

断点回归设计：理论前沿进展与新应用场景¹

刘冲² 诸宇灵³ 李皓宇⁴

摘要 近年来,作为最重要的准实验方法之一,断点回归设计(RDD)在经济学研究中的重要性不仅体现在应用实践中的快速发展,更体现在一系列前沿理论的突破与完善。本文系统梳理了连续性和局部随机性两种断点回归设计的分析框架,总结了两种框架的关键假设、估计方法、统计推断、实现方式及主要操作步骤,在此基础上对比了两者在理论和应用场景方面的差异。本文进一步阐述了多重配置变量、多重断点和拐点回归设计等三种特殊情形下的RDD估计方法,并提供了具体的实现方式,为丰富RDD方法的应用场景、拓展实证研究的可行性提供了有力支撑。

关键词 断点回归设计;连续性框架;局部随机性框架;理论进展;应用场景

0 引言

在经济学研究中,研究者十分关心变量之间的因果关系,以此解释经济现象背后的规律,或者评估特定事件冲击及经济政策带来的影响。类似于自然科学研究中使用的控制变量法,在社会科学领域中可以使用随机控制实验(Randomized Controlled Trial, RCT)进行研究。但是,囿于人力、物力、财力或伦理等方面的限制,并非所有实证研究都能够通过RCT方法进行。鉴于此,基于反事实的准实验(Quasi-experimental)方法在社会科学研究中越来越受到重视(Gangl, 2010; Morgan and Winship, 2015),工具变量法(Instrumental Variable, IV)、双重差分法(Difference-in-Difference, DID)、断点回归设计(Regression Discontinuity Design, RDD)等因果识别方法(Angrist and Pischke, 2010)在实证

¹ 作者感谢国家社会科学基金重大项目(项目批准号:21&ZD097)、国家自然科学基金重大项目(项目批准号:72192844)的资助。文责自负。

² 刘冲,北京大学经济学院长聘副教授,E-mail: pkuliuchong@pku.edu.cn。

³ 诸宇灵(通讯作者),北京大学光华管理学院博士研究生,E-mail: zhuyuling@pku.edu.cn。

⁴ 李皓宇,北京大学经济学院博士研究生,E-mail: Haoyu_li@pku.edu.cn。

研究中的应用越来越广泛。

在众多准实验方法中,断点回归设计(Regression Discontinuity Design, RDD)是一种非常重要的因果识别方法,具有一些独特的优势。一般而言,断点回归设计比其他方法更接近于随机实验,可以得到与RCT类似的估计结果(Shadish et al., 2002; Lee and Lemieux, 2010; 罗胜, 2016; Athey and Imbens, 2017; 刘生龙, 2021),它能够从实验基准中还原因果效应(Recovering Experimental Benchmarks, Green et al., 2009; Hyytinen et al., 2018),具有更强的因果推断力,能够避免因果估计的内生性问题,反映变量之间真实的因果关系(Lee, 2008),因而在进行因果推断和政策评估时,RDD是最可信的准实验方法之一(Cattaneo and Titiunik, 2022)。此外,RDD能够在较弱的假设下识别因果效应,且假设易于被检验(Cattaneo et al., 2020d; Valentim et al., 2021),还可以灵活地通过使用参数和非参数等不同估计方法、调整带宽等方式对局部平均处理效应进行估计、推断和稳健性检验,从而增强RDD估计结果的可信度(Cattaneo et al., 2020a, 2022)。

Hahn et al. (2001)从理论上对断点回归设计的模型识别和模型估计做出了严格证明,并提出了相应的估计方法,进而发展为最常用的基于连续性的RDD框架,并逐渐得到了广泛应用。以Angrist and Pischke (1999)为起点,2000年以来,RDD被越来越多地被应用于经济学的不同领域。2000—2019年间,使用RDD方法进行实证研究的英文文献达200余篇(Villamizar-Villegas et al., 2021),涉及教育(Jacob and Lefgren, 2004; Oreopoulos, 2006; Pop-Eleches and Urquiola, 2013; Cook and Kang, 2016)、劳动(Battistin et al., 2009; Jepsen et al., 2016)、财税(Meng, 2013; Li et al., 2021)、医疗(Almond et al., 2010; Bernal et al., 2017)、环境(Chay and Greenstone, 2005; Greenstone and Gallagher, 2008)、社会治安(Hjalmarsson, 2009; Depew and Eren, 2016)等不同细分领域。

2010年后,随着可供使用的中国微观数据逐渐丰富,国内学界开始较多地使用RDD方法进行实证研究(刘生龙, 2021)。根据在中国知网上的检索,2010—2021年间,至少有300余篇文章在实证研究中使用了RDD方法,图1给出了历年使用RDD方法进行实证研究的文章数量变化趋势,这些研究涵盖了健康、劳动、金融、教育、财税等诸多领域。例如,雷晓燕等(2010)研究了退休制度对老年人健康的影响,刘生龙等(2016)研究了义务教育的回报率,李明等(2018)研究了减税对企业效益的影响,李芳华等(2020)研究了精准扶贫政策对劳动收入和劳动力供给的影响,陆蓉和谢晓飞(2020)研究了指数成分股交换对收益率的影响,杜鹏程等(2021)研究了税收征管改革对企业劳动收入份额的影响。

近年来,作为一种重要的因果识别方法,RDD的快速发展不仅体现在应用

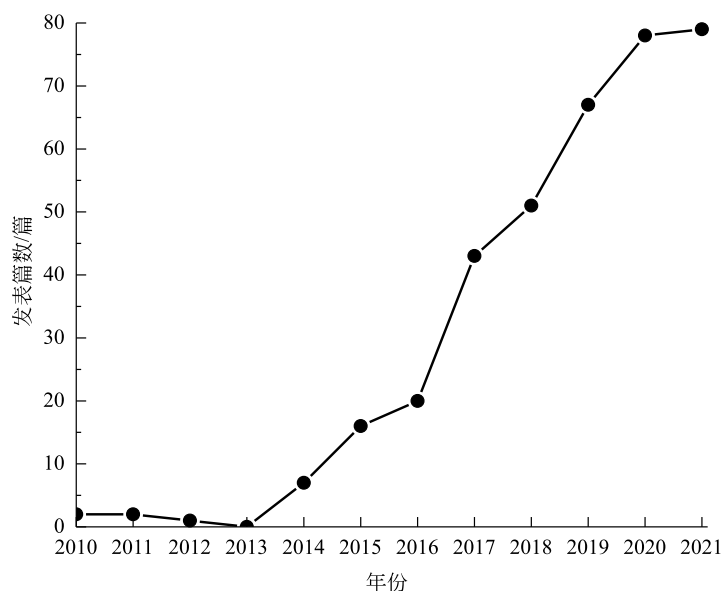


图1 近年来使用 RDD 方法进行实证研究的中文文章数量

数量和范围的拓展(Villamizar-Villegas et al., 2021),更体现在一系列理论前沿进展的出现。基于在断点附近可以近似认为在进行随机实验的思想,Cattaneo et al. (2015, 2017)发展出了另一种基于局部随机性的 RDD 框架,这一框架与基于连续性的 RDD 框架适用于不同的应用场景,连续性的 RDD 框架更适用于大样本,而局部随机性的 RDD 框架则适用于样本量较小和离散配置变量等情形下的研究。在最优带宽的选择(Imbens and Kalyanaraman, 2012; Calonico et al., 2020)、协变量的引入(Calonico et al., 2019)、分位数处理效应的估计(Frandsen et al., 2012)、离散型配置变量的 RDD 推断(Kolesár and Rothe, 2018)等方面,RDD 方法也经历了诸多理论发展。此外,RDD 的前沿进展还体现在应用场景的丰富,例如从单一断点拓展到多重断点、从数值的跳跃拓展到变量斜率的弯折等等,一系列理论突破发展迅速(Choi and Lee, 2021),应用范围也不断拓宽。

本文首先梳理了 RDD 方法的两种分析框架,即基于连续性的断点回归设计和基于局部随机性的断点回归设计,并从关键假设和估计方法等方面比较了两种框架的异同,总结了两种框架的适用场景。在此基础上,本文介绍了 RDD 方法的前沿进展,重点关注多重配置变量、多重断点和拐点回归设计这几种特殊应用场景,并结合 Stata 软件为 RDD 方法的实证应用提供参考。本文的工作总结了近年来 RDD 方法在理论突破、应用范围上的快速发展以及在实证操作层面的具体步骤,对现有中文综述性文章进行了拓宽,同时也为 RDD 实证研究

提供了应用操作层面的参考。

1 RDD 方法的两种分析框架

在精确断点回归设计中(Sharp RDD)^①,我们记配置变量^②(Assignment Variable)为 X ,配置变量的断点(Cutoff 或 Threshold)为 c ,个体 i 的处理状态为

$$D_i = \begin{cases} 0, & \text{if } X_i < c \\ 1, & \text{if } X_i \geq c \end{cases}.$$

RDD 分析框架主要有两种,一是基于连续性的 RDD 框架(Continuity-Based Framework),二是基于局部随机性的 RDD 框架(Local Randomization Framework),两者都要求位于断点 c 两侧且配置变量 X 得分接近的个体之间具备“可比性”。两种分析框架的主要差异在于对“可比性”的设定:在基于连续性的 RDD 框架中,可比性指的是潜在结果变量(Outcome Variable)须满足连续性;而在基于局部随机性的 RDD 框架中,可比性指的则是断点附近需要满足类似于随机实验设计的条件。同时,两种方法对“可比性”也有共同要求,即被观测对象不能通过操纵自己的配置变量来有意识地控制自己的处理状态,这也是使用 RDD 方法的前提。目前针对此要求,可以使用配置变量的概率密度函数进行检验,即如果在断点附近,配置变量的概率密度函数连续,则可以认为配置变量没有被操纵。在实证研究中,通常采用 Stata 软件中的 `rddensity` 命令进行上述检验^③。

由于两种分析框架依赖于不同的识别假设,使用的估计和推断方法也有所不同,因此在特定应用场景中的有效性也存在差异。

1.1 基于连续性的 RDD 框架(The Continuity-Based Framework)

1.1.1 基本假设

连续性的 RDD 框架需要满足以下两个关键假设:

假设 1: 用以分析的观测样本是从无穷大的整体中随机抽取的。

具体而言,在这一框架下,观察到的数据 $\{Y_i(1), Y_i(0), X_i, D_i\}$ ($i=1, 2, \dots, n$)是整体的一个随机样本,研究者主要关注的是结果变量的条件期望函数

① 受篇幅所限,本文中我们重点讨论精确断点回归设计,模糊断点回归设计较为类似。

② 文献中也会采用驱动变量(Running Variable 或 Forcing Variable)的说法。

③ 根据 Cattaneo et al. (2020b),原有的 `rddensity` 命令在 2020 年进行了更新,目前在使用 `rddensity` 命令前需要先安装 `lpdensity` 命令。`lpdensity` 是基于局部多项式的、数据驱动的非参数估计,可以得到断点附近连续的概率密度函数;而原有的 `rddensity` 命令使用的计算方法是将断点附近的配置变量划分为宽度预先给定的若干区间,用每个区间内的样本数量除以区间宽度得到概率密度,实际上得到的是频率分布直方图。

$$\mu(x) = E[Y_i | X_i = x] = \begin{cases} \mu_0(x) = E[Y_i(0) | X_i = x], & \text{if } x < c \\ \mu_1(x) = E[Y_i(1) | X_i = x], & \text{if } x \geq c \end{cases}。记 X_i = x 处的平均处$$

理效应为 $\tau(x)$, 有 $\tau(x) = E[Y_i(1) - Y_i(0) | X_i = x] = \mu_1(x) - \mu_0(x)$ 。

假设 2: 在断点 c 处, $\mu(x)$ 是连续的。

由于可能不存在 $X_i = c$ 的观测样本(简称观测,下同),或者该处的观测量很少,为了得到控制组和处理组在断点处结果变量的均值($E[Y_i(0) | X_i = c]$ 和 $E[Y_i(1) | X_i = c]$), 需要利用断点 c 附近的观测来近似未知的回归函数 $\mu(x)$ 。这一假设也意味着,在断点 c 处, $\mu_1(x)$ 和 $\mu_0(x)$ 是关于配置变量 X 的连续函数,影响结果变量的其他潜在变量都不存在跳跃,断点两侧附近的个体之间唯一的区别就是其处理状态 D_i ,从而确保断点附近的个体具有“可比性”。因此,在精确断点回归设计中,研究者感兴趣的是断点处的平均处理效应 $\tau(c) = E[Y_i(1) - Y_i(0) | X_i = c]$ 。基于 $\mu_1(x)$ 和 $\mu_0(x)$ 的连续性假设,有 $\tau(c) = \lim_{x \downarrow c} E[Y_i | X_i = x] - \lim_{x \uparrow c} E[Y_i | X_i = x]$ (Hahn et al., 2001)。

需要明确的是,断点处的平均处理效应 $\tau(c)$ 不同于平均处理效应 $ATE = E[Y_i(1) - Y_i(0)]$, $\tau(c)$ 只是 $x = c$ 处的个体的平均处理效应,即, RDD 估计的实际上是“局部平均处理效应”(Local Average Treatment Effect, LATE),这在一定程度上限制了断点回归估计的外部有效性。

1.1.2 断点回归处理效应的估计:基于局部多项式的点估计

在 RDD 中,通常使用局部多项式来估计未知的函数 $E[Y_i(t) | X_i = c]$, $t = 0, 1$, 这一方法强调在断点 c 附近对回归函数的拟合,而不考虑与断点距离较远的观测,因此得到的估计结果更加稳健,对离群值等数据问题的敏感度更低。局部多项式法和普通最小二乘法(OLS 法)的主要区别在于,OLS 法假定用以进行估计的多项式是真实的函数形式,而局部多项式法只是将其视为真实函数形式的近似,因此得到的估计中包含近似误差或模型错误设定导致的偏差。

在利用局部多项式进行估计前,首先需要确定多项式的阶数 p 以及核函数 $K(x)$ 。给定带宽 h , 多项式的阶数 p 越高,对回归函数拟合的准确性越高,但同时也会提高处理效应估计的方差。具体而言,局部常数拟合($p = 0$)容易产生欠拟合问题,高阶多项式($p > 2$)则可能导致过拟合问题,在估计边界点时不稳定,出现龙格现象(Runge Phenomenon) (Calonico et al., 2015),同时可能带来估计噪声(Noisy Estimates)、对多项式次数敏感(Sensitivity to the Degree of the Polynomial)、置信区间覆盖率低(Poor Coverage of Confidence Intervals)等问题,因此在实证研究中一般建议采用 $p = 1$ (线性)或 $p = 2$ (二阶)进行估计(Gelman and Imbens, 2019)。

核函数 $K(x)$ 的作用是根据配置变量 X_i 和断点值 c 之间的距离赋予观测非

负权重。具体而言,在利用局部多项式进行点估计时,只使用给定带宽(Bandwidth)范围内的观测,即配置变量 X 的值处于 $[c-h, c+h]$ 范围内的观测,更接近断点 c 的观测在估计中通常会被赋予更高的权重,这一权重的分配方式是由核函数 $K(x)$ 决定的。常用的核函数包括三角核函数(Triangular Kernel)、伊番科尼可夫核函数(Epanechnikov Kernel)、均匀核函数(Uniform Kernel)等,图2直观地刻画了不同核函数在赋权时的差异^①。不过在实证研究中,核函数 $K(x)$ 的选择对于估计和推断结果的影响往往较小。

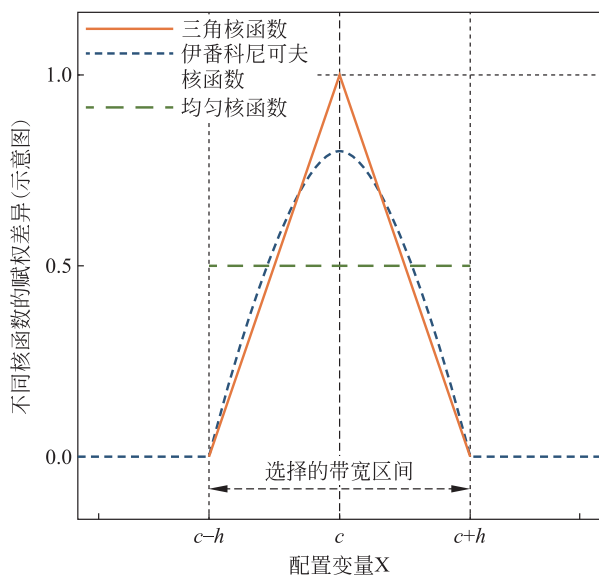


图2 RDD估计时可以采用的不同核函数

给定带宽 h 、多项式阶数 p 以及核函数 $K(x)$,在基于连续性的RDD框架中,局部多项式估计主要包含以下三个步骤:

(1) 对于 $X_i \geq c$ 的处理组样本,利用加权最小二乘法估计 $\mu_1 = E[Y_i(1) | X_i = c]$, $\hat{\mu}_1$ 即下述局部加权最小二乘法估计得到的截距项: $\hat{\beta}_+ = \arg \min_{\beta_{+,0}, \dots, \beta_{+,p}} \sum_{i=1}^n \mathbb{I}(X_i \geq c) (Y_i - \beta_{+,0} - \beta_{+,1}(X_i - c) - \dots - \beta_{+,p}(X_i - c)^p)^2 K\left(\frac{x_i - c}{h}\right)$,

^① 这几种核函数对带宽区间范围以外的观测,赋予的权重都为0。三角核函数的表示形式为 $K(x) = (1-|x|)\mathbb{I}(|x| \leq 1)$,其中 $\mathbb{I}$ 表示若括号内不等式成立则取1,否则取0,即以对称、线性的方式对断点两侧的观测赋权, $X_i = c$ 处权重最大;伊番科尼可夫核函数的表示形式为 $K(x) = \frac{3}{4}(1-x^2)\mathbb{I}(|x| \leq 1)$;均匀核函数的表示形式为 $K(x) = a \cdot \mathbb{I}(|x| \leq 1)$,其中 a 为大于0的常数,即赋予带宽范围内所有观测相同的权重。

其中 p 是选定的多项式阶数, $K\left(\frac{x_i - c}{h}\right)$ 是权重。

(2) 对于 $X_i < c$ 的控制组样本, 利用加权最小二乘法估计 $\mu_0 = E[Y_i(0) | X_i = c]$, $\hat{\mu}_0$ 即下述局部加权最小二乘法估计得到的截距项: $\hat{\beta}_- = \arg \min_{\beta_{-,0}, \dots, \beta_{-,p}} \sum_{i=1}^n \mathbb{I}(X_i < c) (Y_i - \beta_{-,0} - \beta_{-,1}(X_i - c) - \dots - \beta_{-,p}(X_i - c)^p)^2 K\left(\frac{x_i - c}{h}\right)$,

其中, p 是选定的多项式阶数, $K\left(\frac{x_i - c}{h}\right)$ 是权重。

(3) 根据估计的 $\hat{\mu}_1$ 和 $\hat{\mu}_0$, 计算局部平均处理效应: $\hat{\tau}_{\text{SRD}} = \hat{\mu}_1 - \hat{\mu}_0$ 。

1.1.3 带宽 h 的选择方法: 数据驱动的非参数方法

带宽 h 决定了断点 c 两侧能够用以进行断点回归估计和推断的样本量。在早期的实证研究中, 研究者往往根据先验知识事先设定带宽 h , 由于断点回归估计结果对带宽 h 的选择很敏感, 因此这一做法已逐渐被一些数据驱动的非参数方法所取代。

带宽 h 的选择面临“偏差-方差”的权衡问题, 给定多项式的阶数 p , 带宽 h 越小, 由模型错误设定导致的偏差就越小, 断点回归处理效应估计的方差越大, 数据驱动的带宽选择方法就在试图解决这一问题。在基于连续性的 RDD 框架中, 最常用的带宽选择方法是 MSE 标准 (Imbens and Kalyanaraman, 2012; Calonico et al., 2014, 2019; Arai and Ichimura, 2018), 即选择带宽 h_{MSE} , 使得断点回归估计的均方误差 (Mean Square Error, MSE) 最小化, 在应用中, 对于断点两侧的样本, 也可以分别选择不同的带宽。

根据 MSE 标准得到的 h_{MSE} 对断点回归的点估计而言是最优的, 但在构造估计的置信区间时, 更好的带宽选择方法是 CER 标准, 即选择带宽 h_{CER} , 使得覆盖误差率 (Coverage Error Rate, CER) 最小化^① (Calonico et al., 2018, 2021)。需要注意的是, 使用带宽 h_{CER} 进行点估计往往不是 MSE 最优的。所以在实际应用中, 更适合的方式是使用带宽 h_{MSE} 对断点回归进行 MSE 最优点估计, 然后研究者可以选择使用带宽 h_{MSE} 或使用 h_{CER} 来构建置信区间^②。

1.1.4 断点回归处理效应的统计推断

利用 OLS 法进行统计推断依赖于 t 统计量在大样本中近似标准正态分布

① 记 $IC(\hat{\tau})$ 是基于估计值 $\hat{\tau}$ 的断点回归参数 $\tau(c)$ 的 α 水平 (α -level) 的置信区间, CER 最优的带宽 h_{CER} 由下式得到: $h_{\text{CER}} = \arg \min_{h>0} P[\tau(c) \in IC(\hat{\tau})] - (1-\alpha)$ 。

② 因此在本文后续的实证示例中, 我们在非参数方法中没有报告 CER 标准下的点估计结果。

这一性质,但在基于 h_{MSE} 的断点回归估计中并未考虑相关统计量的分布特性,因此直接通过 OLS 法来估计断点回归处理效应的置信区间可能会导致无效的统计推断。

常见的构造置信区间的方法有两种:一种方法是在估计置信区间时选择一个比断点回归点估计使用的带宽 h_{MSE} 更小的带宽 h_{CER} ,但这种方法会使得用以进行断点回归点估计的观测量比进行统计推断时多,从而损失统计效力。另一种更常用的方法是稳健偏差校正法(Robust Bias Correction Method),即在断点回归点估计和推断时使用相同的带宽 h_{MSE} ,对估计的偏差进行校正(这一偏差项形成于 MSE 最优的带宽选择过程),消除由于带宽较大而导致的模型错误设定的影响,同时对标准误进行调整以处理偏差校正导致的其他样本误差,构造以偏差校正的点估计为中心的置信区间,这一方法的优点是在估计和推断中使用相同的观测,从而统计效力更高(Calónico et al., 2014)。

1.1.5 模型拓展:引入协变量

在实证中,研究者可能希望在 RDD 中进一步引入协变量。记个体 i 的协变量向量为 $\mathbf{Z}_i$, $\mathbf{Z}_i(1)$ 和 $\mathbf{Z}_i(0)$ 分别代表处理组和控制组样本的潜在协变量向量,有 $\mathbf{Z}_i = \begin{cases} \mathbf{Z}_i(0), & \text{if } X_i < c \\ \mathbf{Z}_i(1), & \text{if } X_i \geq c \end{cases}$ 。如果协变量的值是在个体受处理前给定的(Pre-determined),则对于所有观测样本 i ,应该满足 $\mathbf{Z}_i(1) = \mathbf{Z}_i(0)$ 。常用的引入协变量的方法有两种:一种方法适用于仅在模型中加入几个离散协变量的情况,只需根据使用的协变量将数据分成不同子样本,然后沿用前述提到的方法进行断点回归估计。另一种方法是加入协变量调整的断点回归估计,在这一方法中,协变量可以是离散的,也可以是连续的,如果协变量 $\mathbf{Z}_i$ 是在个体受处理前给定的,那么经协变量调整的断点回归估计值 $\hat{\tau}_{SRD}$ 就是 τ_{SRD} 的一致估计。

1.1.6 实证应用

综上,连续性的 RDD 框架在应用时主要分为参数和非参数两种方法。在参数方法中,带宽 h 和多项式阶数 p 由研究者基于经验、研究问题特点和数据特性等情况事先给定;而在非参数方法中,给定阶数 p ,带宽 h 是通过 MSE 标准或 CER 标准等最优带宽选择标准得到的。应用基于连续性的 RDD 分析框架,除了报告系数的点估计值及其显著性、观测量 N 等常规指标外,往往还需要报告以下几个参数:带宽 h 、多项式阶数 p 、使用的核函数 $K(x)$ 类型等,使用非参数方法时还需要报告带宽选择标准。

此外,研究者需要进行以下几个步骤的工作:第一,通过绘制散点图和拟合线等方式观察结果变量在断点处是否存在跳跃;第二,使用 `rddensity` 命令检验

配置变量的概率密度函数,判断配置变量是否被操纵;第三,进行局部平滑性检验,确认除结果变量外,其他前定协变量在断点附近不存在跳跃,研究者可以通过绘制散点图加以检验,也可以将其他前定协变量作为结果变量,通过断点回归方法进行检验;第四,选取虚拟断点进行安慰剂检验,如果虚拟断点处检验的结果是连续的,则能够更好地表明原断点的真实性。

1.2 基于局部随机性的 RDD 框架(The Local Randomization Framework)

1.2.1 基本假设

不同于基于连续性的 RDD 框架,基于局部随机性的 RDD 框架不要求观测样本是从无穷大的整体中随机抽取的,因此这一方法特别适用于对小样本的研究。此外,这一方法也不要求回归函数或配置变量在断点处的连续性,因此适用于配置变量为离散值的情况。

局部随机性的 RDD 分析框架的基本思想是,在断点附近的一个小区间内(称为“观测窗口”,Window),一个观测对象是否受到处理是完全随机的,结果变量仅受处理状态的影响,而与配置变量无关。为了保证这一随机性,样本需要满足以下两个假设:

假设 1: 在观测窗口 W_0 内,样本的分布函数 $F_{X_i|X_i \in (W_0)}(X)$ 已知,对所有样本点都相同,与结果变量 Y_i 无关,即分布函数可以被表示为 $F_{X_i|X_i \in (W_0)}(X) = F(X)$ 。

假设 2: 在观测窗口 W_0 内,配置变量 X_i 仅通过处理状态影响结果变量 Y_i ,即 $Y_i(X_i, D_i(X_i)) = Y_i(D_i)$ 。

上述两个假设成立的条件下,在观测窗口 W_0 内,个体是否受到处理与其结果变量的值无关,保证了“受处理”本身的外生性,因此接近于随机实验。在观测窗口内,结果变量的期望可以视为常数,即有 $E(Y_i|X_i \geq c) = Y(1)$, $E(Y_i|X_i < c) = Y(0)$,而在观测窗口外,函数可以是任意形式,不过这一方法无需对观测窗口外的函数做出估计。

如图 3 所示, W_1 是满足两条基本假设的最大观测窗口,即最优观测窗口;当观测窗口为 W_2 时,不满足“结果变量仅由处理状态决定,与配置变量无关”的基本假设;当观测窗口为 W_3 时,虽然满足两条基本假设,但样本量相比于 W_1 更少,会降低估计效力^①。

局部随机性的 RDD 分析框架无须估计结果变量与配置变量之间的函数关

^① 由于估计方法使用的是有限样本的费雪推断法,样本量减少后,计算原假设成立概率的公式中的分母会变小,使得原假设成立概率变大,详见 1.2.2 部分。

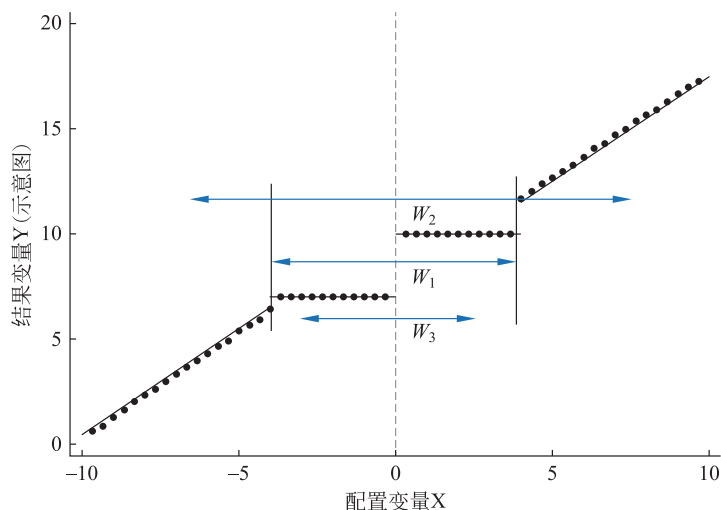


图3 基于局部随机性的 RDD 框架的观测窗口示意图

系,可以直接得出处理效应 τ 的估计结果。在实证应用中,直接将处理组和控制组的结果变量平均值相减,就可以得到使用这一方法估计出的处理效应。

基于局部随机性的 RDD 框架的第二条假设可以进一步放松,允许 Y_i 与 X_i 有关,但要求在观测窗口内存在一个变换 ϕ ,使得 $\phi(Y_i(X_i, D_i), D_i, X_i) = \tilde{Y}_i(D_i)$,而处理效应 $\tilde{\tau} = \tilde{Y}(1) - \tilde{Y}(0)$ 。在实际应用中,通常使用关于 X_i 的 p 阶多项式 $\varphi(X_i)$ 来构造变换 ϕ ,即 $\tilde{Y}_i(D_i) = Y_i + \varphi(X_i)$,而且断点两侧的函数可以不同,这使得本方法在形式上接近于基于连续性的 RDD 框架。尽管形式上类似,但在估计方法上,两种框架是截然不同的。

1.2.2 估计方法

基于局部随机性的 RDD 框架要求在观测窗口内观测样本的结果变量与配置变量无关,只与处理状态 D_i 有关,往往满足这一假设的观测窗口很小,窗口内样本量较少,因此,在估计时通常使用适用于有限样本的费雪推断法 (Fisherian Inference Approach)。

费雪推断法认为,结果变量的期望不是随机变量,而是只与处理状态有关的固定值 $Y_i(0)$ 和 $Y_i(1)$,由此提出不存在处理效应的原假设:

$$H_0^F: Y_i(0) = Y_i(1)$$

如果原假设成立,则观测窗口内的样本是否受处理都不会影响观测值,因此,将样本以任何方式分为两组,处理组与控制组结果变量的平均值都应当相等。如果在观测窗口 W_0 内,样本数为 n_{W_0} ,其中处理组样本数为 $n_{W_0,+}$,控制组

样本数为 $n_{w_0,-} = n_{w_0} - n_{w_0,+}$, 则一共有 $B = \binom{n_{w_0}}{n_{w_0,+}} = \frac{n_{w_0}!}{n_{w_0,+}! \times n_{w_0,-}!}$ 种分组方法, 每

种分组方法被取到的概率为 $\frac{1}{B}$ 。如果实际分组得到的处理组与控制组结果变量的平均值之差在所有可能的分组方法中排序第 m 大, 则原假设成立的概率为 $p^F = \frac{m}{B}$, 那么如果单边拒绝原假设, 则有 $Y_i(1) > Y_i(0)$ 的置信度为 $\left(1 - \frac{m}{B}\right)$ ①。

例如, 如果观测窗口内共有 5 个观测值, 分别记为 $a_1, a_2, \dots, a_5$, 其中第 1、2、3 个为控制组, 第 4、5 个为处理组, 结果变量分别为 1, 1, 1, 2, 2。则如果随机分组, 分组方法一共有 $\binom{5}{2} = 10$ 种, 分别记为 $t_1, t_2, \dots, t_{10}$, 其中实际分组方法为 t_{10} , 每种分组方法下处理组与控制组结果变量的平均值之差如表 1 所示。

表 1 10 种分组方法及结果变量平均值之差

观测值	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
$a_1 = 1$	1	1	1	1	0	0	0	0	0	0
$a_2 = 1$	1	0	0	0	1	1	1	0	0	0
$a_3 = 1$	0	1	0	0	1	0	0	1	1	0
$a_4 = 2$	0	0	1	0	0	1	0	1	0	1
$a_5 = 2$	0	0	0	1	0	0	1	0	1	1
$Y_i(1) - Y_i(0)$	-2/3	-2/3	1/6	1/6	-2/3	1/6	1/6	1/6	1/6	1

从表 1 中可以看出, 实际分组方法 t_{10} 得到的差值在 10 种分组方法中最大, 从而原假设成立的概率 $p^F = \frac{1}{10}$, 因此以 $1 - \frac{1}{10} = 90\%$ 的置信度认为 $Y_i(1) > Y_i(0)$ 。

在实际应用中, 估计方法又可分为小样本、较大样本和大样本三种。小样本估计方法与上述例子相同, 分为以下两步:

(1) 按照观测窗口 W_0 内处理组和控制组的数量, 将样本按照全部可能的组合分组, 所有分组方法 t_{w_0} 的集合记为 T_{w_0} , 计算每个分组方法中处理组与控制组结果变量的平均值之差 $S(t_{w_0}, Y_{w_0}) = \bar{Y}_i(1) - \bar{Y}_i(0)$, 实际分组方法的差值记为 S^{obs} 。

① 同理, 若实际分组得到的处理组与控制组结果变量的平均值之差在所有可能的分组方法中排序第 m 小, 则原假设成立的概率为 $p^F = \frac{m}{B}$, 那么如果单边拒绝原假设, 则有 $Y_i(1) < Y_i(0)$ 的置信度为 $\left(1 - \frac{m}{B}\right)$ 。

(2) 若预期处理效应为正, 计算原假设成立的概率 $p^F = P(S(T_{w_0}, Y_{w_0}) \geq S^{obs}) = \frac{N(S(t_{w_0}, Y_{w_0}) \geq S^{obs})}{\binom{n_{w_0}}{n_{w_0,+}}}$, 其中分子 $N(S(t_{w_0}, Y_{w_0}) \geq S^{obs})$ 表示平均值之差大于实际分组方法的个数, 由此可得到处理效应大于 0 的置信度 $1-p^F$; 预期处理效应为负时, 不等号的方向相反。

当观测窗口内样本量较大时, 随着样本量的增加, 计算所有可能的分组方法结果的计算量将以阶乘的速度增加, 因此常常使用随机抽样模拟的方法来减小计算量, 具体有以下三步:

(1) 计算实际分组方法中处理组与控制组结果变量的平均值之差 S^{obs} 。

(2) 随机选取 B 种分组方法, 分组方法 j 中处理组与控制组的差值记为 $S(t_{w_0}^j, Y_{w_0})$ 。

(3) 若预期处理效应为正, 原假设成立的概率 $p^F = P(S(T_{w_0}, Y_{w_0}) \geq S^{obs}) = \frac{\sum_{j=1}^B \mathbb{I}(S(t_{w_0}^j, Y_{w_0}) \geq S^{obs})}{B}$, 其中 $\mathbb{I}$ 表示若括号内不等式成立则取 1, 否则取 0。由此可得到处理效应大于 0 的置信度 $1-p^F$; 预期处理效应为负时, 不等号的方向相反。

在大样本的情况下, 设实际处理效应为 τ , 可以认为处理组与控制组结果变量的平均值之差 $\hat{\tau}$ 服从均值为 τ 、标准差为 σ 的正态分布, 标准差 σ 的估计量为 $\hat{\sigma}$, 有 $\hat{\sigma} = \sqrt{\frac{\hat{\sigma}_+^2}{n_{w_0,+}} + \frac{\hat{\sigma}_-^2}{n_{w_0,-}}}$, 其中 $\hat{\sigma}_+$ 为处理组的样本结果变量方差, $\hat{\sigma}_-$ 为控制组的样本结果变量方差, 由此可以计算原假设 $H_0^F: \tau=0$ 成立的概率。

1.2.3 置信区间

在统计推断中, 除原假设是否成立外, 研究者往往还关心处理效应置信度为 $(1-\alpha)$ 的置信区间。原假设为 $H_{\tau_0}^F: \tau=\tau_0$, 设调整后的结果变量 $\tilde{Y}_i = Y_i - \tau_0 D_i$ 。在有限样本的情况下, 分别使用调整后的结果变量按照 1.2.2 中的方法计算每个分组方法中处理组与控制组结果变量的平均值之差 $S(t_{w_0}, \tilde{Y}_{w_0})$ 和实际分组方法得到的差值 $\tilde{S}^{obs}$ 。置信区间的下界 τ_{min} 为满足 $p^F = P(S(T_{w_0}, \tilde{Y}_{w_0}) \geq \tilde{S}^{obs}) \leq \alpha$ 的 τ_0 最大值, 上界 τ_{max} 为满足 $p^F = P(S(T_{w_0}, \tilde{Y}_{w_0}) \leq \tilde{S}^{obs}) \leq \alpha$ 的 τ_0 最小值。

在大样本的情况下, 由于平均值之差 $\hat{\tau}$ 服从正态分布, 可以直接使用正态分布表得到置信区间。例如, 置信度为 95% 的置信区间为 $[\hat{\tau} \pm 1.96\hat{\sigma}]$ 。

1.2.4 选择合适的观测窗口

从1.2.2和1.2.3中可以看出,局部随机性的RDD框架非常依赖于观测窗口的具体范围:窗口太大,很有可能不满足本方法的两个基本假设;窗口太小,则会导致观测值过少。研究者除了根据经验在断点附近取一个合适大小的窗口进行估计外,还可以利用协变量得到合适大小的观测窗口。

为了能够利用协变量得到合适大小的观测窗口,被选择的协变量 Z 需要满足以下两个条件:第一, Z 的取值不受处理状态影响,即 Z 为前定变量;第二,在用以估计的全样本中,至少某些观测的 Z 与 X 需要存在相关性。

只有当协变量满足上述两个条件时,研究者才能利用它们寻找合适大小的观测窗口。由于 Z 是前定变量,因此在断点处的处理状态不会影响 Z 的取值,即有 $Z_i(0) = Z_i(1)$ 。但是,因为 Z 需要满足第二个条件,所以当观测窗口足够大时,一定有某些观测的 Z 与 X 之间存在相关性,而由于处理组与控制组的 X 显著不同,使得两组的 Z 也产生显著差别,从而拒绝原假设 $H_0^F: Z_i(0) = Z_i(1)$ 。因此,只有当选择的观测窗口足够小,使得处理组与控制组的 X 没有显著差别,即使 Z 与 X 存在相关性,也不会影响两组中 Z 的均值,从而保证 $Z_i(0) = Z_i(1)$ 。由此可以进一步推测,在这一足够小的观测窗口中,即使 X 与结果变量 Y 存在相关性,由于处理组与控制组的 X 没有显著差别,这一相关性也不会显著影响两组 Y 的取值,因而可以认为在该观测窗口内, $Y_i(0)$ 与 $Y_i(1)$ 的差异仅由处理状态引起,满足1.2.1中的假设2,即 $Y_i(X_i, D_i(X_i)) = Y_i(D_i)$ 。

在实际应用中,选取协变量后,在观测窗口 W_0 内对原假设 $H_0^F: Z_i(0) = Z_i(1)$ 使用1.2.2中的方法进行检验。如果无法拒绝原假设,表明所选的观测窗口能够使用基于局部随机性的RDD框架进行研究;如果拒绝原假设,则表明在该观测窗口内, Z_i 的取值受到 X_i 的影响,局部随机性的假设不成立,选定的观测窗口可能过大。正如1.2.1中所述,观测窗口过小会降低估计效力,因此最合适的窗口大小是刚好不能拒绝原假设的临界值(Cattaneo et al., 2022)^①。

1.2.5 配置变量为离散变量的情况

当配置变量为离散变量时,由于函数 $\mu(x)$ 的连续性不再成立,更适合使用基于局部随机性的RDD框架。为了保证局部随机性框架的两条假设成立,在

^① 在Stata中可以使用rdlocrand命令包中的rdwinselect命令来寻找最优观测窗口。在该命令中,指定配置变量和协变量后,有两种寻找方法:(1)规定初始观测值数量和每次尝试增加的观测值数量,不断增加观测数量,直至拒绝原假设;(2)规定初始观测窗口大小和每次尝试增加的窗口宽度,不断增加窗口宽度,直至拒绝原假设。

选择观测窗口时,通常选择最小的观测窗口,即只包含控制组配置变量的最大值和处理组配置变量的最小值(即断点两侧,距离断点最近的观测的取值)。通常而言,当配置变量为离散变量时,不同取值的配置变量都能够对应一定数量的观测样本,而且基于局部随机性的 RDD 框架在估计时所需的样本量较少,因此即使观测窗口只包含控制组配置变量的最大值和处理组配置变量的最小值,一般也能够满足估计所需的样本量要求^①。如果存在一个满足假设的观测窗口,那么最小观测窗口(该窗口本身或该窗口的子集)一定满足假设。选择最小的观测窗口能够尽可能地减少 X_i 对 Y_i 的影响,保证在观测窗口内 $Y_i(0)$ 与 $Y_i(1)$ 的差异是由处理状态引起的。

在观测窗口内部,如果局部随机性假设成立,即,一个观测值是否受到处理与其配置变量值无关,使得离散型配置变量的不连续跳跃不会影响估计结果。因此,需要按照 1.2.4 中的方法,使用协变量来检验最小观测窗口是否满足局部随机性。

1.2.6 实证应用

综上,在汇报结果时,建议汇报处理效应 $\hat{\tau}$ 、原假设成立概率 p^F 、95%置信区间及观测窗口 W_0 。如果没有对结果变量进行变换,则最好通过散点图等形式表明,在观测窗口内结果变量与配置变量无关;如果对结果变量进行了变换,则需要汇报变换阶数及核函数形式。如果使用了协变量来确定观测窗口,还需要汇报不同观测窗口下对协变量进行费雪推断时原假设成立的概率 p^F 。

此外,可以进行以下四项证伪(falsification)和稳健性(robustness)检验,来检验数据是否符合本方法的基本假设:第一,选择与处理状态无关的协变量,使用费雪推断法或直接画图来说明,在观测窗口内,处理组与控制组的协变量值没有显著差别;第二,检验观测窗口内,被观测对象能否通过操纵自己的配置变量来改变自己的处理状态,通常检验观测窗口内样本点的处理状态(0或1)是否符合成功概率为 0.5 的伯努利分布,如果不能拒绝符合成功概率为 0.5 的伯努利分布的原假设,则通过该检验;第三,在观测窗口外,选择虚拟断点和相同宽度的虚拟观测窗口,检验是否有处理效应,预期结果为没有处理效应;第四,选择宽度小于最优观测窗口的多个观测窗口,检验估计结果对观测窗口大小的敏感性。

^① 当配置变量为连续变量时,如果仅取断点两侧最近的观测的取值,控制组和处理组可能各自仅有一个或有限的几个样本,很难进行统计推断,因此最小观测窗口不适用于配置变量为连续变量的情况。

1.3 两种 RDD 分析框架的对比

表 2 给出了两种分析框架在理论和应用上的差别。值得注意的是,尽管目前的实证研究较少单独使用基于局部随机性的 RDD 框架,但是已有文献在稳健性检验中使用这一分析框架来增强结论的可信性。例如, Santoleri et al. (2021) 在稳健性检验部分使用了局部随机性的 RDD 框架,发现与使用基于连续性的 RDD 框架相比,大部分处理效应的估计值具有相同的显著性。

表 2 两种分析框架在理论和应用方面的差别

	基于连续性的 RDD	基于局部随机性的 RDD
基本假设	观测样本是从无穷大的整体中随机抽取的; 在断点处,回归函数是连续的; 允许配置变量除影响处理状态外,对结果变量也产生影响。	观测窗口内,样本点的分布是随机的; 结果变量仅与处理状态有关,与配置变量无关; 观测窗口内样本是否受处理是完全随机的,而且结果变量的变化完全由处理状态决定。
估计方法	多项式参数估计; 局部多项式非参数估计 ^①	控制组和处理组均值之差; 使用变换来排除 X_i 对 Y_i 的影响
统计推断	选择更小带宽,如 h_{CER} ; 稳健偏差校正法	费雪推断法(小样本); 依分布收敛于正态分布(大样本)
带宽/观测窗口选择	选择带宽 h_{MSE} ,使得断点回归估计的均方误差最小化; 选择带宽 h_{CER} ,使得覆盖误差率最小化。	选择在观测窗口外与配置变量相关且不受处理状态影响的协变量,寻找令协变量与配置变量无关的最大观测窗口。
适用场景	断点附近的样本量较大且连续。	断点附近的样本量较少; 配置变量为离散变量; 作为基于连续性的 RDD 的稳健性检验。

为了进一步比较两种方法的异同,本文以 Nekoei and Weber(2017)的数据为例,分别应用两种分析框架进行了估计。该研究的背景是奥地利通过了一项关于延长失业补助发放周期的政策,在这一政策下,对于年龄在 40 岁及以上且近 10 年内工作总时长大于等于 6 年的失业者,其可以领取失业补助的时长将从 30 周延长至 39 周。我们以这一政策为例,首先基于连续性的 RDD 分析框

① 1.1 中的估计方法先使用非参方法确定了最优带宽,然后在最优带宽内进行加权局部线性回归,为非参数估计的一般步骤。参数估计法一般直接使用全样本数据或指定带宽,在断点两侧进行多项式估计。

架,复现了 Nekoei and Weber(2017)的研究,表 3 报告了这一结果。其中,表 3 (a)及表 3(b)的第(2)列复现了 Nekoei and Weber(2017)的基准回归结果^①,即延长失业补助时间会延长失业者的失业时长,但也会提高其下一份工作的薪酬。表 3 的其他部分则报告了我们根据 Nekoei and Weber(2017)的方法,进一步在给定不同的阶数 p 和带宽 h 下的估计结果。在此基础上,我们采用了非参数方法,表 4 给出了在 Stata 软件中使用 `rdrobust` 命令,给定不同阶数 p 、使用不同核函数 $K(x)$ 进行非参估计的结果。

表 3 应用基于连续性的 RDD 框架,使用参数方法得到的估计结果

表 3(a) 结果变量:失业时长						
变量	(1)	(2)	(3)	(4)	(5)	(6)
age>40	1.332 *** (0.352)	1.932 *** (0.526)	1.798 ** (0.699)	1.818 *** (0.495)	1.661 ** (0.742)	1.109 (0.989)
阶数 p	1	2	3	1	2	3
带宽 h	10	10	10	5	5	5
观测量 N	1589178	1589178	1589178	828019	828019	828019
R^2	0.002	0.002	0.002	0.001	0.001	0.001
表 3(b) 结果变量:薪酬变化						
变量	(1)	(2)	(3)	(4)	(5)	(6)
age>40	0.00426 *** (0.00114)	0.00449 *** (0.00170)	0.00411 * (0.00226)	0.00397 ** (0.00159)	0.00572 ** (0.00238)	0.00415 (0.00318)
阶数 p	1	2	3	1	2	3
带宽 h	10	10	10	5	5	5
观测量 N	1187476	1187476	1187476	617874	617874	617874
R^2	0.000	0.000	0.000	0.000	0.000	0.000

注:括号内为标准误,***、**、* 分别代表在 1%、5%、10% 的显著性水平。

对比表 3 和表 4 可以看出,非参数方法确定的带宽值远小于 Nekoei and Weber(2017)使用的带宽值。由于更小的带宽使得真正进入回归的样本数变少,所得结果的显著性也有所降低。因此,在实证分析中,往往需要根据数据的具体情况 and 研究背景来选择合适的带宽,对于样本量不够大的数据样本,使用数据驱动的非参数方法不一定能得到更准确的结果。

^① 对应的原始结果详见 Nekoei and Weber(2017)原文中的表 2 第(1)列及第(4)列。

表 4 应用基于连续性的 RDD 框架,使用非参数方法得到的估计结果

(a) 结果变量:失业时长						
变量	(1)	(2)	(3)	(4)	(5)	(6)
age>40	1.475 ** (0.701)	1.609 ** (0.659)	1.118 * (0.644)	1.261 (0.932)	1.321 (0.925)	1.630 ** (0.874)
阶数 p	1	1	1	2	2	2
核函数	三角核函数 (Triangular)	伊番科尼可夫 核函数 (Epanechnikov)	均匀核函数 (Uniform)	三角核函数 (Triangular)	伊番科尼可夫 核函数 (Epanechnikov)	均匀核函数 (Uniform)
带宽选 择标准	MSE	MSE	MSE	MSE	MSE	MSE
带宽 h	2.887	3.054	2.839	3.527	3.405	3.486
观测量 N	482756	510363	474851	588876	568398	581763
(b) 结果变量:薪酬变化						
变量	(1)	(2)	(3)	(4)	(5)	(6)
age>40	0.00469 * (0.00247)	0.00491 ** (0.00242)	0.00398 * (0.00206)	0.00408 (0.00319)	0.00445 (0.00316)	0.00648 ** (0.00309)
阶数 p	1	1	1	2	2	2
核函数	三角核函数 (Triangular)	伊番科尼可夫 核函数 (Epanechnikov)	均匀核函数 (Uniform)	三角核函数 (Triangular)	伊番科尼可夫 核函数 (Epanechnikov)	均匀核函数 (Uniform)
带宽选 择标准	MSE	MSE	MSE	MSE	MSE	MSE
带宽 h	2.461	2.410	2.940	3.177	3.079	2.957
观测量 N	307020	301048	366818	394895	383787	368570

注:括号内为标准误,***、**、* 分别代表在 1%、5%、10% 的显著性水平。

在 Stata 软件中进行基于局部随机性的 RDD 估计,可以使用 rdlocrand 命令包^①中的 rdrandinf 命令(Cattaneo et al., 2016a)。表 5 给出了使用局部随机性的 RDD 分析框架得到的结果。由于局部随机性的分析框架主要适用于小样本,当选择的观测窗口较大、样本数较多时,现有程序将花费较长的时间处理数据。

表 5(a)和(b)的第(1)列使用全样本数据进行估计,从图 4 可以看出,全样本并不满足局部随机性框架下“结果变量的变化完全由处理状态决定”的假设,因此直接使用全样本数据进行估计得到的结果并不可靠,需要对 Y_i 进行变换。表 5(a)和(b)的第(2)及第(3)列对结果变量进行了 $\tilde{Y}_i(D_i) = Y_i + \varphi(X_i)$ 的变

^①除本文提及的 rdrandinf 和 rdwinselect 命令外,该命令包中还有 rdsensitivity 和 rdrbounds 命令,均用于检验估计结果对观测窗口大小的敏感性。

表 5 局部随机性的 RDD 框架的估计结果

(a) 结果变量:失业时长				
变量	(1)	(2)	(3)	(4)
age>40	9.516 *** [0.000]	1.332 *** [0.000]	1.932 *** [0.000]	-5.201 [0.669]
95%置信区间	[9.281,9.847]	[1.117,1.615]	[1.717,2.215]	[-26.83,18.22]
对 y 多项式变换阶数	不作变换	1	2	不作变换
核函数	均匀核函数 (Uniform)	均匀核函数 (Uniform)	均匀核函数 (Uniform)	均匀核函数 (Uniform)
左窗口	30	30	30	39.997
右窗口	50	50	50	40.001
左侧 N	839876	839876	839876	150
右侧 N	749302	749302	749302	139
(b) 结果变量:薪酬变化				
变量	(1)	(2)	(3)	(4)
age>40	-0.006 *** [0.000]	0.004 *** [0.000]	0.004 *** [0.000]	-0.020 [0.662]
95%置信区间	[-0.008,-0.005]	[0.003,0.006]	[0.003,0.006]	[-0.116,0.072]
对 y 多项式变换阶数	不作变换	1	2	不作变换
核函数	均匀核函数 (Uniform)	均匀核函数 (Uniform)	均匀核函数 (Uniform)	均匀核函数 (Uniform)
左窗口	30	30	30	39.997
右窗口	50	50	50	40.001
左侧 N	638623	638623	638623	80
右侧 N	548853	548853	548853	74

注:方括号内为有限样本条件下,费雪推断法下原假设成立的概率,***、**、* 分别代表在 1%、5%、10%的显著性水平。

换,其中第(2)列中 $\varphi(X_i)$ 为1阶多项式,第(3)列中 $\varphi(X_i)$ 为2阶多项式。如前文所述,对 Y_i 进行多项式变换,形式上与基于连续性的断点分析框架较为相似,即类似于表3(a)和(b)中的第(1)和第(2)列。在估计结果上,变换后的结果也与表3(a)和(b)的第(1)和第(2)列较为接近。

在 Nekoei and Weber(2017)的研究中,年龄是一个具体到日的离散变量,因此表5(a)和(b)的第(4)列使用前文所述的对离散变量的处理方法,即选择只包含控制组的最大值和处理组最小值的观测窗口,将两组的均值之差作为估计结果,发现估计结果是负向的,且不显著。图5刻画了观测窗口内的结果变量分布状况,从图5可以看出,在断点两侧,结果变量的取值较为分散,极端值对均值的影响可能会掩盖处理状态对处理组均值的影响。由此可见,基于局部随机性的断点分析框架的两条基本假设对数据本身提出了较高要求,这也制约了这一分析框架的应用范围。

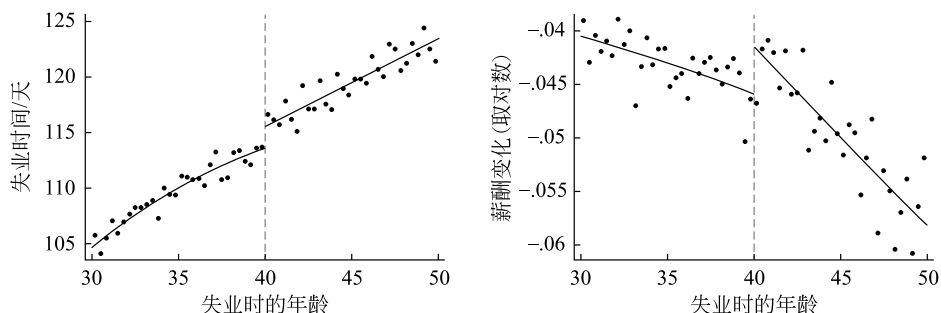


图4 全样本年龄与结果变量的关系图

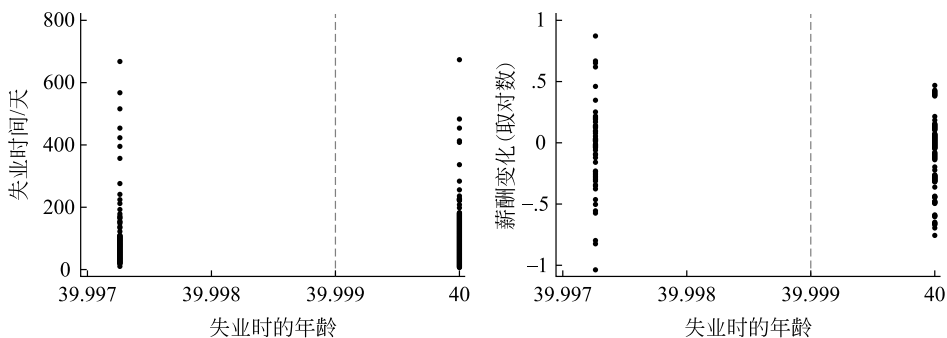


图5 最小观测窗口内结果变量的分布

2 理论拓展与新的应用场景

在使用 RDD 方法时,经常会遇到上述一般分析框架无法解决的问题。例如,配置变量有多个、断点有多个(Cattaneo et al., 2020c);抑或,在断点前后,结果变量没有发生跳跃,但结果变量关于配置变量的变化斜率发生弯折等。本部分将介绍利用 RDD 解决这些场景的前沿理论进展。

2.1 多重配置变量

在实证研究中,处理状态可能由多个配置变量共同决定。例如,在研究奖学金对学生未来表现的影响时,获得奖学金的条件是各科成绩都达到优秀。此时,控制组与处理组的分界是由各科成绩共同决定的,断点由点变为连续的边界。假设语文、数学、英语三科都是成绩达 80 分为优秀(满分为 100 分),将三科成绩分别设为 X_1 、 X_2 、 X_3 ,则在三维坐标系中,边界由三个小平面对组成: $(X_1 = 80, 80 \leq X_2 \leq 100, 80 \leq X_3 \leq 100)$, $(80 \leq X_1 \leq 100, X_2 = 80, 80 \leq X_3 \leq 100)$,

$(80 \leq X_1 \leq 100, 80 \leq X_2 \leq 100, X_3 = 80)$, 即图6中的斜线、浅色和深色部分。

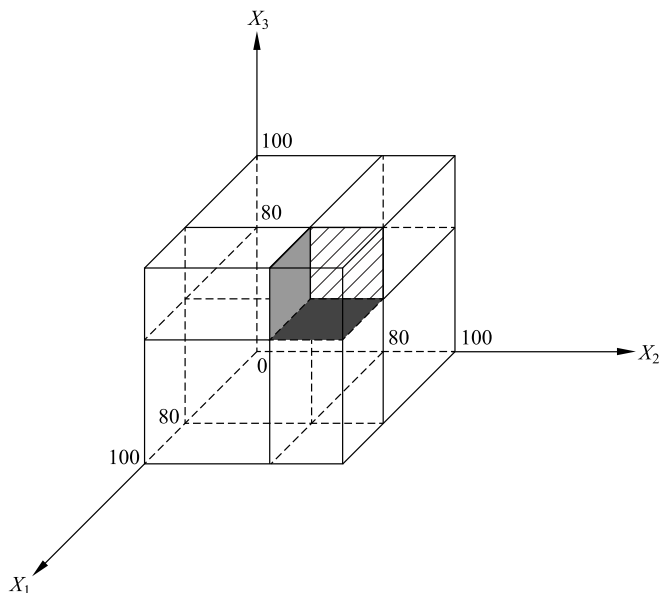


图6 多重配置变量 RDD 的边界示意图

类似于单一配置变量的断点回归设计,在多重配置变量的情形下,研究者仍然关注处理状态对结果变量的影响。设边界 $B \in \mathbf{R}^d$, 其中 $\mathbf{R}^d$ 表示由 d 个配置变量组成的 d 维向量空间,处理组与控制组分别记为 B_i 和 B_c ,则对于边界上的点 b ,处理效应为:

$$E(Y(D_i = 1) - Y(D_i = 0) | X_i = b), \quad b \in B$$

在精确断点回归设计的情形下,处理效应可以表示为:

$$\tau_{\text{SRD}}(b) = \lim_{x \rightarrow b, x \in B_i} E(Y_i | X_i = x) - \lim_{x \rightarrow b, x \in B_c} E(Y_i | X_i = x), \quad b \in B$$

由于对边界上任何一点 b ,都可以计算出对应的 $\tau_{\text{SRD}}(b)$ 值,因此,得到的处理效应 $\tau_{\text{SRD}}(b)$ 是关于 b 的函数。在实证研究中,Cattaneo et al. (2020c)在 Stata 软件中开发了 rdms 命令,可以估计配置变量数量为 2 个时,不同 b 的取值下的处理效应,在使用时需输入两个配置变量名,及需要估计的边界上的点的二维坐标值。不过 rdms 存在一定的局限性:通常情况下,研究者希望得到平均处理效应,但是 rdms 命令无法得到平均处理效应;此外,rdms 命令只能应用于配置变量个数为 2 的情形^①。

^① rdms 命令不适用于配置变量大于 2 个的情况,不过在涉及多维配置变量的实证研究中,以地理边界作为断点的情形比较常见,此时配置变量为 2 维。在其他涉及多维配置变量的研究中,目前研究者更多采用降维的方式进行估计。

在单一配置变量的情形下,可以通过观测值到断点的距离($X_i - c$)来进行局部多项式估计,从而得到局部平均处理效应。在多重配置变量的情形下,可以计算各观测点到边界的距离,以距离值进行局部多项式估计,来得到局部平均处理效应。Feigenbaum et al. (2017)将计算观测点到边界距离的方法归纳为以下两种:

(1) 欧几里得距离,即计算到边界的最短直线距离。设观测值 $X_i \in \mathbf{R}^d$, 则 $x_i^E = \sqrt{\sum_{j=1}^k x_{ij}^2}$, 其中 k 表示使观测值由控制组变为处理组(反之亦然)所需的最少的配置变量个数。假设一名学生的成绩分别为 76、77、90, 则他到边界的欧几里得距离为 $\sqrt{(76-80)^2 + (77-80)^2} = 5$; 假设一名学生的成绩分别为 86、87、90, 则他到边界的欧几里得距离为 $\sqrt{(86-80)^2} = 6$ 。该方法的优点是较为简单, 但缺点是难以解释距离的直观含义, 反事实对照组不明确。

(2) 最小垂直距离, 即将每个配置变量到边界距离的绝对值相加。设观测值 $X_i \in \mathbf{R}^d$, 则 $x_i^R = \sum_{j=1}^k |x_{ij}|$, k 的含义同上。假设一名学生的成绩分别为 76、77、90, 则他到边界的最小垂直距离为 $|76-80| + |77-80| = 7$; 假设一名学生的成绩分别为 86、87、90, 则他到边界的最小垂直距离为 $|86-80| = 6$ 。该方法可以较为直观地说明距离的含义, 如在这一例子中, 到边界的距离即为学生改变分组所需的最小总分的变化。

如果各个配置变量的单位不同, 例如配置变量分别为某家公司的资产负债率、营业收入、净利润和现金短债比, 在使用上述方法计算距离之前, 通常还需要在带宽内对配置变量进行标准化, 将各个配置变量都调整成断点为 0、取值范围为 $[-1, 1]$ 的连续变量。

计算到边界的距离本质上是多重维度的配置变量降低到一维^①。因此, 除计算距离外, 也可以通过降维的方法使多重配置变量变为单一配置变量。Choi and Lee (2017) 归纳了三种常用的降维方法:

(1) 构建新的变量。如果配置变量的断点值之间存在相关关系, 则可以利用相关关系构建新变量。例如, 在一项马萨诸塞州公立大学的奖学金政策中,

^① Keele and Titiunik (2015) 认为, 多维配置变量时, 如果直接使用距离作为一维配置变量, 可能存在以下这种情况: 两个在边界附近的观测值分属处理组与控制组, 两者到边界的距离相同, 但实际位置相距很远, 由于结果变量与实际位置有关, 使得两者受处理前后的结果变量期望 $\mu_0(x)$ 和 $\mu_1(x)$ 不同, 难以满足“在断点 c 处, 影响结果变量的其他潜在变量都不存在跳跃, 断点两侧附近的个体之间唯一的区别就是其处理状态 D_i ”这一假设。因此, Keele and Titiunik (2015) 建议在进行地理断点的研究时, 只对边界上感兴趣的点附近的处理效应进行估计, 而非对整个边界进行估计, 这一思路与前文所述的 *rdms* 命令类似。但是, 由于 *rdms* 命令需要对边界上的点逐个汇报, 且不能得出整个样本的平均处理效应, 因此该方法尚未得到广泛应用, 研究者在实证中更倾向于使用降维的方法。

高中生获得奖学金需满足三个条件:数学和英语成绩至少有一门在 260 分以上,另一门在 250 分以上,且在本学区排名达到前 25%。Hinnerich and Pettersson-Lidbom(2014)在研究奖学金对学生的影响时,构建了一个新的变量:给定学生的学区和英语成绩,他的数学成绩与能拿到奖学金的数学成绩之间的差值。

(2) 当断点值为固定值时,可以使用标准化后的配置变量的最大值(或最小值)作为新变量。例如,英国证监会要求上市公司在进行并购时,被并购公司与原公司的资产比、被并购公司与原公司的利润比、被并购公司的收购价与原公司市值的比值、被并购公司与原公司的资本比,任意一个达到 25%,需提交股东大会表决,Becht et al. (2016)在研究中使用了标准化后的四个配置变量值的最大值 M 作为配置变量,研究召开股东大会对股票收益率的影响。

(3) 对子样本进行研究。当处理状态由两个配置变量 $X_1 \geq b_1$ 、 $X_2 \geq b_2$ 决定时,可以在 $X_2 \geq b_2$ 和 $X_2 < b_2$ 这两个子样本中研究 $X_1 = b_1$ 的断点导致的处理状态变化的影响;同样,也可以在 $X_1 \geq b_1$ 和 $X_1 < b_1$ 这两个子样本中研究 $X_2 = b_2$ 的断点带来的影响(Schmieder et al., 2012; Caliendo et al., 2013)。当配置变量较多时,对子样本逐个进行研究可能较为烦琐,这时可以根据研究需要,选择几个子样本进行研究。例如,德国认定受控外国公司(CFC)的四个条件为外国税率低于 30%、消极投资所得超过 10%、消极投资总额超过 80000 欧元、德国母公司所有海外子公司的消极投资超过 80000 欧元,同时满足税率条件和至少一个投资条件即为 CFC。Egger and Wamser(2015)按照各公司满足的条件,将全部样本分为 $2^4 = 16$ 组,其中 7 组为处理组,然后研究了其中几个处理组被认定为 CFC 后企业固定资产规模的变化。

除了以上方法之外,在特定情境下,研究者也可以使用特殊的降维方法。例如,西方国家选举中,每个选区选出一个议会议席,而控制议会半数以上议席的党派成为多数党,成为多数党这一“处理状态”由 n 个选区的选举结果,即 n 个配置变量共同决定。Fiva et al. (2018)假定同一党派在各个选区选举中的得票率具有高度相关性,若某党派差 s 个议席成为多数党,则直接将选举失败选区中得票率第 s 高的选区结果作为配置变量。

目前,国内利用多重配置变量进行断点回归分析的研究还比较少。鄢伟波等(2019)在研究中没有使用上文所述的降维方法,而是参照 Becker et al. 的研究(2013),直接使用线性空间中的多项式函数进行局部多项式估计,该方法的思路如下:

在精确断点回归设计中,若配置变量为 n 维向量 x_i ,协变量为 m 维向量 z_i ,断点向量为 c_0 ,直接在 n 维和 m 维向量空间中,选取带宽 h 进行线性回归(或使用 1.1 中的非参数估计方法):

$$y_i = \alpha + f_0(x_i - c_0) + g_0(z_i - E(z)) + D_i [\beta + f_1(x_i - c_0) + g_1(z_i - E(z))] + \varepsilon_i$$

其中, f_0 和 f_1 是 n 维向量空间中的 k 阶多项式, g_0 和 g_1 是 m 维向量空间中的 k 阶多项式, $E(z)$ 是协变量在带宽内样本中的均值。

这一方法的优点在于无需对观测值数据进行降维等预处理,但缺点是,该方法实际上估计的是在点 $x_i = c_0$ 周围 h 范围内的处理效应,而非整个边界上的处理效应。因此,在进行多重配置变量的断点回归设计时,本文更推荐采用降维的方法。

多重配置变量 RDD 的一个常见应用情形是使用地理边界作为断点。在利用地理断点的研究中,边界往往是地图上的地理分界线,如行政边界、河流、山脉等,由于地图上的每个点都可以用二维坐标的经纬度表示,因此类似于二维配置变量的情形。例如, Dell (2010) 使用行政边界作为地理断点,以经纬度作为配置变量,研究历史上强迫当地劳动力采矿的制度对经济发展的长期影响;除经纬度外, Dell (2010) 还使用到边界的距离作为配置变量。在计算距离时,通常直接使用最短地理距离,即欧几里得距离,许多学者将欧几里得距离作为配置变量进行了研究。如 Young et al. (2016) 以美国的州边界作为地理断点,以到边界距离作为配置变量,研究不同州税率差距对富豪在各州之间迁居的影响; Rozenas et al. (2017) 以苏联行政边界作为地理断点,研究了历史政策对当今政党支持率的影响; Giua (2017) 利用地区行政边界估计了欧盟区域政策对就业的影响。也有学者基于中国的行政边界,利用地理断点进行了一系列研究(黄新飞等, 2014; Lu et al., 2019; 田文佳等, 2019)。除行政边界外,山脉和河流等自然边界也可以作为地理断点。例如,许多研究使用了我国特有的“秦岭-淮河”南北方分界线作为地理断点(Chen et al., 2013; 晋晶等, 2020; Ito and Zhang, 2020; 杨金玉, 2021), 或者使用澜沧江作为地理断点(李楠和林友宏, 2016)。

2.2 多重断点

在多重断点的情形下,配置变量只有一个,但断点值有多个。例如, Cerqua and Pellegrini (2014) 的研究中,意大利对企业的补贴方式为:政府给每个地区分配补贴配额,并对每个提交申请的企业进行综合评分,按照得分从高到低向企业分配定额补贴,直至所在地区的补贴用尽。因此,在每个地区内,都有受补贴和不受补贴两类企业,但划分这两类企业的得分值因地区而异,这种情形被称为“非累积型多重断点”。再如,席鹏辉和梁若冰(2015)的研究中,我国的空气质量指数被 50、100、150、200、300 等断点分为若干等级,前一个断点的处理效应可能会累积到后一个断点的处理效应上,这种情形被称为“累积型多重断点”。

2.2.1 非累积型多重断点

非累积型多重断点指的是,处理状态 D 只有 0 与 1 这两个取值,但样本被分为 n 组,在不同组中,断点值 C_i 不相同,即 $D_i = \mathbb{I}(X_i \geq C_i)$,其中 $\mathbb{I}$ 表示若括号内不等式成立则取 1,否则取 0; C_i 表示样本所属组的断点值。此时可以计算汇总 (Pooled) 的平均处理效应 (Cattaneo et al., 2016b):

假设各组的断点组成集合 C ,对于任意断点 $c \in C$,首先,在每个组内,计算处理效应:

$$\tau(c) = \lim_{x \downarrow c} E(Y | X_i = x) - \lim_{x \uparrow c} E(Y | X_i = x)$$

然后,根据每个组内带宽中的样本数量,给每个组赋权重:

$$\omega(c) = \frac{N(C_i = c, c - h \leq X_i \leq c + h)}{\sum_{c \in C} N(C_i = c, c - h \leq X_i \leq c + h)}$$

汇总的平均处理效应 τ_p 为:

$$\tau_p = \sum_{c \in C} \tau(c) \omega(c)$$

Cattaneo et al. (2020c) 在 Stata 软件中开发了 *rdmc* 命令,用于非累积型多重断点的估计。非累积型多重断点已经广泛应用于教育 (Francis-Tan and Tannuri-Pianto, 2018; Zimmerman, 2019; Aguirre and Matta, 2021)、商业分析 (Liu et al., 2019)、医疗 (Fort et al., 2020) 等领域的研究中。

在更深入的研究中, Cattaneo et al. (2021) 拓展了非累积型多重断点的应用场景。当不同断点组的样本在处理前或处理后的回归函数满足“平行性”时,可以借鉴双重差分法的思想,估计远离断点样本的平均处理效应。如图 7 所示,若样本存在两组断点 c_1 和 c_2 ,各个样本组的回归函数 $\mu_{d,c}(x)$ 是连续的, $\mu_{d,c}(x) = E[Y_i(d) | X_i = x, C_i = c]$,其中 $d=0,1$ 表示处理状态, $c=c_1, c_2$ 。待估计的 c_1 样本组中配置变量取值为 $\bar{x}$ 处的平均处理效应记为 $\tau_{c_1}(\bar{x})$,即 $E[Y_i(1) - Y_i(0) | X_i = \bar{x}, C_i = c_1]$ 。当 $X_i = \bar{x}$ 时,可以观测到 c_1 样本组的结果变量期望位于 e 点。 $X < c_1$ 时,两组样本都没有受到处理,如果 $\mu_{0,c_1}(x)$ 和 $\mu_{0,c_2}(x)$ 满足平行性,则有 $|ab| = |cd| = |fg|$,从而得到 e 点的反事实 f 点,由于 a, b, e, g 四点都可被观测到,从而 $\tau_{c_1}(\bar{x}) = |ef| = |eg| - |ab|$ 。

同理,如图 8 所示, $X \geq c_2$ 时,如果 $\mu_{1,c_1}(x)$ 和 $\mu_{1,c_2}(x)$ 满足平行性,对于 c_2 样本组中远离断点的控制组样本 $\bar{x}$ (结果变量的期望位于图 8 中 g 点),可得到 g 点的反事实 f 点,进而计算出该点样本的平均处理效应 $\tau_{c_2}(\bar{x}) = |fg| = |eg| - |ab|$ 。

从图 7 和图 8 中可以看出, Cattaneo et al. (2021) 提出的估计方法需要对不

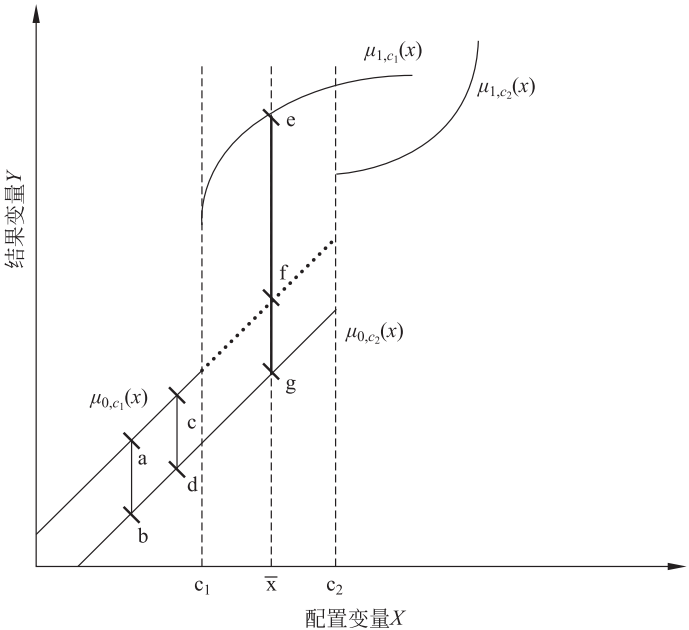


图 7 利用非累积型多重断点的处理前平行性进行推断示意图

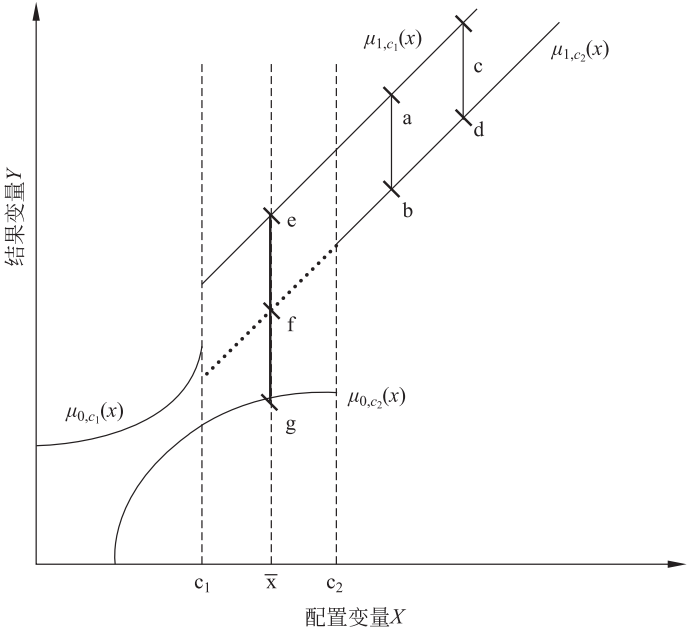


图 8 利用非累积型多重断点的处理后平行性进行推断示意图

同断点组样本处理前或处理后回归函数的平行性进行检验,类似于双重差分法的平行趋势检验。通常采用局部多项式回归进行检验,方法如下:

以处理前平行性为例,选取 $X < c_1$ 的样本,进行局部多项式回归:

$$Y_i = \alpha + \beta \mathbb{I}(C_i = c_1) + r_p(X_i) \gamma + r_p(X_i) \mathbb{I}(C_i = c_1) \delta + \mu_i$$

其中, $r_p(X_i)$ 为 X_i 的 p 阶多项式。如果无法拒绝 $\delta = 0$ 的原假设,则可以认为满足处理前的平行性。

2.2.2 累积型多重断点

累积性多重断点是指,处理状态有 n 个取值,被 $n-1$ 个断点分隔。在该情形下,可以直接对 $n-1$ 个断点进行 $n-1$ 次断点回归设计,得到 $n-1$ 个处理效应值。例如,图9中,共有 $X=4$ 、 $X=8$ 、 $X=11$ 、 $X=15$ 、 $X=19$ 五个断点,将样本分为六组,每个断点处都有对应的处理效应,在研究中往往需要逐个进行分析。

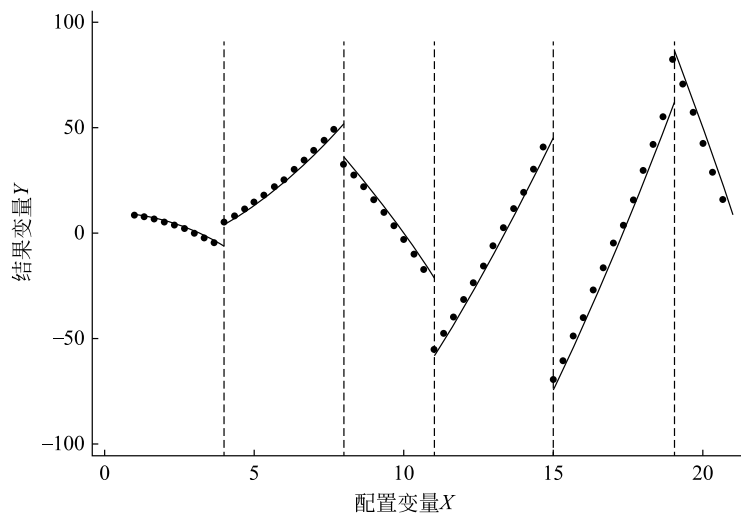


图9 累积型多重断点示意图

值得注意的是,在累积型多重断点的情形下,一个观测值可能既作为前一个断点回归设计的处理组,又作为后一个断点回归设计的控制组,使得相邻两个断点的处理效应具有累积性。在实证研究中,可以通过缩小各断点回归的带宽,使得相邻两个断点回归设计的样本不发生重叠,以消除估计结果的累积性,保证各个断点回归分析估计结果的独立性。

前文所述的 Stata 软件中的 `rdms` 命令 (Cattaneo et al., 2020c) 也可以用于累积型多重断点的估计,在使用时,只输入一个配置变量及断点向量,即可估计各个断点处的处理效应。累积型多重断点在教育 (Smith et al., 2017)、劳动 (Dube et al., 2019)、住房 (Eerola and Lyytikäinen, 2021)、金融 (Chen et al.,

2019)等领域的实证研究中也已经得到了较为广泛的应用。

2.3 拐点回归设计(Regression Kink Design, RKD)

在断点回归设计(RDD)中,我们主要利用结果变量和处理状态 D 在断点 c 处的跳跃进行因果推断,但是当两者在断点 c 处是连续变量时,RDD就难以识别因果效应,这种问题在研究现实问题时经常会出现,导致无法使用RDD来估计因果效应。在这种情况下,如果两者在断点 c 处存在拐点/弯折(Kink),即斜率存在变化,则可以通过与RDD类似的思路,利用拐点来识别因果效应,这一方法就是拐点回归设计(Regression Kink Design, RKD)。例如,在美国的一项政策中,失业救济金水平是关于个体申领失业救济金前一年收入的函数,该政策并不具备使用RDD所需的变量在断点处的跳跃,但是根据失业救济金总额的公式,个体可以领取的失业救济金总额在断点处存在斜率的变化(即存在弯折),因此研究者可以使用RKD来研究失业救济金水平对失业持续时间的影响(Card et al., 2015a)。

在拐点回归设计(RKD)中,结果变量和处理状态 D 在断点 c 处可以存在跳跃,也可以是连续的,但要求它们相对于配置变量 X 的导数在断点 c 处存在跳跃,从而RKD就可以利用断点 c 处的导数变化(即斜率变化)来进行因果识别。图10提供了RKD的基本原理示意图。简言之,RKD就是将结果变量在断点 c 处的弯折归因于处理状态 D 在断点 c 处的弯折,相当于结果变量和处理状态 D 相对于配置变量 X 导数的断点回归设计。

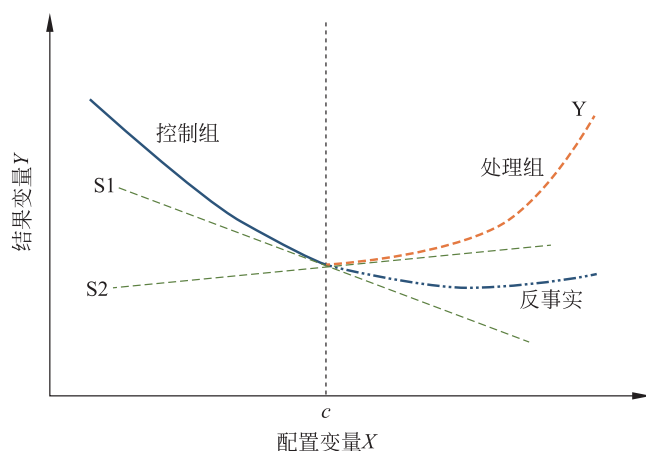


图10 RKD的基本原理示意图

在RKD中,处理状态 D 可以是二元变量(Dong, 2018),也可以是关于配置变量 X 的连续函数(Card et al., 2015a),后者具有更一般化的形式,因此在本

文中,我们主要介绍当处理状态 D 为关于配置变量 X 的连续函数时的 RKD 情况。

我们首先考虑精确拐点回归设计(Sharp RKD):

记 $Y \equiv y(D, X, Z)$, $D \equiv d(X)$, $y_1(d, x, z) \equiv \frac{\partial y(d, x, z)}{\partial d}$, $y_2(d, x, z) \equiv \frac{\partial y(d, x, z)}{\partial x}$, 其中, Y 是结果变量, D 是处理状态, X 是配置变量, Z 是其他影响结果变量的未观测因素, $d(X)$ 是处理状态 D 关于配置变量 X 的已知的连续函数, 函数 $y(D, X, Z)$ 是潜在结果方程。拐点回归设计(RKD)需要满足以下几个基本假设:

假设 1(平滑性假设): $y(d, x, z)$ 是 x 的连续可导函数, 且 $y_2(d, x, z) \equiv \frac{\partial y(d, x, z)}{\partial x}$ 在断点 c 处是连续的, 即有 $\lim_{x \rightarrow c^+} y_2(d, x, z) = \lim_{x \rightarrow c^-} y_2(d, x, z)$ 。

假设 2(密度平滑性假设): 条件密度函数 $f_{X|Z}(x|z)$ 及其偏导数 $\frac{\partial f_{X|Z}(x|z)}{\partial x}$ 在断点 c 处是连续的。

假设 3(弯折性假设): $d(X)$ 是已知的连续函数, 在断点 c 两侧连续可导, 但处理状态 D 在断点 c 处存在拐点/弯折(kink), 即有 $\lim_{x \rightarrow c^+} d'(x) \neq \lim_{x \rightarrow c^-} d'(x)$ 。

在上述假设 1~3 成立的情况下, 其他影响结果变量的未观测因素的分布在断点 c 处连续可导, 即 $Pr(Z \leq z | X=x)$ 在断点 c 处关于 x 连续可导, 从而断点 c

处的因果效应可以表示为: $\tau = \frac{\lim_{x \rightarrow c^+} \mu'(x) - \lim_{x \rightarrow c^-} \mu'(x)}{\lim_{x \rightarrow c^+} d'(x) - \lim_{x \rightarrow c^-} d'(x)} = E[y_1(d(c), c, Z) | X=c]$, 其中 $\mu(x) = E[Y | X=c]$, $\mu'(x) = \frac{\partial \mu(x)}{\partial x}$ 。

与基于连续性的 RDD 分析框架类似, 在精确拐点回归设计中也可以采用局部多项式进行估计。给定带宽 h 、多项式阶数 p 以及核函数 $K(x)$, 精确 RKD 的局部多项式估计主要也包含以下三个步骤:

(1) 对于 $X_i \geq c$ 的处理组样本, 求解下述最优化问题, 得到 $\hat{\beta}_1^+$: $\min_{\beta_{+,0}, \dots, \beta_{+,p}} \sum_{i=1}^n \mathbb{I}(X_i \geq c) (Y_i - \beta_{+,0} - \beta_{+,1}(X_i - c) - \dots - \beta_{+,p}(X_i - c)^p)^2 K\left(\frac{X_i - c}{h}\right)$, 其中 p 是选定的多项式阶数, $K\left(\frac{x_i - c}{h}\right)$ 是权重。

(2) 对于 $X_i < c$ 的控制组样本, 求解下述最优化问题, 得到 $\hat{\beta}_1^-$: $\min_{\beta_{-,0}, \dots, \beta_{-,p}} \sum_{i=1}^n \mathbb{I}(X_i < c)$

$(Y_i - \beta_{-,0} - \beta_{-,1}(X_i - c) - \cdots - \beta_{-,p}(X_i - c)^p)^2 K\left(\frac{X_i - c}{h}\right)$, 其中 p 是选定的多项式阶数, $K\left(\frac{X_i - c}{h}\right)$ 是权重。

(3) 在精确拐点回归设计(Sharp RKD)中, 由于处理状态的分配规则 $d(X)$ 是已知的, 因此其在断点 c 两侧的导数也是已知的, 记 $\kappa_1^+ = \lim_{x \rightarrow c^+} d'(x)$, $\kappa_1^- = \lim_{x \rightarrow c^-} d'(x)$,

从而精确 RKD 的估计量可以表示为: $\hat{\tau}_{SRKD} = \frac{\hat{\beta}_1^+ - \hat{\beta}_1^-}{\kappa_1^+ - \kappa_1^-}$ 。

在实证研究中, 观测到的处理状态分配规则可能是由配置变量和其他未观测因素共同决定的, 因此, 与模糊断点设计(Fuzzy RDD)类似, 也存在模糊拐点设计(Fuzzy RKD)。Card et al. (2015a) 在精确 RKD 假设的基础上, 引入了单调性假设, 证明通过拐点可以识别局部平均处理效应。

由于在模糊 RKD 中, 处理状态的分配规则 $d(\cdot)$ 不是已知确定的, 因此需要估计在断点 c 两侧分配函数斜率的变化, 即需要进一步在断点 c 两侧分别求解以下两个最优化问题:

$$\begin{aligned} \min_{\kappa_{+,0}, \dots, \kappa_{+,p}} \sum_{i=1}^n \mathbb{I}(X_i \geq c) (D_i - \kappa_{+,0} - \kappa_{+,1}(X_i - c) - \cdots - \kappa_{+,p}(X_i - c)^p)^2 K\left(\frac{X_i - c}{h}\right) \\ \min_{\kappa_{-,0}, \dots, \kappa_{-,p}} \sum_{i=1}^n \mathbb{I}(X_i < c) (D_i - \kappa_{-,0} - \kappa_{-,1}(X_i - c) - \cdots - \kappa_{-,p}(X_i - c)^p)^2 K\left(\frac{X_i - c}{h}\right) \end{aligned}$$

由此得到的分配函数斜率变化的估计量为 $\hat{\kappa}_1^+ - \hat{\kappa}_1^-$, 从而模糊 RKD 的估计量可以表示为: $\hat{\tau}_{FRKD} = \frac{\hat{\beta}_1^+ - \hat{\beta}_1^-}{\hat{\kappa}_1^+ - \hat{\kappa}_1^-}$ 。

Dong(2018) 讨论了当处理状态 D 为二元变量时对处理效应的识别, Card et al. (2015a) 则讨论了处理状态 D 是关于配置变量 X 的连续函数时的情况, 在这两篇文章的基础上, Chen et al. (2020) 和 Chiang and Sasaki(2019) 分别探讨了当处理状态 D 为二元变量和连续时对因果的分位数效应(Quantile-wise effects)的识别。RKD 模型的进一步拓展包括当断点未知时的 RKD(Hansen et al., 2017) 以及当断点是关于协变量的函数时的情况(Yang et al., 2021)。

在 RDD 估计中可能存在由于模型错误设定导致的偏差(Calónico et al., 2014), 与之类似, RKD 估计中也可能出现这一问题。在 RKD 的诸多应用场景中, 结果变量和配置变量之间普遍存在非线性关系, 而这一非线性关系可能导致基于常规标准误的 RKD 估计是有偏的(Ando, 2017), 从而估计值在统计意义上的显著性不可信(Ganong and Jäger, 2018)。Ganong and Jäger(2018) 提出

了一种置换检验(Permutation Test)方法来处理这一模型错误设定问题,补充现有的 RKD 和 RDD 推断方法。研究者可以通过分别对存在和不存在政策弯折的地区进行安慰剂检验,构造安慰剂估计值的分布,并利用这一分布来衡量基准模型系数在统计意义上的显著性(Ganong and Jäger, 2018),对 RKD 估计进行稳健性检验。

在利用 Stata 软件进行 RKD 估计时,首先可以通过绘制散点图及拟合线来确认被干预可能性的斜率是否在拐点处发生变化、结果变量关于配置变量的弯折是否在同一点发生。在此基础上,可以使用 `rdrobust` 命令进行 RKD 估计,还可以通过调整 `rdrobust` 可选项中的 `scalear` 来指定感兴趣的参数的调整项。其他关于假设的检验和稳健性检验与 RDD 估计基本类似。

在现实生活中,一些政策设计和数据结构虽然在断点处不存在跳跃,但正如我们在本部分开头提到的失业救济金政策(Card et al., 2015a),这些数据很多都在断点处存在弯折。因此,自被提出以来,拐点回归设计(RKD)已被应用于经济学实证研究的诸多领域,例如公共债务与经济增长(Yang and Su, 2018; Bentour, 2021)、税收(Dobridge, 2015; Engström et al., 2015; Jacob, 2016; Paetzold, 2019)、社会保险(Gelber et al., 2017; Johnston, 2021; Lurie et al., 2021)、失业救济金(Card et al., 2015b; Landais, 2015; Kyrrä and Pesola, 2020)、医疗保险(Menezes-Filho and Politi, 2020)、养老金(Messacar, 2018)、教育(Manoli and Turner, 2018; Marx and Turner, 2018; Sohn and Lee, 2019; Eng and Matsudaira, 2021)、IPO 定价(Busaba and Restrepo, 2022)、处方药需求的价格敏感性(Simonsen et al., 2016)等等。

3 总结与展望

RDD 方法在各种因果推断方法中最接近随机实验,能够得出更可靠的结论(Shadish et al., 2002; Lee and Lemieux, 2010; Athey and Imbens, 2017),因此近年来在经济学实证研究中得到了广泛应用。RDD 方法最常用的估计框架是基于连续性的分析框架,另一个分析框架是基于局部随机性的分析框架。在使用两种框架时,需要特别注意所使用的数据能否支持所用到的 RDD 方法的基本假设,使断点两侧的观测样本满足“可比性”,否则得到的估计结果将缺乏可信性。

在两种估计框架的基础上,一系列前沿理论已经将多重配置变量、多重断点、拐点回归等特殊场景纳入了 RDD 分析框架。除本文给出的这三种特殊场景外,计量经济学界仍在不断探索 RDD 方法的新应用场景,如模糊断点回归下的外部有效性(Bertanha and Imbens, 2020)、离散配置变量(Kolesár and Rothe,

2018)、结果变量取值不同时的异质性处理效应(Shen and Zhang, 2016)等,为拓展实证研究的可行性边界提供理论支撑。

在RDD理论发展的基础上,研究者还在实证分析中探索将断点回归方法与其他因果识别方法结合使用,其中较为常见的是RDD-DID法(Fremigacci, 2010; Bilotkach et al., 2019; Fisher and Zhu, 2019; Barnes et al., 2020; Brodeur et al., 2021)及RKD-DID法(Landais, 2015; Messacar, 2018),将断点回归法与双重差分法结合。例如, Fisher and Zhu(2019)利用RDD-DID法探究了改变财务激励对重新合伙的影响; Messacar(2018)将拐点回归模型扩展为双重差分形式,利用RKD-DID研究了雇主养老金缴费对工人退休储蓄和总财富积累的影响。RDD-DID法也在国内的研究中得到应用,如梁平汉等(2020)研究了无纸化申报改革对企业出口行为的影响。

目前,国内部分实证研究在使用RDD方法时,存在缺乏细致讨论识别假设能否成立,以及带宽选取方式误用等问题,导致由此得出的结论缺乏可靠性。本文的意义在于,借由对不同应用场景下RDD方法的基本假设和原理的介绍,提供规范化使用RDD方法的一般步骤,以期提高研究者利用RDD方法进行因果推断的严谨性,从而进行科学的政策评估,形成高质量的学术论文或研究报告,讲好中国故事、服务科学决策。

参考文献

- 杜鹏程,王姝勋,徐舒. 2021. 税收征管、企业避税与劳动收入份额——来自所得税征管范围改革的证据[J]. 管理世界, 37(7): 105-118.
- Du P C, Wang S X, Xu S. 2021. Tax administration, tax avoidance and labor income share: Evidence from the corporate tax collection reform in China [J]. *Journal of Management World*, 37(7): 105-118. (in Chinese)
- 黄新飞,陈珊珊,李腾. 2014. 价格差异、市场分割与边界效应——基于长三角15个城市的实证研究[J]. 经济研究, 49(12): 18-32.
- Huang X F, Chen S S, Li T. 2014. Price differences, market segmentation and border effect—Based on empirical analysis of 15 cities in Yangtze River Delta Region [J]. *Economic Research Journal*, 49(12): 18-32. (in Chinese)
- 晋晶,王宇澄,郑新业. 2020. 集中供暖要跨过淮河吗?——基于中国家庭能源消费数据的估计[J]. 经济学(季刊), 19(2): 685-708.
- Jing J, Wang Y C, Zheng X Y. 2020. Should district heating cross the Huai River? —Estimation based on Chinese residential energy consumption survey [J]. *China Economic Quarterly*, 19(2): 685-708. (in Chinese)
- 雷晓燕,谭力,赵耀辉. 2010. 退休会影响健康吗? [J]. 经济学(季刊), 9(4):

1539-1558.

Lei X Y, Tan L, Zhao Y H. 2010. Does retirement affect health? [J]. *China Economic Quarterly*, 9(4): 1539-1558. (in Chinese)

李芳华, 张阳阳, 郑新业. 2020. 精准扶贫政策效果评估——基于贫困人口微观追踪数据[J]. *经济研究*, 55(8): 171-187.

Li F H, Zhang Y Y, Zheng X Y. 2020. Impact of targeted poverty alleviation: Micro tracking data of poor populations[J]. *Economic Research Journal*, 55(8): 171-187. (in Chinese)

李明, 李德刚, 冯强. 2018. 中国减税的经济效应评估——基于所得税分享改革“准自然试验”[J]. *经济研究*, 53(7): 121-135.

Li M, Li D G, Feng Q. 2018. The effect of tax changes in China: Evidence from the income tax sharing reform as a quasi-experiment [J]. *Economic Research Journal*, 53(7): 121-135. (in Chinese)

李楠, 林友宏. 2016. 管治方式转变与经济发展——基于清代西南地区“改土归流”历史经验的考察[J]. *经济研究*, 51(7): 173-188.

Li N, Lin Y H. 2016. Governance and economic development: A historical investigation on the southwest China[J]. *Economic Research Journal*, 51(7): 173-188. (in Chinese)

梁平汉, 邹伟, 胡超. 2020. 时间就是金钱: 退税无纸化改革、行政负担与企业出口[J]. *世界经济*, 43(10): 52-73.

Liang P H, Zou W, Hu C. 2020. Paperless reform of tax rebates, administrative burden and firm exports[J]. *The Journal of World Economy*, 43(10): 52-73. (in Chinese)

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[J]. *经济研究*, 51(2): 154-167.

Liu S L, Zhou S J, Hu A G. 2016. Compulsory education law and economic return to education in Urban China: Based on regression discontinuity design [J]. *Economic Research Journal*, 51(2): 154-167. (in Chinese)

刘生龙. 2021. 社会科学研究中的断点回归设计: 最新代表性研究及其展望[J]. *公共管理评论*, 3(2): 140-159.

Liu S L. 2021. Regression discontinuity design in social science research: Analysis of the latest representative research and prospects for development[J]. *China Public Administration Review*, 3(2): 140-159. (in Chinese)

陆蓉, 谢晓飞. 2020. 凤尾变鸡头: 被忽视的指数成分股交换[J]. *金融研究*(6): 171-187.

Lu R, Xie X F. 2020. When the “Tail of the Lion” becomes the “Head of the

- Dog”: The ignored switching of CSI 300/500 member stocks [J]. *Journal of Financial Research*, (6): 171-187. (in Chinese)
- 罗胜. 2016. 断点回归设计: 基本逻辑、方法、应用述评[J]. *统计与决策* (10): 78-80.
- Luo S. 2016. Regression discontinuity design: Basic thoughts, methods and applications[J]. *Statistics & Decision*, (10): 78-80. (in Chinese)
- 田文佳, 余靖雯, 龚六堂. 2019. 晋升激励与工业用地出让价格——基于断点回归方法的研究[J]. *经济研究*, 54(10): 89-105.
- Tian W J, Yu J W, Gong L T. 2019. Promotion incentives and industrial land leasing prices: A regression discontinuity design[J]. *Economic Research Journal*, 54(10): 89-105. (in Chinese)
- 席鹏辉, 梁若冰. 2015. 空气污染对地方环保投入的影响——基于多断点回归设计[J]. *统计研究*, 32(9): 76-83.
- Xi P H, Liang R B. 2015. Can air pollution influence the local environmental protection expenditure—Based on the multiple-cutoffs regression discontinuity design [J]. *Statistical Research*, 32(9): 76-83. (in Chinese)
- 谢谦, 薛仙玲, 付明卫. 2019. 断点回归设计方法应用的研究综述[J]. *经济与管理评论*, 35(2): 69-79.
- Xie Q, Xue X L, Fu M W. 2019. A literature review of the application of regression discontinuity design[J]. *Review of Economy and Management*, 35(2): 69-79. (in Chinese)
- 鄢伟波, 王小华, 温军. 2019. 分层制度提升新三板流动性了吗? ——来自多维断点回归的经验证据[J]. *金融研究* (5): 170-189.
- Yan W B, Wang X H, Wen J. 2019. Has the tiered system enhanced the liquidity of the NEEQ? Empirical evidence from the multivariate regression discontinuity design[J]. *Journal of Financial Research*, (5): 170-189. (in Chinese)
- 杨金玉. 2021. 空气污染与高管流动研究——基于秦岭-淮河线的断点回归设计[J]. *南开经济研究* (2): 143-163.
- Yang J Y. 2021. Air pollution and executive migration: Based on regression discontinuity design of China's Huai River Policy [J]. *Nankai Economic Studies*, (2): 143-163. (in Chinese)
- Aguirre J, Matta J. 2021. Walking in your footsteps: Sibling spillovers in higher education choices[J]. *Economics of Education Review*, 80: 102062.
- Almond D, Doyle J J Jr, Kowalski A E, et al. 2010. Estimating marginal returns to medical care: Evidence from at-risk newborns [J]. *The Quarterly Journal of Economics*, 125(2): 591-634.

- Ando M. 2017. How much should we trust regression-kink-design estimates? [J]. *Empirical Economics*, 53(3): 1287-1322.
- Angrist J D, Lavy V. 1999. Using Maimonides' rule to estimate the effect of class size on scholastic achievement[J]. *The Quarterly Journal of Economics*, 114(2): 533-575.
- Angrist J D, Pischke J S. 2010. The credibility revolution in empirical economics: How better research design is taking the con out of econometrics[J]. *Journal of Economic Perspectives*, 24(2): 3-30.
- Arai Y, Ichimura H. 2018. Simultaneous selection of optimal bandwidths for the sharp regression discontinuity estimator[J]. *Quantitative Economics*, 9(1): 441-482.
- Athey S, Imbens G W. 2017. The state of applied econometrics: Causality and policy evaluation[J]. *Journal of Economic Perspectives*, 31(2): 3-32.
- Barnes S R, Beland L P, Huh J, et al. 2020. The effect of COVID-19 lockdown on mobility and traffic accidents: Evidence from Louisiana [R]. GLO Discussion Paper, No. 616.
- Battistin E, Brugiavini A, Rettore E, et al. 2009. The retirement consumption puzzle: Evidence from a regression discontinuity approach[J]. *American Economic Review*, 99(5): 2209-2226.
- Becht M, Polo A, Rossi S. 2016. Does mandatory shareholder voting prevent bad acquisitions? [J]. *The Review of Financial Studies*, 29(11): 3035-3067.
- Becker S O, Egger P H, von Ehrlich M. 2013. Absorptive capacity and the growth and investment effects of regional transfers: A regression discontinuity design with heterogeneous treatment effects[J]. *American Economic Journal: Economic Policy*, 5(4): 29-77.
- Bentour E M. 2021. On the public debt and growth threshold: One size does not necessarily fit all[J]. *Applied Economics*, 53(11): 1280-1299.
- Bernal N, Carpio M A, Klein T J. 2017. The effects of access to health insurance: Evidence from a regression discontinuity design in Peru [J]. *Journal of Public Economics*, 154: 122-136.
- Bertanha M, Imbens G W. 2020. External validity in fuzzy regression discontinuity designs[J]. *Journal of Business & Economic Statistics*, 38(3): 593-612.
- Bilotkach V, Kawata K, Kim T S, et al. 2019. Quantifying the impact of low-cost carriers on international air passenger movements to and from major airports in Asia [J]. *International Journal of Industrial Organization*, 62: 28-57.
- Brodeur A, Clark A E, Fleche S, et al. 2021. COVID-19, lockdowns and well-being: Evidence from Google Trends[J]. *Journal of Public Economics*, 193: 104346.
- Busaba W Y, Restrepo F. 2022. The "7% solution" and IPO (under) pricing [J].

- Journal of Financial Economics*, 144(3): 953-971.
- Caliendo M, Tatsiramos K, Uhlenborff A. 2013. Benefit duration, unemployment duration and job match quality: A regression-discontinuity approach[J]. *Journal of Applied Econometrics*, 28(4): 604-627.
- Calonico S, Cattaneo M D, Titiunik R. 2014. Robust nonparametric confidence intervals for regression-discontinuity designs[J]. *Econometrica*, 82(6): 2295-2326.
- Calonico S, Cattaneo M D, Titiunik R. 2015. Optimal data-driven regression discontinuity plots[J]. *Journal of the American Statistical Association*, 110(512): 1753-1769.
- Calonico S, Cattaneo M D, Farrell M H. 2018. On the effect of bias estimation on coverage accuracy in nonparametric inference[J]. *Journal of the American Statistical Association*, 113(522): 767-779.
- Calonico S, Cattaneo M D, Farrell M H, et al. 2019. Regression discontinuity designs using covariates[J]. *The Review of Economics and Statistics*, 101(3): 442-451.
- Calonico S, Cattaneo M D, Farrell M H. 2020. Optimal bandwidth choice for robust bias-corrected inference in regression discontinuity designs[J]. *The Econometrics Journal*, 23(2): 192-210.
- Calonico S, Cattaneo M D, Farrell M H. 2021. Coverage error optimal confidence intervals for local polynomial regression[Z]. arXiv: 1808.01398.
- Card D, Johnston A, Leung P, et al. 2015a. The effect of unemployment benefits on the duration of unemployment insurance receipt: New evidence from a regression kink design in Missouri, 2003—2013 [J]. *American Economic Review*, 105(5): 126-130.
- Card D, Lee D S, Pei Z, et al. 2015b. Inference on causal effects in a generalized regression kink design[J]. *Econometrica*, 83(6): 2453-2483.
- Cattaneo M D, Frandsen B R, Titiunik R. 2015. Randomization inference in the regression discontinuity design: An application to party advantages in the U. S. Senate[J]. *Journal of Causal Inference*, 3(1): 1-24.
- Cattaneo M D, Titiunik R, Vazquez-Bare G. 2016a. Inference in regression discontinuity designs under local randomization [J]. *The Stata Journal: Promoting communications on statistics and Stata*, 16(2): 331-367.
- Cattaneo M D, Titiunik R, Vazquez-Bare G, et al. 2016b. Interpreting regression discontinuity designs with multiple cutoffs[J]. *The Journal of Politics*, 78(4): 1229-1248.
- Cattaneo M D, Titiunik R, Vazquez-Bare G. 2017. Comparing inference approaches for RD designs: A reexamination of the effect of head start on child mortality[J].

- Journal of Policy Analysis and Management*, 36(3): 643-681.
- Cattaneo M D, Idrobo N, Titiunik R. 2020a. A practical introduction to regression discontinuity designs: Foundations[M]. Cambridge: Cambridge University Press.
- Cattaneo M D, Jansson M, Ma X W. 2020b. Simple local polynomial density estimators [J]. *Journal of the American Statistical Association*, 115(531): 1449-1455.
- Cattaneo M D, Titiunik R, Vazquez-Bare G. 2020c. Analysis of regression-discontinuity designs with multiple cutoffs or multiple scores[J]. *The Stata Journal: Promoting communications on statistics and Stata*, 20(4): 866-891.
- Cattaneo M D, Titiunik R, Vazquez-Bare G. 2020d. The regression discontinuity design [M]//The SAGE Handbook of Research Methods in Political Science and International Relations. New York: Sage Publications, 835-857.
- Cattaneo M D, Keele L, Titiunik R, et al. 2021. Extrapolating treatment effects in multi-cutoff regression discontinuity designs [J]. *Journal of the American Statistical Association*, 116(536): 1941-1952.
- Cattaneo M D, Titiunik R. 2022. Regression discontinuity designs[Z]. arXiv preprint arXiv: 2108.09400.
- Cattaneo M D, Idrobo N, Titiunik R. 2022. A practical introduction to regression discontinuity designs: Extensions[M]. Cambridge: Cambridge University Press, to appear.
- Cerqua A, Pellegrini G. 2014. Do subsidies to private capital boost firms' growth? A multiple regression discontinuity design approach[J]. *Journal of Public Economics*, 109: 114-126.
- Chay K Y, Greenstone M. 2005. Does air quality matter? Evidence from the housing market[J]. *Journal of political Economy*, 113(2): 376-424.
- Chen H, Huynh K P, Shy O. 2019. Cash versus card: Payment discontinuities and the burden of holding coins[J]. *Journal of Banking & Finance*, 99: 192-201.
- Chen H, Chiang H D, Sasaki Y. 2020. Quantile treatment effects in regression kink designs[J]. *Econometric Theory*, 36(6): 1167-1191.
- Chen Y Y, Ebenstein A, Greenstone M, et al. 2013. Evidence on the impact of sustained exposure to air pollution on life expectancy from China's Huai River policy [J]. *Proceedings of the National Academy of Sciences of the United States of America*, 110(32): 12936-12941.
- Chiang H D, Sasaki Y. 2019. Causal inference by quantile regression kink designs[J]. *Journal of Econometrics*, 210(2): 405-433.
- Choi J Y, Lee M J. 2017. Regression discontinuity: Review with extensions [J]. *Statistical Papers*, 58(4): 1217-1246.

- Choi J Y, Lee M J. 2021. Basics and recent advances in regression discontinuity: Difference versus regression forms[J]. *Journal of Economic Theory and Econometrics*, 32(3): 1-68.
- Cook P J, Kang S M. 2016. Birthdays, schooling, and crime: Regression-discontinuity analysis of school performance, delinquency, dropout, and crime initiation [J]. *American Economic Journal: Applied Economics*, 8(1): 33-57.
- Dell M. 2010. The persistent effects of Peru's mining *Mita* [J]. *Econometrica*, 78(6): 1863-1903.
- Depew B, Eren O. 2016. Born on the wrong day? School entry age and juvenile crime [J]. *Journal of Urban Economics*, 96: 73-90.
- Dobridge C L. 2015. Fiscal stimulus and firms: A tale of two recessions [C]// Proceedings of the 108th Annual Conference on Taxation and Minutes of the Annual Meeting of the National Tax Association. National Tax Association, 108: 1-61.
- Dong Y Y. 2018. Jump or kink? Regression probability jump and kink design for treatment effect evaluation[R]. Working Paper. University of California, Irvine.
- Dube A, Giuliano L, Leonard J. 2019. Fairness and frictions: The impact of unequal raises on quit behavior[J]. *American Economic Review*, 109(2): 620-663.
- Eerola E, Lyytikäinen T. 2021. Housing allowance and rents: Evidence from a stepwise subsidy scheme[J]. *The Scandinavian Journal of Economics*, 123(1): 84-109.
- Egger P H, Wamser G. 2015. The impact of controlled foreign company legislation on real investments abroad. A multi-dimensional regression discontinuity design [J]. *Journal of Public Economics*, 129: 77-91.
- Eng A, Matsudaira J. 2021. Pell grants and student success: Evidence from the universe of federal aid recipients[J]. *Journal of Labor Economics*, 39(S2): S413-S454.
- Engström P, Nordblom K, Ohlsson H, et al. 2015. Tax compliance and loss aversion [J]. *American Economic Journal: Economic Policy*, 7(4): 132-164.
- Feigenbaum J J, Fournaies A, Hall A B. 2017. The majority-party disadvantage: Revising theories of legislative organization [J]. *Quarterly Journal of Political Science*, 12(3): 269-300.
- Fisher H, Zhu A N. 2019. The effect of changing financial incentives on repartnering [J]. *The Economic Journal*, 129(623): 2833-2866.
- Fiva J H, Folke O, Sørensen R J. 2018. The power of parties: Evidence from close municipal elections in Norway[J]. *The Scandinavian Journal of Economics*, 120(1): 3-30.
- Fort M, Ichino A, Zanella G. 2020. Cognitive and noncognitive costs of day care at age 0-2 for children in advantaged families[J]. *Journal of Political Economy*, 128(1):

- 158-205.
- Francis-Tan A, Tannuri-Pianto M. 2018. Black movement: Using discontinuities in admissions to study the effects of college quality and affirmative action[J]. *Journal of Development Economics*, 135: 97-116.
- Frandsen B R, Frölich M, Melly B. 2012. Quantile treatment effects in the regression discontinuity design[J]. *Journal of Econometrics*, 168(2): 382-395.
- Fremigacci F. 2010. Maximum benefits duration and older workers' transitions out of unemployment: A regression discontinuity approach[R]. CEE Working Papers 1012.
- Gangl M. 2010. Causal inference in sociological research [J]. *Annual Review of Sociology*, 36: 21-47.
- Ganong P, Jäger S. 2018. A permutation test for the regression kink design[J]. *Journal of the American Statistical Association*, 113(522): 494-504.
- Gelber A, Moore T J, Strand A. 2017. The effect of disability insurance payments on beneficiaries' earnings[J]. *American Economic Journal: Economic Policy*, 9(3): 229-261.
- Gelman A, Imbens G. 2019. Why high-order polynomials should not be used in regression discontinuity designs [J]. *Journal of Business & Economic Statistics*, 37(3): 447-456.
- Giua M. 2017. Spatial discontinuity for the impact assessment of the EU regional policy: The case of Italian objective 1 regions[J]. *Journal of Regional Science*, 57(1): 109-131.
- Green D P, Leong T Y, Kern H L, et al. 2009. Testing the accuracy of regression discontinuity analysis using experimental benchmarks [J]. *Political Analysis*, 17(4): 400-417.
- Greenstone M, Gallagher J. 2008. Does hazardous waste matter? Evidence from the housing market and the superfund program[J]. *The Quarterly Journal of Economics*, 123(3): 951-1003.
- Hahn J, Todd P, Van der Klaauw W. 2001. Identification and estimation of treatment effects with a regression-discontinuity design[J]. *Econometrica*, 69(1): 201-209.
- Hansen B, Nguyen T, Waddell G R. 2017. Benefit generosity and injury duration: Quasi-experimental evidence from regression kinks [R]. IZA Discussion Paper 10621. Bonn, Germany: Institute of Labor Economics.
- Hinnerich B T, Pettersson-Lidbom P. 2014. Democracy, redistribution, and political participation: Evidence from Sweden 1919—1938 [J]. *Econometrica*, 82(3): 961-993.
- Hjalmarsson R. 2009. Juvenile jails: A path to the straight and narrow or to hardened

- criminality? [J]. *The Journal of Law and Economics*, 52(4): 779-809.
- Hyytinen A, Meriläinen J, Saarimaa T, et al. 2018. When does regression discontinuity design work? Evidence from random election outcomes[J]. *Quantitative Economics*, 9(2): 1019-1051.
- Imbens G, Kalyanaraman K. 2012. Optimal bandwidth choice for the regression discontinuity estimator[J]. *The Review of Economic Studies*, 79(3): 933-959.
- Ito K, Zhang S. 2020. Willingness to pay for clean air: Evidence from air purifier markets in China[J]. *Journal of Political Economy*, 128(5): 1627-1672.
- Jacob B A, Lefgren L. 2004. Remedial education and student achievement: A regression-discontinuity analysis [J]. *The Review of Economics and Statistics*, 86(1): 226-244.
- Jacob M. 2016. Cross-base tax elasticity of capital gains[J]. *Applied Economics*, 48(28): 2611-2624.
- Jepsen C, Mueser P, Troske K. 2016. Labor market returns to the GED using regression discontinuity analysis[J]. *Journal of Political Economy*, 124(3): 621-649.
- Johnston A C. 2021. Unemployment insurance taxes and labor demand: Quasi-experimental evidence from administrative data [J]. *American Economic Journal: Economic Policy*, 13(1): 266-293.
- Keele L J, Titiunik R. 2015. Geographic boundaries as regression discontinuities [J]. *Political Analysis*, 23(1): 127-155.
- Kolesár M, Rothe C. 2018. Inference in regression discontinuity designs with a discrete running variable[J]. *American Economic Review*, 108(8): 2277-2304.
- Kyyrä T, Pesola H. 2020. The effects of UI benefits on unemployment and subsequent outcomes: Evidence from a kinked benefit rule [J]. *Oxford Bulletin of Economics and Statistics*, 82(5): 1135-1160.
- Landais C. 2015. Assessing the welfare effects of unemployment benefits using the regression kink design [J]. *American Economic Journal: Economic Policy*, 7(4): 243-278.
- Lee D S. 2008. Randomized experiments from non-random selection in U. S. House elections[J]. *Journal of Econometrics*, 142(2): 675-697.
- Lee D S, Lemieux T. 2010. Regression discontinuity designs in economics[J]. *Journal of Economic Literature*, 48(2): 281-355.
- Li B, Liu C, Sun S T. 2021. Do corporate income tax cuts decrease labor share? Regression discontinuity evidence from China [J]. *Journal of Development Economics*, 150: 102624.
- Liu X, Lee D, Srinivasan K. 2019. Large-scale cross-category analysis of consumer

- review content on sales conversion leveraging deep learning [J]. *Journal of Marketing Research*, 56(6): 918-943.
- Lu Y, Wang J, Zhu L M. 2019. Place-based policies, creation, and agglomeration economies: Evidence from China's economic zone program[J]. *American Economic Journal: Economic Policy*, 11(3): 325-360.
- Lurie I Z, Sacks D W, Heim B. 2021. Does the individual mandate affect insurance coverage? Evidence from tax returns[J]. *American Economic Journal: Economic Policy*, 13(2): 378-407.
- Manoli D, Turner N. 2018. Cash-on-hand and college enrollment: Evidence from population tax data and the earned income tax credit [J]. *American Economic Journal: Economic Policy*, 10(2): 242-271.
- Marx B M, Turner L J. 2018. Borrowing trouble? Human capital investment with opt-in costs and implications for the effectiveness of grant aid [J]. *American Economic Journal: Applied Economics*, 10(2): 163-201.
- Menezes-Filho N, Politi R. 2020. Estimating the causal effects of private health insurance in Brazil: Evidence from a regression kink design[J]. *Social Science & Medicine*, 264: 113258.
- Meng L S. 2013. Evaluating China's poverty alleviation program: A regression discontinuity approach[J]. *Journal of Public Economics*, 101: 1-11.
- Messacar D. 2018. Crowd-out, education, and employer contributions to workplace pensions: Evidence from Canadian tax records[J]. *The Review of Economics and Statistics*, 100(4): 648-663.
- Morgan S L, Winship C. 2015. Counterfactuals and causal inference[M]. Cambridge: Cambridge University Press.
- Nekoei A, Weber A. 2017. Does extending unemployment benefits improve job quality? [J]. *American Economic Review*, 107(2): 527-561.
- Oreopoulos P. 2006. Estimating average and local average treatment effects of education when compulsory schooling laws really matter[J]. *American Economic Review*, 96(1): 152-175.
- Paetzold J. 2019. Do commuting subsidies increase commuting distances? Evidence from a Regression Kink Design [J]. *Regional Science and Urban Economics*, 75: 136-147.
- Pop-Eleches C, Urquiola M. 2013. Going to a better school: Effects and behavioral responses[J]. *American Economic Review*, 103(4): 1289-1324.
- Rozenas A, Schutte S, Zhukov Y. 2017. The political legacy of violence: The long-term impact of Stalin's repression in Ukraine [J]. *The Journal of Politics*, 79(4):

- 1147-1161.
- Santoleri P, Mina A, Di Minin A, et al. 2021. The causal effects of R&D grants: Evidence from a regression discontinuity[R]. LEM Working Paper.
- Schmieder J F, von Wachter T, Bender S. 2012. The effects of extended unemployment insurance over the business cycle: Evidence from regression discontinuity estimates over 20 years[J]. *The Quarterly Journal of Economics*, 127(2): 701-752.
- Shadish W R, Cook T D, Campbell D T. 2002. Experimental and quasi-experimental designs for generalized causal inference [M]. Boston, MA: Houghton Mifflin Company.
- Shen S, Zhang X H. 2016. Distributional tests for regression discontinuity: Theory and empirical examples[J]. *Review of Economics and Statistics*, 98(4): 685-700.
- Simonsen M, Skipper L, Skipper N. 2016. Price sensitivity of demand for prescription drugs: Exploiting a regression kink design[J]. *Journal of Applied Econometrics*, 31(2): 320-337.
- Smith J, Hurwitz M, Avery C. 2017. Giving college credit where it is due: Advanced placement exam scores and college outcomes [J]. *Journal of Labor Economics*, 35(1): 67-147.
- Sohn H, Lee S W. 2019. Causal impact of having a college degree on women's fertility: Evidence from regression Kink designs[J]. *Demography*, 56(3): 969-990.
- Valentin V, Núñez A R, Dinas E. 2021. Regression discontinuity designs: A hands-on guide for practice[J]. *Italian Political Science Review*, 51(2): 250-268.
- Villamizar-Villegas M, Pinzon-Puerto A, Ruiz-Sanchez M A. 2021. A comprehensive history of regression discontinuity designs: An empirical survey of the last 60 years [J]. *Journal of Economic Surveys*, Accepted.
- Yang L X, Su J J. 2018. Debt and growth: Is there a constant tipping point? [J]. *Journal of International Money and Finance*, 87: 133-143.
- Yang L X, Zhang C L, Lee C, et al. 2021. Panel kink threshold regression model with a covariate-dependent threshold[J]. *The Econometrics Journal*, 24(3): 462-481.
- Young C, Varner C, Lurie I Z, et al. 2016. Millionaire migration and taxation of the elite: Evidence from administrative data[J]. *American Sociological Review*, 81(3): 421-446.
- Zimmerman S D. 2019. Elite colleges and upward mobility to top jobs and top incomes [J]. *American Economic Review*, 109(1): 1-47.

Regression Discontinuity Design: Theoretical Developments and Extended Applications

Chong Liu¹ Yuling Zhu² Haoyu Li¹

(1. School of Economics, Peking University; 2. Guanghua School of Management, Peking University)

Abstract As one of the most important quasi-experimental methods, the Regression Discontinuity Design (RDD) has been widely used in empirical researches in recent years, and it has been undergoing a series of rapid theoretical developments and breakthroughs. This paper provides systematic introduction and guidance of two analytical frameworks of RDD, the Continuity-Based Framework and the Local Randomization Framework, including their basic assumptions, estimation methods, statistical inference, implementations and main steps of empirical analyses. This paper then compares the differences of these two analytical frameworks. Furthermore, this paper expounds RDD methods in three special circumstances, including RDD with multiple running variables, RDD with multiple cutoffs, and Regression Kink Design, and offers detailed approaches to their code implementations. These theoretical developments and their implementations can extend the application of RDD to a wider range of fields and increase the feasibility of empirical researches using RDD.

JEL Classification B23, C10, C50