

面板数据分位数处理效应的估计

献给林正炎教授 85 寿辰

蔡宗武¹, 方颖^{2,3}, 林明^{2,3*}, 詹铭峰⁴

1. Department of Economics, University of Kansas, Lawrence, KS 66045, USA;

2. 厦门大学王亚南经济研究院, 厦门 361005;

3. 厦门大学经济学院, 厦门 361005;

4. 首都经济贸易大学统计学院, 北京 100070

E-mail: caiz@ku.edu, yifst1@xmu.edu.cn, linming50@xmu.edu.cn, m.zhan@cueb.edu.cn

收稿日期: 2023-10-08; 接受日期: 2025-02-07; 网络出版日期: 2025-03-18; * 通信作者

国家自然科学基金(批准号: 72033008, 72133002 和 71988101) 和首都经济贸易大学新入职青年教师科研启动基金(批准号: XRZ2023077) 资助项目

摘要 本文提出一种适用于面板数据的分位数处理效应模型, 以刻画处理的分布效应。本文还提出识别性条件并估计被处理个体的反事实分位数, 证明所提出的分位数处理效应估计量的渐近性质。同时, 本文考虑高维的情形, 讨论如何通过 LASSO (least absolute shrinkage and selection operator) 方法选择对照个体和协变量。本文通过数值模拟来说明所提出的模型和方法的有限样本性质。最后, 本文的方法被用于估计引入沪深 300 指数期货交易对中国股市的对数收益率和波动率的分位数处理效应。

关键词 LASSO 方法 面板数据 非参数估计 分位数回归 处理效应

MSC (2020) 主题分类 62F12, 62P20, 62G05

1 引言

自从 Hsiao 等 [26] 的开创性工作以来, 利用面板数据方法(panel data approach, PDA) 估计平均处理效应(average treatment effect, ATE) 在实证经济学中得到了广泛的应用。他们提出通过利用所有个体公因子结构驱动的横截面依赖性来构建被处理个体的反事实结果。在假设政策干预不会影响对照组结果的识别假设下, 他们通过简单的线性回归方法估计了反事实结果, 而无需估计因子和因子载荷。此外, 他们的方法(以下简称 HCW(Hsiao-Chin-Wan) 方法) 具有最少数据需求的优势, 即反事实结果可以使用横截面个体的结果变量来估计, 而不一定需要使用其他协变量, 这在一些实证应用中是一个非常理想的特性。例如, 他们仅利用 22 个国家和地区的季度实际 GDP 增长率估计了香港与内地的政治经济一体化对经济增长的影响。此后, Bai 等 [4]、Ouyang 和 Peng [39] 以及 Li 和 Bell [34] 放宽了 Hsiao

英文引用格式: Cai Z W, Fang Y, Lin M, et al. Estimating quantile treatment effects for panel data (in Chinese). Sci Sin Math, 2026, 56: 1~22, doi: [10.1360/SSM-2023-0271](https://doi.org/10.1360/SSM-2023-0271)

等 [26] 隐含的一些假设; 在选择对照个体时, Li 和 Bell [34] 以及 Carvalho 等 [11] 建议在对照个体数量较大时使用 LASSO (least absolute shrinkage and selection operator) 类型的方法. 此外, Li 和 Bell [34] 推导出了受处理个体 ATE 估计量的渐近分布, 而 Carvalho 等 [11] 则提供了 LASSO 类型 ATE 估计量的理论证明.

凭借诸如假设条件弱、最少数据需求和易于实施等理想的特性, Hsiao 等 [26] 的 PDA 方法越来越受到政策评估文献的关注. 例如, 一些有影响力的应用, 包括但不限于, Chen 等 [12] 研究了引入股指期货交易对中国股市现货价格波动的影响. Bai 等 [4] 探讨了财产税对房价的影响. Fujiki 和 Hsiao [21] 衡量了 1995 年日本阪神大地震的净经济影响. Du 和 Zhang [16] 评估了购房限制的影响以及中国征收房产税对房价的影响. Ouyang 和 Peng [39] 研究了 2008 年中国经济刺激计划的宏观经济影响. Bove 等 [6] 估计了内战对经济的影响. Ke 等 [30] 探讨了中国高铁项目对目标城市经济增长的影响. Li 和 Long [35] 研究了 1995 年 1 月 1 日在美国弗吉尼亚州颁布的司法改革的效应. Carvalho 等 [11] 评估了巴西实施的反逃税计划对通货膨胀和其他宏观经济变量的影响. 最近, Ke 和 Hsiao [31] 应用 Hsiao 等 [26] 的 PDA 方法来评估在全球经济活动受到限制期间, 湖北省的管控政策对经济带来的影响.

然而, Hsiao 等 [26] 的方法是为估计 ATE 而设计的, 在某些应用中 ATE 可能不足以充分刻画政策效应, 特别是当结果分布不对称、异质或厚尾时. 例如, 期货交易的引入如何影响现货股票波动性 (volatility index, VIX) 是一个重要且仍存在很大争议的问题. Chen 等 [12] 使用 Hsiao 等 [26] 的 PDA 方法估计了引入股指期货交易对股票 VIX 的影响. 但是, 大量的文献表明现货股票 VIX 的分布是倾斜且厚尾的, 如图 1 所示. 该图给出了 2002 年 1 月至 2021 年 2 月中国股市沪深 300 指数的引入股指期货交易前、引入股指期货交易后和全样本 VIX 的分布密度估计. 从该图中可以看出, VIX 的分布具有明显的不对称性和厚尾性, 这些现象也可以从第 4 节的最后一个表中得到有力的支持. 为了正确评估引入期货交易对股票 VIX 的影响, 本文将研究政策对感兴趣的结果变量分布的影响, 而不仅仅是对平均值的影响.

受上述实证问题的启发, 本文考虑通过分位数处理效应 (quantile treatment effect, QTE) 来描述政策对观测的结果变量分布的影响. 尽管关于 QTE 的识别和估计的文献越来越多, 如文献 [7, 14, 15, 19, 20, 27, 41] 等, 但是正如 Cai [7] 所指出的, 文献中关于面板数据的 QTE 估计方面所作的努力非常有

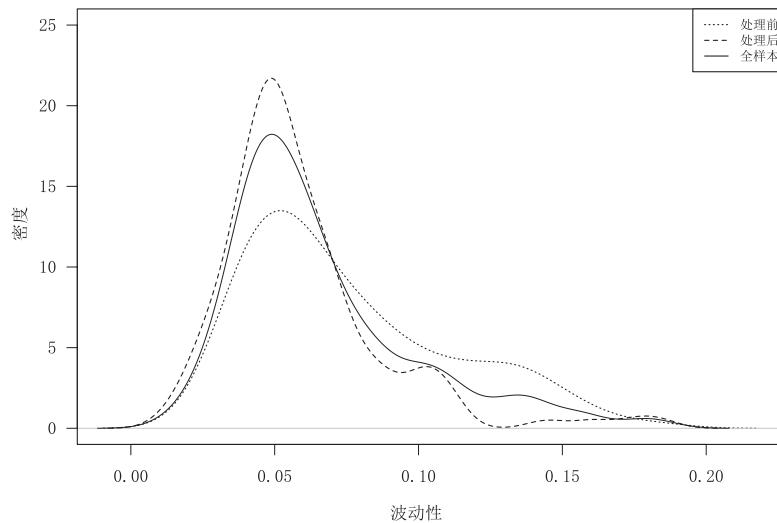


图 1 沪深 300 指数处理前、处理后和全样本月度波动率的密度估计图

限. Callaway 等 [10] 以及 Callaway 和 Li [9] 都在双重差分的设定下考虑了 QTE 的估计问题. 据我们所知, 目前还没有考虑使用文献 [26] 中的大型面板数据估计 QTE 的文献. 本文的贡献在于将 HCW 方法推广到对面板数据的 QTE 的估计, 以获得对分布处理效应的了解. 为了估计处理个体的反事实分位数, 与 Hsiao 等 [26] 的做法不同, 本文引入条件累积分布函数 (cumulative distribution function, CDF) 不变性的假设, 利用条件 CDF 和无条件 CDF 之间关系, 首先估计出处理个体反事实结果的条件分布函数, 再进一步估计出处理个体反事实结果的无条件分布函数, 以及其所对应的反事实分位数. 针对不同的控制组个体数量和协变量维数, 本文分别考虑非参数方法 (维数比较低情形)、半参数方法 (降维方法) 和带惩罚项的半参数方法 (变量选择) 来估计处理个体反事实结果的条件分布. 此外, 本文还推导出本文所提出的 QTE 估计量的大样本性质, 以及提出基于分块 Bootstrap 构建置信区间的方法.

正如 Xiao 和 Koenker [45] 所指出的, 分位数估计是金融应用中现代风险管理的重要组成部分. 因此, QTE 技术可用于评估某些新的金融政策是否对金融市场或机构的 VIX 和波动率的波动率 (volatility-in-volatility, VVIX)¹⁾ 产生重大影响. 实际上, 本文的动机是通过 QTE 来考察中国金融期货交易所于 2010 年 4 月 16 日正式推出的沪深 300 指数期货交易对现货市场 VIX 及其 VVIX 是否产生影响. 在相关政策出台后, 有人批评股指期货交易的推出可能会因过度投机而动摇现货市场; 也有人认为股指期货市场可以提高信息流动的速度和质量, 使金融市场更加完善. 通过使用所提出的建模方法, 本文将估计引入沪深 300 指数期货交易对中国股市 VIX 和 VVIX 的影响. 详细的分析结果在第 4 节中陈述.

本文余下内容的安排如下. 第 2 节首先简要回顾有关面板数据处理效应研究的文献, 然后描述所提出模型的设定并给出识别条件. 此外, 本节讨论估计面板数据 QTE 的方法, 并建立估计量的渐近性质, 以及构建基于分块 Bootstrap 的置信区间. 最后, 本节还讨论基于分位数回归 (quantile regression, QR) 方法和降维方法来选择对照个体和协变量的估计方法. 第 3 节进行数值模拟研究, 用于评估所提出的估计量的有限样本性质. 第 4 节将所提出的方法用于估计引入沪深 300 指数期货交易对中国股票市场的对数收益率和波动率的 QTE. 第 5 节给出本文结论.

2 面板数据的分位数处理效应

2.1 HCW 方法

本小节介绍 HCW 方法的框架. 分别用 y_{it}^1 和 y_{it}^0 表示第 i 个个体在第 t 期接受处理和没有接受处理的结果, 其中 $1 \leq i \leq N, 1 \leq t \leq T$. 第 i 个个体在第 t 期的处理效应定义为 $\Delta_{it} = y_{it}^1 - y_{it}^0$. 由于 y_{it}^1 和 y_{it}^0 无法同时被观测到, 实际观测结果 y_{it} 由下式给出:

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

其中, 若第 i 个个体在第 t 期接受处理, 则 $d_{it} = 1$; 否则 $d_{it} = 0$. HCW 方法关注这样的情形: 面板数据中只有一个处理个体, 且该个体在 $T_1 + 1$ 及其以后的时期接受处理, 其中 $1 < T_1 < T$. 不失一般性, 假设接受处理的个体是第一个个体, 即第 $j = 2, \dots, N$ 个个体在所有的 $t = 1, \dots, T$ 期都不接受处理, 而第 1 个个体在 $t = 1, \dots, T_1$ 期也没有接受处理, 处理仅发生在第 1 个个体的 $t = T_1 + 1, \dots, T$ 期.

1) 波动率的波动率的定义可以在文献 [13, 25, 28] 中找到. 事实上, 与 VIX 类似, Hollstein 和 Prokopcuk [25] 以及 Chen 等 [13] 提供了一些经验证据, 表明市场 VVIX 可以预测市场回报并驱动随时间变化的波动率风险.

Hsiao 等 [26] 使用以下因子结构对所有个体的横截面依赖性进行建模:

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

其中, α_i 是第 i 个个体的截距项, b_i 是 $K \times 1$ 维的因子载荷向量, f_t 是 $K \times 1$ 维 (不可观察) 的公共因子, u_{it} 是一个零均值弱相关且平稳的误差项. 如果 T 和 N 都很大, 则可以采用 Bai 和 Ng [5] 的方法或 Pesaran [40] 的方法来估计公因子 f_t . 在 T 和 N 都不大的情形下, Hsiao 等 [26] 提出了一种新颖的方法, 不直接估计公因子 f_t , 而利用 $\tilde{y}_t = (y_{2t}, \dots, y_{Nt})^\top$ 代替 f_t 来估计处理个体在接受处理后的反事实结果 y_{1t}^0 , $t = T_1 + 1, \dots, T$. 具体而言, 在一定的条件下, Hsiao 等 [26] 根据以下回归模型对反事实结果 y_{1t}^0 进行估计:

$$y_{1t}^0 = \beta^\top x_t + u_{1t}, \quad t = 1, \dots, T_1,$$

其中, $x_t = (1, y_{2t}, \dots, y_{Nt})^\top$, $\beta = (\beta_1, \dots, \beta_N)^\top$ 为未知系数, u_{1t} 是零均值且方差有限的误差项. 这样, 利用普通最小二乘法可以得到未知系数 β 的估计量

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

通过假设处理前后数据关系的不变性及其他假设, 详见文献 [26], 反事实结果 y_{1t}^0 , $t = T_1 + 1, \dots, T$ 的预测值可以通过下式估计:

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

最后, 第一个个体的平均处理效应 $\Delta_1 = E(\Delta_{1t}) = E(y_{1t}^1 - y_{1t}^0)$ 的估计量通过对处理后各期观测结果和估计的反事实结果之间的差值的平均来构建

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

其中, $T_2 = T - T_1$. 此外, Hsiao 等 [26] 建议使用 Akaike [2] 的 Akaike 信息准则 (Akaike information criterion, AIC) 或 Hurvich 和 Tsai [29] 的 AICC (corrected Akaike information criterion) 模型选择方法来检验是否需要在该方法中使用所有的横截面个体. Li 和 Bell [34] 以及 Carvalho 等 [11] 提出使用 LASSO 方法替代 AIC 或 AICC 方法来选择 HCW 方法中的对照个体.

2.2 分位数处理效应的模型识别

众所周知, 如果潜在结果的分布是不对称的或厚尾的, ATE 并非是评估处理效应的一个理想指标. 为了掌握某个处理的全局效应, 就要研究处理的分布效应. 本节的重点是提出基于面板数据的 QTE 的模型及识别.

本文关于面板数据的设置与文献 [26] 相同. 假设面板数据集为 $\{(y_{it}, z_t); 1 \leq i \leq N, 1 \leq t \leq T\}$, 其中, z_t 是 $d_z \times 1$ 维的协变量²⁾. 此外, 这里假设面板数据中的时间序列是严平稳的. 对面板数据中的第一个个体, 从 $t = T_1 + 1$ 期到 $t = T$ 期接受处理, 而在 $T_1 + 1$ 期之前不接受处理. 其余的 $N - 1$ 个个体作为对照组, 在整个过程中都不接受处理. 令 $T_2 = T - T_1$. 为简单起见, 令 $\lim_{T \rightarrow \infty} T_2/T_1 = c$, 其中, $c > 0$ 是一个常数. 因此, $\lim_{T \rightarrow \infty} T_1/T = \lambda$, $\lim_{T \rightarrow \infty} T_2/T = 1 - \lambda$, 其中, $\lambda = 1/(1 + c)$. 为了表示方便, 将面板数据分为以下 4 个部分:

2) 例如, 在本文第 4 节的实证研究中, 协变量包括月度 CPI (consumer price index) 增速、月度 M1 增速和月度 M2 增速等 3 个宏观经济变量.

\$Y_1\$	\$X_1\$
\$Y_2\$	\$X_2\$

其中, $Y_1 = \{y_{1t}, t = 1, \dots, T_1\}$ 表示从 $t = 1$ 期到 $t = T_1$ 期的第一个个体的观测结果, $X_1 = \{(y_{it}, z_t); i = 2, \dots, N, t = 1, \dots, T_1\}$ 表示从 $t = 1$ 期到 $t = T_1$ 期的其余个体的观测结果以及协变量, $Y_2 = \{y_{1t}, t = T_1 + 1, \dots, T\}$ 表示从 $t = T_1 + 1$ 期到 $t = T$ 第一个个体的观测结果, $X_2 = \{(y_{it}, z_t); i = 2, \dots, N, t = T_1 + 1, \dots, T\}$ 表示从 $t = T_1 + 1$ 期到 $t = T$ 期的其余个体的观测结果和协变量. 令 $Y_2^0 = \{y_{1t}^0, t = T_1 + 1, \dots, T\}$ 为 Y_2 的反事实结果. 为了方便起见, 观测结果 Y_2 也表示为 $Y_2^1 = \{y_{1t}^1, t = T_1 + 1, \dots, T\}$. 在处理期, 第一个个体的 QTE 定义为

$$\Delta_\tau = q_{1\tau}^1 - q_{1\tau}^0, \quad (2.1)$$

其中, $q_{1\tau}^j$ 表示分布函数 $F_j(y) := P(y_{1t}^j \leq y), t > T_1$ 的 τ 分位数, $j = 0, 1, \tau \in (0, 1)$. 为了估计 Δ_τ , 需要分别估计 $q_{1\tau}^1$ 和 $q_{1\tau}^0$. 由于在 $t > T_1$ 期第一个个体受处理的结果是可观测的, 可以用 Y_2 的样本分位数(记为 $\hat{q}_{1\tau}^1$)对 $q_{1\tau}^1$ 进行估计. 估计 QTE 的困难在于第一个个体的反事实结果 Y_2^0 是不可观测的. 与文献 [26] 中的因子建模思想不同, 本文提出利用条件 CDF 和非条件 CDF 之间的关系来估计第一个个体反事实结果的分位数 $q_{1\tau}^0$, 具体如下. 将 $q_{1\tau}^0$ 表示为

$$q_{1\tau}^0 = \inf\{y : F_{Y_2^0}(y) \geq \tau\} = \inf\{y : E[F_{Y_2^0|X_2}(y | X_{2t})] \geq \tau\},$$

其中, $F_{Y_2^0}(\cdot)$ 是 $t > T_1$ 时 y_{1t}^0 的 CDF, $F_{Y_2^0|X_2}(\cdot | \cdot)$ 是 $t > T_1$ 时给定 X_{2t} 下 y_{1t}^0 的条件 CDF. 利用上式可以给出如下 $q_{1\tau}^0$ 的估计量:

$$\bar{q}_{1\tau}^0 = \inf g \left\{ y : \frac{1}{T_2} \sum_{t=T_1+1}^T F_{Y_2^0|X_2}(y | X_{2t}) \geq \tau \right\}. \quad (2.2)$$

一般情形下, 条件 CDF $F_{Y_2^0|X_2}(y | x)$ 是未知的, 上述的 $q_{1\tau}^0$ 估计量无法直接使用. 为了从观测数据中获得 $q_{1\tau}^0$ 的可行估计, 首先需要估计 $F_{Y_2^0|X_2}(y | x)$. 为此, 本文引入如下不变性识别假设, 这类似于 Hsiao 等 [26] 在 ATE 设置中施加的假设, 以及 Callaway 等 [10] 关于分位数设置施加的假设.

假设 2.1 $Y_1 | X_1$ 和 $Y_2^0 | X_2$ 的条件 CDF 保持不变, 即 $F_{Y_1|X_1}(\cdot | \cdot) = F_{Y_2^0|X_2}(\cdot | \cdot) \equiv F(\cdot | \cdot)$.

假设 2.1 假定了面板数据中的某种结构不变性, 该假设保证了在给定对照组结果和协变量的情形下, 第一个个体未受处理的潜在结果的条件分布在处理前和处理后保持相同. Rothe [41] 在数据独立同分布的设定下作出了类似的假设. 在该假设下, 可以利用处理前观测到的数据来估计第一个个体处理后的反事实结果的条件 CDF. 由于该假设涉及反事实结果, 无法通过观测数据直接验证. 如果能找到一组符合文献 [8, 假设 2] 的工具变量, 则可以用 Cai 等 [8] 提出的方法对假设 2.1 进行检验. 在实际应用中, 如何寻找符合条件的工具变量也是一个具有挑战性的任务.

2.3 估计方法

根据假设 2.1 中的识别条件, 如果 X_{1t} 的维数 $d_x = N - 1 + d_z$ 不是很大, 则可以使用核方法来估计 $F(y | x)$. 具体而言, $F(y | x)$ 可以使用处理前的观测数据进行估计, 如下所示:

$$\tilde{F}(y | x) = \frac{\sum_{t=1}^{T_1} I(Y_{1t} \leq y) K_h(X_{1t} - x)}{\sum_{t=1}^{T_1} K_h(X_{1t} - x)}, \quad (2.3)$$

其中, $I(\cdot)$ 是示性函数, $K_h(x) = h^{-d_x} K((x - \bar{x})/h)$, $K(\cdot)$ 是假设 2.4 中定义的高阶核函数³⁾, h

³⁾ 有关高阶核定义的详细信息, 可参见文献 [22].

是窗宽. 将估计的条件 CDF 代入 (2.2), 可以得到

$$\tilde{q}_{1\tau}^0 = \inf \left\{ y : \frac{1}{T_2} \sum_{t=T_1+1}^T \tilde{F}_{Y_2^0|X_2}(y | X_{2t}) \geq \tau \right\} = \inf \{y : \tilde{F}_{Y_2^0}(y) \geq \tau\},$$

其中,

$$\tilde{F}_{Y_2^0}(y) := \frac{1}{T_2} \sum_{t=T_1+1}^T \tilde{F}_{Y_2^0|X_2}(y | X_{2t}).$$

注意, 当 (2.3) 中使用高阶核时, 函数 $\tilde{F}_{Y_2^0}(y)$ 可能是非单调的或取区间 $[0, 1]$ 之外的值. 这种情形可以使用 Rothe [41] 的再加权的方法或 Hsu 等 [27] 法人单调化的方法将 $\tilde{F}_{Y_2^0}(y)$ 转换为单调非递减 CDF. 这里, 本文参照 Hsu 等 [27] 的作法, 令

$$\hat{F}_{Y_2^0}(y) = \frac{\sup_{u \leq y} \tilde{F}_{Y_2^0}(u)}{\sup_{-\infty < u < \infty} \tilde{F}_{Y_2^0}(u)},$$

则 $\hat{F}_{Y_2^0}(y)$ 以概率 1 是一个 CDF. 从而, (2.1) 中第一个个体的 QTE 可以由下式进行估计:

$$\hat{\Delta}_\tau = \hat{q}_{1\tau}^1 - \hat{q}_{1\tau}^0, \quad (2.4)$$

其中, $\hat{q}_{1\tau}^0 = \inf \{y : \hat{F}_{Y_2^0}(y) \geq \tau\}$.

2.4 演近理论

本小节将推导 (2.4) 中 QTE 估计量的演近理论. 为了建立演近理论, 本文引入如下假设.

假设 2.2 面板数据中的时间序列 $\{(y_{it}, z_t); 1 \leq i \leq N, 1 \leq t \leq T\}$ 是严平稳且是 α -混合的. α -混合系数满足 $\alpha(s) = O(s^{-\varepsilon_0})$, 其中, ε_0 是某个大于 1 的常数. 另外, 假设 $\lim_{T \rightarrow \infty} T_2/T_1 = c$, $0 < c < \infty$, $\lambda = \lim_{T \rightarrow \infty} T_2/T \in (0, 1)$.

- 假设 2.3** (i) X_{1t} 和 X_{2t} 的支撑集 (分别记为 \mathcal{X}_1 和 \mathcal{X}_2) 是紧的, 并且满足 $\mathcal{X}_2 \subset \mathcal{X}_1$.
(ii) X_{1t} 的密度函数 (记为 $f_{X_1}(x)$) 在 \mathcal{X}_1 上是一致连续的并且下界不为零.
(iii) 对于任意 $x \in \mathcal{X}_2$, $F(y | x)$ 关于 y 连续, 并且其条件密度函数 $f(y | x)$ 有界.

假设 2.4 核函数 $K(\cdot)$ 满足如下条件:

- (i) $K(u)$ 的支撑集是有界的;
- (ii) $\int K(u)du = 1$;
- (iii) 对于某个 $\ell > d_x$, 当 $1 \leq |j| \leq \ell$ 时, $\int u^j K(u)du = 0$, 并且当 $|j| = \ell$ 时, $\int |u^j K(u)|du < \infty$, 其中, $j = (j_1, \dots, j_{d_x})$ 是 d_x 维的非负整数向量, $|j| := j_1 + \dots + j_{d_x}$, $u^j := \prod_{k=1}^{d_x} u_k^{j_k}$;
- (iv) $K(u)$ 是 ℓ 次可导的, 其导数是一致连续且有界的.

假设 2.5 对于 $j = 0, 1$, $\{y_{1t}^j, t > T_1\}$ 的密度函数 $f_{Y_2^j}(y)$ 在该分布的分位数点 $q_{1\tau}^j$ 处连续且严格大于零.

假设 2.6 (i) 密度函数 $f_{X_j}(x)$ 在 \mathcal{X}_j 的内部是 ℓ 次可微的, 其偏导数是一致连续且有界的, $j = 1, 2$;

- (ii) 条件 CDF $F(y | x)$ 在 \mathcal{X}_1 的内部关于 x 是 ℓ 次可微的, 其偏导数是一致连续且有界的.

假设 2.7 当 $T \rightarrow \infty$ 时, $T^{1/2}h^{d_x}/\log(T) \rightarrow \infty$ 且 $T^{1/2}h^\ell \rightarrow 0$.

在介绍本文的主要定理之前, 先证明两个引理.

引理 2.1 在假设 2.1–2.7 下, 对 $y \in (-\infty, \infty)$, 有

$$\begin{aligned} \sqrt{T_2}[\tilde{F}_{Y_2^0}(y) - F_{Y_2^0}(y)] &= \frac{1}{\sqrt{T_2}} \sum_{t=T_1+1}^T [F(y | X_{2t}) - F_{Y_2^0}(y)] \\ &\quad + \frac{1}{\sqrt{T_1}} \sum_{t=1}^{T_1} \sqrt{\frac{T_2}{T_1}} \cdot \frac{f_{X_2}(X_{1t})}{f_{X_1}(X_{1t})} \cdot [I(Y_{1t} \leq y) - F(y | X_{1t})] + o_p(1). \end{aligned}$$

证明 我们有

$$\begin{aligned} &\sqrt{T_2}[\tilde{F}_{Y_2^0}(y) - F_{Y_2^0}(y)] \\ &= \sqrt{T_2} \left\{ \frac{1}{T_2} \sum_{t=T_1+1}^T [\tilde{F}(y | X_{2t}) - F(y | X_{2t})] + \frac{1}{T_2} \sum_{t=T_1+1}^T [F(y | X_{2t}) - F_{Y_2^0}(y)] \right\} \\ &= \sqrt{T_2} \left\{ \frac{1}{T_2} \sum_{t=T_1+1}^T [\tilde{F}(y | X_{2t}) - F(y | X_{2t})] - \int [\tilde{F}(y | x) - F(y | x)] dF_{X_2}(x) \right\} \\ &\quad + \sqrt{T_2} \left[\int \tilde{F}(y | x) dF_{X_2}(x) - \int F(y | x) dF_{X_2}(x) \right] + \frac{1}{\sqrt{T_2}} \sum_{t=T_1+1}^T [F(y | X_{2t}) - F_{Y_2^0}(y)] \\ &=: S_1 + S_2 + \frac{1}{\sqrt{T_2}} \sum_{t=T_1+1}^T [F(y | X_{2t}) - F_{Y_2^0}(y)]. \end{aligned}$$

参见文献 [41, 引理 1] 的证明, 可以得到 $S_1 = o_p(1)$. 考虑 S_2 , 其可被分解为

$$\begin{aligned} S_2 &= \sqrt{T_2} \left[\int \tilde{F}(y | x) dF_{X_2}(x) - \int F(y | x) dF_{X_2}(x) \right] \\ &= \sqrt{T_2} \left[\int \frac{\frac{1}{T_1} \sum_{t=1}^{T_1} K_h(X_{1t} - x) I(Y_{1t} \leq y)}{\tilde{f}_{X_1}(x)} dF_{X_2}(x) - \int F(y | x) dF_{X_2}(x) \right] \\ &= \sqrt{T_2} \left[\int \frac{\frac{1}{T_1} \sum_{t=1}^{T_1} K_h(X_{1t} - x) I(Y_{1t} \leq y)}{\tilde{f}_{X_1}(x)} dF_{X_2}(x) \right. \\ &\quad \left. - \int \frac{\frac{1}{T_1} \sum_{t=1}^{T_1} K_h(X_{1t} - x) F(y | x)}{\tilde{f}_{X_1}(x)} dF_{X_2}(x) \right] \\ &= \sqrt{T_2} \left\{ \int \frac{1}{T_1} \sum_{t=1}^{T_1} \frac{K_h(X_{1t} - x)}{\tilde{f}_{X_1}(x)} [I(Y_{1t} \leq y) - F(y | X_{1t})] dF_{X_2}(x) \right. \\ &\quad \left. + \int \frac{1}{T_1} \sum_{t=1}^{T_1} \frac{K_h(X_{1t} - x)}{\tilde{f}_{X_1}(x)} [F(y | X_{1t}) - F(y | x)] \right\} dF_{X_2}(x) \\ &=: \sqrt{T_2}(A + B), \tag{2.5} \end{aligned}$$

其中,

$$\tilde{f}_{X_1}(x) = \frac{1}{T_1} \sum_{t=1}^{T_1} K_h(X_{1t} - x).$$

由文献 [37, 定理 2], 可以得到

$$\sup_{x \in \mathcal{X}_2} |\tilde{f}_{X_1}(x) - f_{X_1}(x)| = O_p \left\{ \left(\frac{\log T_1}{T_1 h^{d_x}} \right)^{1/2} + h^\ell \right\}.$$

在假设 2.7 下, $\sup_{x \in \mathcal{X}_2} |\tilde{f}_{X_1}(x) - f_{X_1}(x)| = o_p(T_1^{-1/4})$. 将 $1/\tilde{f}_{X_1}(x)$ 在 $1/f_{X_1}(x)$ 处进行二阶 Taylor 展开, 可以得到

$$\begin{aligned} A &= \frac{1}{T_1} \sum_{t=1}^{T_1} [I(Y_{1t} \leq y) - F(y | X_{1t})] \int \frac{f_{X_2}(x)}{\tilde{f}_{X_1}(x)} K_h(X_{1t} - x) dx \\ &= \frac{1}{T_1} \sum_{t=1}^{T_1} [I(Y_{1t} \leq y) - F(y | X_{1t})] \int \frac{f_{X_2}(x)}{f_{X_1}(x)} K_h(X_{1t} - x) dx \\ &\quad - \frac{1}{T_1} \sum_{t=1}^{T_1} [I(Y_{1t} \leq y) - F(y | X_{1t})] \int \frac{f_{X_2}(x)}{f_{X_1}^2(x)} [\tilde{f}_{X_1}(x) - f_{X_1}(x)] K_h(X_{1t} - x) dx + o_p(T_1^{-1/2}) \\ &=: A_1 - A_2 + o_p(T_1^{-1/2}). \end{aligned}$$

由于 $K(\cdot)$ 是高阶核, 因此,

$$\begin{aligned} A_1 &= \frac{1}{T_1} \sum_{t=1}^{T_1} [I(Y_{1t} \leq y) - F(y | X_{1t})] \frac{f_{X_2}(X_t)}{f_{X_1}(X_t)} + O_p(h^\ell) \\ &= \frac{1}{T_1} \sum_{t=1}^{T_1} [I(Y_{1t} \leq y) - F(y | X_{1t})] \frac{f_{X_2}(X_t)}{f_{X_1}(X_t)} + o_p(T^{-1/2}) \end{aligned}$$

且

$$\begin{aligned} A_2 &= \frac{1}{T_1} \sum_{t=1}^{T_1} [I(Y_{1t} \leq y) - F(y | X_{1t})] \int \frac{f_{X_2}(x)}{f_{X_1}^2(x)} [\tilde{f}_{X_1}(x) - f_{X_1}(x)] K_h(X_{1t} - x) dx \\ &= \frac{1}{T_1} \sum_{t=1}^{T_1} [I(Y_{1t} \leq y) - F(y | X_{1t})] \int \left[\frac{1}{T_1} \sum_{s=1}^{T_1} K_h(X_{1s} - x) - f_{X_1}(x) \right] K_h(X_{1t} - x) \frac{f_{X_2}(x)}{f_{X_1}^2(x)} dx \\ &= \frac{1}{T_1^2} \sum_{s,t=1}^{T_1} [I(Y_{1t} \leq y) - F(y | X_{1t})] \left[\int K_h(X_{1s} - x) K_h(X_{1t} - x) \frac{f_{X_2}(x)}{f_{X_1}^2(x)} dx \right. \\ &\quad \left. - \int K_h(X_{1t} - x) \frac{f_{X_2}(x)}{f_{X_1}(x)} dx \right] \\ &= \frac{1}{T_1^2} \sum_{s,t=1}^{T_1} [I(Y_{1t} \leq y) - F(y | X_{1t})] \left[K_h(X_{1s} - X_{1t}) \frac{f_{X_2}(X_{1t})}{f_{X_1}^2(X_{1t})} - \frac{f_{X_2}(X_{1t})}{f_{X_1}(X_{1t})} \right] + o_p(T_1^{-1/2}) \\ &= \frac{1}{T_1^2} \sum_{s \neq t} [I(Y_{1t} \leq y) - F(y | X_{1t})] \left[K_h(X_{1s} - X_{1t}) \frac{f_{X_2}(X_{1t})}{f_{X_1}^2(X_{1t})} - \frac{f_{X_2}(X_{1t})}{f_{X_1}(X_{1t})} \right] + o_p(T_1^{-1/2}). \end{aligned}$$

注意到, 当 $T_1 \rightarrow \infty$ 和 $|s - t| \rightarrow \infty$ 时,

$$\begin{aligned} &E \left\{ [I(Y_{1t} \leq y) - F(y | X_{1t})] \left[K_h(X_{1s} - X_{1t}) \frac{f_{X_2}(X_{1t})}{f_{X_1}^2(X_{1t})} - \frac{f_{X_2}(X_{1t})}{f_{X_1}(X_{1t})} \right] \middle| X_{1t}, Y_{1t} \right\} \\ &\rightarrow g[I(Y_{1t} \leq y) - F(y | X_{1t})] \left[\frac{f_{X_2}(X_{1t})}{f_{X_1}^2(X_{1t})} \cdot \int K_h(x - X_{1t}) f_{X_1}(x) dx - \frac{f_{X_2}(X_{1t})}{f_{X_1}(X_{1t})} \right] \\ &= o_p(T_1^{-1/2}). \end{aligned}$$

因此, A_2 的主项是一个退化的二阶 U 统计量且 $A_2 = O_p(T_1^{-1} h^{-d_x}) = o_p(T_1^{-1/2})$. 从而,

$$A = A_1 - A_2 + o_p(T_1^{-1/2}) = \frac{1}{T_1} \sum_{t=1}^{T_1} [I(Y_{1t} \leq y) - F(y | X_{1t})] \frac{f_{X_2}(X_t)}{f_{X_1}(X_t)} + o_p(T^{-1/2}). \quad (2.6)$$

类似地，可以证明 $B = o_p(T^{-1/2})$ 。由 (2.5), (2.6) 和 $\lim_{T \rightarrow \infty} T_2/T_1 = c$ ，可以得到

$$\begin{aligned} S_2 &= \sqrt{T_2}[A_1 - A_2 + B_1 - B_2 + o_p(T_1^{-1/2})] \\ &= \frac{1}{\sqrt{T_1}} \sum_{t=1}^{T_1} \sqrt{\frac{T_2}{T_1}} \frac{f_{X_2}(X_t)}{f_{X_1}(X_t)} \cdot [I(Y_{1t} \leq y) - F(y | X_{1t})] + o_p(1). \end{aligned}$$

所以，

$$\begin{aligned} \sqrt{T_2}[\tilde{F}_{Y_2^0}(y) - F_{Y_2^0}(y)] &= S_1 + S_2 + \frac{1}{\sqrt{T_2}} \sum_{t=T_1+1}^T [F(y | X_{2t}) - F_{Y_2^0}(y)] \\ &= \frac{1}{\sqrt{T_1}} \sum_{t=1}^{T_1} \sqrt{\frac{T_2}{T_1}} \frac{f_{X_2}(X_t)}{f_{X_1}(X_t)} \cdot [I(Y_{1t} \leq y) - F(y | X_{1t})] \\ &\quad + \frac{1}{\sqrt{T_2}} \sum_{t=T_1+1}^T [F(y | X_{2t}) - F_{Y_2^0}(y)] + o_p(1). \end{aligned}$$

至此完成了引理 2.1 的证明。 \square

引理 2.2 在假设 2.1–2.7 下，对 $y \in (-\infty, \infty)$ ，有 $\hat{F}_{Y_2^0}(y) - \tilde{F}_{Y_2^0}(y) = o_p(T_2^{-1/2})$ 和

$$\begin{aligned} \sqrt{T_2}[\hat{F}_{Y_2^0}(y) - F_{Y_2^0}(y)] &= \frac{1}{\sqrt{T_2}} \sum_{t=T_1+1}^T [F(y | X_{2t}) - F_{Y_2^0}(y)] \\ &\quad + \frac{1}{\sqrt{T_1}} \sum_{t=1}^{T_1} \sqrt{\frac{T_2}{T_1}} \frac{f_{X_2}(X_{1t})}{f_{X_1}(X_{1t})} \cdot [I(Y_{1t} \leq y) - F(y | X_{1t})] + o_p(1). \end{aligned}$$

证明 首先证明 $\sup_{u \leq y} \{\tilde{F}_{Y_2^0}(u)\} - \tilde{F}_{Y_2^0}(y) = o_p(T_2^{-1/2})$ 。由引理 2.1，对 $r > 0$ ，有

$$\begin{aligned} &\sqrt{T_2}[\tilde{F}_{Y_2^0}(y-r) - \tilde{F}_{Y_2^0}(y)] - \sqrt{T_2}[\tilde{F}_{Y_2^0}(y) - \tilde{F}_{Y_2^0}(y)] \\ &= -\frac{1}{\sqrt{T_2}} \sum_{t=T_1+1}^T [P(y-r < Y_{2t}^0 \leq y | X_{2t}) - P(y-r < Y_{2t}^0 \leq y)] \\ &\quad - \frac{1}{\sqrt{T_1}} \sum_{t=1}^{T_1} \sqrt{\frac{T_2}{T_1}} \frac{f_{X_2}(X_{1t})}{f_{X_1}(X_{1t})} \cdot [I(y-r < Y_{1t} \leq y) - P(y-r < Y_{1t} \leq y | X_{1t})] + o_p(1) \\ &=: G(r) + o_p(1). \end{aligned} \tag{2.7}$$

易知 $E[G(r)] = 0$ 和 $\lim_{r \rightarrow 0^+} \text{Var}[G(r)] = 0$ 。由 Markov 不等式可知，对任意的 $\varepsilon > 0$ 和 $\epsilon > 0$ ，存在一个充分小的 $\delta > 0$ 使得 $P(\sup_{0 \leq r < \delta} \{G(r)\} > \varepsilon) < \epsilon$ 。从而，由 (2.7) 可知，对任意的 $\varepsilon > 0$ 和 $\epsilon > 0$ ，存在 $\delta > 0$ 和 $N_1 > 0$ 使得对任意的 $T_2 > N_1$ 有

$$P\left(\sqrt{T_2} \left\{ \sup_{y-\delta < u \leq y} \{\tilde{F}_{Y_2^0}(u) - F_{Y_2^0}(u)\} - [\tilde{F}_{Y_2^0}(y) - F_{Y_2^0}(y)] \right\} > \varepsilon\right) < \epsilon. \tag{2.8}$$

注意到，对任意的 $u \in (y-\delta, y)$ ，有 $F_{Y_2^0}(u) \leq F_{Y_2^0}(y)$ 。从而，由 (2.8) 可以得到

$$P\left(\sqrt{T_2} \left\{ \sup_{y-\delta < u \leq y} \{\tilde{F}_{Y_2^0}(u)\} - \tilde{F}_{Y_2^0}(y) \right\} > \varepsilon\right) < \epsilon. \tag{2.9}$$

同时, 我们有

$$\begin{aligned}
& P\left(\sqrt{T_2}\left[\sup_{u \leq y-\delta}\{\tilde{F}_{Y_2^0}(u)\} - \tilde{F}_{Y_2^0}(y)\right] > \varepsilon\right) \\
& \leq P\left(\sup_{u \leq y-\delta}\{\tilde{F}_{Y_2^0}(u)\} - \tilde{F}_{Y_2^0}(y) > 0\right) \\
& = P\left(\sup_{u \leq y-\delta}\{\tilde{F}_{Y_2^0}(u) - F_{Y_2^0}(u) + F_{Y_2^0}(u)\} - [\tilde{F}_{Y_2^0}(y) - F_{Y_2^0}(y)] - F_{Y_2^0}(y) > 0\right) \\
& \leq P\left(\sup_{u \leq y-\delta}\{\tilde{F}_{Y_2^0}(u) - F_{Y_2^0}(u)\} + F_{Y_2^0}(y-\delta) - [\tilde{F}_{Y_2^0}(y) - F_{Y_2^0}(y)] - F_{Y_2^0}(y) > 0\right) \\
& = P\left(\sup_{u \leq y-\delta}\{\tilde{F}_{Y_2^0}(u) - F_{Y_2^0}(u)\} - [\tilde{F}_{Y_2^0}(y) - F_{Y_2^0}(y)] > F_{Y_2^0}(y) - F_{Y_2^0}(y-\delta)\right).
\end{aligned}$$

注意到, 由引理 2.1 有

$$\sup_{u \leq y-\delta}\{\tilde{F}_{Y_2^0}(u) - F_{Y_2^0}(u)\} - [\tilde{F}_{Y_2^0}(y) - F_{Y_2^0}(y)] \xrightarrow{P} 0, \quad F_{Y_2^0}(y) - F_{Y_2^0}(y-\delta) > 0.$$

从而, $P(\sqrt{T_2}[\sup_{u \leq y-\delta}\{\tilde{F}_{Y_2^0}(u)\} - \tilde{F}_{Y_2^0}(y)] > \varepsilon) \rightarrow 0$, 这意味着存在一个 N_2 使得对任意的 $T_2 > N_2$ 有

$$P\left(\sqrt{T_2}\left\{\sup_{u \leq y-\delta}\{\tilde{F}_{Y_2^0}(u)\} - \tilde{F}_{Y_2^0}(y)\right\} > \varepsilon\right) < \epsilon. \quad (2.10)$$

对 $T_2 > \max\{N_1, N_2\}$, 由 (2.9) 和 (2.10) 可以得到

$$\begin{aligned}
& P\left(\sqrt{T_2}\left\{\sup_{u \leq y}\{\tilde{F}_{Y_2^0}(u)\} - \tilde{F}_{Y_2^0}(y)\right\} > \varepsilon\right) \\
& \leq P\left(\sqrt{T_2}\left\{\sup_{y-\delta < u \leq y}\{\tilde{F}_{Y_2^0}(u)\} - \tilde{F}_{Y_2^0}(y)\right\} > \varepsilon\right) + P\left(\sqrt{T_2}\left\{\sup_{u \leq y-\delta}\{\tilde{F}_{Y_2^0}(u)\} - \tilde{F}_{Y_2^0}(y)\right\} > \varepsilon\right) \\
& \leq 2\epsilon,
\end{aligned}$$

这意味着

$$\sup_{u \leq y}\{\tilde{F}_{Y_2^0}(u)\} - \tilde{F}_{Y_2^0}(y) = o_p(T_2^{-1/2}). \quad (2.11)$$

容易证明

$$\sup_u\{\tilde{F}_{Y_2^0}(u)\} = \lim_{y \rightarrow \infty} \sup_{u \leq y}\{\tilde{F}_{Y_2^0}(u)\} = \lim_{y \rightarrow} \tilde{F}_{Y_2^0}(y) + o_p(T_2^{-1/2}) = 1 + o_p(T_2^{-1/2}). \quad (2.12)$$

由 (2.11) 和 (2.12) 可以得到

$$\begin{aligned}
\tilde{F}_{Y_2^0}(y) - \tilde{F}_{Y_2^0}(y) &= \frac{\sup_{u \leq y}\{\tilde{F}_{Y_2^0}(u)\}}{\sup_u\{\tilde{F}_{Y_2^0}(u)\}} - \tilde{F}_{Y_2^0}(y) \\
&= \sup_{u \leq y}\{\tilde{F}_{Y_2^0}(u)\}[1 + o_p(T_2^{-1/2})] - \tilde{F}_{Y_2^0}(y) \\
&= o_p(T_2^{-1/2}).
\end{aligned}$$

再由引理 2.1 可以得到

$$\begin{aligned}
\sqrt{T_2}[\hat{F}_{Y_2^0}(y) - F_{Y_2^0}(y)] &= \sqrt{T_2}[\hat{F}_{Y_2^0}(y) - \tilde{F}_{Y_2^0}(y)] + \sqrt{T_2}[\tilde{F}_{Y_2^0}(y) - F_{Y_2^0}(y)] \\
&= \sqrt{T_2}[\tilde{F}_{Y_2^0}(y) - F_{Y_2^0}(y)] + o_p(1) \\
&= \frac{1}{\sqrt{T_2}} \sum_{t=T_1+1}^T [F(y | X_{2t}) - F_{Y_2^0}(y)] \\
&\quad + \frac{1}{\sqrt{T_1}} \sum_{t=1}^{T_1} \sqrt{\frac{T_2}{T_1}} \cdot \frac{f_{X_2}(X_{1t})}{f_{X_1}(X_{1t})} \cdot [I(Y_{1t} \leq y) - F(y | X_{1t})] + o_p(1).
\end{aligned}$$

至此完成了引理 2.2 的证明. \square

为了证明如下定理, 定义

$$\sigma_\tau^2 = \sum_{h=-\infty}^{\infty} \text{Cov}(\xi_t, \xi_{t-h}) + \sum_{h=-\infty}^{\infty} \text{Cov}(\eta_t, \eta_{t-h}),$$

其中, $\xi_t = \psi_{1\tau}(Y_{2t}) - \psi_{0\tau}(X_{2t})$, $\eta_t = \psi_{2\tau}(X_{1t}, Y_{1t})$, $\psi_{0\tau}(X_{2t}) = [\tau - F(q_{1\tau}^0 | X_{2t})]/f_{Y_2^0}(q_{1\tau}^0)$, $\psi_{1\tau}(Y_{2t}^1) = [\tau - I(Y_{2t}^1 \leq q_{1\tau}^1)]/f_{Y_2^1}(q_{1\tau}^1)$, $\psi_{2\tau}(X_{1t}, Y_{1t}) = \sqrt{c} f_{X_2}(X_{1t}) [I(Y_{1t} \leq q_{1\tau}^0) - F(q_{1\tau}^0 | X_{1t})]/[f_{Y_2^0}(q_{1\tau}^0) f_{X_1}(X_{1t})]$. 根据 α -混合的 Davydov 不等式 [23] 和假设 2.2, 可知 σ_τ^2 存在. 下面的定理给出了 (2.4) 中 QTE 估计量的渐近性质.

定理 2.1 在假设 2.1–2.7 下, 有 $\sqrt{T_2}(\hat{\Delta}_\tau - \Delta_\tau) \xrightarrow{d} N(0, \sigma_\tau^2)$, 其中, σ_τ^2 是渐近方差.

证明 首先, 由文献 [46, 定理 2.2] 和假设 2.2 可得, $\hat{q}_{1\tau}^1$ 的 Bahadur 表示为

$$\sqrt{T_2}(\hat{q}_{1\tau}^1 - q_{1\tau}^1) = \frac{1}{\sqrt{T_2}} \sum_{t=T_1+1}^T \psi_{1\tau}(Y_{2t}^1) + o_p(1). \quad (2.13)$$

为了证明本定理, 只需得到 $\hat{q}_{1\tau}^0$ 的 Bahadur 表示. 定义 $Z_{T_2} = \sqrt{T_2}[\tau - \hat{F}_{Y_2^0}(q_{1\tau}^0)]/f_{Y_2^0}(q_{1\tau}^0)$. 注意到, $\hat{F}_{Y_2^0}(q)$ 是一个 CDF. 对任意的 $u \in (-\infty, \infty)$, $\hat{q}_{1\tau}^0 \leq q_{1\tau}^0 + u/\sqrt{T_2}$ 可以推导出 $\hat{F}_{Y_2^0}(q_{1\tau}^0 + \frac{u}{\sqrt{T_2}}) \geq \tau$. 从而, 对任意的 $\epsilon > 0$, 有

$$\begin{aligned}
P(\sqrt{T_2}(\hat{q}_{1\tau}^0 - q_{1\tau}^0) \leq u, Z_{T_2} > u + \epsilon) \\
&= P(\hat{q}_{1\tau}^0 \leq q_{1\tau}^0 + u/\sqrt{T_2}, Z_{T_2} > u + \epsilon) \\
&\leq P(\tau - \hat{F}_{Y_2^0}(q_{1\tau}^0 + u/\sqrt{T_2}) \leq 0, Z_{T_2} > u + \epsilon) \\
&= P\left(\frac{\sqrt{T_2}[\tau - \hat{F}_{Y_2^0}(q_{1\tau}^0 + \frac{u}{\sqrt{T_2}})]}{f_{Y_2^0}(q_{1\tau}^0)} \leq 0, Z_{T_2} > u + \epsilon\right) \\
&= P(Z_{T_2} + R_{T_2} \leq u, Z_{T_2} > u + \epsilon),
\end{aligned} \quad (2.14)$$

其中,

$$\begin{aligned}
R_{T_2} &= \left[\frac{\hat{F}_{Y_2^0}(q_{1\tau}^0) - \hat{F}_{Y_2^0}(q_{1\tau}^0 + \frac{u}{\sqrt{T_2}})}{f_{Y_2^0}(q_{1\tau}^0)/\sqrt{T_2}} - \frac{F_{Y_2^0}(q_{1\tau}^0) - F_{Y_2^0}(q_{1\tau}^0 + \frac{u}{\sqrt{T_2}})}{f_{Y_2^0}(q_{1\tau}^0)/\sqrt{T_2}} \right] \\
&\quad + \left[\frac{F_{Y_2^0}(q_{1\tau}^0) - F_{Y_2^0}(q_{1\tau}^0 + \frac{u}{\sqrt{T_2}})}{f_{Y_2^0}(q_{1\tau}^0)/\sqrt{T_2}} + u \right] \\
&=: R_{T_2,1} + R_{T_2,2}.
\end{aligned}$$

由引理 2.2, 有

$$\begin{aligned} R_{T_2,1} = & -\frac{1}{f_{Y_2^0}(q_{1\tau}^0)} \left\{ \frac{1}{\sqrt{T_2}} \sum_{t=T_1+1}^T [P(q_{1\tau}^0 < Y_{2t}^0 \leq q_{1\tau}^0 + u/\sqrt{T_2} | X_{2t}) - P(q_{1\tau}^0 < Y_{2t}^0 \leq q_{1\tau}^0 + u/\sqrt{T_2})] \right. \\ & + \frac{1}{\sqrt{T_1}} \sum_{t=1}^{T_1} \sqrt{\frac{T_2}{T_1}} \cdot \frac{f_{X_2}(X_{1t})}{f_{X_1}(X_{1t})} \cdot [I(q_{1\tau}^0 < Y_{1t} \leq q_{1\tau}^0 + u/\sqrt{T_2}) \\ & \left. - P(q_{1\tau}^0 < Y_{1t} \leq q_{1\tau}^0 + u/\sqrt{T_2} | X_{1t})] \right\} + o_p(1). \end{aligned}$$

容易知道, $R_{T_2,1}$ 的主项是零均值的且当 $T_2 \rightarrow \infty$ 时其方差趋于 0. 从而, $R_{T_2,1} = o_p(1)$. 同时, 我们有 $R_{T_2,2} \rightarrow \frac{-uf_{Y_2^0}(q_{1\tau}^0)}{f_{Y_2^0}(q_{1\tau}^0)} + u = 0$, 从而, $R_{T_2} = o_p(1)$. 由 (2.14) 可得

$$P(\sqrt{T_2}(\hat{q}_{1\tau}^0 - q_{1\tau}^0) \leq u, Z_{T_2} > u + \epsilon) \leq P(Z_{T_2} + R_{T_2} \leq u, Z_{T_2} > u + \epsilon) \rightarrow 0. \quad (2.15)$$

类似地, 对任意的 $u \in (-\infty, \infty)$ 和 $\epsilon > 0$, 可以得到

$$P(\sqrt{T_2}(\hat{q}_{1\tau}^0 - q_{1\tau}^0) \geq u, Z_{T_2} < u - \epsilon) \rightarrow 0. \quad (2.16)$$

由 (2.15) 和 (2.16) 可得 $\sqrt{T_2}(\hat{q}_{1\tau}^0 - q_{1\tau}^0) = Z_{T_2} + o_p(1)$. 从而, 由引理 2.2 可得

$$\begin{aligned} \sqrt{T_2}(\hat{q}_{1\tau}^0 - q_{1\tau}^0) &= -\sqrt{T_2}[\hat{F}_{Y_2^0}(q_{1\tau}^0) - \tau]/f_{Y_2^0}(q_{1\tau}^0) + o_p(1) \\ &= \frac{1}{\sqrt{T_2}} \sum_{t=T_1+1}^T \frac{-[F(q_{1\tau}^0 | X_{2t}) - \tau]}{f_{Y_2^0}(q_{1\tau}^0)} \\ &\quad - \frac{1}{\sqrt{T_1}} \sum_{t=1}^{T_1} \frac{1}{f_{Y_2^0}(q_{1\tau}^0)} \cdot \sqrt{\frac{T_2}{T_1}} \cdot \frac{f_{X_2}(X_{1t})}{f_{X_1}(X_{1t})} \cdot [I(Y_{1t} \leq q_{1\tau}^0) - F(q_{1\tau}^0 | X_{1t})] + o_p(1) \\ &= \frac{1}{\sqrt{T_2}} \sum_{t=T_1+1}^T \psi_{0\tau}(X_{2t}) - \frac{1}{\sqrt{T_1}} \sum_{t=1}^{T_1} \psi_{2\tau}(X_{1t}, Y_{1t}) + o_p(1). \end{aligned} \quad (2.17)$$

由 (2.13) 和 (2.17), 有

$$\begin{aligned} \sqrt{T_2}(\hat{\Delta}_\tau - \Delta_\tau) &= \frac{1}{\sqrt{T_2}} \sum_{t=T_1+1}^T [\psi_{1\tau}(Y_{2t}) - \psi_{0\tau}(X_{2t})] + \frac{1}{\sqrt{T_1}} \sum_{t=1}^{T_1} \psi_{2\tau}(X_{1t}, Y_{1t}) + o_p(1) \\ &= \frac{1}{\sqrt{T_2}} \sum_{t=T_1+1}^T \xi_t + \frac{1}{\sqrt{T_1}} \sum_{t=1}^{T_1} \eta_t + o_p(1) \\ &= K_1 + K_2 + o_p(1), \end{aligned}$$

根据平稳性, 对 $s > 0$, 定义 $C(s) = \text{Cov}(\eta_t, \xi_{t+s})$. 通过简单的计算可以得到

$$\text{Cov}(K_1, K_2) = \sqrt{\frac{T_1}{T_2}} \frac{1}{T_1} \sum_{s=1}^{T_1} \min\{s, T_2\} C(s) + \sqrt{\frac{T_2}{T_1}} \frac{1}{T_2} \sum_{s=T_1+1}^{T-1} \min\{T_1, T-s\} C(s). \quad (2.18)$$

由 Kronecker 引理、Davydov 不等式和假设 2.2 可知, 当 $T_1 \rightarrow \infty$ 时, (2.18) 中的第一项收敛于 0, 而第二项有界且被 $\sum_{s=T_1+1}^\infty |C(s)| \rightarrow 0$ 所控制. 由 α -混合序列的中心极限定理, 有

$$\sqrt{T_2}(\hat{\Delta}_\tau - \Delta_\tau) \xrightarrow{d} N(0, \sigma_\tau^2).$$

至此完成了定理 2.1 的证明. □

定理 2.1 意味着 $\hat{\Delta}_\tau$ 是一致的估计量. 尽管估计步骤中的条件 CDF 是使用核方法估计的, 但 $\hat{\Delta}_\tau$ 仍然可以达到 $\sqrt{T_2}$ 收敛速度. 由定理 2.1 还知道, 对于给定的 τ , 如果 σ_τ^2 已知, 则可以构造 Δ_τ 的 $(1 - \alpha)100\%$ 置信区间 $\hat{\Delta}_\tau \pm z_{\alpha/2}/\sqrt{T_2}\sigma_\tau$, 其中, $\alpha > 0$, $z_{\alpha/2} := \Phi^{-1}(1 - \alpha/2)$ 是标准正态分布 $\alpha/2$ 水平的上分位数. 当 σ_τ^2 未知时, 一致估计 σ_τ^2 的一种方法是使用 Newey 和 West [38] 的异方差和自相关一致 (heteroskedasticity autocorrelation consistent, HAC) 估计方法. 然而, 由于 σ_τ^2 的复杂结构, 获得 σ_τ^2 的一致估计并不容易. 因此, 在下一小节使用 Bootstrap 方法来构造 Δ_τ 的置信区间.

2.5 Bootstrap 推断

由于考虑的是面板数据, 这里使用 Künsch [33] 的分块 Bootstrap 方法来构造 Δ_τ 的置信区间, 具体描述如下. 首先, 对样本 $W_{1t} := (Y_{1t}, X_{1t})$, $t = 1, 2, \dots, T_1$, 构造分块 $V_{1t} = \{W_{1t}, W_{1,t+1}, \dots, W_{1,t+b_1-1}\}$, $t = 1, 2, \dots, T_1 - b_1 + 1$, 其中, $b_1 = \lfloor \sqrt[3]{T_1} \rfloor$ 为每个分块的长度, $\lfloor x \rfloor$ 表示小于等于 x 的最大整数. 对样本 $W_{2t} = (Y_{2t}, X_{2t})$, $t = T_1 + 1, T_1 + 2, \dots, T$, 同样构造分块 $V_{2t} = \{W_{2t}, W_{2,t+1}, \dots, W_{2,t+b_2-1}\}$, $t = T_1 + 1, T_1 + 2, \dots, T - b_2 + 1$, 其中, $b_2 = \lfloor \sqrt[3]{T_2} \rfloor$. 其次, 有放回地从样本 $V_1 = \{V_{11}, V_{12}, \dots, V_{1,T_1-b_1+1}\}$ 中抽取 $l_1 = \lfloor T_1/b_1 \rfloor$ 个分块, 记得到的分块为 $V_1^* = \{V_{11}^*, V_{12}^*, \dots, V_{1l_1}^*\}$; 有放回地从样本 $V_2 = \{V_{2,T_1+1}, V_{2,T_1+2}, \dots, V_{2,T-b_2+1}\}$ 中抽取 $l_2 = \lfloor T_2/b_2 \rfloor$ 个分块, 记得到的分块为 $V_2^* = \{V_{21}^*, V_{22}^*, \dots, V_{2l_2}^*\}$, 其中, V_1 和 V_2 的长度分别为 $b_1 l_1 \approx T_1$ 和 $b_2 l_2 \approx T_2$. 再次, 基于重抽样获得样本 $W_1^* = \{W_{11}^*, W_{12}^*, \dots, W_{1,b_1 l_1}^*\} = \{V_{11}^*, V_{12}^*, \dots, V_{1l_1}^*\}$ 和 $W_2^* = \{W_{21}^*, W_{22}^*, \dots, W_{2,b_2 l_2}^*\} = \{V_{21}^*, V_{22}^*, \dots, V_{2l_2}^*\}$, 根据第 2.3 小节所提出的方法计算分位数处理效应的估计量. 最后, 重复 B (如 $B = 1000$) 次以上的步骤, 获得 B 个分位数处理效应的估计量, 基于这些估计量即可构造 Δ_τ 的 $(1 - \alpha)100\%$ 置信区间. 该分块 Bootstrap 方法将被用于本文第 4 节的实证研究.

2.6 高维数情形

本小节考虑有较多对照个体和协变量的情形, 即 d_x 较大的情形. 对于这种情形, 由于所谓的“维数灾难”, 通过核回归方法估计条件 CDF $F(\cdot | x)$ 并不理想. 为了避免“维数灾难”问题, 可以使用 Aït-sahali 和 Brandt [1] 以及 Hall 和 Yao [24] 的思想, 采用指数方法 (index approach), 利用 $\beta^\top x$ 来估计 $F(\cdot | \beta^\top x)$, 其中 $\beta^\top x$ 是一个标量或低维向量. 指数方法将多变量问题简化为单变量/低维变量问题, 从而避免了“维数灾难”. 本文使用另一种方法, 通过 QR [32] 来估计条件 CDF. 具体而言, 令 $q(\tau, x)$ 为 $F(y | x)$ 的 τ 条件分位数, 即 $q(\tau, x) = F^{-1}(\tau | x)$. 若可以对 $q(\tau, x)$ 进行估计, 则可以用 $q(\tau, x)$ 和 $F(y | x)$ 的关系来恢复条件 CDF. 事实上, 条件 CDF 和条件分位数之间有如下关系:

$$F(y | x) = \int_0^1 I(q(u, x) \leq y) du \approx \varepsilon + \int_\varepsilon^{1-\varepsilon} I(q(u, x) \leq y) du, \quad (2.19)$$

其中, $\varepsilon > 0$ 是一个取值较小的常数. 上式使用近似值是考虑了低分位数时的计算稳定性. 假设给定 X_1, Y_1 的 τ 条件分位数具有参数形式 $q(\tau, x) = \beta_\tau^\top x$, 则条件分位数函数的估计量为 $\hat{q}(\tau, x) = \hat{\beta}_\tau^\top x$, 其中,

$$\hat{\beta}_\tau = \arg \min_{\beta_\tau} \sum_{t=1}^{T_1} \rho_\tau(y_{1t} - \beta_\tau^\top X_{1t}), \quad (2.20)$$

$\rho_\tau(v) = v[\tau - I(v < 0)]$. 因此, 根据 (2.19), 可以得到条件 CDF 的一个估计量为

$$\hat{F}(y | x) = \varepsilon + \int_\varepsilon^{1-\varepsilon} I(\hat{\beta}_\tau^\top x \leq y) du \approx \varepsilon + \sum_{j=1}^m \delta_j I(\hat{\beta}_{\tau_j}^\top x \leq y), \quad (2.21)$$

其中, $\varepsilon \leq \tau_0 < \dots < \tau_m \leq 1 - \varepsilon$, 并且当 $T_1 \rightarrow \infty$ 时, 有 $m \rightarrow \infty$ 和 $\delta_j = \tau_j - \tau_{j-1} \rightarrow 0$. 如果面板数据中的时间序列是平稳并且 α -混合的, 类似于文献 [14, 命题 5], 可以证明 $\hat{F}(y | x)$ 是一致的并且具有渐近正态性, 如下所示:

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

其中, $V(y | x) > 0$ 是渐近方差, 它的取值取决于 x 和 y .

值得一提的是, 我们也可以利用机器学习的方法 (如随机森林) 对条件分布函数 $F(y | x)$ 进行非参数估计, 感兴趣的读者可参见文献 [3, 17]. 基于机器学习的估计方法虽然往往在预测性能上表现优异, 但缺少相应的理论基础, 可以作为未来的研究方向.

2.7 选择对照个体和协变量

在实际应用中, 若横截面和协变量的维数很大并且存在稀疏性, 则使用所有对照个体和协变量可能会导致估计不稳定. 对于这种情形, 本文建议使用类似于文献 [18, 43] 中带有惩罚的 QR 方法来估计条件分位数函数. 具体而言, 在 (2.20) 中添加惩罚项

$$\hat{\beta}_{\text{pen}, \tau} = \arg \min_{\beta_\tau} \sum_{t=1}^{T_1} \rho_\tau(y_{1t} - \beta_\tau^\top X_{1t}) + \sum_{j=1}^{d_x} \psi_{\lambda^*}(\beta_{\tau, j}), \quad (2.22)$$

其中, $\psi_{\lambda^*}(\cdot)$ 是某个惩罚函数, 如 Tibshirani [43] 的绝对值惩罚函数, 或者 Fan 和 Li [18] 的平滑剪切绝对偏差 (smoothly clipped absolute deviation) 惩罚函数. 这可以通过 R 软件中的 rqPen 包 [42] 或 hreg 包 [47] 实现. 带惩罚的条件 CDF 的估计量为

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

其中, $\varepsilon > 0$ 是一个取值较小的常数. 与 Li 和 Zhu [36] 的结果类似, 可以针对相依数据获得估计量的渐近正态性, 特别是对于充分满足某些强混合条件的序列. 通过第 3 节中的 Monte Carlo 模拟研究, 本文验证了基于惩罚 QR 方法的估计量的有效性.

最后, 使用 (2.22) 中给出的惩罚 QR 方法的一个优点是, 通过一些简单的扩展, 我们甚至可以处理横截面和协变量的维数远大于时间序列维数的情形, 详情参见文献 [42, 44, 47]. 这种对超高维情形的扩展有待于今后进一步的研究.

3 Monte Carlo 模拟

本节进行一系列 Monte Carlo 模拟实验来验证本文提出的面板数据 QTE 估计量的有限样本性质. 第一个模拟实验是针对低维情形设计的, 可以通过核方法估计条件 CDF. 第二个模拟实验针对的是对照个体或协变量数量适中时, 使用 QR 模型估计条件 CDF. 第三个模拟实验展示带惩罚的 QR 方法在面对稀疏的高维情形时的表现. 对于前两个模拟, 时间长度设置为 $T = 200, 400, 800$, 以及 $\lambda = T_1/T = 1/2$; 而对于第三个模拟, 时间长度设置为 $T = 100, 200, 400$, 以及 $\lambda = T_1/T = 1/2$, 对每种设定进行 1000 次 Monte Carlo 模拟. 为了评估本文提出的 QTE 估计量的表现, 本文计算了 1000 个 QTE 估计量绝对误差的中位数 (median absolute error, MAE), 即 $\text{MAE} = \text{median}\{|\hat{\Delta}_\tau^j - \Delta_\tau| : j = 1, \dots, 1000\}$ 及其标准差 (standard deviation, SD, 在表内括号中给出), 其中, $\tau = 0.25, 0.50, 0.75$. 下面在生成数据

中所使用的 AR(1) 过程为 $\xi_t = \varphi \xi_{t-1} + v_t$, 其中, 具体系数 φ 在稍后给出, $v_t \sim \text{Exponential}(1) - 1$ 是均值为 0、方差为 1 的白噪声序列.

例 3.1 假设 $N = 3$, 即存在两个对照个体, $d_x = 2$. 令

$$\begin{aligned} y_{1t} &= \frac{1}{\sqrt{5}}y_{2t} + \frac{2}{\sqrt{5}}\sin(y_{3t}) + \sqrt{y_{2t}^2 + y_{3t}^2} \cdot \varepsilon_t, \quad 1 \leq t \leq T_1, \\ y_{1t}^0 &= \frac{1}{\sqrt{5}}y_{2t} + \frac{2}{\sqrt{5}}\sin(y_{3t}) + \sqrt{y_{2t}^2 + y_{3t}^2} \cdot \varepsilon_t, \quad T_1 + 1 \leq t \leq T, \\ y_{1t}^1 &= \rho_t + y_{1t}^0, \quad T_1 + 1 \leq t \leq T, \end{aligned}$$

其中, $\{y_{2t}\}_{t=1}^T$, $\{y_{3t}\}_{t=1}^T$, $\{\rho_t\}_{t=T_1+1}^T$ 和 $\{\varepsilon_t\}_{t=1}^T$ 相互独立, $\{y_{2t}\}_{t=1}^{T_1}$ 和 $\{y_{3t}\}_{t=1}^{T_1}$ 是系数 $\varphi = 0.6$ 的 AR(1) 过程, $\{y_{2t}\}_{t=T_1+1}^T$ 和 $\{y_{3t}\}_{t=T_1+1}^T$ 是系数 $\varphi = 0.4$ 的 AR(1) 过程, $\{\rho_t\}_{t=T_1+1}^T$ 是系数 $\varphi = 0.5$ 的 AR(1) 过程, $\{\varepsilon_t\}_{t=1}^T$ 是与 $\{v_t\}_{t=1}^T$ 同分布且独立的白噪声序列. $Y_1 | X_1$ 和 $Y_2^0 | X_2$ 满足假设 2.1 的识别性条件. 不同 τ 处的真实 QTE 可以通过大样本量数值模拟获得. 由于维数 $d_x = 2$ 较小, 所以这里 $Y_1 | X_1$ 的条件 CDF 使用第 2.3 小节描述的非参数方法进行估计. 为了构造一个满足假设 2.4 的高阶核 $K(\cdot)$, 令 $K(u) = \prod_{j=1}^{N-1} e_1^\top S^{-1}(1, u_j, u_j^2, u_j^3)^\top k(u_j)$, 其中, $e_1 = (1, 0, \dots, 0)^\top$ 是单位向量, $S = (\mu_{i+r})_{0 \leq i, r \leq p}$ 是一个矩阵, $\mu_j = \int u^j k(u) du$, $k(u)$ 是一元 Epanechnikov 核函数, 这样构造的 $K(u)$ 是一个 4 阶核. 窗宽取为 $h = 3.12\sigma_{X_1}T_1^{-1/6}$, 其中, 3.12 是带有 2 维协变量 X_1 的 4 阶 Epanechnikov 核的基于经验法则选取的窗宽, σ_{X_1} 是协变量 X_1 的样本标准差. 表 1 列出了本例的模拟结果. 从表中可以看出, 本文所提出的 QTE 估计量的有限样本性质表现良好, MAE 和 SD 随着 T 增大而迅速减小. 另外注意到, 当 T_2 从 100 增加到 400 时, MAE 值减少了将近一半, 这表明估计量的收敛速率为 $\sqrt{T_2}$.

例 3.2 假设 $N = 8$, 即考虑 7 个对照个体的情形. 令

$$\begin{aligned} y_{1t} &= \frac{1}{\sqrt{7}} \sum_{i=2}^8 y_{it} + \varepsilon_t, \quad 1 \leq t \leq T_1, \\ y_{1t}^0 &= \frac{1}{\sqrt{7}} \sum_{i=2}^8 y_{it} + \varepsilon_t, \quad T_1 + 1 \leq t \leq T, \\ y_{1t}^1 &= \rho_t + y_{1t}^0, \quad T_1 + 1 \leq t \leq T, \end{aligned}$$

其中, $\{y_{2t}\}_{t=1}^T, \dots, \{y_{8t}\}_{t=1}^T, \{\rho_t\}_{t=T_1+1}^T$ 与 $\{\varepsilon_t\}_{t=1}^T$ 相互独立, $\{y_{2t}\}_{t=1}^{T_1}, \dots, \{y_{8t}\}_{t=1}^{T_1}$ 是系数 $\varphi = 0.6$ 的 AR(1) 过程, $\{y_{2t}\}_{t=T_1+1}^T, \dots, \{y_{8t}\}_{t=T_1+1}^T$ 是系数 $\varphi = 0.4$ 的 AR(1) 过程, $\{\rho_t\}_{t=T_1+1}^T$ 是系数 $\varphi = 0.5$ 的 AR(1) 过程, $\{\varepsilon_t\}_{t=1}^T$ 是与 $\{v_t\}_{t=1}^T$ 同分布且独立的白噪声序列. 不同 $\tau \in (0, 1)$ 处的真实 QTE 同样通过大样本量数值模拟获得. 由于协变量维数中等, 所以这里采用参数 QR 方法 (2.21) 对 $Y_1 | X_1$ 的条件 CDF 进行估计. 计算 1000 次模拟中 QTE 估计量 $\hat{\Delta}_\tau$ 的 MAE 和 SD, 结果列在表 2 中. 从表中可以看出, MAE 和 SD 基本都很小, 并且随着样本量的增大而变小. 当样本量 T_2 从 100 增加到 400 时, MAE 值同样减少了近一半, 这表明估计量的收敛速率达到了 $\sqrt{T_2}$.

表 1 例 3.1 模拟的 MAE 和 SD (括号中)

(T_1, T_2)	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$
(100, 100)	0.189 (0.180)	0.194 (0.184)	0.263 (0.239)
(200, 200)	0.141 (0.135)	0.145 (0.137)	0.192 (0.176)
(400, 400)	0.095 (0.089)	0.096 (0.094)	0.134 (0.130)

表 2 例 3.2 模拟的 MAE 和 SD (括号中)

(T_1, T_2)	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$
(100, 100)	0.290 (0.240)	0.307 (0.254)	0.357 (0.311)
(200, 200)	0.194 (0.179)	0.214 (0.180)	0.267 (0.219)
(400, 400)	0.146 (0.131)	0.147 (0.137)	0.191 (0.176)

表 3 例 3.3 模拟的 MAE 和 SD (括号中)

方法	(T_1, T_2)	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$
QR	(50, 50)	0.976 (0.978)	0.868 (0.826)	1.009 (1.130)
	(100, 100)	0.296 (0.270)	0.314 (0.269)	0.373 (0.340)
	(200, 200)	0.185 (0.163)	0.195 (0.171)	0.242 (0.220)
PQR	(50, 50)	0.357 (0.340)	0.378 (0.351)	0.467 (0.432)
	(100, 100)	0.242 (0.222)	0.262 (0.225)	0.304 (0.298)
	(200, 200)	0.181 (0.153)	0.187 (0.165)	0.233 (0.205)

例 3.3 考察带惩罚的分位数回归 (penalized quantile regression, PQR) 方法在提出的 QTE 估计量中的表现. 假设有 $d_x = 40$ 个对照个体, 并且模型具有较大的稀疏性. 令

$$\begin{aligned} y_{1t} &= \frac{5}{2} \cdot y_{2t} + \frac{7}{2} \cdot y_{3t} + 0 \cdot \sum_{i=4}^{41} y_{it} + \varepsilon_t, \quad 1 \leq t \leq T_1, \\ y_{1t}^0 &= \frac{5}{2} \cdot y_{2t} + \frac{7}{2} \cdot y_{3t} + 0 \cdot \sum_{i=4}^{41} y_{it} + \varepsilon_t, \quad T_1 + 1 \leq t \leq T, \\ y_{1t}^1 &= \rho_t + y_{1t}^0, \quad T_1 + 1 \leq t \leq T, \end{aligned}$$

其中, $\{y_{2t}\}_{t=1}^T, \dots, \{y_{41,t}\}_{t=1}^T, \{\rho_t\}_{t=T_1+1}^T$ 与 $\{\varepsilon_t\}_{t=1}^T$ 相互独立, $\{y_{2t}\}_{t=1}^{T_1}, \dots, \{y_{41,t}\}_{t=1}^{T_1}$ 是系数 $\varphi = 0.6$ 的 AR(1) 过程, $\{y_{2t}\}_{t=T_1+1}^T, \dots, \{y_{41,t}\}_{t=T_1+1}^T$ 是系数 $\varphi = 0.4$ 的 AR(1) 过程, $\{\rho_t\}_{t=T_1+1}^T$ 是系数 $\varphi = 0.5$ 的 AR(1) 过程, $\{\varepsilon_t\}_{t=1}^T$ 是与 $\{\nu_t\}_{t=1}^T$ 同分布的白噪声序列. 不同 $\tau \in (0, 1)$ 处的真实 QTE 通过大样本量数值模拟获得. 本文分别使用 (2.21) 中的 QR 方法和 (2.22) 中 PQR 方法来计算 QTE 估计量 $\hat{\Delta}_\tau$. 1000 次模拟的 QTE 估计量的 MAE 和 SD 列在表 3 中. 从表中可以看出, 当 N 较大且 T 较小时, 基于 PQR 方法的 MAE 和 SD 数值均显著小于基于 QR 方法的数值, 这符合我们的预期. 此外, 随着 T_2 从 50 增加到 200, 基于 PQR 方法的 MAE 减少了大约一半, 这表明估计量的收敛速度大致是 $\sqrt{T_2}$.

4 实证研究

4.1 数据和描述性统计

本节使用所提出的方法来估计引入沪深 300 指数 (CSI 300) 期货交易对中国股市现货价格及其波动率的分位数处理效应. 作为金融改革的一部分, 2010 年 4 月 16 日, 中国金融期货交易所正式推出沪深 300 指数期货合约. 期货交易的引入是否会对股票市场产生积极影响是金融文献中一个有争议的问题. 有人批评股指期货交易的推出可能会因过度投机而动摇现货市场; 也有人认为股指期货市场可以提高信息流动的速度和质量, 使金融市场更加完善. 为了更全面地考察引入沪深 300 指数

期货交易对我国股票市场波动性的影响,本节综合考虑地理联系和贸易关系,选取了 13 个国家或地区的金融市场指数作为对照个体,其中包括恒生指数 (Hang Seng Index, HSI)、恒生香港中资企业指数 (Hang Seng China Affiliated Corporation Index, HSCCI)、韩国综合股价指数 (Korea's Composite Stock Price Index, KOSPI)、日本日经 225 指数 (Japan's Nikkei 225 Index, Nikkei 225)、新加坡海峡时报指数 (Singapore's Strait Times Index, STI)、中国台湾综合指数 (Taiwan Province's Composite Index, TPCI)、英国 FTSE 100 指数 (the United Kingdom's FTSE 100 Index, FTSE 100)、美国标普 500 指数 (the United States' S&P 500 Index, SP 500)、法国 CAC 40 指数 (France's CAC 40 Index, CAC 40)、德国法兰克福 DAX 指数 (Germany's Frankfurt DAX Index, DAX)、巴西 Bovespa 指数 (Brazil's Bovespa Index, BBI)、加拿大标普/TSX 综合指数 (Canada's S&P/TSX Composite Index, SPTSX) 和澳大利亚所有普通股指数 (Australia's All Ordinaries Index, AOI). 此外,还选取了 3 个宏观经济变量:月度 CPI 增长率、月度 M1 增长率和月度 M2 增长率. 所有市场指数均来自 Resset 金融研究数据库 (<http://www.resset.cn/enddatabases>), 宏观经济数据来自 CEIC (China economic information center) 数据库 (<https://www.ceicdata.com>). 14 个指数的每月股票对数回报率是通过一个月里的最后一天与第一天的对数价格之差计算的. 数据时间段为 2002 年 1 月至 2021 年 2 月, 总样本量为 $T = 230$. 14 个指数和 3 个宏观经济变量的对数收益的描述性统计见表 4. 从表 4 中可以看出,对于大多数市场指数,它们的对数收益分布几乎是对称的.

参照文献 [12], 14 个指数的月度股票波动率通过当月日度指数收益率的标准差乘以当月交易天数的平方根来计算. 表 5 报告了 14 个指数波动率的描述统计量,从表中可以看出,所有 14 个指数的波

表 4 月度收益率和宏观经济变量的描述统计量

指数	均值	标准差	中位数	最小值	最大值	峰度	偏度
CSI 300	-0.052	1.056	-0.064	-3.807	4.544	6.350	0.424
HSI	-0.088	0.867	-0.065	-2.535	2.303	3.295	0.106
HSCCI	-0.113	1.063	-0.100	-3.426	2.993	3.317	-0.022
KOSPI	-0.051	0.827	-0.065	-2.730	2.498	3.982	-0.189
Nikkei 225	-0.077	0.949	-0.097	-2.650	3.146	4.128	0.464
STI	-0.023	0.800	-0.090	-2.340	5.368	13.436	1.760
TPCI	0.037	0.737	-0.037	-2.452	2.921	4.171	0.211
FTSE 100	-0.160	0.813	-0.254	-1.686	4.224	7.182	1.325
SP 500	-0.085	0.821	-0.148	-2.656	4.673	9.084	1.371
CAC 40	-0.075	0.927	-0.106	-2.262	4.199	5.728	0.972
DAX	-0.106	0.943	-0.122	-2.423	4.337	6.298	1.023
BBI	-0.141	1.083	-0.280	-3.380	3.600	4.030	0.430
SPTSX	0.009	0.767	-0.006	-1.513	5.016	12.538	2.042
AOI	-0.012	0.603	-0.012	-1.933	2.187	4.533	0.190
CPI 增长率	0.001	0.028	0.002	-0.095	0.058	4.106	-0.615
M1 增长率	0.134	0.075	0.127	0.000	0.390	3.053	0.554
M2 增长率	0.149	0.047	0.142	0.080	0.297	3.861	0.800

注: 每月股票对数回报率的计算方法为 100 乘以一个月最后一天与第一天对数价格之差. CPI, M1 和 M2 增长率分别指各月与上年同月相比的涨幅.

动率分布都是非对称的(参见表 5 中的第 8 列)和厚尾的(参见表 5 中的第 7 列). 从图 1 中也可以得到类似的结论. Chen 等 [12] 以 2002 年 1 月至 2011 年 6 月为样本期, 采用 Hsiao 等 [26] 的面板数据政策评估方法, 并基于中国与国际股票市场的相关性构建现货市场波动率的反事实, 得出股指期货交易的引入可以显著降低中国股市现货市场波动率均值的结论. 然而, 与 Chen 等 [12] 不同, 本节考虑股指期货交易的引入对中国股市的对数收益和波动率的分位数处理效应. 根据沪深 300 指数期货推出日期, 本节将数据集的整个时间段分为两部分: 2002 年 1 月至 2010 年 4 月为处理前期, 观测值的样本量为 $T_1 = 100$; 2010 年 5 月到 2021 年 2 月为处理后期, 观测值的样本量为 $T_2 = 130$. 最后, 考虑到对照个体和宏观经济变量的总数量为 $d_x = 16$, 本实证研究中选择使用 QR 方法估计条件 CDF.

4.2 引入股指期货交易对股票收益率的 QTE

本小节考察引入股指期货交易对我国股票市场的月度对数收益率的 QTE. 图 2 展示了根据本文提出的估计方法得到的引入沪深 300 指数期货交易对中国股市对数收益率的 QTE 估计值, 以及基于分块 Bootstrap 的各分位数的 95% 置信区间(红色阴影区域), 其中, Bootstrap 重复次数 $B = 1000$. 此外, 我们还绘制了通过 HCW 方法计算的 ATE 与它的 95% 置信区间(蓝色阴影区域). 从图 2 可以看出, 首先, ATE 的 95% 置信区间基本为零, 这意味着平均收益不受沪深 300 指数期货交易推出的影响, 这与人们的预期相符合; 其次, 估计的 QTE 随 τ 发生变化, 并且在较低分位数处显著为正(在 10% 分位数处约为 0.008), 而在较高分位数处显著为负(在 90% 分位数处约为 -0.01). 这表明沪深 300 指数期货交易的推出对中国股市不同分位数的对数收益有不同的影响. Xiao 和 Koenker [45] 指出, 对数收益率的分位数可以用来表征对数收益率的风险. 例如, 下分位数对应于风险价值(value-at-risk, VaR), 这是金融文献中众所周知的下行风险衡量标准. 下分位数处的正 QTE 表明, 沪深 300 期货交易的引入可以通过使负对数收益率绝对值减小来降低 VaR. 同时, 较高分位数处的负 QTE 表明, 沪深 300 期货交易的引入还可以通过使正对数收益率减小来降低市场波动. 与 Chen 等 [12] 的结论类似, 本小节的

表 5 月度波动率的描述统计量

指数	均值	标准差	中位数	最小值	最大值	峰度	偏度
CSI 300	0.066	0.033	0.057	0.013	0.184	4.568	1.334
HSI	0.056	0.032	0.047	0.020	0.325	26.522	3.777
HSCCI	0.068	0.034	0.060	0.026	0.317	15.987	2.712
KOSPI	0.053	0.030	0.046	0.017	0.249	12.208	2.381
Nikkei 225	0.059	0.031	0.053	0.019	0.318	24.820	3.284
STI	0.044	0.030	0.037	0.013	0.256	18.776	3.284
TPCI	0.049	0.025	0.042	0.016	0.142	4.812	1.401
FTSE 100	0.047	0.029	0.039	0.012	0.231	13.735	2.711
SP 500	0.046	0.034	0.037	0.011	0.276	18.382	3.287
CAC 40	0.058	0.034	0.049	0.017	0.248	10.028	2.240
DAX	0.059	0.034	0.050	0.018	0.239	9.239	2.165
BBI	0.073	0.037	0.065	0.028	0.360	28.565	4.116
SPTSX	0.040	0.031	0.032	0.011	0.290	29.700	4.398
AOI	0.040	0.024	0.035	0.012	0.222	20.896	3.370

注: 月度股指波动率的计算方法为当月每日指数收益率的标准差乘以当月交易天数的平方根.

结果表明引入沪深 300 期货交易使得中国股市更加稳定。然而, Chen 等 [12] 没有发现这种分布上的不对称效应。

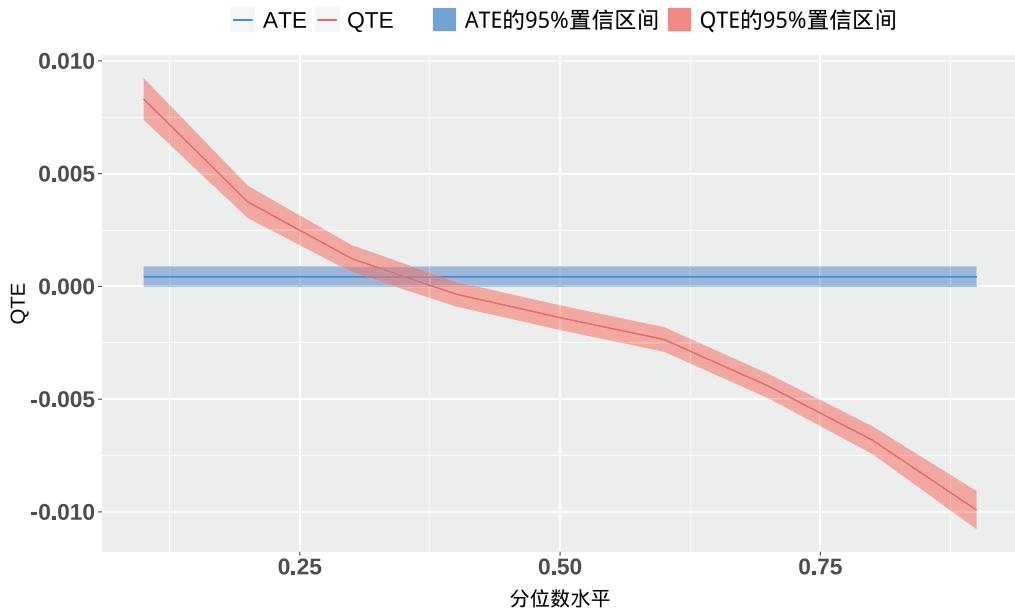


图 2 (网络版彩图) 沪深 300 指数月度对数收益率在不同 τ 处的 QTE 估计值 (红色实线), 以及基于分块 Bootstrap 计算的 95% 置信区间 (红色阴影区域); 基于 HCW 方法的 ATE 估计值 (蓝色实线) 及 95% 置信区间 (蓝色阴影区域)

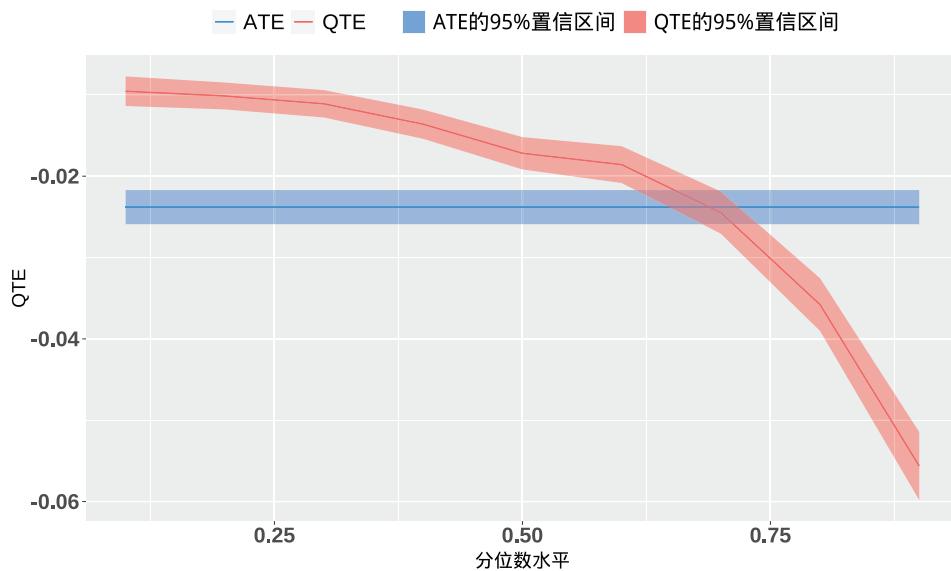


图 3 (网络版彩图) 沪深 300 指数月度波动率在不同 τ 处的 QTE 估计值 (红色实线), 以及基于分块 Bootstrap 计算的 95% 置信区间 (红色阴影区域); 基于 HCW 方法的 ATE 估计值 (蓝色实线) 及 95% 置信区间 (蓝色阴影区域)

4.3 引入股指期货交易对股市波动率的 QTE

本小节研究引入股指期货交易对我国股票市场波动率的 QTE. 与上一小节不同的是, 本小节中的 y_{1t} 是月度波动率而不是月度对数收益率. 我们通过本文所提出的方法来估计引入沪深 300 股指期货交易对中国股市波动率的 QTE. 图 3 绘制出了 QTE 的估计值, 以及基于分块 Bootstrap 的各分位数的 95% 置信区间 (红色阴影区域), 其中, Bootstrap 重复次数 $B = 1000$. 此外, 图 3 还绘制了基于 HCW 方法计算出 ATE 及其 95% 置信区间 (蓝色阴影区域). 首先, 由于波动率的分布是不对称且厚尾的, 反映到图 3 可以看出, 中位数效应 $\hat{\Delta}_{\tau=1/2}$ 与平均处理效应有明显差别. 其次, 可以看到, 估计的 QTE 大致分成两段呈线性下降趋势, 显著为负, 并且较高分位数的波动率比较低分位数的波动率降低得更多, 例如, 在 90% 分位数为 -0.054 , 而在 10% 分位数为 -0.010 . HCW 方法得到的 ATE 给出了引入沪深 300 指数期货交易对波动率变化的总体影响, 而本文估计的 QTE 可以提供引入沪深 300 指数期货交易对波动率变化影响的更多细节. 总体来看, 引入沪深 300 期货交易可以降低股票市场的波动率, 且较高分位数的波动率降低得更加明显. 结果与第 4.2 小节中风险价值的研究一致, 都表明引入沪深 300 期货交易可以使中国股市更加稳定.

5 结论

为了更全面地衡量某个处理造成的影响, 本文提出针对面板数据 QTE 的估计方法, 该方法将 HCW 方法从均值估计推广到分位数估计的框架下. 为了能够使用可观测的对照个体的数据来估计处理个体的无法观测的反事实结果变量的分布特征, 本文引入了条件 CDF 不变性假设. 利用该假设, 本文提出了一种简单的方法来估计面板数据中处理个体的 QTE. 该方法的主要难点在于估计条件 CDF. 本文考虑了估计条件 CDF 的非参数方法和参数方法. 此外, 当对照个体的数量大于面板数据的时间长度时, 本文建议使用 LASSO 类型的 QR 方法来估计条件 CDF. 在实证分析中, 本文采用所提出的面板数据 QTE 估计方法来估计沪深 300 指数期货市场的推出对我国股票市场对数收益率和波动率的 QTE. 最后, 本文所提到的一些扩展可以作为未来的研究主题.

最后, 本文仍然有许多拓展的空间, 第一, 如正文所提到超高维问题, 值得进一步研究; 第二, 仍然在非参数的框架下, 如何应用机器学习方法对条件分布函数 $F(y | x)$ 进行非参数估计是非常有趣的研究课题, 特别地, 其相应的理论问题可以作为未来的研究方向.

致谢 作者们非常感谢邵启满副主编的邀请, 本文献给浙江大学林正炎教授 85 寿辰. 同时, 作者们感谢审稿人的宝贵意见和建议, 使本文的质量得到显著提高.

参考文献

- 1 Aït-sahali Y, Brandt M W. Variable selection for portfolio choice. *J Finance*, 2001, 56: 1297–1351
- 2 Akaike H. A new look at the statistical model identification. *IEEE Trans Automat Control*, 1974, 19: 716–723
- 3 Athey S, Tibshirani J, Wager S. Generalized random forests. *Ann Statist*, 2019, 47: 1148–1178
- 4 Bai C E, Li Q, Ouyang M. Property taxes and home prices: A tale of two cities. *J Econometrics*, 2014, 180: 1–15
- 5 Bai J, Ng S. Determining the number of factors in approximate factor models. *Econometrica*, 2002, 70: 191–221
- 6 Bove V, Elia L, Smith R P. On the heterogeneous consequences of civil war. *Oxf Econ Pap*, 2017, 69: 550–568
- 7 Cai Z. Recent developments in estimating treatment effects for panel data. *China J Econometrics*, 2021, 1: 232–249
- 8 Cai Z, Fang Y, Lin M, et al. Testing conditional independence in causal inference for time series data. *Stat Neerl*, 2024, 78: 397–426
- 9 Callaway B, Li T. Quantile treatment effects in difference in differences models with panel data. *Quant Econ*, 2019, 10: 1579–1618

- 10 Callaway B, Li T, Oka T. Quantile treatment effects in difference in differences models under dependence restrictions and with only two time periods. *J Econometrics*, 2018, 206: 395–413
- 11 Carvalho C, Masini R, Medeiros M C. ArCo: An artificial counterfactual approach for high-dimensional panel time-series data. *J Econometrics*, 2018, 207: 352–380
- 12 Chen H, Han Q, Li Y, et al. Does index futures trading reduce volatility in the Chinese stock market? A panel data evaluation approach. *J Futures Markets*, 2013, 33: 1167–1190
- 13 Chen T F, Chordia T, Chung S L, et al. Volatility-of-volatility risk in asset pricing. *Rev Asset Pricing Studies*, 2022, 12: 289–335
- 14 Chernozhukov V, Fernández-Val I, Galichon A. Quantile and probability curves without crossing. *Econometrica*, 2010, 78: 1093–1125
- 15 Chernozhukov V, Fernández-Val I, Melly B. Inference on counterfactual distributions. *Econometrica*, 2013, 81: 2205–2268
- 16 Du Z, Zhang L. Home-purchase restriction, property tax and housing price in China: A counterfactual analysis. *J Econometrics*, 2015, 188: 558–568
- 17 Elie-Dit-Cosaque K, Maume-Deschamps V. Random forest estimation of conditional distribution functions and conditional quantiles. *Electron J Stat*, 2022, 16: 6553–6583
- 18 Fan J, Li R. Variable selection via nonconcave penalized likelihood and its oracle properties. *J Amer Statist Assoc*, 2001, 96: 1348–1360
- 19 Firpo S. Efficient semiparametric estimation of quantile treatment effects. *Econometrica*, 2007, 75: 259–276
- 20 Firpo S, Fortin N M, Lemieux T. Unconditional quantile regressions. *Econometrica*, 2009, 77: 953–973
- 21 Fujiki H, Hsiao C. Disentangling the effects of multiple treatments—measuring the net economic impact of the 1995 great Hanshin-Awaji earthquake. *J Econometrics*, 2015, 186: 66–73
- 22 Gasser T, Müller H, Mammitzsch V. Kernels for nonparametric curve estimation. *J Royal Statist Soc Ser B*, 1985, 47: 238–252
- 23 Hall P, Heyde C C. Martingale Limit Theory and Its Applications. New York: Academic Press, 1980
- 24 Hall P, Yao Q. Approximating conditional distribution functions using dimension reduction. *Ann Statist*, 2005, 33: 1404–1421
- 25 Hollstein F, Prokopczuk M. How aggregate volatility-of-volatility affects stock returns. *Rev Asset Pricing Studies*, 2018, 8: 253–292
- 26 Hsiao C, Steve Ching H, Wan S K. A panel data approach for program evaluation: Measuring the benefits of political and economic integration of Hong Kong with the mainland. *J Appl Econometrics*, 2012, 27: 705–740
- 27 Hsu Y C, Lai T C, Lieli R P. Counterfactual treatment effects: Estimation and inference. *J Business Economic Statist*, 2022, 40: 240–255
- 28 Huang D, Schlag C, Shaliastovich I, et al. Volatility-of-volatility risk. *J Financ Quant Anal*, 2019, 54: 2432–2452
- 29 Hurvich C M, Tsai C L. Regression and time series model selection in small samples. *Biometrika*, 1989, 76: 297–307
- 30 Ke X, Chen H, Hong Y, et al. Do China's high-speed-rail projects promote local economy?—New evidence from a panel data approach. *China Economic Rev*, 2017, 44: 203–226
- 31 Ke X, Hsiao C. Economic impact of the most drastic lockdown during COVID-19 pandemic—The experience of Hubei, China. *J Appl Econometrics*, 2022, 37: 187–209
- 32 Koenker R, Bassett G. Regression quantiles. *Econometrica*, 1978, 46: 33–50
- 33 Künsch H R. The jackknife and the bootstrap for general stationary observations. *Ann Statist*, 1989, 17: 1217–1241
- 34 Li K T, Bell D R. Estimation of average treatment effects with panel data: Asymptotic theory and implementation. *J Econometrics*, 2017, 197: 65–75
- 35 Li Q, Long W. Do parole abolition and Truth-in-Sentencing deter violent crimes in Virginia? *Empirical Economics*, 2018, 55: 2027–2045
- 36 Li Y, Zhu J. L_1 -norm quantile regression. *J Comput Graph Statist*, 2008, 17: 163–185
- 37 Masry E. Multivariate local polynomial regression for time series: Uniform strong consistency and rates. *J Time Series Anal*, 1996, 17: 571–599
- 38 Newey W K, West K D. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 1987, 55: 703–708
- 39 Ouyang M, Peng Y. The treatment-effect estimation: A case study of the 2008 economic stimulus package of China. *J Econometrics*, 2015, 188: 545–557
- 40 Pesaran M H. Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica*, 2006, 74: 967–1012
- 41 Rothe C. Nonparametric estimation of distributional policy effects. *J Econometrics*, 2010, 155: 56–70

- 42 Sherwood B, Maidman A. rqPen: Penalized quantile regression. R package version 1.4, 2016
- 43 Tibshirani R. Regression shrinkage and selection via the lasso. *J R Stat Soc Ser B Stat Methodol*, 1996, 58: 267–288
- 44 Wang L, Wu Y, Li R. Quantile regression for analyzing heterogeneity in ultra-high dimension. *J Amer Statist Assoc*, 2012, 107: 214–222
- 45 Xiao Z, Koenker R. Conditional quantile estimation for generalized autoregressive conditional heteroscedasticity models. *J Amer Statist Assoc*, 2009, 104: 1696–1712
- 46 Xing G, Yang S, Liu Y, et al. A note on the Bahadur representation of sample quantiles for α -mixing random variables. *Monatsh Math*, 2012, 165: 579–596
- 47 Yi C, Huang J. Semismooth Newton coordinate descent algorithm for elastic-net penalized Huber loss regression and quantile regression. *J Comput Graph Statist*, 2017, 26: 547–557

Estimating quantile treatment effects for panel data

Zongwu Cai, Ying Fang, Ming Lin & Mingfeng Zhan

Abstract In this paper, we propose a quantile treatment effect model for panel data to characterize the distributional effects of a treatment. We provide an identification condition to estimate the counterfactual quantile for the treated unit. In addition, we establish the asymptotic properties of the proposed quantile treatment effect estimator, together with an extension to the high-dimensional case, and discuss how to choose the control units and covariates using a LASSO (least absolute shrinkage and selection operator) type method. A simulation study is conducted to illustrate our proposed model and modeling method. Finally, we apply the proposed method to estimate the quantile treatment effects of introducing CSI 300 index futures trading on both the log-return and volatility of the stock market in China.

Keywords LASSO method, panel data, nonparametric estimation, quantile regression, treatment effects

MSC(2020) 62F12, 62P20, 62G05

doi: 10.1360/SSM-2023-0271