

# 国内双重差分法的研究现状与潜在问题<sup>①</sup>

陈 林<sup>1,2</sup> 伍海军<sup>3</sup>

(1. 暨南大学产业经济研究院; 2. 中国社会科学院经济研究所;  
3. 暨南大学统计学系)

**【摘要】**在国内政策效果评估领域,基于自然实验的双重差分法是应用最广的研究工具,并在近年呈现加速发展的趋势。但通过归纳双重差分法的基本应用条件与国内研究现状,本文发现部分研究出现了内生性、控制组受影响、样本异质性问题。其中尤以内生性问题较为严重,甚至使部分研究产生了一定的缺陷,所谓的自然实验往往难以成立。本文根据实验理论与相关理论基础,对国内主要研究成果进行分类与评述,并针对所发现的问题提出改良办法。

**关键词** 双重差分 自然实验 准实验 内生性问题

**中图分类号** F224.0 **文献标识码** A

DOI:10.13653/j.cnki.jqte.2015.07.010

## 引 言

改革开放 30 多年来,我国仍处在渐进体制改革的进程中,新制度、新政策的出台显得必不可少。如此一来,我国经济学界亟需一门能够定量考察制度绩效与客观评估政策效果的研究工具,而客观度量新政策实施对经济体影响的因果动态检验却又非常困难。20 世纪 80 年代以来,国外经济学界兴起了一种专门用于分析政策效果的计量方法——双重差分法(Differences-in-Differences Method)。这种将制度变迁和新政策视为一次外生于经济系统的“自然实验”的研究方法,源于自然科学且思路简洁,估计方法日趋成熟,被西方学界广泛应用于诸多领域之中。

在我国,周黎安和陈烨(2005)最早引入了双重差分法。该文巧妙利用我国农村税费改革最早在安徽省进行试点,其后在江苏、湖南、湖北的部分县进行试点,最后才推向全国的事实,将试点期的税费改革视为一次“准实验”,从而对农村税费改革的政策效果进行了系统的双重差分实证检验。自此,我国基于双重差分法的研究成果如雨后春笋般出现。2005~2015 年 3 月,关于双重差分的学术论文和研究生学位论文分别已有 345 篇和 138 篇<sup>②</sup>,并呈现加速增长趋势——2008 年和 2009 年均只新增了 8 篇论文,而 2010 年~2012 年则分别新增 36 篇、37 篇和 55 篇,2013 年和 2014 便达到了新增 76 篇和 108 篇的规模,博士、硕士论文也从 2008 年和 2009 年的每年新增 5 篇左右上升至 2012 年的 25 篇和 2013 年的 26 篇,2014 年则高达 45 篇。

<sup>①</sup> 本文获得国家自然科学基金项目(71203078、71333007、41371174)、教育部哲学社会科学重大课题攻关项目(14JZD021)、国家社会科学基金项目(13CJL065)、广东产业发展粤港澳台区域合作研究中心及暨南大学“跨越计划”的资助。

<sup>②</sup> 数据来源为 CNKI 中国知网数据库中篇名或摘要或关键词含有“双重差分”或“倍差”的文献数量。

双重差分法的以下几个优点是其被广泛应用的主要原因。计量模型简单易用,回归估计方法成熟。单纯从计量角度来看,双重差分法其实就是将两个虚拟变量及其交乘项增加进回归方程,方法简单而有效。相对于静态比较法,双重差分法不是直接对比样本在政策前后的均值变化,而是使用个体数据进行回归,从而判断政策的影响是否具有显著的统计意义。相对于传统办法,双重差分法能够避免政策作为解释变量所存在的内生性问题,即有效控制了被解释变量和解释变量之间的相互影响效应。如果样本是面板数据,那么双重差分模型不仅可以利用解释变量的外生性,而且可以控制不可观测的个体异质性对被解释变量的影响。自然实验利用外生事件的冲击进行研究,其样本分组和处理变量均独立于个体异质性。因此,双重差分法既能控制样本之间不可观测的个体异质性,又能控制随时间变化的不可观测总体因素的影响,因而能得到对政策效果的无偏估计。

既然双重差分法是无可多得科学评估政策效果的研究工具,且具有上述诸多优点,为何我国经济学界近三年才出现相关研究成果的井喷?其实,西方经济学界从自然科学界引入自然实验和双重差分法早在20世纪70年代末(Ashenfelter, 1978),而理工学科研究对实验样本进行分组并观察实验前后的差异,更是延绵数百年的科学传统。周黎安和陈烨(2005)是一次经过深思熟虑的开拓性尝试,该文从各方面保证了使用双重差分法的必要条件,从各个维度考察了实证结果的稳健性。而这次尝试“姗姗来迟”的根本原因正是双重差分法的使用具有极其严苛的前提条件,而我国绝大多数的政策实施并非一次严谨的自然实验,根本不满足进行双重差分研究所需的基本条件与前提假设。

20世纪末的中国农村税费改革具有很多自然实验所必备的基本条件,周黎安和陈烨(2005)使用双重差分法进行研究具有很强的说服力。即便如此,该文作者也只是谨慎地将农村税费改革视为一次“准自然实验(Quasi-natural Experiment)”,并对双重差分法的适用性和稳健性进行细致的“实验前测”,最终才得出如下判断——“本文是国内文献中首次利用双重差分模型的思路对一项重大政策改革的效果进行评估”。

然而,当我国的双重差分法研究于近三四年进入井喷期后,学界似乎已经把双重差分法等同于两个虚拟变量的简单相乘,而忘却了自然实验才是双重差分的根本,亦忽视了使用该计量模型的诸多基本前提,将一些完全不符合自然实验基本条件的政策冲击视为自然实验,将一些本不适用于双重差分法的样本数据进行回归估计。

近年来的双重差分研究发展迅猛,但却存在诸多不符合学术研究要求的缺陷,主要包括内生性、控制组受影响、样本异质性及概念混淆等问题。如果研究者继续忽略现有研究成果所存在的缺陷,而没有采取有效的方法缓解或消除这些缺陷,那么就会对今后的经验研究产生一定的负面影响,甚至使政策评估出现偏差乃至完全相反的结论。有鉴于此,本文认为有必要严谨而全面地讨论双重差分法的基本应用条件与前提假设,归纳当前我国这方面研究的基本情况,指出其存在的主要问题并尽可能提供解决问题的建议。本文的分析或许不仅有助于今后的研究者了解双重差分法的应用条件、研究现状和未来方向,还能有助于他们更准确地使用双重差分法,从而推进政策评估与双重差分领域的研究发展。

## 一、基于自然实验的双重差分法

### 1. 自然实验:双重差分法的理论基础

双重差分法的理论框架建立在“自然实验(Natural Experiment)”基础上,没有自然实验,所谓的“双重差分法”研究只能称作针对两个虚拟变量交叉相乘项的实证研究。

自然实验不同于“真实实验”。真实实验指的是研究者在可以控制一切无关因素的情况下进行实验,从而观察“实验变项”的变化对“因变项”的影响效果<sup>①</sup>。相对于真实实验,自然实验无法控制一切无关因素。因此,自然实验的基本特征是,存在随机扰动和无关因素的影响,所以需要进行统计分析。

此外,研究者还需注意自然实验外部效度<sup>②</sup>较差的问题。因为真实实验可以控制一切无关因素,并随机抽取实验对象,所以它一般是可重复的并具有一般意义的。自然实验则不然,尤其政策冲击实验更是难以进行完全一模一样的重复实验。正因如此,实验理论强调指出:自然实验的外部效度较差,即基于自然实验的研究结果并不一定可推论到研究对象以外的其他受试者,或该次研究情境以外的其他情境。比如说,用市级面板数据得出的政策效果可能不适用于其他省份及全国,那么经济学界常用的以点及面、管中窥豹的研究推论在双重差分研究中尤需谨慎。

## 2. 双重差分计量模型

要进行自然实验,首先要进行随机或近似于随机的样本分组。第一组是进行了“实验处理”的实验组(Treatment Group,亦可翻译为“处理组”),第二组是不实施实验变项的控制组(Control Group,亦可翻译为“对照组”)。在政策效果评估的研究中,实验处理或实验变项即指政策实施,受到新政策影响的样本即为实验组,否则为控制组。

记样本的分组虚拟变量为 $d^j$ ,实验组的 $d^1=1$ ,表明该组样本受到了政策冲击;控制组的 $d^0=0$ ,表明没有受到政策冲击。记所有样本的时间虚拟变量为 $d_t$ ,实验处理前 $d_t=0$ ,实验处理后 $d_t=1$ ,表明政策冲击已发生。

时间虚拟变量 $d_t$ 与分组虚拟变量 $d^j$ 的乘积即为双重差分估计量 $d_t^j$ (Difference in Differences Estimator,  $d_t^j=d_t \cdot d^j$ )。 $d_t^j$ 是实验处理(政策实施)对实验因变项(被解释变量)的影响是否显著的判别依据。

双重差分计量模型一般可表达为:

$$y_{it}^j = \alpha_0 + \alpha_1 d_t + \alpha_2 d^j + \beta d_t^j + \alpha_3 x_{it}^j + \epsilon_{it}^j \quad (1)$$

其中, $y_{it}^j$ 为第 $i$ 个个体的被解释变量, $x_{it}^j$ 为考察个体差异的控制变量, $\epsilon_{it}^j$ 为随机扰动项, $\alpha$ 、 $\beta$ 为回归系数。 $\beta$ 正是研究者最关心的实验变项(双重差分估计量)对实验因变项(被解释变量)的影响效果,即所谓的政策效应。

双重差分估计量应具备外生性,即以下基本条件必须成立<sup>③</sup>:

$$E[\epsilon_{it}^j | d_t^j] = 0 \quad (2)$$

要保证式(2)成立,就必须实现分组随机与实验时间随机。 $E[\epsilon_{it}^j | d_t^j] = 0$ 将保证分组事件的随机性,使分组虚拟变量为外生变量,是自然实验完成了有效样本分组的必要条

① 用计量经济学的术语表达实验理论中的变项和因变项,则为自变量(解释变量)和因变量(被解释变量)。

② 在实验理论中,自然实验的内部效度天生地比外部效度高。所谓外部效度(External Validity)是指实验结果的概括性和代表性,而内部效度(Internal Validity)则指实验者所操纵的实验变项对因变项所造成的影响的真正程度,即实验处理是否确实造成有意义的差异。

③ 当计量方程存在控制个体异质性的控制变量时,该条件应为 $E[\epsilon_{it}^j | d_t^j, x_{it}^j] = 0$ 且 $Cov[\epsilon_{it}^j, d_t^j] = 0$ ,但该条件常常被简化为 $E[\epsilon_{it}^j | d_t^j] = 0$ 。因为,只要实验分组 $d^j$ 与实验时间 $d_t$ 是随机的,OLS估计或固定效应面板估计对 $\beta$ 的估计便是一致的,即使对 $x_{it}^j$ 回归系数的估计可能出现有偏。

件。 $E[\epsilon_i | d_i] = 0$  将保证实验发生时间的随机性,使时间虚拟变量为外生变量,是实验发生时间与实验因变量(被解释变量)无关的必要条件。上述条件满足,且自然实验的其他条件成立时,可得实验处理的真实效应 $\hat{\beta}$ :

$$\hat{\beta} = (\bar{y}_1^1 - \bar{y}_0^1) - (\bar{y}_1^0 - \bar{y}_0^0) = E[y | d_i=1] - E[y | d_i=0] \quad (3)$$

$\beta = E[y | d_i] > 0$  表明政策对被解释变量的影响是正向的,反之则反之。 $\beta$  是实验变量对实验组产生的实验效果,如果政策不出现, $\beta$  的期望值是 0。

### 3. 自然实验与准实验的异同

我国经济学界常用的自然实验概念,其实往往是所谓的准实验(Quasi-experiment)。二者具有着细微却显著的区别——准实验的实验样本挑选与分组是实验者人为进行的,而自然实验则是完全随机的“自然事件”(Dinardo, 2008)。准实验无法在实验环境中完全随机地选择并分组实验对象,所以其样本选择及分组只能尽力迫近随机。从下文的文献综述中,可以看出国内外对自然实验和准实验界定的严谨程度差异,我国常常出现概念混淆的误区。

为保证农村税费改革实验的分组具备如式(2)所示的随机性,周黎安和陈烨(2005)进行了深入的稳健性分析(即实验前测)。首先,虽然安徽是农业大省,但周黎安和陈烨(2005)指出中央政府对试点省份的选择可以近似为随机事件,即此次准实验的样本选择接近于随机。其次,安徽省 21 个试点县的选择,并非是抓阄式的随机抽取,反而是经过了精心的挑选,省政府有可能刻意选择与被解释变量相关的县进行实验。因此, $E[\epsilon_i | d^i]$  看似可能不等于 0,分组事件的严格外生性无法保证。为此,该文使用两种方法检验了分组事件的内生性问题。稳健性检验结果表明,实验组的样本选择倾向与被解释变量农民收入增长率不显著相关。因此, $E[\epsilon_i | d^i]$  可近似为 0,从而判断农村税费改革试点实验具有较强的“政策外生性”。

农村税费改革是一次典型的准实验,因为其样本分组是实验者(政府)随机抽选的。而 Meyer 等(1995)则认为 20 世纪 80 年代初的美国工伤补助标准提升改革是一次自然实验。1980~1982 年美国两个州政府相继提高了高收入人群的工伤补助标准,该文假设:州政府对高收入的界定标准(实验分组依据)与受伤工人领取补偿的意愿无关,即高收入界定标准实际上是政府“拍脑袋”随机划分出来(或者说出于其他考虑而确定),并不会说据此界线划分出来的高收入人群就会特别贪婪或懒惰,特别乐于拉长工伤休养期(被解释变量,即领取工伤补助的时间)。因此可保证  $E[\epsilon_i | d^i] = 0$ ,该实验的分组可视为严格外生和随机的。

此外,自然实验与准实验又同时是受控实验(Controlled Experiment)<sup>①</sup>与现场实验(Field Experiment)的一种。

自然实验中的“自然”一词意指,实验样本挑选与分组为完全随机的“自然事件”,而非指研究对象必须为自然现象,亦非指自然实验多用于自然科学。因此,自然实验的主要应用领域涵盖了自然科学与人文社科领域,如生物学、流行病学、经济学、社会学、管理学等。绝大多数自然实验并非研究自然现象,因此自然实验和准实验实际上是受控实验的一种。

<sup>①</sup> 受控实验即通过对某些影响实验结果的无关因素加以控制,系统地操纵某些实验条件,然后观测与这些实验条件相伴随现象的变化,从而确定条件与现象之间的因果关系的一种研究方法。

而受控实验又可分为两类：实验室实验（Laboratory Experiment）和现场实验，前者主要用于自然科学，后者主要应用在人文社会学科。现场实验是在社会情境中展开的，非所有研究条件和无关因素都在实验者的控制之下，而是由社会情境所确定。实验室实验多数可以控制所有无关因素，因而多为真实实验；现场实验几乎都是自然实验或准实验<sup>①</sup>。

以双重差分法进行政策评估的研究框架多为受控实验中的现场实验。一个受控的实验室实验一般是真实实验，一般可保证  $\sum \epsilon_{it}^j \cong 0$ ；而受控的现场实验则一般是自然实验或准实验。由于无法准确控制所有无关因素与误差，所以现场实验下的双重差分研究只需保证无关因素和误差导致的随机扰动期望值等于零（ $E[\epsilon_{it}^j] = 0$ ）即可。

#### 4. 基于自然实验的双重差分研究的基本假设

综合实验理论与前人研究，可知双重差分研究必须满足以下基本假设。这些基本假设的满足，亦是在经济学研究中应用双重差分法的基本条件。

基本假设①：随机分组，保证每个样本有同等机会接受同一实验处理，即  $E[\epsilon_{it}^j | d^j] = 0$ 。

基本假设②：随机事件，保证实验发生时间的随机性，即  $E[\epsilon_{it}^j | d_t] = 0$ <sup>②</sup>。

基本假设③：控制组不受实验变项的任何影响，即  $E[y_{it}^0 | d_t^i = 1] - E[y_{it}^0 | d_t^i = 0] = 0$ 。

基本假设④：同质性，实验组与控制组样本是统计意义上的同质个体。

基本假设⑤：实验处理（政策实施）的唯一性，实验期间应保证实验变项只出现一次。一旦出现多次的实验处理，实验效果便无法从中分离出来。而且，只要在实验处理前后发生了与被解释变量高度相关的事件，就会导致有未考察到的变量对被解释变量产生较大影响。因此，双重差分研究所要考察的政策冲击，应该是单次的，而非连续的事件。

五条基本假设可归纳为以下两类。第一，随机性假设。双重差分研究必须通过随机化排除那些无法控制因素的影响，从而控制所有可能影响实验结果的无关因素。

实验前测与实验后测均是保证随机化的常用手段。例如，周黎安和陈烨（2005）对改革前的1999年农民收入增长率与是否试点县进行回归，实质上是应用了实验理论中的实验前测法，试图保证实验分组的随机性。又如，比较OLS估计和固定效应面板模型估计的结果， $\beta$ 的差异不大即表明该次自然实验基本成立，这样的实验后测可以保证样本分组满足无条件随机。

第二，同质性假设。实验组和控制组样本除实验者所操纵的实验变项（政策冲击）不同外，其余各方面都应达到近乎相等或完全相似的程度。

同质性一般通过随机抽样和大样本进行保证。检验异方差亦是保证同质性的手段。由于无关因素对个体的影响是一致的，实验组和控制组应具有同方差性： $\text{Cov}[\epsilon_{it}^j, \epsilon_{kt}^j | d^j] = 0, \forall i \neq k$ 。同质性还意味着：如果政策不出现， $\beta$ 的期望值应该是0。这表明实验组和控制组样本在实验前的被解释变量具有同样的时间趋势。周黎安和陈烨（2005）、郑新业等（2011）、张霞和毕毅（2013）进行的实验前测正是这方面的努力。

① 真实实验与受控实验之间的关系较复杂：一次真实实验既可以是受控实验，也可以是非受控实验，如一些宇宙学的实验研究，即使不能人为控制无关因素，宇宙环境自身便已提供很好的实验环境；受控实验亦不一定是真实实验，尤其以人文社科领域的实验为主。例如，伽利略自由落体实验为受控实验和真实实验；研究宇宙红移和黑洞现象的实验为非受控实验和真实实验。孟德尔豌豆杂交实验为受控实验、真实实验和现场实验；物理与工程类实验多为受控实验、真实实验和实验室实验。

② 只要基本假设①和②满足，即可保证式（2）的双重差分估计量外生性（ $E[\epsilon_{it}^j | d_t^i] = 0$ ）。

自然实验研究还强调,控制无关因素,减低实验误差。进入实验系统内部的无关因素往往是由实验者没有控制好的实验环境所造成的。因此,自然实验应设法控制实验变项(政策冲击)以外的所有无关因素<sup>①</sup>,使其对实验效果的影响减至最低程度。自然实验的内部效度,或者说政策效果评估的客观性高低,视研究者对无关因素与误差控制的好坏而定。控制越好则因变项的差异越能被解释为由实验处理所造成,回归结果也是无偏的;控制越差,实验结果越无法解释,回归估计也越不可靠。

控制无关因素与误差的主要手段有:第一,随机抽样与大样本;第二,选择合理的回归估计方法,如存在显著组间差异的就应该使用固定效应面板模型而非 OLS 估计;第三,合理增加控制变量 $x_i$ ;第四,同时事件,即实验同时作用于所有个体( $d_i = d_i$ );等等。

总之,一旦基本假设①~④得不到满足,双重差分估计量 $d_i$ 将是有偏的。这样的双重差分研究难以得出政策实施的真实效果。而基本假设⑤则相对较弱,国内外部分研究或多或少都会不满足这一假设。

## 二、国内的研究现状、主要问题与改良办法

### 1. 基于自然实验的双重差分研究典型

纵观国内的现有研究成果,齐良书和赵俊超(2012)是我国迄今为止最典型的一篇双重差分研究文献,甚至可以视为一篇基于自然实验的双重差分研究范文。该文直接构造了一次针对贫困地区中小学寄宿生的营养干预实验,由来自中国发展基金会的 CDRF 实验项目及其数据支撑起整个自然实验与计量回归的开展。通过对比同一地区有否参与营养补助实验的青少年样本,研究发现营养补贴对青少年的智力与体力存在边际递减的积极作用,进而提出了一些关于营养补贴政策后续开展的具体建议。

在成功的案例里,部分双重差分研究使用了宏观或产业层面数据。姚永玲和汤学兵(2008)、张德荣和郑晓婷(2013)使用双重差分模型分别研究了北京奥运效应和大城市住宅限购政策对房地产价格的影响,结果发现:奥运和限购令其实对房价波动的影响不显著。赵峦和孙文凯(2010)利用我国农信社体制改革的分省分批试点特性,以双重差分法考察了组建省级农信社实验对农户扶持力度的影响。研究结果发现,改革的作用是负面的,改革后农信社的农户贷款覆盖面反而在减少。李楠和乔榛(2010)利用 1999~2006 年中国工业行业数据进行双重差分研究,将国有企业产权改革视为非同时实施的自然实验,结果发现以劳均营业收入和劳均利润为代表的企业绩效受到了国企改革的正面影响。自 1999 年以来的国有企业改革,使国有企业的整体业绩获得了明显提高。但在控制了国企垄断程度自 2003 年开始不断加强的因素后,研究发现国企的部分超额利润其实来自于行政垄断。

亦有研究使用了微观个体数据。周晓艳等(2011)利用 2003~2006 年农业部的农村固定观察点数据,将 2003 年以来的新型农村合作医疗制度改革作为一次准实验,采用双重差分模型来考察政策实施对中国农村居民储蓄行为的影响。研究发现:新政策的实施显著减少了实验对象家庭的家庭储蓄,从而间接地刺激了农村消费与全国内需,因此政策效果是积极的。邢春冰和聂海峰(2010)对 1998 年户籍制度改革进行了一次较严谨的双重差分研究,结果发现:户籍改革实验使得一个城市男性娶农村女性的概率增加了 5% 左右。虽然该文使用的是截面数据,但却巧妙地使用结婚年份是否处于户籍改革后作为时间差分变量 $d_i$ ,分组

① 当然,部分无关因素与其他实验结果的交互作用,可将其列入残差从而忽略掉这些无关因素的负面作用。

变量 $d^i$ 则选用是否农业户口。范子英和李欣(2014)以2003年政府交替期间的部长更换作为一次自然实验,利用双重差分法研究了地方官员晋升部级干部后,政治关联对其原来所在地区获得的转移支付产生的积极影响。研究发现:新上任部长的政治关联效应会导致原地方政府获得额外的转移支付,但却仅体现在正部长,副部长对其来源地获得的财政资金影响不大。

邢春冰和李实(2011)以1999年的扩招政策作为一次自然实验,利用双重差分模型研究该政策对就业市场的冲击,结果发现扩招使大学毕业生的失业率显著提高9个百分点左右。从各个维度来看,该文都很好地满足了基于自然实验的双重差分研究的基本假设。略显不足的地方是,虽然老一辈大学生与扩招后的大学毕业生具有明显的共性,但进行同质性检验的实验前测还是很有必要的。此外,值得改进的是,被忽略的组间相互影响问题(Meyer, 1995)可能存在,因为老牌大学生保住了岗位可能就意味着新毕业的大学生找不到工作,组间样本的被解释变量存在挤出效应。

部分研究虽存在不可控的系统性误差,但仍不失为双重差分研究的成功案例。郑新业等(2011)以双重差分法研究“省直管县”改革,其河南省县级面板数据模型的回归结果表明,体制改革的效果是积极的。该文的研究步骤严谨地按照周黎安和陈烨(2005)的设置,对分组随机性及同质性都进行了实验前测。但实验前测的结果表明,经济总量大的县相对容易被挑选出来进行实验。虽然分组变量 $d^i$ 与被解释变量经济增长率在统计意义上不相关,但分组随机条件还是可能不满足,即 $E[\epsilon_i^j | d^i] \neq 0$ 。尽管如此,正如内生性问题似乎无法在OLS回归中完全克服,由分组非随机导致的政策内生性问题同样难以在双重差分研究中彻底消失。因此,进行了实验前测的郑新业等(2011)仍不失为一次成功的双重差分研究。

庄毓敏等(2012)以双重差分法研究了美国1992~2007年银行业信用衍生品交易市场兴起对银行业系统风险的影响,在一定程度上解释了2007年次贷危机诞生的原因。该文基本满足自然实验研究的各项基本假设,只有 $d_i$ 的确定存在着无法消除的系统性误差。虽然1993年是信用衍生品交易诞生的时间,但由于1993~1996年受其信息披露规则限制,这方面的数据完全缺失。根据美国信用衍生品市场信息披露规则,可知1996年前的信用衍生品市场交易量极小。据此,庄毓敏等(2012)以1996年取代了1993年作为信用衍生品交易实验的发生时间。这种难以彻底消除的系统性误差,对回归估计产生的副作用自然是可以接受的。

聂辉华等(2009)、张霞和毕毅(2013)以2004年的东北地区增值税改革试点为自然事件,对增值税改革的政策效果进行了评估,结果发现:增值税改革明显提升企业的资本密度、劳动生产率、研发密度及以资产负债率为代表的非债税盾。增值税改革的政策效果是积极的,值得向全国推广。张霞和毕毅(2013)的主要瑕疵为实验组与控制组样本量差距过远,前者样本数为7.8万家,后者为115.0万家,二者比例低于聂辉华等(2009)的研究,将对双重差分估计量 $d^i$ 的显著性造成影响。本文认为,这个问题其实可以很简单地通过随机抽样解决。此外,为改进聂辉华等(2009)的研究,张霞和毕毅(2013)还对回归结果的稳健性进行了实验前测——尝试证明实验组(东北企业)与控制组(非东北企业)样本之间不存在显著的异质性,尽管该实验前测的结果不是那么理想。

从各个维度来看,上述文献都基本上满足了自然实验与双重差分研究的各项基本假设,是较典型的双重差分研究论文。

## 2. 不满足基本假设①（随机分组）的研究现状及其改进可能

Meyer (1995) 的研究综述发现, 国外双重差分研究较常犯的错误便是违背基本假设①, 即实验分组不满足随机性假设—— $E[\epsilon_i^j | d^j] \neq 0$ 。该文将违背基本假设①的问题归纳为“选择偏误”。本文认为, 它是双重差分研究中“政策内生性”问题的一种, 在我国尤为显著。

叶青等 (2012) 以企业进入胡润百富榜作为自然实验, 考察该事件对企业会计信息质量的影响。但该文不满足双重差分研究的基本假设①——随机分组, 样本抽样亦不具有随机性。进入实验组的富豪榜企业自然是胡润挑选的业绩好或规模大的企业, 控制组的非富豪榜企业则是研究者根据行业、规模及业绩人工选择出来的样本。这样无法保证此次实验的内部效度, 因为样本选择不具有一般性和随机性。另外, 对实验组和控制组进行样本选择的并非同一主体 (前者为胡润, 后者为研究者), 加剧了实验分组的内生性问题。

这样处理的选择偏误或者说政策内生性问题相对较为严重, 正如 Meyer (1995) 的观点——实验组分组存在偏误, 将导致对被解释变量产生的影响是来自分组, 而非政策实施; 即使政策没有实施, 也会导致双重差分估计量显著不为零。也就是说, 富豪榜企业被分到实验组 (上榜) 是因为它们在实验前的被解释变量就已经差别于非富豪榜企业。事实上, 实验组的两个被解释变量累积收益率和非经常性损益的均值分别是控制组的 3.5 倍和 5 倍 (叶青等, 2012)。

王孝松等 (2010)、白重恩等 (2011) 关于出口退税政策效果的研究同样出现了政策内生性问题。正如两文所述, 近年来出口退税率调整的政策目标明确指向——控制出口过快增长和调整“高污染、高耗能、资源型”产品的出口量, 因此政府对实验组样本选择的的目的性很强——专门挑选劳动和资源密集型产业。在 2007 年 7 月的退税政策调整中, “三高”产品的出口退税率平均降低了 11.06%, 远高于平均值 5.87%。换言之, 实验分组明显不满足随机分组假设, 即基本假设①的随机分组不满足。

两文同时使用的被解释变量为某类商品出口额的增长率, 而政府调整出口退税率的目标恰恰是要“控制出口过快增长”的出口品, 因此  $E[y_i^j | d^j = 1] > E[y_i^j | d^j = 0]$ 。正是由于如此的政策目标, 基本假设②的随机事件 ( $E[\epsilon_i^j | d_i] = 0$ ) 同样不满足。范子英和田彬彬 (2014) 也出现了类似的轻微问题。

杨阳和万迪昉 (2010) 的股指期货研究、翟爱梅和钟山 (2012) 的 H 股上市卖空机制研究、许红伟和陈欣 (2012) 及肖浩和孔爱国 (2014) 的融资融券研究对证券市场的系列体制改革进行了双重差分研究。但由于政策试点往往实施在权重股、蓝筹股, 因此, 上述研究的样本选择与分组目的性同样过强, 在一定程度上偏离了自然实验的分组随机基本假设①。

一旦样本选择与分组的目的性很强, 以随机性为前提的自然实验就可能不会成立。只有有效控制选择偏误, 双重差分回归的结果才会可靠 (Heckman 等, 1998)。关于研究样本的选择偏误问题及其对回归结果的影响, 更细致的计量技术方面探讨可见于 Heckman 和 Robb (1985)、Manski (1989) 的研究。

本文认为: 分组非随机的问題, 除了要进行相关的实验前测外, 还可以通过随机抽样的办法进行缓解。比如, 出口退税政策研究在选取控制组样本的时候, 可以使用同样是“三高”的, 但退税率却没有被政府调整的出口品。这种偶发的遗漏对于政府而言是“随机误差”, 但对于研究者来说, 却是不可多得的随机抽样行为。至于证券市场研究, 只要保证实验组和控制组样本同为蓝筹股或权重股, 或选择周边证券市场的股票作为控制组, 研究就可以视相关改革为随机分组的自然实验。



### 3. 不满足基本假设②（随机事件）的研究现状及其改进可能

储著贞和梁权熙（2012）以2009年“四万亿”财政扩张政策为一次自然实验事件，使用双重差分法检验了这次政策实施对银行贷款增长率的正向冲击效应。由于该文选取的因变量、自变量均多与宏观经济环境相关，而“四万亿”财政扩张政策又在一定程度内生于国际金融危机和我国的宏观经济指标下滑，因此该次自然实验同样出现了政策内生性问题。但与分组非随机问题不同的是，造成政策内生的主要原因是，该研究无法保证实验发生时间的随机性（ $E[\epsilon_{it} | d_i] = 0$ ），即不满足基本假设②。

曾海舰和苏冬蔚（2010）认为，1998年的信贷扩张和2004年的信贷紧缩可视为一次政府实施的自然实验，进而将小规模、民营化程度高以及担保能力弱的公司列为实验组，将大规模、国有化程度高以及担保能力强的公司列为控制组。研究表明，政策冲击对中国上市公司资本结构产生了重大影响。但由于中央的货币政策与宏观经济环境高度相关，导致实验发生时间不具有随机性，即不能保证  $E[\epsilon_{it} | d_i] = 0$ 。

此外，“四万亿”财政扩张政策以及调低或调高准备金率的货币政策往往是政府的连续行为，即在实验处理后的年份，中央会陆续地实施相似政策。因此，曾海舰和苏冬蔚（2010）的实验事件从时间点变为连续事件，有可能使研究不满足基本假设⑤——实验处理（政策冲击）的唯一性，从而影响了政策效果评估的客观性。相似的问题还出现在王孝松等（2010）的出口退税政策调整研究中，不过该文对此问题的影响进行了评估与控制。从另一个角度看来，不满足基本假设⑤实际导致的是“遗漏关键变量”问题（Meyer, 1995）。但正如上文所言，不满足基本假设⑤是国内外双重差分研究的常见现象，这或许并不会大幅削弱上述研究结果的有效性。

本文认为：实验时间非随机的问题，可以通过更换与政策目标不直接相关的被解释变量、控制变量进行缓解，从而减轻  $d_i$ ， $x_{it}$  与  $y_{it}$  之间的内生性问题。当然，进行实验前测检验政策发生的时间点与被解释变量无关，也是可行办法之一。

### 4. 不满足基本假设③（控制组不受影响）的研究现状及其改进可能

左翔等（2011）以2005年河南省农业税费改革为自然实验，以农业占GDP比重的均值对样本进行分组，低于均值的为控制组样本，高于均值的为实验组样本。实验分组依据不再是离散变量，而是连续变量，该文认为“其基本思路和回归结果与普通的双重差分方法在本质上并无二致”。然而，河南省农村税费改革的实验对象是全省所有县——2004年全省的农业税率下调到4%，2005年进一步下调至2%。也就是说，该文所用的控制组样本全部受到了政策实验的冲击，因而不满足基本假设③。

Meyer等（1995）的实验组样本为高于某一收入水平的工伤者，但其划分界线是由政府的改革文件确定的，因此其分组依据是一个离散变量，并伴随政府实施政策而产生。左翔等（2011）则混淆了研究者与实验者的身份——实施政策的实验者是政府，而进行实验分组（是否高于均值）却是由研究者决定的。而且，两个被解释变量农业支出占一般预算支出比重、单位耕地农业财政支出与分组变量  $d_i$  农业占GDP比重高度相关，随机分组的基本假设①不满足。解释变量与被解释变量高度相关，导致超过一半的回归方程拟合优度高于0.8，可能出现了Meyer（1995）提出的双重差分研究常见的控制变量或解释变量与被解释变量高度相关的同步性（Simultaneity）问题。

本文认为：实验者（政府）与样本分组者（研究者）不是同一主体的问题，是必须克服的。在选取研究视角时，研究者应该站在实验者（实施新政策的政府）的立场去进行样本选

择与分组,切忌过度人为干预实验分组的最终结果。否则,从严格意义上来看,出现类似问题的上述研究其实只是基于单虚拟变量(时间差分变量)的实证研究。

上述问题同样出现于王鑫和吴斌珍(2011)、汪伟等(2013)的税费改革研究中,其“连续型双重差分估计”<sup>①</sup>研究实际上并未对样本进行严格的实验分组。其 $d^i$ 是一个连续的人均所得税和农业税费减免变量,减免额大为实验组,减免额小为控制组。所以该文的控制组样本同样受到了税费改革的影响, $d^i$ 更与被解释变量城市居民消费和农村居民消费率高度相关,即不满足基本假设③——控制组不受影响。而余森杰和梁中华(2014)对于关税下调的研究,同样可能出现了上述问题——加入WTO带来关税下调的主要受益者当然是实验组的“一般进口企业”,但控制组的“加工进口企业”也可能受到其他WTO配套政策的冲击。这是因为,2002年的这个虚拟变量既可以代表WTO带来的关税下调冲击,同样也可以代表WTO带来的林林种种的贸易政策改变,总有一些影响着加工进口企业。控制组受到实验变项的其他效应影响,或许会削弱时间差分变量 $d_t$ 的解释力。

至于控制组受到政策影响的问题,则可以通过引入外地样本作为控制组而实现改良。

#### 5. 并非基于自然实验的双重差分研究

国内部分双重差分研究不满足基本假设④——样本具有同质性,甚至出现了实验组样本并非同一个体的案例。袁渊和左翔(2011)以双重差分法研究浙江省2002年进行的“省直管县”改革,试图检验这次自然实验是否提升了实验组样本浙江县级企业的经营绩效。但该文受制于面板数据缺乏等因素,无法构造时间差分变量 $d_t$ ,遂使用福建省县级企业作为 $d_t=0$ , $d^i=1$ 的实验组样本,浙江县级企业作为 $d_t=1$ , $d^i=1$ 的实验组样本,两省的市级企业为控制组样本。但这样的数据结构和研究设定似乎无法构成一次自然实验或准实验,因为既没有“省直管县”实验事件的发生(使用的数据全为2003年的实验处理后数据),实验前后的实验组样本亦根本不是同一个实验对象。双重差分估计量 $d_t^i$ 实际上只是“是否浙江县级企业”的虚拟变量,其回归系数显著为正只表明:浙江县级企业比浙江市级、福建市级及县级企业的业绩都好。或者说,该文提取的实验组样本不能代表其试图要研究的对象,实验设计与研究者的本意不符,这是一个较严重的外部效度问题(Meyer, 1995)。因此,袁渊和左翔(2011)的主要结论“‘扩权强县’对浙江省县辖企业的发展有显著地促进作用”,似乎难以得到有效的支撑。但换个角度来看,这样的研究虽不能算作基于自然实验的双重差分研究,却也可以视为针对两个虚拟变量交乘项的实证研究。

张川川等(2014)以双重差分法研究新型农村社会养老保险政策(简称“新农保”)对老年人(55~64岁)的冲击,其特色也是试图使用截面数据构造自然实验。但由于缺乏面板数据,其 $d_t$ 差分变量实则为“是否大于60岁”。也就是说,此次实验设定:参与“新农保”社区中55~59岁的老年人(实验组样本)在2012年北京大学国家发展研究院进行普查的时间点上,并没有受到来自2010年就开始实施的“新农保”的冲击,其收入、消费、幸福感等被解释变量都不会受到新政策的影响。首先,与袁渊和左翔(2011)类似,该文实验前后的实验组样本——55~59岁老人和60~64岁老人其实不是严格意义上的同一个实验对象。其次,参加了“新农保”的55~59岁老人一定会预期到5~1年后的退休以及“新农

<sup>①</sup> 在部分国外研究中,交互项中有两个连续变量或者其中一个为连续变量,仍可视作双重差分研究,如Rajan和Zingales(1998)关于金融与经济发展的论文中所示。但“连续型双重差分估计”无疑会损害到控制组不受影响的基本假设。至于如何进行更深入的处理,还有待学界讨论。

保”所带来的额外福利；59岁的人也应该不会因为今年还没退休拿社保，而导致其消费行为、幸福感与同一地区的60岁人出现截然不同的差异。

上述两文的实验设定似乎无法支持一次自然实验或准实验，时间差分变量 $d_i$ 的构建相对不成功，又或者说它们只是一次基于分组差分变量 $d_i'$ 的实证研究。能够成功使用截面数据构建自然实验的国内研究似乎只有邢春冰和聂海峰（2010）。

还有部分研究对自然实验的界定出现了偏差，其实验事件的发生并非外生于样本个体，亦非由实验者（政府）实施的，而是被实验对象自行选择的结果。如廖理和张学勇（2008）的股权分置改革研究，其中家族上市公司实施股改的实验事件完全内生于企业自身的经营与战略，因而似乎构成不了一次自然实验，而仅是一次双虚拟变量实证研究。类似的问题还见于李新义和汪浩瀚（2010）的企业横向兼并研究、蒋冠宏和蒋殿春（2014a, 2014b）的中国企业OFDI研究。这个问题可以很容易解决——在进行自然实验研究前，经济学者可以对照医学和生物学的小白鼠实验。如果研究样本（小白鼠）可以自行选择实验时间和分组的话，那么这次自然实验显然是无法成立的，因为聪明的小白鼠永远不会启动实验，而强壮的小白鼠也永远不会进入实验组。

#### 6. 其他问题

（1）实验为非同时事件与同步性问题。陈刚（2012）研究了2008年高级人民法院院长异地交流轮岗改革对司法效率的影响，研究结果表明：这次改革明显提高了司法效率，异地调动法官所在地的结案率比其他地区高2个百分点左右。该文特色为对个体的实验处理为非同时事件（ $d_i \neq d_i'$ ），样本个体出现异地交流时 $d_i'$ 才等于1。

非同时的实验处理将加剧数据的序列自相关性，以至于为控制好无关因素与误差，陈刚（2012）增加了较多控制变量。但由于部分控制变量与被解释变量结案率可能存在一定内生性，又导致了回归结果的拟合优度高达0.99，以致于产生了多重共线性问题。 $d_i \neq d_i'$ 的假定还见于刘佳等（2011）的省直管县改革研究。这导致部分成果呈现出一定的同步性问题，而序列自相关的问题亦往往会未被检验。

本文认为，只要双重差分研究的实验处理（政策实施）为非同时事件（ $d_i \neq d_i'$ ），那么在添加控制变量时就须更为谨慎，因为这时的序列自相关问题可能已然出现。不过，上述问题可通过进一步细分实验组样本来解决。比如，将同一年实施改革的实验组样本划归一组，进行分组别的双重差分回归。

（2）遗漏变量、序列自相关以及政治影响问题。Meyer（1995）指出，在实验前后发生了研究者未观察到的事件，会导致双重差分研究遗漏了关键的解释变量或控制变量，进而影响政策评估结果的客观性。事实上，不满足基本假设⑤（政策冲击唯一性）的研究（储著贞和梁权熙，2012；曾海舰和苏冬蔚，2010；王孝松等，2010），其实就是出现了遗漏变量——实验项（政策）实际上是多次发生，而研究却未能考察到。

此外，Bertrand等（2004）认为数据的序列自相关问题会导致双重差分估计的标准误差偏低，导致对原假设的过度拒绝，使双重差分估计量的显著性虚高。该文指出，解决序列自相关问题最简单的办法便是——将所有时间序列数据简单划分为政策冲击前后两个时间段，即使研究满足“同时事件”假设——实验同时作用于所有个体（ $d_i = d_i'$ ）。另一个常用的解决办法便是“随机推断（Randomization Inference）”，这种实验后测型的稳健性检验曾被廖理和张学勇（2008）使用。这是一种有效的值得推广的双重差分法稳健性检验。

从一般均衡和政治经济学的角度看,政策实施与政治问题戚戚相关。因此,政策实施很可能影响了实验者(政府)或被实验者(研究样本)的后期行为模式。这种政治影响问题在社会实验理论中,被归纳为“实验者偏误(Experimenter Bias)”问题(威勒和沃克,2010)。

上述三个问题或多或少会出现在基于自然实验的双重差分研究中,其负面影响难以完全避免,使用上文提及的实验前测或实验后测等稳健性检验是较理想的解决办法。

(3) 双重差分研究的称谓问题。学界还有部分双重差分研究并非构建在自然实验的基础上。王永进和盛丹(2013)使用“双重差分法”研究了地理集聚对企业商业信用的影响。其时间差分变量 $d_i$ 是由研究者所发现的2001年企业搬迁较多这一现象而确定,并非源于外生的政策实施。严格意义上来说,这只是一次以双虚拟变量来控制样本组间差异及内生性的实证研究。许斌和韩高峰(2009)的贸易出口配额研究、包群等(2011)的出口收入分配效应研究、曾海舰(2012)的房地产市场研究、项松林和马卫红(2013)的出口企业学习效应研究均与此类似,尝试以双重差分变量控制计量模型的内生性与组间差异。

正因如此,上述论文均没有明确指出自身为一次自然实验或准实验研究。它们实际上是基于双虚拟变量的实证研究。是否双重差分研究,其实只是称谓问题,上面的讨论不涉及以上研究的有效性。

除此以外,自然实验与准实验的概念混淆,被忽略的组间相互影响问题,实验组与控制组样本量差距过远,统计或抽样时出现误差等问题也常见于我国的双重差分研究。一部分问题难以完全克服,且在国外关于双重差分的专著中有所深入分析,另一部分问题则在上文有所提及,在此不赘述。

### 三、结论与建议

#### 1. 简要结论

本文根据实验理论与国外双重差分法的研究基础,剖析了自然实验和双重差分计量模型的主要特征,归纳出双重差分研究所需的基本假设,并提出了经济学研究应用双重差分法时值得注意的系列问题。在此基础上,本文回顾了近年我国双重差分研究的“井喷式”发展历程,并对部分代表性研究成果进行了评述。研究结果发现:我国当前的双重差分研究出现了大面积的实验设定偏差等较严重的问题,主要包括内生性、控制组受影响、样本异质性、自然实验与准实验混淆等。甚至有部分所谓的自然实验研究或许根本构成不了一次自然实验,而只是使用到虚拟变量的实证研究。

其中,尤以政策内生性问题最为严重。政府在选择政策实施对象和时间之际,往往是有目的而非随机的。这势必导致实验分组和实验时间的非随机性,从而使双重差分估计量(实验变项)与被解释变量(实验因变项)之间产生了较严重的内生性问题<sup>①</sup>。那么,双重差分研究的基本假设 $E[\epsilon_i^j | d_i] = 0$ 就根本不成立。只要实验样本分组的目的性很强,以随机性为前提的自然实验就可能不会成立(Meyer, 1995)。因此,政策内生性问题对于部分出

<sup>①</sup> 部分研究者往往只看到,仅依靠计量技术本身来消除实证研究中的内生性问题并不容易,而双重差分法的比较优势恰恰在于可克服政策作为解释变量时产生的内生性问题,但却忽略了另一内生性问题——政策内生性。不顾双重差分法的理论基础——自然实验——是否适用的实证研究,视双重差分法等同于双虚拟变量法,并不能减轻研究的内生性问题,反而会使问题加剧。

现较大偏误的研究来说,甚至是致命的<sup>①</sup>。

## 2. 对策建议

基于自然实验的双重差分模型的应用条件较为严苛。正因如此,双重差分法被引入经济学领域(Ashenfelter, 1978)并稳步发展了20多年后,国内经济学界才出现了第一篇成功的双重差分研究(周黎安和陈烨, 2005)。而其后数年内,每年出现的双重差分研究也不超过10篇,近三年才出现了相关研究的井喷式增长,而存在问题的研究又占其大部分。

综上所述,本文认为:并非所有的政策冲击都可以用双重差分法进行效果评估。

研究者在选择研究对象(具体的某项政策)时必须进行深入的事前分析,以考察这次政策实施是否满足自然实验或准实验的各项基本特征与假设。例如,基于自然实验的双重差分研究应该保证基本假设①~④中半数以上成立,否则该次研究便可能无法构建起一次自然实验。另一方面,研究者还须对样本异质性、序列自相关、组间相互影响等问题进行稳健性检验。

倘若今后的论文写作与期刊审稿工作,一直未能意识到上述问题的严重性,长此以往势必对我国的经验研究产生较大的负面影响。甚至使政策效果评估陆续出现偏差乃至完全相反的结论,间接损害了政府的宏观调控效率,使相关的经济学研究失去了其应有的实践指导意义。为此,本文分析了部分代表性文献的潜在缺陷,并提出了可行的改良办法。尽管所选文献未能不一而足,但本文或许还是有助于学术界了解自然实验与双重差分法的应用条件和主要问题,有助于更准确地使用双重差分法,从而推进政策评估与双重差分领域的研究发展。

从国外引进前沿方法来研究中国问题,是我国经济学界的长处。一旦出现突破性成果,其他的国内研究者往往会趋之若鹜。如此现象的成因或许可以由阿施的“从众实验”(Asch, 1958)解释。该实验结果表明,人们会迫于其他人的选择或者说社会压力,而作出显然违背事实的判断与行为,这种人类行为方式即为“从众行为”。本文认为,在未能透彻掌握一种前沿经济学方法前,从众地生搬硬套进行经验研究,对我国经济学发展似乎并无益处。

当然,本文仅为引玉之砖,还存在诸多不足之处,后续研究有待学界进一步拓展。

## 参考文献

[1] Asch S. E., 1958, *Interpersonal Influence* [A], in Eleanor Maccoby, Theodore Newcomb, Eugene Hartley (eds), *Reading in Social Psychology* [M], New York: Holt, Rinehart and Winston Publishing.

[2] Ashenfelter Orley C., 1978, *Estimating the Effect of Training Programs on Earnings* [J], *Review of Economics and Statistics*, 60 (1), 47~57.

① 为更好地展示政策内生性问题的重要性,本文尝试用作者求学阶段曾从事的土木工程实验来进行类比说明。研究者考察预应力混凝土梁的抗裂缝性能而构建如下实验。要合理评估预应力的实验作用,在营造实验对象——混凝土梁的时候,实验者必须使用同样标号和强度的水泥与钢筋。如果所使用的水泥或钢筋出现差异,那么实验就违反了关于实验对象同质性的基本假设④,随机误差必不为零,实验效度大幅受损。如果研究者刻意证明预应力梁的优势,专门挑选由高标号水泥及大口径钢筋生产的梁来作为实验对象,而以质量相对较差的梁作为控制组。那么,这次实验就会违反关于分组随机的基本假设①,造成分组变量内生于被解释变量混凝土抗裂能力。这时,违背基本假设①就不再是影响实验效度的问题,而是使整个实验失败——预应力梁抗裂缝能力强不能被证明是实验变项(预应力)的原因,而是因为非随机地选取了更高强度的实验样本。要证明预应力梁抗裂缝能力差也很简单,只需刻意选取水泥未干等不适当的时间点进行实验处理,或偷偷对控制组实施预应力作用,使实验违背时间随机的基本假设②和控制组不受影响的基本假设③。那么,这样的实验同样无效。由此可见,关于政策外生性的两大基本假设①和②——分组随机和事件随机,对于自然实验研究是至关重要的。不难看出,国内部分双重差分研究所犯的失误似乎正如上所述。

- [3] Ashenfelter Orley C., David Card, 1985, *Using the Longitudinal Structure of Earnings to Estimate the Effect of Training Programs* [J], *Review of Economics and Statistics*, 67 (4), 648~660.
- [4] Bertrand Marianne, Esther Duflo, Sendhil Mullainathan, 2004, *How Much Should We Trust Differences-in-Differences Estimates?* [J], *Quarterly Journal of Economics*, 119 (1), 249~275.
- [5] Card David, Alan B. Krueger, 1994, *Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania* [J], *American Economic Review*, 84 (4), 772~793.
- [6] Dinardo J., 2008, *Natural Experiments and Quasi-natural Experiments* [A], in Steven N. Durlauf, E. Lawrence (eds), *New Palgrave Dictionary of Economics* [M], London: Palgrave Macmillan Publishing.
- [7] Heckman James, Hidehiko Ichimura, Jeffrey Smith, Petra Todd, 1998, *Characterizing Selection Bias Using Experimental Data* [J], *Econometrica*, 66 (5), 1017~1098.
- [8] Heckman James, R. Jr. Robb, 1985, *Alternative Methods for Evaluating the Impact of Interventions* [A], in J. Heckman, B. Singer (eds), *Longitudinal Analysis of Labor Market Data* [M], Cambridge: Cambridge University Publishing.
- [9] Imbens Guido, Jeffrey Wooldridge, 2007, *What's New in Econometrics?* [C], NBER Summer Institute Minicourse.
- [10] Manski C. F., 1989, *Anatomy of Selection Problem* [J], *Journal of Human Resources*, 24 (3), 341~360.
- [11] Meyer Bruce D., 1995, *Natural and Quasi-experiments in Economics* [J], *Journal of Business and Economic Statistics*, 13 (2), 151~161.
- [12] Meyer Bruce D., W. Kip Viscusi, David L. Durbin, 1995, *Workers' Compensation and Injury Duration: Evidence from a Natural Experiment* [J], *American Economic Review*, 85 (3), 322~340.
- [13] Rajan R. G., L. Zingales, 1998, *Which Capitalism? Lessons from the East Asian Crisis* [J], *Journal of Applied Corporate Finance*, 11 (3), 40~48.
- [14] 白重恩、王鑫、钟笑寒:《出口退税政策调整对中国出口影响的实证分析》[J],《经济学(季刊)》2011年第4期。
- [15] 包群、邵敏、侯维忠:《出口改善了员工收入吗?》[J],《经济研究》2011年第9期。
- [16] 陈刚:《法官异地交流与司法效率——来自高院院长的经验证据》[J],《经济学(季刊)》2012年第4期。
- [17] 储著贞、梁权熙:《宏观调控、所有权结构与商业银行信贷扩张行为》[J],《国际金融研究》2012年第3期。
- [18] 狄恩:《实验设计和分析》[M],世界图书出版社,2010。
- [19] 范子英、李欣:《部长的政治关联效应与财政转移支付分配》[J],《经济研究》2014年第6期。
- [20] 范子英、田彬彬:《出口退税政策与中国加工贸易的发展》[J],《世界经济》2014年第4期。
- [21] 蒋冠宏、蒋殿春:《中国工业企业对外直接投资与企业生产率进步》[J],《世界经济》2014a年第9期。
- [22] 蒋冠宏、蒋殿春:《中国企业对外直接投资的“出口效应”》[J],《经济研究》2014b年第5期。
- [23] 李楠、乔榛:《国有企业改制政策效果的实证分析——基于双重差分模型的估计》[J],《数量经济技术经济研究》2010年第2期。
- [24] 李新义、汪浩瀚:《双边市场横向兼并的定价及福利研究——以中国网络传媒业为例》[J],《财经研究》2010年第1期。
- [25] 廖理、张学勇:《全流通纠正终极控制者利益取向的有效性——来自中国家族上市公司的证据》[J],《经济研究》2008年第8期。
- [26] 刘佳、马亮、吴建南:《省直管县改革与县级政府财政解困——基于6省面板数据的实证研究》[J],《公共管理学报》2011年第3期。
- [27] 聂辉华、方明月、李涛:《增值税转型对企业行为和绩效的影响——以东北地区为例》[J],《管

理世界》2009年第5期。

[28] 齐良书、赵俊超：《营养干预与贫困地区寄宿生人力资本发展——基于对照实验项目的研究》[J]，《管理世界》2012年第2期。

[29] 汪伟、艾春荣、曹晖：《税费改革对农村居民消费的影响研究》[J]，《管理世界》2013年第1期。

[30] 王孝松、李坤望、包群、谢申祥：《出口退税的政策效果评估：来自中国纺织品对美出口的经验证据》[J]，《世界经济》2010年第4期。

[31] 王鑫、吴斌珍：《个人所得税起征点变化对居民消费的影响》[J]，《世界经济》2011年第8期。

[32] 王永进、盛丹：《地理集聚会促进企业间商业信用吗？》[J]，《管理世界》2013年第1期。

[33] 威勒、沃克：《实验设计原理：社会科学理论验证的一种路径》[M]，重庆大学出版社，2010。

[34] 项松林、马卫红：《出口企业具有学习效应吗？——基于中国企业微观数据的经验分析》[J]，《世界经济研究》2013年第10期。

[35] 肖浩、孔爱国：《融资融券对股价特质性波动的影响机理研究——基于双重差分模型的检验》[J]，《管理世界》2014年第8期。

[36] 邢春冰、李实：《扩招“大跃进”、教育机会与大学毕业生就业》[J]，《经济学（季刊）》2011年第4期。

[37] 邢春冰、聂海峰：《城里小伙儿遇到农村姑娘：婴儿户口、户籍改革与跨户籍通婚》[J]，《世界经济文汇》2010年第4期。

[38] 许斌、韩高峰：《配额、汇率和中国纺织品出口价格》[J]，《世界经济》2009年第6期。

[39] 许红伟、陈欣：《我国推出融资融券交易促进了标的股票的定价效率吗？——基于双重差分模型的实证研究》[J]，《管理世界》2012年第5期。

[40] 杨旭武：《实验误差原理与数据处理》[M]，科学出版社，2009。

[41] 杨阳、万迪昉：《股指期货真的能稳定市场吗？》[J]，《金融研究》2010年第12期。

[42] 姚永玲、汤学兵：《奥运与北京房地产价格的变化——基于DD法的分析》[J]，《财贸经济》2008年第8期。

[43] 叶青、李增泉、李光青：《富豪榜会影响企业会计信息质量吗？——基于政治成本视角的考察》[J]，《管理世界》2012年第1期。

[44] 余森杰、梁中华：《贸易自由化与中国劳动收入份额——基于制造业贸易企业数据的实证分析》[J]，《管理世界》2014年第7期。

[45] 袁渊、左翔：《“扩权强县”与经济增长：规模以上工业企业的微观证据》[J]，《世界经济》2011年第3期。

[46] 曾海舰：《房产价值与公司投融资变动——抵押担保渠道效应的中国经验证据》[J]，《管理世界》2012年第5期。

[47] 曾海舰、苏东蔚：《信贷政策与公司资本结构》[J]，《世界经济》2010年第8期。

[48] 翟爱梅、钟山：《卖空机制对股票价格波动的影响：基于A+H股公司的实证研究》[J]，《南方经济》2012年第8期。

[49] 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年第1期。

[50] 张德荣、郑晓婷：《“限购令”是抑制房价上涨的有效政策工具吗？——基于70个大中城市的实证研究》[J]，《数量经济技术经济研究》2013年第11期。

[51] 张霞、毕毅：《增值税改革对企业资本结构的影响研究》[J]，《中央财经大学学报》2013年第8期。

[52] 赵峦、孙文凯：《农信社改革对改善金融支农的政策效应评估——基于全国农户调查面板数据的倍差法分析》[J]，《金融研究》2010年第3期。

[53] 郑新业、王晗、赵益卓：《“省直管县”能促进经济增长吗？——双重差分方法》[J]，《管理世界》2011年第8期。

[54] 周黎安、陈烨：《中国农村税费改革的政策效果：基于双重差分模型的估计》[J]，《经济研究》

2005年第8期。

[55] 周晓艳、汪德华、李钧鹏：《新型农村合作医疗对中国农村居民储蓄行为影响的实证分析》[J]，《经济科学》2011年第2期。

[56] 庄毓敏、孙安琴、毕毅：《信用风险转移创新与银行（体系）的稳定性——基于美国银行数据的实证研究》[J]，《金融研究》2012年第6期。

[57] 左翔、殷醒民、潘孝挺：《财政收入集权增加了基层政府公共服务支出吗？——以河南省减免农业税为例》[J]，《经济学（季刊）》2011年第4期。

## Research Status and Potential Problems of Differences-in-differences Method in China

Chen Lin<sup>1,2</sup> Wu Haijun<sup>3</sup>

(1. Institute of Industrial Economics, Jinan University;

2. Institute of Economics, Chinese Academy of Social Sciences;

3. Department of Statistics, Jinan University)

**Abstract:** Differences-in-differences method (DID) based on the natural experiment is the most widely used research tool in the academic literature about effect of policy, and the trend of domestic literatures using DID method has been increasing rapidly in recent years. However, through concluding the basic application conditions and domestic research actuality of DID method, this paper finds that domestic research literatures are intermingled with the good and bad, and there are some serious problems, such as the endogeneity problem, control group related and sample heterogeneity. Particularly, the problem of endogeneity is more serious, even making the academic research shows some certain defect, in which the hypothesis of natural experiment is often difficult to establish. According to related experimental research and theoretical study abroad, we classify and comment the major domestic research literatures using DID method, and raise improvement measures for the potential problems identified.

**Key Words:** Differences-in-Differences Method; Natural Experiment; Quasi-experiment; Endogeneity Problem

**JEL Classification:** C9

(责任编辑:王喜峰)