,仅有曹静等(2014)<sup>[40]</sup>、王骏和孙志军(2015)<sup>[32]</sup>使用了似不相关回归。再次,使用了参数估计的 35 篇文章中,只有杨小聪(2017)<sup>[41]</sup>、张英和陈绍志(2015)<sup>[42]</sup>、曹静等(2014)<sup>[40]</sup>3 篇文章尝试到配置变量八九次的多项式。最后,在使用非参数估计且说明了所用核密度函数的 11 篇文章中,只使用矩形核密度函数、同时使用矩形和三角形核密度函数以及只使用三角形核密度函数的分别为 2 篇、5 篇和 4 篇。

由于运用 RDD 涉及众多技术环节,国内对 RDD 的运用还存在不少问题,主要表现在以下几个方面:首先,5 篇文章未做适用性检验,2 篇文章在适用性检验未通过时依然使用 RDD。张明(2017)<sup>[34]</sup>既没有用配置变量也没有用前定变量做适用性检验。陈强等(2017)<sup>[43]</sup>、黄新飞和杨丹(2017)<sup>[44]</sup>、高彦彦和王逸飞(2017)<sup>[45]</sup>以及王旭光(2017)<sup>[46]</sup>使用的配置变量不可能被操控,无须检验,但他们都没有检验前定变量的连续性。邹红和喻开志(2015)<sup>[47]</sup>、邓婷鹤和何秀荣(2016)<sup>[48]</sup>使用的配置变量“年龄”均在断点(60岁)处不连续,但依然运用 RDD。邹红

<sup>①</sup>运用 IV 必需的技术环节是报告第一阶段回归的 F 统计量取值,运用 DID 必需的技术环节是做平行趋势假设检验。

和喻开志(2015)<sup>[47]</sup>做回归分析时去掉了60岁断点上的样本,但是出于男性实际停止工作的时间普遍是61岁这个考虑,并没有从RDD适用性这个角度深化、细化研究内容,譬如检验样本是否在50岁、60岁、70岁等年龄上存在堆积问题。其次,对描述统计图的运用不规范。如前所述,画图在RDD应用研究中十分必需。然而,有11篇没有画描述统计图。根据Lee和Lemieux(2010)<sup>[2]</sup>和英文顶级期刊的实际,画拟合曲线图以多项式方法居多。然而,在画有描述统计图的35篇文章中,15篇没有说明拟合曲线图所采用的方法,14篇采用局部多项式方法,6篇采用了平滑、多元回归等其他方法。未画描述统计图和未说明画拟合曲线所采用方法的文章合计达26篇(占56.52%),表明国内学界对画图重要性的认识不够。最后,使用非参数估计时的关键技术细节未说明。在29篇使用了非参数估计的文章中,5篇没有报告带宽的确定方法,18篇(占62.07%)未说明使用了哪个核密度函数。

### 五、总结和展望

RDD在断点附近的局部等价于一个随机试验,这一特点使得其备受经验研究人员的青睐。在国外,RDD从20世纪90年代末期开始被大量应用于经济学研究。Lee和Lemieux(2010)<sup>[2]</sup>提出了运用RDD的必备技术环节,相当于给出了RDD应用的规范。但是,由于RDD的理论研究仍在不断深入,至今尚未定型, Lee和Lemieux(2010)<sup>[2]</sup>之后应用RDD的文献出现了新的技术环节,譬如新的最优带宽确定方法、甜甜圈RDD和参数估计中最高只能使用二次多项式等。国内运用RDD做研究始于2010年,最近两年大幅增加。虽然国内熟悉Lee和Lemieux(2010)<sup>[2]</sup>提出的规范,但运用RDD时存在不少问题,譬如不做适用性检验、不重视描述统计图、不交代清楚非参数估计的关键细节等。为了使得研究结论更稳健,基于五大英文顶级期刊发表的论文,我们建议,国内运用RDD时,在完成Lee和Lemieux(2010)<sup>[2]</sup>提出的技术动作后,需要增加如下动作:考虑到学术界在最优带宽确定方法上尚未达成共识<sup>[26][49]</sup>,非参数估计时应该同时尝试CV、IK和CCT三种方法;在配置变量存在堆积现象或被操纵的可能时,使用甜甜圈RDD。尝试到四次多项式。

RDD从理论上讲是更好的因果识别方法,但运用此方法要求在断点附近有较多观测值、对数据要求很高。可喜的是,最近几年,我国经济社会的数字化进程加速,越来越多的经济活动开始数据化,越来越多的中央政府部门和地方政府公开了行政管理数据,为运用RDD做研究提供了契机。特别值得指出的是,运用医疗卫生行业行政管理数据的研究已经成批量涌现。

### 参考文献:

- [1] Angrist J D, Pischke J. The credibility revolution in empirical economics: How better research design is taking the con out of econometrics[J]. Journal of Economic Perspectives, 2010, 24(2): 3-30.
- [2] Lee D S, Lemieux T. Regression discontinuity designs in economics[J]. Journal of Economic Literature, 2010, 48(2): 281-355.
- [3] Thistlethwaite D L, Campbell D T. Regression-discontinuity analysis: An alternative to the ex post facto experiment[J]. Journal of Educational Psychology, 1960, 51(6): 309-317.
- [4] 雷晓燕, 谭力, 赵耀辉. 退休会影响健康吗? [J]. 经济学(季刊), 2010, (04): 1539-1558.
- [5] 余静文, 赵大利. 城市群落的崛起、经济绩效与区域收入差距——基于京津冀、长三角和珠三角城市圈



的分析[J].中南财经政法大学学报,2010,(04):15-20,142.

[6]Imbens G W, Lemieux T. Regression discontinuity designs: A guide to practice[J]. Journal of Econometrics, 2008, 142(2): 615-635.

[7]Der Klaauw W V. Regression-discontinuity analysis: A survey of recent developments in economics[J]. Labour, 2008, 22(2): 219-245.

[8]Nyhan B, Skovron C, Titiunik R. Differential registration bias in voter file data: A sensitivity analysis approach[J]. American Journal of Political Science, 2017, (61): 744-760.

[9]Hausman C, David S. Rapson. Regression discontinuity in time: Considerations for empirical applications [J] Annual Review of Resource Economics, 2018,10(1): 533 - 552.

[10]余静文,王春超.新“拟随机实验”方法的兴起——断点回归及其在经济学中的应用[J].经济学动态,2011,(02):125-131.

[11]Papay J P, Willett J B, Murnane R J. Extending the regression-discontinuity approach to multiple assignment variables[J]. Journal of Econometrics, 2011, 161(2):203-207.

[12]Frandsen B R, Frolich M, Melly B, Quantile treatment effects in the regression discontinuity design[J]. Journal of Econometrics, 2012, 168(2): 382-395.

[13]Card D, Giuliano L. Can tracking raise the test scores of high-ability minority students? [J]. The American Economic Review, 2016, 106(10): 2783-2816.

[14]Cattaneo M D, Keele L, Titiunik R. Interpreting regression discontinuity designs with multiple cutoffs[J]. The Journal of Politics, 2016, 78(4): 1229-1248.

[15]Wing C, Cook T D. Strengthening the regression discontinuity design using additional design elements: A within-study comparison[J]. Journal of Policy Analysis and Management, 2013, 32(4): 853-877.

[16]Angrist J D, Rokkanen M. Wanna get away? Regression discontinuity estimation of exam school effects away from the cutoff[J]. Journal of the American Statistical Association, 2015, 110(512): 1331-1344.

[17]Kolesar M, Rothe C. Inference in regression discontinuity designs with a discrete running variable[J]. The American Economic Review, 2018, 108(8): 2277-2304.

[18]Imbens G W, Kalyanaraman K. Optimal bandwidth choice for the regression discontinuity estimator[J]. The Review of Economic Studies, 2012, 79(3): 933-959.

[19]Calonico S, Cattaneo M D, Titiunik R. Robust nonparametric confidence intervals for regression-discontinuity designs[J]. Econometrica, 2014, 82(6): 2295-2326.

[20]Justin McCrary. Manipulation of the running variable in the regression discontinuity design: A density test [J]. Journal of Econometrics, 2008, (142): 698-714.

[21]Frandsen, Brigham R. Party bias in union representation elections: Testing for manipulation in the regression discontinuity design when the running variable is discrete[Z]. Unpublished Manuscript, Brigham Young University, Provo Utah, USA, 2013.

[22]Barreca A I, Guldi M, Lindo J M. Saving Babies? Revisiting the effect of very low birth weight classification[J]. Quarterly Journal of Economics, 2011, 126(4): 2117-2123.

[23]Gelman A, Imbens G W. Why High-order Polynomials Should not be Used in Regression discontinuity designs[J]. Journal of Business & Economic Statistics, 2017,(05): 1-10.

[24]Dell M. Trafficking networks and the Mexican drug war[J]. The American Economic Review, 2015, 105

(6): 1738-1779.

[25] Pinotti P. Clicking on heaven's door: The effect of immigrant legalization on crime[J]. The American Economic Review, 2017, 107(1): 138-168.

[26] Malamud O, Popeleches C. Home computer use and the development of human capital[J]. Quarterly Journal of Economics, 2011, 126(2): 987-1027.

[27] Chetty R, Friedman J N, Lethpetersen S. Active vs. passive decisions and crowd-out in retirement savings accounts: Evidence from Denmark[J]. Quarterly Journal of Economics, 2014, 129(3): 1141-1219.

[28] Meng K C. Using a free permit rule to forecast the marginal abatement cost of proposed climate policy [J]. The American Economic Review, 2017, 107(3): 748-784.

[29] Clark D, Martorell P. The signaling value of a high school diploma[J]. Journal of Political Economy, 2014, 122(2): 282-318.

[30] Deshpande, Manasi. Does welfare inhibit success? The long-term effects of removing low-income youth from the disability rolls[J]. The American Economic Review, 2016, 106(11): 3300-3330.

[31] Auffhammer M, Kellogg R. Clearing the air? The effects of gasoline content regulation on air quality [J]. The American Economic Review, 2011, 101(6): 2687-2722.

[32] 王骏, 孙志军. 重点高中能否提高学生的学业成绩——基于 F 县普通高中的断点回归设计研究[J]. 北京大学教育评论, 2015, (04): 82-109, 186.

[33] 李宏彬, 施新政, 吴斌珍. 中国居民退休前后的消费行为研究[J]. 经济学(季刊), 2015, (01): 117-134.

[34] 张川川, John Giles, 赵耀辉. 新型农村社会养老保险政策效果评估——收入、贫困、消费、主观福利和劳动供给[J]. 经济学(季刊), 2015, (01): 203-230.

[35] 张明. 税收征管与企业全要素生产率——基于中国非上市公司的实证研究[J]. 中央财经大学学报, 2017, (01): 11-20.

[36] 马光荣, 郭庆旺, 刘畅. 财政转移支付结构与地区经济增长[J]. 中国社会科学, 2016, (09): 105-125.

[37] 何文剑, 徐静文, 张红霄. 森林采伐限额管理制度能否起到保护森林资源的作用[J]. 中国人口·资源与环境, 2016, (07): 128-136.

[38] 李江一, 李涵. 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[J]. 经济学动态, 2017, (03): 62-73.

[39] 王家庭, 毛文峰, 臧家新, 等. 节约集约用地政策对城市蔓延的遏制效应[J]. 城市问题, 2017, (05): 9-16.

[40] 曹静, 王鑫, 钟笑寒. 限行政策是否改善了北京市的空气质量? [J]. 经济学(季刊), 2014, (03): 1091-1126.

[41] 杨小聪, 彭飞, 康丽丽. 绿色地铁: 轨道交通对空气污染的净化效果评估——基于南京地铁 3 号线的实证研究[J]. 甘肃行政学院学报, 2017, (04): 82-94, 129-130.

[42] 张英, 陈绍志. 产权改革与资源管护——基于森林灾害的分析[J]. 中国农村经济, 2015, (10): 15-27.

[43] 陈强, 孙丰凯, 徐艳嫻. 冬季供暖导致雾霾? 来自华北城市面板的证据[J]. 南开经济研究, 2017, (04): 25-40.

[44] 黄新飞, 杨丹. 产业转移促进区域经济协调发展了吗——基于广东省县域数据的断点回归分析[J]. 国际经贸探索, 2017, (02): 101-112.

- [45]高彦彦,王逸飞.熔断制度可以降低中国股市波动吗?——基于断点回归设计的实证分析[J].华东经济管理,2017,(06):104-112.
- [46]王旭光.新型农村养老保险政策提升农民消费水平了吗——来自CFPS数据的实证研究[J].南方经济,2017,(01):1-12.
- [47]邹红,喻开志.退休与城镇家庭消费:基于断点回归设计的经验证据[J].经济研究,2015,(01):124-139.
- [48]邓婷鹤,何秀荣.退休对男性老年人健康的影响——基于断点回归的实证研究[J].人口与经济,2016,(06):82-91.
- [49]Crost B, Felter J H, Johnston P B, et al. Aid under fire: Development projects and civil conflict[J]. The American Economic Review, 2014, 104(6): 1833-1856.

(责任编辑:宋敏)

## A Literature Review of the Application of Regression Discontinuity Design

XIE Qian<sup>1</sup>, XUE Xianling<sup>2</sup>, FU Mingwei<sup>1</sup>

(1. Institute of Economics, Chinese Academy of Social Sciences, Beijing 100836, China;  
2. Graduate School, Chinese Academy of Social Sciences, Beijing 102488, China)

**Abstract:** Recent years has witnessed increasing interest of China's economic scholars on Regression Discontinuity Design (RDD) and fast expansion of empirical papers using RDD. This paper summarizes the roadmap of using RDD advocated by Lee and Lemieux (2010) at first, and then extracts three advancements from the empirical papers using RDD published on the top five economic journals between 2011 and 2017: new methods of selecting optimal bandwidth, donut-hole RDD, and using up to second-order polynomials in parametric estimation. Compared to literature in English, we find that Chinese economic scholars are familiar with the roadmap advocated by Lee and Lemieux (2010) and some of them have used the advancements stated above in their research. However, generally speaking, Chinese papers are characterized by three obvious drawbacks: lack of validity test, neglect of descriptive figure, and ambiguity on the key details when using non-parametric estimation. Our suggestion for using RDD are as follows: try different methods such as CV, IK, and CCT to select optimal bandwidth when employing non-parametric estimation; use donut-hole RDD when the assignment variable is possibly heaped or precisely manipulated; try up to quartic order polynomials when employing parametric estimation.

**Key Words:** Regression discontinuity design; Parametric estimation; Non-parametric estimation; Optimal bandwidth; Donut-hole RDD