

doi: 10.12012/CJoE2021-0016

# 基于面板数据的处置效应估计的计量方法 最新进展

蔡宗武

(堪萨斯大学经济系, 堪萨斯劳伦斯 66045, 美国)

**摘要** 本文着重介绍了基于面板数据的处置效应估计的计量方法最新进展。首先, 简要介绍了面板数据的项目评估或经济政策评价的现代计量分析的基本模型及其性质和应用。其次, 主要关注估计面板数据的平均和分位数处置效应的最新方法和模型。最后, 从理论、方法和实证方面探讨了基于面板数据的经济政策评价的计量经济学和统计学未来可能的研究方向, 特别地, 如何结合三者: 机器学习、经济政策评价和面板数据的未来的研究方向展开讨论, 为年轻学者提供参考。

**关键词** 平均处置效应; 双重差分; 机器学习; 面板数据; 分位数处置效应; 双因素固定效应

## Recent Developments in Estimating Treatment Effects for Panel Data

CAI Zongwu

(Department of Economics, University of Kansas, Lawrence, KS 66045, USA)

**Abstract** This survey paper highlights some recent developments in estimating treatment effects for panel data. First, this paper begins with a brief introduction of the basic model setup in modern econometric analysis of program or economic policy evaluation for panel data. Second, the primary attention goes to the focus on estimating both the average and quantile treatment effects for panel data. Finally, it concludes the paper by addressing theoretically, methodologically and empirically some possible

收稿日期: 2021-01-16

基金项目: 国家自然科学基金 (71631004, 72033008)

Supported by National Natural Science Foundation of China (71631004, 72033008)

作者简介: 蔡宗武, 美国堪萨斯大学经济系 Charles Oswald 讲席教授, 研究方向: 计量经济学、计量金融学、数据科学、经济政策评估、面板数据分析、时间序列分析、应用宏观分析与预测、微观计量经济学、统计学, E-mail: caiz@ku.edu.

致谢: 感谢洪永淼教授和汪寿阳教授的宝贵意见, 并特别感谢詹铭峰提供图, 陆凤彬、包皓文、崔如鸿的中文翻译及纠正原文的错误。不过, 所有错误都属于我本人。

future research directions for young scholars in econometrics and statistics, particularly, some interesting and challenging research topics related to a combination of machine learning and casual inference for panel data.

**Keywords** average treatment effects; difference-in-differences; machine learning; panel data; quantile treatment effects; two-way fixed effects

## 1 引言

为了评估或评价培训计划和实施政治或社会或者经济政策或事件对结果的影响, 人们需要估计其处置效应 (treatment effect (或者处理效应)), 这在经济、金融等许多应用领域都引起了极大的兴趣和关注, 在估计处置效应的过程中, 往往需要高级计量经济学知识或统计分析工具. 估计处置效应的难度在于没有处置的单元的结果未被直接观测到, 通常称为反事实 (counterfactual). 为了克服这一困难, 已有文献中提出了许多方法, 如回归调整、匹配、逆概率 (倾向得分函数, propensity score function) 加权、双重差分 (difference-in-differences, DiD)、断点回归方法等等, 来估计平均处置效应 (ATE) 和分位数处置效应 (QTE). 然而, 上述方法大多依赖于估计倾向得分函数, 比如, 关于横截面数据可参考 Imbens and Wooldridge (2009) 的文章, 时间序列数据可参考 Liu, Cai and Fang et al. (2020) 的综述文章, QTE 方面可参考 Tang (2020) 的综述文献, 更多细节的内容可见参考书 Cerulli (2015).

在上述的解决处置效应的计量经济学方法中, 在应用中, 比较流行的方法是经典 DiD 方法, 被要求有两个时间段 (处置前后) 和两组 (处置组和对照组) 的情况下. 并且如果平行趋势假设成立, 那么处置组的 ATE 可以简单地通过处置组的平均结果随时间的变化和对照组的平均结果随时间的变化的差异来估计. 众所周知, DiD 类型方法避免估计倾向得分函数, 以避免倾向得分函数可能的错误假设. 因此, 对于大面板数据, Hsiao, Ching and Wan (2012) (以下简称 HCW) 把经典的 DiD 方法做了创造性拓展, 可以用来处理大面板数据, 称为 HCW 方法.

事实上, HCW 提出了一个基于因子模型来估计面板数据的平均处置效应. 与面板数据的经典 DiD 方法不同, HCW 考虑了具有多个单元和多个时间段的面板数据, 其中处置发生在特定时间和之后. 他们的重点是通过将其他单位作为对照组来观测特定单位的处置效应. 如果 HCW 的面板数据的横截面截面和时间序列的维度较大, 可以采用 Bai and Ng (2002), Pesaran (2006), Bai (2009) 等方法来估计影响因子. 然而, 在许多应用中, 研究者经常遇到维度不大的情况. 针对这一问题, HCW 提出了一种不考虑因子的反事实结果估计方法. 在某些简单假设条件下, HCW 证明了利用普通最小二乘回归仍然可以得到其估计量, 从而使计算变得简单. 此外, 正如 Li and Bell (2017) 所指出的, HCW 中的方法优点之一是不需要假设没有样本选择效应. 换句话说, 它绕过了虚拟处置和结果之间的相关性问题. 此外, 它不要求处置单位和控制单位在没有处置的时间遵循平行路径.

近年来, 研究人员从理论和实证的角度对 HCW 的研究方法进行了广泛的拓展和应用. 例如, 首先, 通过应用 HCW 的方法, Chen, Han and Li et al. (2013) 构建现货市场波动率的反事实以

及主要基于中国与国际股票市场的横截面相关性, 研究了引入沪深 300 (CSI300)<sup>1</sup> 指数期货交易对中国股市现货价格波动率 (VIX) 的影响. 其次, Bai, Li and Ouyang (2014) 将 HCW 的方法扩展到相关的时间系列是非平稳单位根过程的情况, 然后, 探讨房产税对房价的影响, 利用 2011 年 1 月开始在上海和重庆进行的房产税政策试验, 并利用其他省市的房价, 采用 HCW 的方法估计上海和重庆在不征收房产税的情况下的假设房价. 第三, Ouyang and Peng (2015) 放宽了 HCW 中的线性回归函数假设, 允许存在非参数回归函数, 并将 HCW 模型推广到半参数设置, 并以此方法研究 2008 年中国的经济刺激方案. 第四, Du and Zhang (2015) 建议使用“多个剔除交叉验证”(leave-many-out cross validation) 准则代替 HCW 中的 Akaike 信息准则 (AIC) 来选择最优控制单元, 并将其应用于对中国的购房限制、房产税和房价的反事实分析. 此外, Ke, Chen and Hong et al. (2017) 利用 1990 年到 2013 年中国地级市数据, 通过在没有高铁项目的情况下利用选定的非高铁城市的结果, 构建高铁城市人均实际 GDP 的假设反事实, 评估了高铁项目对目标城市节点 (高铁城市) 经济增长的影响. 此外, Li and Bell (2017) 认为 HCW 的方法可以在限制性较小的假设下仍然是有效, 实际上, Li and Bell (2017) 去掉了 HCW 中施加的一些假设, 然后导出了 HCW 平均处置效应估计量的渐近性质. 另外, Li and Bell (2017) 将最小绝对收缩和选择算子 (least absolute shrinkage and selection operator, LASSO) 方法引入 HCW 的方法中使得控制元选择在横截面尺寸很大的情况下更加有效, 而 Carvalho, Masini and Medeiros (2018) 证明了 LASSO 估计量是相合的并且渐近正态分布的. 更重要的是, Carvalho et al. (2018) 将 HCW 的方法推广为一种灵活且易于实施的方法, 称为 ArCo, 用于在没有对照组的情况下, 估计干预对单个处置单元的因果效应, 这与文献中之前的建议一致, 此外, 他们还考虑了多种干预措施和受污染效应的测试. 最后, Carvalho et al. (2018) 应用他们的模型评估了 2007 年巴西实施的反逃税计划对通货膨胀和其他宏观经济变量 (比如: 经济增长、零售和信贷) 的影响.

最近, 为了利用面板数据或纵向数据 (longitudinal data) 进行因果推断, 以捕捉异质处置效应, 双因素线性固定效应回归 (two-way linear fixed effect, 记为 2FE) 已成为面板数据估计因果效应的默认方法. 使用 2FE 估计量的主要目的是同时调整未观测到的特定单位和/或特定群体和/或特定时间的混杂因素. 事实上, 根据 De Chaisemartin and D'Haultfoeuille (2020) 的一项调查, 在 2010 年至 2012 年期间, 顶尖经济学杂志《美国经济评论》发表的所有实证文章中, 有 19% 使用了双因素固定效应回归来估计处置对结果的影响. 例如, 为了估计 ATE, Imai and Kim (2019, 2020) 提出了与 2FE 模型匹配的方法, 以捕获未观测到的特定单位和特定时间的混杂因素, 而 De Chaisemartin and D'Haultfoeuille (2020) 提出了另外一种新的估计方法, 以适应未观测到的特定群体和特定时间的混杂因素. 最后, Sun and Abraham (2020) 提出了一种替代方法来估计一个有趣处置的动态效应, 而不是包括处置的领先和滞后的双向固定效应回归, 当平行趋势假设只有在观测到的协变量的条件作用下才可能成立时, Callaway and Sant'Anna (2020) 表明, 即使观测到的特征的差异导致组间的非平行结果动态变化, 在交错 DiD 设置中也确定了一系列因果效应参数. 根据他们的鉴定结果, 他们提出了不同的聚类方案, 可以用来突出不同维度的处置效应异质性, 并总结参与处置的整体效应.

尽管 ATE 方法在应用很受欢迎, 但当潜在结果的真实分布是不对称或异质性或厚尾时, ATE

<sup>1</sup> 沪深 300 指数是一个市值加权的股市指数, 旨在复制在上海和深圳交易所交易的前 300 只股票的表现.

可能不能很好地代表政策的效果。例如, 让我们观察 Chen et al. (2013) 所考虑的沪深 300 指数现货价格波动率在 2002 年 1 月至 2020 年 10 月的分布估计, 如图 1 所示, 从图 1 可以清楚地看到三个样本的密度分布: 处置前 (点线)、处置后 (虚线) 和整个样本 (实线) 的分布是不对称的, 并且是严重向右倾斜的。因此, ATE 可能不适用于这个应用问题来刻画政策效应。相反, 我们需要考虑处置效应的分布影响。在文献中, 分位数处置效应 (QTE) 可以用来识别政策在对应于观测结果和反事实结果的整个分布的任何分位数上的效应。

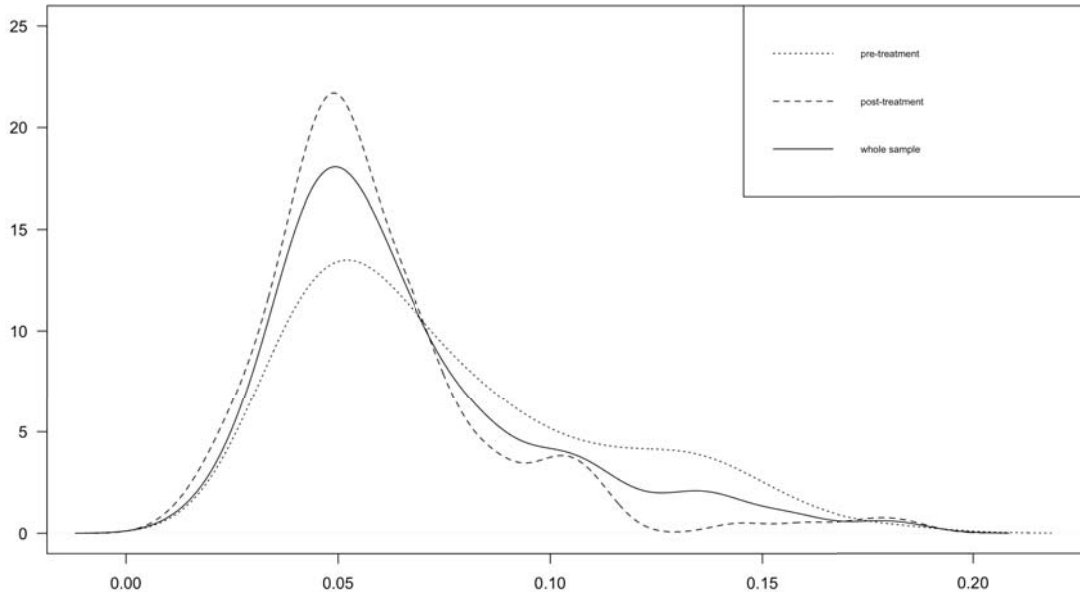


图 1 2002 年 1 月至 2020 年 10 月沪深 300 指数的估计波动率密度函数估计:  
处置前 (点线) 和处置后 (虚线) 以及整个样本 (实线)

然而, 关于用面板数据估计 QTE 的文献非常有限。大概, 最近的文献包括 Callaway, Li and Oka (2018) 以及 Callaway and Li (2019), 他们考虑在 Copula 不变性的假设下 (见后面的假设 Q2), 在固定时间段 (有限面板) 的 DiD 设置下使用面板数据估计 QTE (见后面的假设 Q2), 以及 Cai, Fang and Lin et al. (2021) 的文章, 该文将 HCW 的方法推广到面板数据的 QTE 设置, 从而对大面板数据的处置效果进行了综合刻画和考察。估计反事实的分位数的处置单位, 不同于 HCW 的方法及其它的扩展, Cai et al. (2021) 引入了条件累积分布函数 (CDF) 不变性的假设 (见后面的假设 Q3), 提出了一个简单的方法来使用条件 CDF 和无条件的 CDF 之间的关系。通过这种方法, 可以使用非参数和半参数方法来估计条件 CDF。此外, Cai et al. (2021) 导出了所提出的 QTE 估计量的大样本性质, 以及一种基于区块 (blockwise) Bootstrap 的易于实现的构造置信区间的方法。

最后, 使用 QTE 面板数据方法, Cai et al. (2021) 通过分位数处置效果分析, 研究引入沪深 300 指数期货交易, 2010 年 4 月 16 日由中国金融期货交易所正式推出, 是否对会现货市场波动率以及像 Huang, Schlag and Shaliastovich et al. (2019) 那样的波动率之波动率 (volatility-in-

volatility, VVIX)<sup>2</sup> 产生影响. 引入后, 有人批评引入指数期货交易可能会因过度投机而动摇现货市场, 也有人认为指数期货市场可以提高信息流的速度和质量, 使金融市场更加完善. Cai et al. (2021) 的研究有趣发现是, 引入沪深 300 指数期货交易不仅对中国金融市场的 VIX 有显著影响, 而且对 VVIX 也有显著影响.

本文的其余部分组织如下. 第 2 节致力于回顾经典的 DiD 方法及其变体, 如用于估计多个时间段的 ATE 的双因素固定效应. 第 3 节详细阐述了 HCW 的方法及其扩展, 第 4 节讨论了有限面板数据和大型面板数据的 QTE 估计. 第 5 节是本文的结束语, 同时, 讨论了一些非常有趣而且挑战性极强的未来研究问题, 尤其和机器学习相关的未来研究方向.

## 2 面板数据的双重差分方法

### 2.1 两期的双重差分方法

首先, 让我回顾一下经典的 DiD 方法, 尽管它在计量经济学文献中众所周知, 例如, 见 Cerulli (2015). 事实上, 经典的 DiD 需要从一个实验组和一个控制组在两个或两个以上不同时期 (比如,  $T$  个时期, 且  $T \geq 2$ ) 测量的数据, 具体来说, 处置前至少一个时间段, 处置后至少一个时间段. 为了评估一个项目或处置对一个个体群组的结果的影响, 通常假设有两组处置状态  $D = 0$  或  $1$ , 其中  $0$  表示未接受处置的个体, 即对照组,  $1$  表示接受处置的个体, 即处置组. 同时, 假设个体在两个时间段内被观测 ( $T = 2$ ),  $t = 0$  和  $1$ ,  $0$  表示处置组接受处置前的一段时间, 即处置前,  $1$  表示处置组接受处置后的一段时间, 即处置后. 每一次观测都以字母  $i = 1, \dots, N$  为指标, 每个个体通常有两个观测, 一个在处置前, 一个在处置后. 为结果  $y_{it}$  建模, 通常考虑以下回归模型

$$y_{it} = \beta_0 + \beta_1 D_i + \beta_2 t + \Delta_{\text{DiD}}(D_i \cdot t) + \varepsilon_i, \quad (1)$$

其中系数是用希腊字母  $\{\beta_j\}_{j=0}^2$  表示的, 还有  $\Delta_{\text{DiD}}$  也是未知参数, 以及  $\varepsilon_i$  是一个随机未观测到的误差项. 由 (1) 式, 我们可以清楚地看到, 系数有如下的解释:  $\beta_0$  是常数项,  $\beta_1$  表示处置组的特定效应 (来解释处置和控制之间的平均永久性差异),  $\beta_2$  呈现控制组和处置组常见的时间趋势,  $\Delta_{\text{DiD}}$  是处置的真实效应. 所以, DiD 方法的重点在于估计  $\Delta_{\text{DiD}}$ .

为了估计  $\Delta_{\text{DiD}}$ , 从 (1) 式, 我们可以很容易地得到  $\Delta_{\text{DiD}} = [E(y_1^1) - E(y_0^1)] - [E(y_1^0) - E(y_0^0)]$ , 其中  $y_0^1$  和  $y_1^1$  分别是处置组处置前后的结果,  $y_0^0$  和  $y_1^0$  是对照组相应结果的样本平均数. 因此, 差分 - 差分 (或双重差分) 估计值定义为处置组处置前后平均结果的差值减去对照组处置前后平均结果的差值, 这是一种差分中的差分, 定义为

$$\hat{\Delta}_{\text{DiD}} = [\bar{y}_1^1 - \bar{y}_0^1] - [\bar{y}_1^0 - \bar{y}_0^0], \quad (2)$$

其中  $\bar{y}_0^1$  和  $\bar{y}_1^1$  分别是处置组处置前后结果的样本平均数,  $\bar{y}_0^0$  和  $\bar{y}_1^0$  是对照组相应结果的样本平均数. 我们可以很容易地证明  $\hat{\Delta}_{\text{DiD}}$  是  $\Delta_{\text{DiD}}$  的无偏和相合的估计. 显然, 这个估计量可以被看作是两个前后估计量 (pre-versus-post estimators) 的差值, 即 (2) 右侧的第一部分, 减去对照组的估计量, 即 (2) 右侧的第二部分. 当然, 上述设置可以很容易地推广到有  $T \geq 2$  但固定的情况, 更多细节读者可以参考 Cerulli (2015).

<sup>2</sup>波动率之波动率的定义可以在 Huang et al. (2019) 中找到.

值得一提的是,最近, Henderson and Sperlich (2021) 成功地把 (2) 推广到带有 confounder (混杂变量) 的非参数 DiD 情况, 也就是考虑处置效应函数  $\Delta_{\text{DiD}}(x)$  依赖于一些混杂变量  $x$ , 然后利用非参数估计方法进行估计  $\Delta_{\text{DiD}}(x)$ , 同时, 利用他们提出的模型来研究和评价美国著名的移民政策 deferred action for childhood arrivals (DACA, “童年抵美者暂缓遣返”)<sup>3</sup> 对那些非法进入美国的小孩教育 (中小学和高等) 情况的影响。

最后, 我深信国内学者对模型 (1) 有许多实证应用, 通常采用 (2) 式中的估计方法  $\hat{\Delta}_{\text{DiD}}$ . 在此, 深表歉意的是, 由于篇幅有限, 没有对国内学者的中文文献进行综述。

## 2.2 双因素固定效应双重差分方法

近年来, 为了利用面板数据或纵向数据进行因果推断, 捕捉异质和动态的处置效应, 双因素线性固定效应回归已成为一种流行的方法, 用于从面板数据估计因果效应, 以调整未观测到的单位特定和/或群体特定和/或时间特定混杂因素. 例如, 参见近期和新兴的关于 DiD 和/或事件研究中不同处置时间的异质处置效应的文献, 包括但不限于, Imai and Kim (2019, 2020), De Chaisemartin and D'Haultfoeuille (2020), Sun and Abraham (2020), Callaway and Sant'Anna (2020) 以及其中的参考文献. 实际上, Imai and Kim (2020) 考虑了以下双因素线性固定效应回归模型

$$y_{it} = \alpha_i + \gamma_t + \beta^\top Z_{it} + \varepsilon_{it}.$$

对于  $i = 1, 2, \dots, N$  以及  $t = 1, 2, \dots, T$ , 其中  $\alpha_i$  和  $\gamma_t$  分别是单位效应和时间固定效应,  $Z_{it}$  是协变量, 然而 Imai and Kim (2019) 考虑了不带  $\gamma_t$  项的上述模型. 此外, Imai and Kim (2020) 表明, 2FE 模型同时对这两种未观测的混杂因素进行调整的能力严重依赖于线性可加效应的假设. 使用 2FE 估计量的另一个常见理由是, 在两组和两个时间段的最简单设置下, 它与 DiD 估计量等价. 此外, 他们认为, 在应用研究中通常遇到的更一般的情况下, 这种等效性并不存在. 相反, 他们证明了多周期的 DiD 估计量等价于加权的 2FE 估计量, 但一些观测值的权重为负. 因此, Imai and Kim (2020) 将 Imai and Kim (2019) 中的单向固定效应回归估计量的匹配表示扩展到 2FE 估计量, 以理解这些广泛使用的估计量在非参数框架内的因果解释. 更多细节读者可参考 Imai and Kim (2020). 此外, Imai and Kim (2019) 应用他们提出的方法来考虑关于关税和贸易总协定成员资格对二元贸易的因果影响的争议。

另外, 在 De Chaisemartin and D'Haultfoeuille (2020) 的这篇文章中, 假设在同一  $(g, t)$  单元格中的所有观测结果有相同的处置方法, 并且处置方法是二元的, 就像处置方法是以县为单位 (county-level law). 考虑  $y_{i,g,t}$  的回归, 即  $t$  期  $g$  组单位  $i$  在对组固定效应、周期固定效应和  $D_{g,t}$  ( $t$  期  $g$  组的处置) 的结果. 令  $\hat{\beta}_{\text{fe}}$  表示  $y_{i,g,t}$  在对组固定效应、周期固定效应和  $D_{g,t}$  上的最小二乘法 (OLS) 回归中  $D_{g,t}$  的系数. 在共同趋势假设下, De Chaisemartin and D'Haultfoeuille (2020) 表明  $\beta_{\text{fe}}$  等于每个处置过的  $(g, t)$  单元格中处置效应的加权总和

$$\beta_{\text{fe}} = E \left( \sum_{(g,t): D_{g,t}=1} W_{g,t} \Delta_{g,t} \right),$$

<sup>3</sup>DACA, “童年抵美者暂缓遣返” 计划, 是由奥巴马于 2012 年发起, 准许在年幼还没有决定权的时候跟着自己的父母或是一级亲属来到美国的人可以继续待在美国合法的工作跟生活, 不必被强硬地遣返回去自己的国家. 但是必须每两年就重新续约一次这个协议

其中  $\Delta_{g,t}$  为  $g$  组和  $t$  期的平均处置效应 (ATE) 并且权重  $W_{g,t}$  相加等于 1 但可能是负的. 正如 De Chaisemartin and D'Haultfoeuille (2020) 所阐述的那样, 负权重的产生是因为  $\beta_{fe}$  是几个 DiD 的加权和, 这些 DiD 比较了各组在连续时间段内结果的演变. 然而, 一些比较中的对照组可能在两个时期都进行处置. 然后, 它在第二个时期的处置效应被 DiD 所差分, 因此为负权重. 由于负权重, 例如线性回归系数可能是负的, 而所有的 ATE 都是真正正的. 因此, De Chaisemartin and D'Haultfoeuille (2020) 提出了另一个估计方法来解决这一问题; 例如, De Chaisemartin and D'Haultfoeuille (2020) 中的定理 1 详细介绍了  $\beta_{fe}$  的显式表达式. 最后, De Chaisemartin and D'Haultfoeuille (2020) 将他们提出的方法应用于两个实际应用: 使用面板数据并控制工人的固定影响, 研究了 1868 年至 1928 年美国选举中报纸对选民投票率的影响, 以及工会成员对工资的影响.

最后, Sun and Abraham (2020) 提出了一种替代方法来估计随时间变化的动态效应, 而不是包括处置的超前和滞后的双向固定效应回归, 并且当平行趋势假设只有在观测到的协变量的条件下才可能成立时, 而 Callaway and Sant'Anna (2020) 考虑了识别和估计问题, 并表明在交错 DiD 设置中识别了一系列因果参数, 即使观测到的特征差异在组间产生了不平行的结果动态变化. Callaway and Sant'Anna (2020) 根据识别结果提出了不同的聚类方案, 可用于突出不同维度的处置效应异质性, 总结参与处置的整体效应. 关于细节, 读者可以参考上述文章和其中的参考文献. Sun and Abraham (2020) 利用他们提出的方法研究了住院治疗的经济后果, 这是美国成年人经济风险的一大来源, 而 Callaway and Sant'Anna (2020) 通过分析 2001 年至 2007 年美国最低工资对青少年就业的影响, 阐述了他们提出的工具的相关性.

在许多使用面板数据或纵向数据的项目评估应用中, 双因素固定效应 (2FE) 方法通常被使用, 如上两节所述. 问题是 2FE 估计量与经典估计量有什么不同. 实际上, 根据上面对两种方法的描述, 人们可能已经看到了它们之间的区别. 然而, 即使使用 DiD 进行估计所需的识别条件比 FE 更少, 由于前期零处置条件, 使用 DiD 减少了估计 ATE 所需的观测次数. 当观测量显著减少时, 可以使用 DiD 方法. 然而, 如果观测数量急剧下降, 则需要使用 FE 估计, 因为 FE 可能对 ATE 产生更稳健的估计. 关于以上两种方法的更多比较, 读者可以参考 Cerulli (2015) 的书. 最后, 正像 Henderson and Sperlich (2021) 那样, 把 (2) 经典 DiD 方法推广到非参数情况, 是否可以把上述 2FE 各种方法推广到非参数情景有待我们进一步探讨和研究, 希望年轻学者可以往这个方向去思考.

### 3 面板数据的 HCW 方法

#### 3.1 模型和估计方法

本节专门介绍 HCW 方法及其扩展和应用. 为此, 对  $1 \leq i \leq N$  和  $1 \leq t \leq T$  ( $T \rightarrow \infty$ ),  $y_{it}^1$  表示 HCW 方法在第  $t$  期内对第  $i$  个单位的应用结果, 而  $y_{it}^0$  表示没有 HCW 方法在第  $t$  期内对第  $i$  个单位的应用结果. HCW 方法在第  $t$  期内对第  $i$  个单位的效果定义为

$$\Delta_{it} = y_{it}^1 - y_{it}^0.$$

由于无法同时观测到  $y_{it}^1$  和  $y_{it}^0$ , 因此观测到的数据具有以下形式:

$$y_{it} = d_{it}y_{it}^1 + (1 - d_{it})y_{it}^0.$$

如果在时间  $t$  内第  $i$  单元正在处置, 则  $d_{it} = 1$ , 否则  $d_{it} = 0$ . HCW 的模型着重于这样一种情况, 即对  $1 < T_1 < T$  时只有一个单位在时间  $T_1 + 1$  后接受处置. 在不失一般性的前提下, 假定它是第一个单元. 换句话说, 对  $j = 2, \dots, N$  和所有的  $t = 1, \dots, T$ , 以及对第一个单元  $y_{1t}$ ,  $t = 1, \dots, T_1$  时,  $y_{jt}$  没有应用 HCW 方法; 只有当  $t = T_1 + 1, \dots, T$  时,  $y_{1t}$  应用了 HCW 方法. HCW 使用以下因子模型来激励他们的估计方法

$$y_{it}^0 = \alpha_i + b_i^\top f_t + u_{it}, \quad i = 1, \dots, N, \quad t = 1, \dots, T, \quad (3)$$

其中  $\alpha_i$  为第  $i$  个个体的特定截距,  $b_i$  为  $K \times 1$  维因子负荷向量,  $f_t$  为  $K \times 1$  维向量 (不可观测) 公共因子,  $u_{it}$  为弱相依平稳误差项. 如果  $T$  和  $N$  都很大, 可以采用 Bai and Ng (2002) 的方法估计  $f_t$  的公因数. 在  $T$  和  $N$  都不很大的情况下, HCW 提出了一种称为控制单元的新方法, 通过使用  $\tilde{y}_t = (y_{2t}, \dots, y_{Nt})^\top$  来代替  $f_t$  预测处置后的反事实结果  $y_{1t}^0$ .

具体来说, 在 HCW 方法中, 尽管驱动所有横截面单元的一些共同因素对每个横截面单元的影响可能不同, 但可以假设横截面单元之间的相关性是由它们导致的. 因此, 根据  $y_{1t}$  和  $\{y_{jt}\}_{j=2}^N$  与 (3) 中的模型的相关性, HCW, Li and Bell (2017)<sup>4</sup>推导出如下回归模型

$$y_{1t} = \beta^\top x_t + u_{1t}, \quad t = 1, \dots, T_1. \quad (4)$$

为估计反事实结果  $y_{1t}^0$ , 其中  $x_t = (1, y_{2t}, \dots, y_{Nt})^\top$ ,  $\beta = (\beta_1, \dots, \beta_N)^\top$ , 和  $u_{1t}$  是一个零均值和有限方差的特殊误差项. 用 OLS 回归法对系数  $\beta$  进行估计,

$$\hat{\beta} = \arg \min_{\beta} \sum_{t=1}^{T_1} (y_{1t} - \beta^\top x_t)^2.$$

通过假设数据结构在处置前后保持不变和其他假设, 如 HCW 中列出的假设 1~6, OLS 估计  $\hat{\beta}$  是相合的, 可以估计  $t = T_1 + 1, \dots, T$  时  $y_{1t}^0$  的反事实预测结果如下:

$$\hat{y}_{1t}^0 = \hat{\beta}^\top x_t, \quad t = T_1 + 1, \dots, T.$$

最后, 对第一个单位的平均处置效果进行估计:

$$\Delta_1 = E(\Delta_{1t}) = E(y_{1t}^1 - y_{1t}^0). \quad (5)$$

可以通过将处置后观测到的结果和估计的反事实结果之间的差异取平均值来构建:

$$\hat{\Delta}_1 = \frac{1}{T_2} \sum_{t=T_1+1}^T (y_{1t}^1 - \hat{y}_{1t}^0), \quad (6)$$

其中  $T_2 = T - T_1$ . 可以证明, 这在一些假设下, 如 HCW 的假设 1~6, 这个估计量是相合的估计.

**备注 1** 我们可以看到, 使用 HCW 方法的一个进步是避免像 Bai and Ng (2002), Pesaran (2006), Bai (2009) 中那样估计公共因子  $\{f_t\}$ . 但是, (4) 中的  $x_t$  应该允许包含一些协变量  $Z_t$ , 它不是控制单元的一部分, 这样在应用中更有吸引力. 这个想法有待我们进一步去探讨.

对于 (3) 中对时间序列  $\{f_t\}$  的要求, HCW 的假设中  $\{f_t\}$  可以是平稳的, 也可以是非平稳的; 从 HCW 的备注 2 中可知, 不过, 还有一个假定  $\sum_{t=1}^T \|f_t\|^2 / T$  收敛于一个常数, 但是, 对于

<sup>4</sup>HCW 在 Li and Bell (2017) 的假设 HCW3 下推导 (4), 该假设被 Li and Bell (2017) 所抛弃.



$f_t$  是一个有或没有漂移的单位根过程, 这个假定显然是不满足的. 因此, Bai et al. (2014) 认为, 如果  $\{f_t\}$  是一个有或没有一个漂移的单位根过程, 不需要做任何线性条件均值函数形式的假设下, Bai et al. (2014) 在命题 2.1 中证明了, 他们仍然可以得到类似于 HCW 的假设 6 的一个结果. 此外, Bai et al. (2014) 表明 (6) 中的  $\hat{\Delta}_1$  是在一些假设下 (见 Bai et al. (2014) 的假设 1~6) 是 (5) 中  $\Delta_1$  的相合估计.

另外, Li and Bell (2017) 放宽了 HCW 中的一些分布假设, 如线性条件均值函数形式假设和 HCW 中的假设 6, 并表明 HCW 的方法确实适用于更广泛的数据生成过程. 然后, Li and Bell (2017) 导出了  $\hat{\Delta}_1$  的渐近分布以便于推断. 此外, Ouyang and Peng (2015) 将 HCW 模型应用于 2008 年中国经济刺激方案的研究, 放宽了 HCW 的线性条件均值假设, 允许使用非参数条件均值函数, 并将 HCW 模型推广到半参数设置. 还有, Carvalho et al. (2018) 将 HCW 的模型扩展到多个处置单元和非线性模型的情况, 而之前的文献只有一个处置单元, 以及应用一个评估巴西反逃税计划对通货膨胀和其他宏观经济变量影响的程序. 最后, 通过应用 HCW 的方法, Chen et al. (2013) 研究了引入指数期货交易对中国股票市场现货价格波动率的影响, Ke et al. (2017) 评价了高铁项目对中国高铁节点目标城市经济增长的影响.

最近, Fujiki and Hsiao (2015) 对 HCW 的方法进行了进一步扩展, 以理清当观测到的结果受两种处置方法的影响时一种方法对另一种的影响. Fujiki and Hsiao (2015) 使用了 1995 年 1 月 17 日发生的 Hanshin-Awaji 大地震作为研究问题的动机, 提出这个新的方法, 他们的研究发现地震没有持续的影响, 观测到的持续效应是兵库县的结构变化造成的.

### 3.2 控制元选择

当有大量的控制单元 ( $N$  固定但很大) 时, 使用所有的控制单元来估计可能不是最好的选择, 因为大量的协变量通常会导致较大的估计方差, 这反过来又会导致较差的样本外预测. 为了选择合适的控制单元, HCW 建议使用 AIC 或其变体, 看看是否需要使用所有的横截面单元. 此外, Li and Bell (2017), Carvalho et al. (2018) 都建议使用 LASSO 方法选择控制单元; 也就是说,

$$\hat{\beta}_{\text{LASSO}} = \arg \min_{\beta} \sum_{t=1}^{T_1} (y_{1t} - \beta^{\top} x_t)^2 + \sum_{j=1}^N \psi_{\lambda}(\beta_j), \quad (7)$$

其中  $\psi_{\lambda}(\cdot)$  是惩罚函数,  $\lambda$  是一个惩罚参数. 至于如何选择惩罚函数, Li and Bell (2017), Carvalho et al. (2018) 取  $\psi_{\lambda}(\cdot)$  为绝对值函数, Wu and Liu (2009) 采用平滑修剪绝对偏差 (SCAD) 罚函数. 特别地, Carvalho et al. (2018) 在 Du and Zhang (2015) 建议下使用交叉验证准则选择  $\lambda$ , 而且为 LASSO 估计提供了理论论证. 正如 Li and Bell (2017) 中提到的, 为了获得  $\beta$  的 LASSO 估计, 需要选择  $\lambda$  的值, 常见的做法包括使用 BIC 或删一交叉验证 (leave-one-out cross validation) 或广义交叉验证方法 (generalized cross validation) 来选择惩罚参数. 如果  $N > T_1$ , 这就是超高维变量选择问题, 可以使用所谓的凹凸方法 (concave convex procedure)-SCAD 算法, 类似于 (7) 的思想来估计  $\beta$ ; 详情可见 Kim, Choi and Oh (2008).

## 4 面板数据的 QTE 方法

众所周知, 如果潜在结果的分布不太集中于均值或存在结果的异质性, 平均处置效果可能不

符合理想方法的描述效果. 要把握一种处置效果的全局效应, 自然要研究处置效果的整体分布. 在本节中, 我们的重点是估计面板数据的分位数处置效果.

#### 4.1 小面板的 QTE 模型

对于两期面板数据, Callaway et al. (2018) 考虑了一个 DiD 框架, 在该框架中, 样本中的所有个体在  $t$  期不接受处置而一小部分接受处置. 设  $D_{it}$  为一个处置指示器, 如果在周期  $t$  内处置单个  $i$ , 则值为 1, 否则为 0. 对于每个个体  $i$ , 在周期  $s \in \{t-1, t\}$  中存在一对潜在结果  $(y_{is}(0), y_{is}(1))$ , 其中  $y_{is}(0)$  和  $y_{is}(1)$  分别表示周期  $s$  中未处置状态和处置状态的潜在结果. 每个人都经历过治疗或未治疗的状态, 但不是两者都有, 因此这对潜在的结果是不可观测到的. 然后, 假设两周期面板数据由观测值  $\{(y_{i,t-1}, y_{it}, Z_i, D_{it})\}_{i=1}^n$  组成, 结构如下:

$$y_{i,t-1} = y_{i,t-1}(0), \quad y_{i,t} = (1 - D_{it})y_{it}(0) + D_{it}y_{it}(1),$$

其中  $Z_i$  是一个协变量. 潜在结果  $y_{i,t-1}(0)$ ,  $y_{i,t-1}(1)$  和  $y_{it}(0)$ ,  $y_{it}(1)$  在横断面上取决于处置状态  $D_{it}$ .

为了衡量处置效果, Callaway et al. (2018) 考虑在给定  $Z_i = z$  的条件分位数下对处置效果 (CQTT) 的识别和估计, 衡量分位数处置在  $D_{it} = 1$ , 共同历史  $Z_i = z$  的情况下对人群中个体的效果. 让  $\mathcal{Z}$  成为  $Z_i$  的未处置和处置组的共同支持. 当  $z \in \mathcal{Z}$  时在  $\tau \in (0, 1)$  处的 CQTT 定义为:

$$\Delta_z^{\text{CQTT}}(\tau) = q_{y_{it}(1)|Z=z, D_{it}=1}(\tau) - q_{y_{it}(0)|Z=z, D_{it}=1}(\tau),$$

其中  $q_{y_{it}(1)|Z=z, D_{it}=1}(\tau)$  是在给定  $Z_i = z$  和  $D_{it} = 1$  的情况下  $y_{it}(1)$  的  $\tau$  条件分位数, 同时  $q_{y_{it}(0)|Z=z, D_{it}=1}(\tau)$  是在给定  $Z_i = z$  和  $D_{it} = 1$  的情况下  $y_{it}(0)$  的  $\tau$  条件分位数. 为了定义  $\Delta_z^{\text{CQTT}}(\tau)$ , 令  $\Delta y_{it}(0) = y_{it}(0) - y_{i,t-1}(0)$ , 同时假设  $\Delta y_{it}(0)$  独立于  $Z_i$  上的条件独立的处置状态  $D_{it}$ , 即经典处置效应文献中所谓的条件独立假设; 详见 Liu et al. (2020), Tang (2020). 关于检验条件独立假设是否成立, 已有一些文献就此提出; 参见 Fang, Tang and Cai et al (2020) 及其参考文献. 此外, 它还需要做出以下假设.

**假设 Q1** (分布 DiD) 对所有  $y \in \text{supp}(\Delta y_{it}(0) | Z_i)$  和  $z \in \mathcal{Z}$ ,

$$P(\Delta y_{it}(0) \leq y | Z_i = z, D_{it} = 1) = P(\Delta y_{it}(0) \leq y | Z_i = z, D_{it} = 0).$$

也就是说, 在给定  $Z_i = z$  和  $D_{it} = 1$  下,  $\Delta y_{it}(0)$  的条件分布和在给定  $Z_i = z$  和  $D_{it} = 0$  下,  $\Delta y_{it}(0)$  的条件分布是完全一样.

**假设 Q2** (Copula 不变性) 对每个  $z \in \mathcal{Z}$  和所有  $u, v \in [0, 1] \times [0, 1]$ ,

$$C_{\Delta y_{it}(0), y_{i,t-1}(0) | Z=z, D_{it}=1}(u, v) = C_{\Delta y_{it}(0), y_{i,t-1}(0) | Z=z, D_{it}=0}(u, v).$$

也就是说, 在给定  $Z = z$  和  $D_t = 1$  下,  $\Delta y_t(0)$  和  $y_{t-1}(0)$  的条件 Copula 和在给定  $Z = z$  和  $D_t = 0$  下,  $\Delta y_t(0)$  和  $y_{t-1}(0)$  的条件 Copula 是完全一样.

在假设 Q1 和 Q2 以及其他一些假设下, Callaway et al. (2018) 表明, 感兴趣变量的反事实分布可以从未处置个体的观测结果中识别出来. 这意味着, 处置组和未处置组在分布意义上必须相似, 不仅是边缘分布, 而且在某些时期的依赖性. 因此假设 Q1 和 Q2 在识别中发挥关键作用, 如 Callaway et al. (2018) 所示. 此外, Callaway et al. (2018) 提供了  $\Delta_z^{\text{CQTT}}(\tau)$  的估计, 由

$\hat{\Delta}_z^{\text{CQTT}}(\tau)$  表示, 并推导了  $\hat{\Delta}_z^{\text{CQTT}}(\tau)$  的渐近结果, 以及一个 Bootstrap 推断. 另外, Callaway et al. (2018) 利用他们提出的方法估计了提高最低工资对由种族、性别和教育程度定义的子群体收入分配的影响.

更进一步, Callaway and Li (2019) 把上述方法推广到至少有三个观察时间点的面板数据情况 ( $T \geq 3$ ). 然后通过假设个体在处置或未处置状态下分别具有潜在结果:  $y_{1t}$  和  $y_{0t}$ , 他们考虑了分位数处置对已处置 (QTT) 的影响,

$$\Delta^{\text{QTT}}(\tau) = q_{y_{1t}|D=1}(\tau) - q_{y_{0t}|D=1}(\tau),$$

其中  $q_{y_{1t}|D=1}(\tau)$  是在给定  $D = 1$  的情况下  $y_{1t}$  的  $\tau$  条件分位数, 而  $q_{y_{0t}|D=1}(\tau)$  是在给定  $D = 1$  的情况下  $y_{0t}$  的  $\tau$  条件分位数. 在假设 Q1 和 Q2 以及一些附加假设下, Callaway and Li (2019) 提供  $\Delta^{\text{QTT}}(\tau)$  的估计值, 记作  $\hat{\Delta}^{\text{QTT}}(\tau)$ , 并推导了  $\hat{\Delta}^{\text{QTT}}(\tau)$  的渐近结果. 此外, Callaway and Li (2019) 利用提出的这个方法估计了提高最低工资对当地劳动力市场失业率分位数的影响, 发现存在显著的异质性.

**备注 2** 对于有限的面板数据, 如何通过使用 2FE 方法来刻画异质性处置的特征来估计 QTE 在文献中似乎没有很好地解决. 因此, 未来有可能对这一课题进行研究.

## 4.2 大面板的 QTE 方法

### 4.2.1 模型和估计方法

不同于 Callaway et al. (2018), Callaway and Li (2019), 假设面板数据是带有协变量  $\{Z_t\}_{t=1}^T$  的  $\{y_{it}, 1 \leq i \leq N\}_{t=1}^T$ , 其中  $Z_t \in R^{d_z}$ , 而且  $T$  可能很大, 比如  $T \rightarrow \infty$ . 所以, 要求面板数据的每个时间序列是严格平稳的. 对第一个单元, 只有  $t = T_1 + 1$  时接受处置, 其余的  $N - 1$  个单位作为对照组, 在整个过程中保持不处置. 定义  $T_2 = T - T_1$ . 为简化表示法, 对于给定的面板数据,  $\{y_{1t}\}_{t=1}^{T_1}$  表示第一个单元且  $\{(y_{it}, Z_t); 2 \leq i \leq N\}_{t=1}^{T_1}$  表示处置前的剩余单位和协变量, 用  $\{X_{1t}\}_{t=1}^{T_1}$  表示. 同时,  $\{y_{1t}\}_{t=T_1+1}^T$  是第一个单元的信息且  $\{(y_{it}, Z_t); 2 \leq i \leq N\}_{t=T_1+1}^T$  为处置后的剩余单位和协变量, 记为  $\{X_{2t}\}_{t=1}^{T_1}$ . 这里,  $Z_t$  是一个  $d_z \times 1$  的协变量向量, 它不包含在 HCW 以及与其的相关文献中. 设  $\{y_{1t}^0\}_{t=T_1+1}^T$  为反事实结果,  $\{y_{1t}^1\}_{t=T_1+1}^T$  为观测结果. 然后, 通过分析如图 1 所示的实例, 为了刻画其异质性和不对称性, Cai et al. (2021) 考虑第一个单元的分位数处置效应, 定义为:

$$\Delta_\tau = q_{1\tau}^1 - q_{1\tau}^0,$$

其中  $q_{1\tau}^j$  是当  $j = 0$  和  $1$  和  $\tau \in (0, 1)$  时  $F_j(y) = P(y_{1t}^j \leq y)$  的第  $\tau$  分位. 为估计分位处置效果  $\Delta_\tau$ , 分别估计  $q_{1\tau}^1$  和  $q_{1\tau}^0$  即可. 由于第一个单元在  $t \geq T_1 + 1$  处置下的结果是可观测的, 所以很容易计算出  $\{y_{1t}^1\}_{t=T_1+1}^T$  的样本分位  $q_{1\tau}^1$ , 记为  $\hat{q}_{1\tau}^1$ . 估计分位数处置效果的困难是由于  $\{y_{1t}^0\}_{t=T_1+1}^T$  是不可观测的, 因此不能直接估计  $q_{1\tau}^0$ . 为了估计第一个单元的反事实分位数, 类似于 HCW 的 ATE 模型, Cai et al. (2021) 提出了一种利用条件 CDF 和无条件 CDF 之间的关系的新方法. 为了估计  $q_{1\tau}^0$ , 根据  $q_{1\tau}^0$  的定义, 我们有如下等式:

$$P(y_{1t}^0 \leq q_{1\tau}^0) = E[F_{y_{1t}^0|X_{2t}}(q_{1\tau}^0|X_{2t})] = \tau,$$

其中  $F_{y_{1t}^0|X_{2t}}(\cdot)$  表示给定  $X_{2t}$  下  $y_{1t}^0$  的条件分布函数. 所以, 通过矩估计的方法可以得到  $q_{1\tau}^0$

的估计量, 求解以下样本形式的方程:

$$\frac{1}{T_2} \sum_{t=T_1+1}^T F_{y_{1t}^0|X_{2t}}(\hat{q}_{1\tau}^0|X_{2t}) = \tau. \quad (8)$$

一般来说, 条件 CDF  $F_{y_{1t}^0|X_2}(y|x)$  是未知的, 因此上述  $q_{1\tau}^0$  的估计是不可行的. 为了从观测数据得到  $q_{1\tau}^0$  的可行估计, 首先需要利用可观测数据来估计  $F_{y_{1t}^0|X_{2t}}(y|x)$ . 为此, 我们需要做出以下假设, 这实际上, 这个假设类似于上文提到的 HCW 和相关参考文献中施加的均值设置的假设, 以及 Callaway et al. (2018) 中的分位数设置的假设 (见假设 Q1 和 Q2).

**假设 Q3** (条件 CDF 不变性) (i) 给定  $X_{1t}$  下的  $y_{1t}^1$  和给定  $X_{2t}$  下的  $y_{1t}^0$  条件 CDFs 的结构相同; 也就是说,  $F_{y_{1t}^1|X_{1t}}(\cdot) = F_{y_{1t}^0|X_{2t}}(\cdot) \equiv F(\cdot)$ , 以及 (ii)  $X_{2t}$  的支撑是  $X_{1t}$  支撑的一个子集.

假设 Q3 假定处置组和对照组之间存在某种结构不变性, 这确保了给定协变量, 处置组和对照组之间未处置的潜在结果的分布具有可比性. 在假设 Q3 下, 可以通过观测到的对照组数据来估计处置组的反事实条件 CDF. 这一假设与 Rothe (2010) 中的非参数结构模型的假设 1 基本上一致, 同时, 类似于 Hsu, Lai and Lieli (2020) 中的假设 2.3. 因此, 可以利用处置前的观测数据估计条件 CDF  $F(y|x)$ , 得到  $\hat{F}(y|x)$ . 为此, 可以使用一种加权的核估计方法, 在 Cai and Wang (2008) 中称为双核估计. 读者可以参考文献 Cai and Wang (2008). 然后将估计的条件 CDF 代入 (8) 中得到:

$$\frac{1}{T_2} \sum_{t=T_1+1}^T \hat{F}(\hat{q}_{1\tau}^0|X_{2t}) = \tau.$$

解上述方程得到  $q_{1\tau}^0$  的估计量. 最后, 第一个单元的分位处置效果的估计由  $\hat{\Delta}_\tau = \hat{q}_{1\tau}^1 - \hat{q}_{1\tau}^0$ . 在一定的条件下, 而且  $\hat{F}(y|x)$  满足一定的渐近性质, Cai et al. (2021) 得到了  $\hat{\Delta}_\tau$  的相合性和渐近正态性.

如果有很多控制单元和协变量个数的时候, 即  $X_{2t}$  的维数  $d_x = N - 1 + d_z$  较大, 对于估计 QTE 方法, 维数诅咒导致采用核回归方法估计条件 CDF  $F(\cdot|x)$  并不理想. 为了克服这一问题, 遵循 Hall and Yao (2005) 的思想, 可以采用指数方法  $\beta^\top x$  来估计  $F(\cdot|\beta^\top x)$ . 通过这种方式, 从统计角度来看, 该指数避免了所谓的维度诅咒, 因为它允许我们将多元问题简化单元, 其中可以实现上述单变量设置中的非参数方法, 因为  $\beta^\top x$  是单变量. 另外, 可以直接使用 Hall and Yao (2005) 提出的估计程序来估计  $\beta$  和  $F(\cdot|\beta^\top x)$ . 然而, 上述的估算方法在计算方面有些复杂. 为了便于计算方便, 另一种方法是通过分位数回归估计条件 CDF, 如 Koenker and Bassett (1978) 所述, 下面将详细描述. 令  $q_\tau(x)$  为  $F(y|x)$  的第  $\tau$  条件分位数, 使  $q_\tau(x) = F^{-1}(\tau|x) \equiv q(\tau, x)$ . 然后一个简单的计算可以得到条件 CDF 和条件分位数之间的关系,

$$F(y|x) = \int_0^1 I(q(u, x) \leq y) du \approx \varepsilon + \int_\varepsilon^{1-\varepsilon} I(q(u, x) \leq y) du, \quad (9)$$

对于一个很小的常数  $\varepsilon > 0$ . 通过假设当  $1 \leq t \leq T_1$  时给定  $X_{1t}$ ,  $y_{1t}$  的第  $\tau$  条件分位函数为  $q_\tau(x) = x^\top \beta(\tau)$ , 那么, 条件分位估计为  $\hat{q}_\tau(x) = x^\top \hat{\beta}(\tau)$ , 其中

$$\hat{\beta}(\tau) = \hat{\beta}_\tau = \arg \min_{\beta_\tau} \sum_{t=1}^{T_1} \rho_\tau(y_{1t} - \beta_\tau^\top X_{1t}). \quad (10)$$

且  $\rho_\tau(v) = v[\tau - I(v < 0)]$ . 因此, 根据 (9), 估计的条件 CDF  $\hat{F}(y|x)$  为:

$$\hat{F}(y|x) \approx \varepsilon + \sum_{j=1}^m \delta_j I(\hat{\beta}_{\tau_j}^\top x \leq y), \quad (11)$$

其中对于任意  $\varepsilon \leq \tau_0 < \dots < \tau_m \leq 1 - \varepsilon$ , 当  $m \rightarrow \infty$  时  $\delta_j = \tau_j - \tau_{j-1} \rightarrow 0$ , 由 (10) 可得  $\hat{\beta}_{\tau_j}$ . 显然, 在实际应用中, (11) 中的最后一个近似可能对  $\{\tau_j\}_{j=0}^m$  的选择很敏感. 不过, 上述想法在实际应用中通常使用, 如 Chernozhukov, Fernández-Val and Galichon (2010), Chernozhukov, Fernández-Val and Melly (2013).

最后, 不同于 Chen et al. (2013) 仅考虑 ATE, Cai et al. (2021) 运用上述所提出方法, 通过估计 QTE 来确定引入沪深 300 指数期货交易对中国股市的 VIX 及其 VVIX 可能产生的影响. 实证结果表明, 引入沪深 300 指数期货交易对我国股市 VIX 及其 VVIX 均有显著的影响.

**备注 3** 需要注意的是, Cai et al. (2021) 仅考虑一个处置, 该假设应该扩展到有多种处置的情况, 如 Fujiki and Hsiao (2015) 估算 ATE 的情况. 这种扩展似乎有助于评估宏观经济政策的有多个处置的效果, 如 Liu et al. (2020) 和其他应用. 此外, 与 Callaway et al. (2018) 类似, 人们可能会有兴趣考虑 (部分) 条件分位数处置效应 (CQTE), 比如,

$$\Delta_\tau(v) = q_\tau^1(v) - q_\tau^0(v),$$

其中  $v$  是  $X$  的一个部分, 这些扩展将作为未来的研究课题.

#### 4.2.2 控制元和协变量选择

如果  $d_x$  很大, Cai et al. (2021) 建议使用分位数回归加惩罚的方法来估计条件 CDF, 类似于 Li and Zhu (2008), Wu and Liu (2009). 具体来说, 与 (7) 类似, 在 (10) 中加入一个惩罚项, 变成如下:

$$\hat{\beta}_{\text{LASSO}}(\tau) = \arg \min_{\beta_\tau} \sum_{t=1}^{T_1} \rho_\tau(y_{1t} - \beta_\tau^\top X_{1t}) + \sum_{j=1}^{d_x} \psi_\lambda(\beta_{\tau,j}), \quad (12)$$

其中  $\psi_\lambda(\cdot)$  是某个罚函数, 如 Li and Zhu (2008) 中的绝对函数或 Wu and Liu (2009) 中的 SCAD 罚函数. 此方法优点之一在于计算简单, 可以通过使用由 Sherwood and Maidman (2016) 创立 R 语言 “rqPen” 包或 Yi and Huang (2017) 的 “hreg” 包, 轻松计算 (12) 中的  $\hat{\beta}_{\text{LASSO}}(\tau)$  值. 然后, 对带有惩罚的条件分布进行近似估计, 对于一个小常数  $\varepsilon > 0$ , 应该有如下表达式:

$$\hat{F}_{\text{LASSO}}(y|x) = \varepsilon + \int_\varepsilon^{1-\varepsilon} I(\hat{\beta}_{\text{LASSO}}(\tau)^\top x \leq y) d\tau,$$

其中  $\hat{\beta}_{\text{LASSO}}(\tau)$  由 (12) 给出.

**备注 4** 首先, 如果面板数据中时间序列是平稳的和  $\alpha$ -混合的, 根据 Chernozhukov, Fernández-Val and Galichon (2010) 命题 5, 可证明 (11) 式中  $\hat{F}(y|x)$  是相合的而且有渐近正态性, 如下:

$$\sqrt{T_1} \left[ \hat{F}(y|x) - F(y|x) \right] \xrightarrow{d} N(0, V(y|x)),$$

其中  $V(y|x) > 0$  是其渐近方差. 同样, 我们也可以遵循 Li and Zhu (2008) 的一个类似的主要结果来获得在相依数据的情况下,  $\hat{F}_{\text{LASSO}}(y|x)$  的渐近性质. 因此, 可以得到  $\hat{\Delta}_\tau$  的渐近性质. 最后, 使用 (12) 中所给出的惩罚分位数回归方法的另外一个优点就是, 通过一些简单的扩展, 甚至

可以处理超高维度 (ultrahigh dimension) 的截面和协变量大于时间序列长度 ( $d_x > T_1$ ) 的情况; 详情见 Wang, Wu and Li (2012), Sherwood and Maidman (2016), 以及 Yi and Huang (2017). 这个研究问题值得我们去进一步探讨.

## 5 结论与未来研究问题探讨

在许多应用领域, 特别是计量经济学, 定量评价经济政策或干预措施的效果是经济研究和政策研究的核心问题之一. 从我个人的角度来看, 本文对面板数据 (包括有限面板和大型面板) 的经济政策评估的最新进展进行了选择性回顾. 综上所述, 经济政策评价仍然是一个充满活力和挑战性的研究领域, 值得进一步研究. 毫无疑问的是, 在不久的将来这一领域将受到极大的关注. 例如, 其重要性和挑战性在于考虑其他方法, 如合成控制方法 (synthetic control method, SCM) 等; 参见 Wan, Xie and Hsiao (2018) 通过模拟来估算 ATE, 以便比较 HCW 方法和 SCM 方法, Cai et al. (2021) 的估算大型面板数据的 QTE 等方法.

随着时间的推移, 经济政策本质上是不同的, 因为同一类型的政策冲击在不同的时期或在不同的经济情景下实施, 相关政策制度的规模是不同的. 因此, 本文的部分讨论了通过控制其他策略场景来确定一个策略影响的一些潜在方法. 当然, 未来关于时变政策影响能否在这种框架下得到解决的研究是非常有趣的, 也会有一些有趣的应用. 除了前面提到的一些开放问题 (如上文所述), 用于估算面板数据的 ATE 和 QTE 的方法不需要倾向评分函数的说明, 以避免可能的错误假设. 另一个避免估计倾向值函数的有趣方法是使用模拟方法来估计反事实的结果, 例如陆昌, 刘祥和杨晓光 (2021) 所提出的方法. 此外, 这些方法之间的一些比较 (理论和实证) 是绝对需要的.

最后, 正如 Carvalho et al. (2018), Athey (2019), 洪永淼和汪寿阳 (2021), 和萧政 (2021) 所提到的, 计量经济学和统计结合人工智能, 机器学习 (ML) 和因果推断在一起进行综合研究, 是一个崭新的研究课题, 以及从 ML 和统计学或者计量经济学的角度提供见解和理论结果的文献, 如何对使用这些工具来估计面板数据的 ATE 和 QTE 的提供了一个很好的参考. 实际上, 对于监督性学习, 文献上已经有很多方法, 比如, 正则回归 (LASSO, 岭回归), 随机森林, 回归树, 支撑向量机器, 神经网络, 矩阵填充 (matrix completion and factorization), 深度学习 (deep learning), 梯度提升 (boosting), 以及模型平均 (model averaging techniques) 等等, 关于经济学和机器学习之间关系的细节, 参见 Athey (2018). 比如, Wager and Athey (2018) 成功地引进所谓因果森林 (causal forest), 是由许多因果树 (causal trees) 来加权而成, 来刻画不同因果树的各种异质性, 而 Carvalho et al. (2018) 提出了人工反事实方法 (ArCo) 来解决高维度面板数据问题. 进一步, 这个 ArCo 方法已经被 Masini and Medeiros (2019) 推广来处理高维度和非平稳 (固定时间趋势或者随机趋势) 问题, 以便估计反事实结果. 不过, 如何把上述提到的机器学习的方法应用到基于面板数据的 ATE 和 QTE 估计问题或者统计推断问题, 建立一套崭新的计量经济学/统计学的方法论和理论体系, 看来不是一件简单的事, 需要国内外学者共同努力去解决. 而解决这些问题对年轻学者来说的确是一个挑战, 希望国内学计量经济学或者统计学学者在这个领域能够开创出一片新的研究天地, 为国内学者在这个研究领域树立一个里程碑.

## 参 考 文 献

- 洪永淼, 汪寿阳, (2021). 大数据、机器学习与统计学: 挑战和机遇 [J]. 计量经济学报, 1(1): 17–35.
- Hong Y M, Wang S Y, (2021). Big Data, Machine Learning and Statistics: Challenges and Opportunities[J]. China Journal of Econometrics, 1(1): 17–35.
- 萧政, (2021). 大数据时代关于预测的几点思考 [J]. 计量经济学报, 1(1): 1–16.
- Hsiao C, (2021). Some Thoughts on Prediction in the Presence of Big Data[J]. China Journal of Econometrics, 1(1): 1–16.
- 陆昌, 刘详, 杨晓光, (2021). 基于反事实模拟的中国股市涨跌停板磁吸效应研究 [J]. 计量经济学报, 1(1): 217–232.
- Lu C, Liu Y, Yang X G, (2021). Magnet Effect of Price Limits in China's Stock Markets Based on Counterfactual Simulation[J]. China Journal of Econometrics, 1(1): 217–232.
- Athey S, (2019). The Impact of Machine Learning on Economics[M]// Agrawal A, Gans J, Goldfarb A. The Economics of Artificial Intelligence: An Agenda. Chicago: University of Chicago Press: 507–547.
- Bai C, Li Q, Ouyang M, (2014). Property Taxes and Home Prices: A Tale of Two Cities[J]. Journal of Econometrics, 180(1): 1–15.
- Bai J, (2009). Panel Data Models with Interactive Fixed Effects[J]. Econometrica, 77(4): 1229–1279.
- Bai J, Ng S, (2002). Determining the Number of Factors in Approximate Factor Models[J]. Econometrica, 70(1): 191–221.
- Cai Z, Fang Y, Lin M, Zhan M, (2021). Estimating Quantile Treatment Effects for Panel Data[R]. Working Paper, Department of Economics, University of Kansas.
- Cai Z, Wang X, (2008). Nonparametric Estimation of Conditional VaR and Expected Shortfall[J]. Journal of Econometrics, 147(1): 120–130.
- Callaway B, Li T, (2019). Quantile Treatment Effects in Difference in Differences Models with Panel Data[J]. Quantitative Economics, 10(4): 1579–1618.
- Callaway B, Li T, Oka T, (2018). Quantile Treatment Effects in Difference in Differences Models Under Dependence Restrictions and With Only Two Time Periods[J]. Journal of Econometrics, 206(2): 395–413.
- Callaway B, Sant'Anna P H C, (2020). Difference-in-Differences With Multiple Time Periods[J]. Journal of Econometrics. doi: <https://doi.org/10.1016/j.jeconom.2020.12.001>.
- Carvalho C, Masini R, Medeiros M C, (2018). ArCo: An Artificial Counterfactual Approach for High Dimensional Panel Time Series Data[J]. Journal of Econometrics, 207(2): 352–380.
- Cerulli G. (2015). Econometric Evaluation of Socio-economic Programs[M]. Berlin: Springer.
- Chen H, Han Q, Li Y, Wu K, (2013). Does Index Futures Trading Reduce Volatility in the Chinese Stock Market? A Panel Data Evaluation Approach[J]. Journal of Futures Markets, 33(12): 1167–1190.
- Chernozhukov V, Fernández-Val I, Galichon A, (2010). Quantile and Probability Curves Without Crossing[J]. Econometrica, 78(3): 1093–1125.
- Chernozhukov V, Fernández-Val I, Melly B, (2013). Inference on Counterfactual Distributions[J]. Econometrica, 81(6): 2205–2268.
- De Chaisemartin C, D'Haultfoeuille X, (2020). Two-way Fixed Effects Estimators with Heterogeneous Treatment Effects[J]. Econometrica, 110(9): 2964–2996.
- Du Z, Zhang L, (2015). Home Purchase Restriction, Property Tax and Housing Price in China: A Counterfactual Analysis[J]. Journal of Econometrics, 188(2): 558–568.

- Fang Y, Tang S, Cai Z, Lin M, (2020). An Alternative Test for Conditional Unconfoundedness Using Auxiliary Variables[J]. *Economics Letters*, 194: 109320.
- Fujiki H, Hsiao C, (2015). Disentangling the Effects of Multiple Treatments — Measuring the Net Economic Impact of the 1995 Great Hanshin-awaji Earthquake[J]. *Journal of Econometrics*, 186(1): 66–73.
- Hall P, Yao Q, (2005). Approximating Conditional Distribution Functions Using Dimension Reduction[J]. *Annals of Statistics*, 33(3): 1404–1421.
- Henderson D J, Sperlich S, (2021). Model-free Difference-in-Differences With Confounders[R]. Working Paper, Department of Economics, Finance and Legal Studies, University of Alabama.
- Hsiao C, Ching S H, Wan K S, (2012). A Panel Data Approach for Program Evaluation: Measuring the Benefits of Political and Economic Integration of Hong Kong With Mainland China[J]. *Journal of Applied Econometrics*, 27(5): 705–740.
- Hsu Y C, Lai T C, Lieli R P, (2020). Counterfactual Treatment Effects: Estimation and Inference[J]. *Journal of Business & Economic Statistics*. doi: <https://doi.org/10.1080/07350015.2020.1800479>.
- Huang D, Schlag C, Shaliastovich I, Thimme J, (2019). Volatility-of-Volatility Risk[J]. *Journal of Financial and Quantitative Analysis*, 54(6): 2432–2452.
- Imai K, Kim I S, (2019). When Should We Use Unit Fixed Effects Regression Models for Causal Inference With Longitudinal Data?[J]. *American Journal of Political Science*, 63(2): 467–490.
- Imai K, Kim I S, (2020). On the Use of Two-way Fixed Effects Regression Models for Causal Inference With Panel Data[J]. *Political Analysis*. doi: <https://doi.org/10.1017/pan.2020.33>.
- Imbens G W, Wooldridge J M, (2009). Recent Developments in the Econometrics of Program Evaluation[J]. *Journal of Economic Literature*, 47(1): 5–86.
- Ke X, Chen H, Hong Y, Hsiao C, (2017). Do China's High-speed-rail Projects Promote Local Economy? — New Evidence From a Panel Data Approach[J]. *China Economic Review*, 44(1): 203–226.
- Kim Y, Choi H, Oh H S, (2008). Smoothly Clipped Absolute Deviation on High Dimensions[J]. *Journal of the American Statistical Association*, 103(484): 1665–1673.
- Koenker R, Bassett G, (1978). Regression Quantiles[J]. *Econometrica*, 46(1): 33–50.
- Li K T, Bell D R, (2017). Estimation of Average Treatment Effects With Panel Data: Asymptotic Theory and Implementation[J]. *Journal of Econometrics*, 197(1): 65–75.
- Li Y, Zhu J, (2008).  $L_1$ -Norm Quantile Regression[J]. *Journal of Computational and Graphical Statistics*, 17(1): 163–185.
- Liu Z, Cai Z, Fang Y, Lin M, (2020). Statistical Analysis and Evaluation of Macroeconomic Policies: A Selective Review[J]. *Applied Mathematics — A Journal of Chinese Universities, Series B*, 35(1): 57–83.
- Masini R, Medeiros M C, (2019). Counterfactual Analysis with Artificial Controls: Inference, High Dimensions and Nonstationarity[R]. Working Paper, Sao Paulo School of Economics, Getulio Vargas Foundation.
- Pesaran M H, (2006). Estimation and Inference in Large Heterogeneous Panels With a Multifactor Error Structure[J]. *Econometrica*, 74(4): 967–1012.
- Ouyang M, Peng Y, (2015). The Treatment Effect Estimation: A Case Study of the 2008 Economic Stimulus Package of China[J]. *Journal of Econometrics*, 188(2): 545–557.
- Rothe C, (2010). Nonparametric Estimation of Distributional Policy Effects[J]. *Journal of Econometrics*, 155(1): 56–70.



- Sherwood B, Maidman A, (2016). RqPen: Penalized Quantile Regression[J]. R package version, 4: 1. <https://cran.r-project.org/web/packages/rqPen/index.html>.
- Sun L, Abraham S, (2020). Estimating Dynamic Treatment Effects in Event Studies With Heterogeneous Treatment Effects[J]. *Journal of Econometrics*. doi: <https://doi.org/10.1016/j.jeconom.2020.09.006>.
- Tang S, (2020). Some Recent Developments in Modeling Quantile Treatment Effects[J]. *Applied Mathematics — A Journal of Chinese Universities, Series B*, 35(2): 220–243.
- Wager S, Athey S, (2018). Estimation and Inference of Heterogeneous Treatment Effects Using Random Forests[J]. *Journal of the American Statistical Association*, 113(523): 1228–1242.
- Wan S K, Xie Y, Hsiao C, (2018). Panel Data Approach vs Synthetic Control Method[J]. *Economics Letters*, 164(3): 121–123.
- Wang L, Wu Y, Li R, (2012). Quantile Regression for Analyzing Heterogeneity in Ultra-high Dimension[J]. *Journal of the American Statistical Association*, 107(497): 214–222.
- Wu Y, Liu Y, (2009). Variable Selection in Quantile Regression[J]. *Statistica Sinica*, 19(2): 801–817.
- Yi C, Huang J, (2017). Semismooth Newton Coordinate Descent Algorithm for Elastic-net Penalized Huber Loss Regression and Quantile Regression[J]. *Journal of Computational and Graphical Statistics*, 26(3): 547–557. <https://cran.r-project.org/web/packages/hqreg/index.html>.