

基于可加逻辑回归的二元内生处理效应估计及其应用*

高 扬 周 玮 钟 威

内容提要：借助观测数据评估经济政策的因果效应时，处理变量通常具有内生性，即处理变量与回归误差项具有相关性，从而导致回归系数不相合。本文提出基于可加逻辑回归的二元内生处理效应估计方法，具体而言，采用可加逻辑回归模型，有效刻画内生变量与工具变量的非线性关系，同时利用组Lasso方法选择重要工具变量。利用所得倾向得分近似二元内生处理变量的最优工具变量，可以有效降低处理效应估计量的方差。此外，为弥补单选法在模型选择上的不足，采用双选法来识别可能因单选法而被遗漏的重要控制变量，以减少遗漏变量偏误。本文证明所提出的二元内生处理效应估计量具有相合性和渐近正态性，蒙特卡洛模拟显示该方法有良好的有限样本表现。在实证分析方面，应用该方法研究身体健康是否会影响个人收入等问题。

关键词：平均处理效应；内生性；高维工具变量；可加逻辑回归；双选法

DOI: 10.19343/j.cnki.11-1302/c.2025.12.010

中图分类号：F224.7 **文献标识码：**A **文章编号：**1002-4565(2025)12-0128-11

Binary Endogenous Treatment Effect Estimation Based on Additive Logistic Regression and Its Application

Gao Yang Zhou Wei Zhong Wei

Abstract: When estimating the causal effects of economic policies on the outcome variable using observational data, treatment variables often exhibit endogeneity, meaning that the treatment variable and the regression error term are correlated, resulting in the inconsistency of regression coefficients. This study proposes a method for estimating binary endogenous treatment effects based on an additive Logistic regression model. Specifically, this study employs the additive Logistic regression model, which effectively characterizes the nonlinear relationship between the binary endogenous variable and instrumental variables in the first stage. Meanwhile, this study utilizes the group Lasso method to select important instrumental variables. This study approximates the optimal instrumental variables for binary endogenous treatment variables using the estimated propensity score, which effectively reduces the variance of the treatment effect estimator. Furthermore, to address the issue of omitted variable bias arising from imperfect model selection under the single variable selection method, the proposed method employs a double-selection approach to identify important control variables. Theoretically, this study demonstrates that the proposed estimator for the binary endogenous treatment effect is consistent and asymptotically normal. Monte Carlo

*基金项目：全国统计科学研究重大项目“基于统计机器学习的因果推断和政策评估”（2022LD08）；国家自然科学基金面上项目“异质性数据的统计建模”（12371265）；国家自然科学基金青年项目“因果链图模型的统计学习方法及其应用”（12501381）；安徽省高校人文社会科学重点项目“时变结构下许多弱工具变量稳健估计方法及应用研究”（2023AH050219）。

simulations illustrate the method's excellent finite sample performance. In empirical analysis, this study applies the proposed method to examine whether physical health affects individual income.

Key words: Average Treatment Effect; Endogeneity; High-dimensional Instrumental Variables; Additive Logistic Regression; Double Selection

一、引言与文献综述

在实证研究中，研究者通常基于大规模的观测数据集，研究作为处理变量的经济政策与感兴趣的目标结果变量之间是否存在因果关系，并评估处理效应。然而，在现实中可能存在混杂变量无法全部观测、测量误差或者样本选择偏差等问题，这会导致处理变量具有内生性。工具变量法（Instrumental Variable, IV）是有效解决内生性问题最常用且最重要的方法之一。该方法通过选择工具变量，使其满足与处理变量相关、不直接影响结果变量且与未观测的混杂变量独立的条件，从而消除混杂偏差。因此，工具变量法能够实现对处理效应的无偏估计，从而实现有效的因果推断，为政策制定者提升决策质量和效率提供科学依据。

（一）高维工具变量的研究进展

为有效估计内生处理效应，学术界开展大量系统性的研究。例如，Angrist和Krueger（1991）研究接受中学教育对收入影响的问题时，提出使用出生季度作为工具变量，以解决家庭或社会背景等一些不可观测的混杂因素导致完成中学教育这一处理变量的内生性问题。Angrist和Imbens（1996）利用工具变量识别依从者所在组的平均处理效应，解决因选择偏差导致的内生性问题。Frölich（2007）进一步提出，工具变量处理效应估计量的有效性取决于对内生处理变量条件期望函数的准确估计。这一条件期望函数本质上是Amemiya（1974）所定义的最优工具变量的一种变形。为更好地逼近最优工具变量，常见的方法是利用原始工具变量的高次方项与交互项在内的大量潜在工具变量，以提高处理效应估计的精确度。然而，不相关工具变量的使用会导致估计结果出现较大偏误。关于如何合理且高效地利用更多工具变量提高处理效应的估计精度问题。早期开创性研究可追溯到Klock和Mennes（1960），运用主成分分析方法降低工具变量维度。近年来，随着高维统计方法的发展，高维工具变量问题得到学界的广泛关注。近代表性工作包括Belloni等（2012）提出post-LASSO方法使用高维工具变量预测最优工具变量；Kueck等（2023）沿袭该方法，利用基于L2-boosting的机器学习方法，分别通过迭代和正交方法从高维工具变量中筛选出有效工具变量，进而估计处理效应。Belloni等（2022）将一维或固定维度内生变量拓展至高维情形。然而针对高维内生变量，模型中通常需要引入大量工具变量，从而导致过度识别问题。为此，Cheng等（2024）提出基于sieve的广义矩估计方法来估计高维面板数据模型。

梳理文献发现，尽管高维工具变量方法已取得一些进展，但上述研究均考虑连续型内生处理变量。然而，在政策评估问题中，作为处理变量的经济政策通常是二元离散型变量。例如，本文实证部分探讨身体健康状态对收入的处理效应，身体是否健康这一处理变量是二元的，且不可观测的个体特征与家庭背景因素会导致该处理变量是内生的；分析举办奥运会对经济发展的处理效应，是否举办奥运会这一处理变量也是二元的，同时奥运会举办城市的选取通常是非随机的，这将不可避免地导致处理变量具有内生性（Zhong等，2022）。因此，本文拟探讨高维工具变量下的二元内生处理效应的有效估计方法，具有一定现实意义。

由于最优工具变量的取值是内生处理变量在给定有效工具变量下的条件期望（Amemiya, 1974），

且此时内生处理效应估计量的渐近方差达到最小。因此,最优工具变量本质上是关于工具变量的函数。一般地,使用非线性模型刻画最优工具变量与工具变量之间的关系是较为合理的做法。已有研究中,Frölich(2007)提出用非参数核方法来估计内生变量的条件期望函数,Fan和Zhong(2018)提出基于非参数可加模型的最优工具变量估计量,Hartford等(2017)提出用深度神经网络来估计最优工具变量以提高估计量的有效性,Bennett等(2019)进一步提出深度广义距估计法。但是基于深度学习的非参数模型存在“黑箱”问题,难以说明每个工具变量的作用和影响,其工具变量的产生不具有可解释性。特别地,在线性关系成立的假设下,Zhong等(2022)考虑二元内生处理变量,通过假设最优工具变量的对数比与工具变量呈现线性关系,采用逻辑回归模型估计最优工具变量。为更加灵活地拟合非线性关系,降低模型错误设定的风险,本文推广Zhong等(2022)的方法,将逻辑回归模型拓展至可加逻辑回归模型(Hastie等,2009)来估计最优工具变量。这一做法的优势在于保留模型可解释性的同时能够捕捉变量之间的非线性关系。

(二) 高维控制变量的研究进展

此外,为得到处理效应的准确估计,研究者通常在模型中引入大量或高维控制变量以捕捉潜在的混杂因素。当这些控制变量能够刻画影响处理效应的分配机制时,则视为强可忽略性(Abrevaya等,2015)成立。然而,在实证研究中,关于如何筛选合适的控制变量,尚未形成一个明确的准则(Donohue和Levitt,2001)。已有文献使用变量选择方法识别有效的控制变量。例如,Belloni等(2014)提出双选法(Double Selection)来估计具有高维控制变量的平均处理效应。该方法结合两个变量选择步骤筛选出与结果变量和处理变量相关的控制变量,利用最小二乘法估计处理效应。该方法的优点在于能有效规避单选法因模型选择不完美而导致的遗漏变量偏误问题,缺点在于没有考虑处理变量的内生性问题。为解决高维控制变量下的内生性问题,Chernozhukov等(2015)假设内生变量与高维工具变量之间满足线性关系,利用高维控制变量以及Neyman正交矩条件估计内生处理效应。本文考虑二元内生处理变量,利用可加逻辑回归模型刻画处理变量与工具变量之间的非线性关系。然而,非线性关系不满足Neyman正交矩条件的设定,如果继续使用该方法会导致估计量的渐近方差增大,从而降低推断效率。因此,建立高维控制变量下二元内生处理效应的有效估计方法是本文重点解决的另一问题。

为有效解决在高维工具变量和控制变量的情况下二元内生处理效应的估计问题,本文创新性地提出采用可加逻辑回归模型来刻画工具变量与内生变量之间的非线性关系,从而能够准确估计二元内生处理变量的最优工具变量,具有提高处理效应估计精度的优点;同时利用双选法从大量潜在控制变量中筛选出重要的控制变量,以减少遗漏变量偏误。具体地,本文提出针对二元内生处理效应的双选可加逻辑工具变量估计量^①(Double Selection plus Additive Logistic Instrumental Variable Estimator, DS-ALIVE)。特别地,若控制变量是已知且有效的,DS-ALIVE退化为可加逻辑工具变量估计量^②(Additive Logistic Instrumental Variable Estimator, ALIVE)。理论上,DS-ALIVE具有相合性和渐近正态性。蒙特卡洛模拟表明,相比较于其他估计量,本文提出的估计量在偏误与均方误方面表现较好。实证分析为印证个人身体健康状况不佳对其收入水平及字词测试表现具有负面影响的说法提供有力证据。

^①本文将基于双选法选择控制变量和基于可加逻辑回归模型估计最优工具变量的方法得到的内生处理效应估计量简称为双选可加逻辑工具变量估计量。

^②本文将基于可加逻辑回归模型估计最优工具变量的方法得到的内生处理效应估计量简称为可加逻辑工具变量估计量。

二、模型设定和估计方法

(一) 模型设定

本文考虑如下结构模型:

$$Y_i = \alpha_0 D_i + \beta_0^T X_i + \varepsilon_i \quad (1)$$

其中, Y_i 代表结果变量, D_i 为处理变量, α_0 为 D_i 的系数, X_i 表示一个 $p \times 1$ 维的控制变量向量, 其对应的真实系数为向量 β_0 , ε_i 代表随机误差项, i 取值从 1 至 n , n 为样本个数。不失一般性, 假设变量 X_i 与扰动项 ε_i 无关, 即 X_i 为外生变量。式 (1) 进一步假设 D_i 是内生变量, 即 $E(\varepsilon_i | D_i) \neq 0$ 。本文聚焦于二元处理变量的情形, 即当个体 i 属于处理组时, 有 $D_i = 1$; 当个体 i 属于控制组时, 有 $D_i = 0$ 。在条件均值独立假设成立的前提下, 即当 $E(\varepsilon_i | D_i, X_i) = E(\varepsilon_i | X_i)$ 时, 给定控制变量 X_i , 平均处理效应可表示为 $\alpha_0 = E(Y_i | D_i = 1, X_i) - E(Y_i | D_i = 0, X_i)$ 。与 Abrevaya 等 (2015) 在研究中采用的条件均值独立假设不同, 本文着重探讨可忽略性假设不成立 (苗旺等, 2018) 且存在大量控制变量时二元内生处理效应的估计方法, 并论证估计量的理论性质。

本文采用工具变量法来解决变量 D_i 的内生性问题。假设存在一个 q 维的 IV 向量 Z_i , $Z_i = (z_{i1}, z_{i2}, \dots, z_{iq})^T$, 同时 X_i 作为外生变量, 也可视为潜在的工具变量。有效的 IV 需满足如下约束: 一是外生性约束, 即给定 X_i 的条件下, IV 与误差项之间是不相关的; 二是排他性约束, 即 X_i 给定时, IV 不直接影响结果变量; 三是相关性约束, 即 Z_i 和 X_i 与内生变量 D_i 存在相关性 (Angrist 等, 1996; Frölich, 2007)。

条件期望 $E(D_i | Z_i, X_i)$ 被视为最优的工具变量, 且此时使用其作为工具变量能使模型的处理效应估计量实现最小渐近方差 (Amemiya, 1974)。式 (1) 中, 二元处理变量 D_i 的最优工具变量可表达为 $E(D_i | Z_i, X_i) = P(D_i = 1 | Z_i, X_i)$, 即为倾向得分。在估计 D_i 的最优工具变量时, 第一阶段往往遵循传统工具变量法的假设, 即事件发生比的对数 $\log(P(D_i = 1 | Z_i, X_i) / P(D_i = 0 | Z_i, X_i))$ 为工具变量 Z_i 和 X_i 的线性模型。然而, 线性模型在捕捉工具变量的非线性影响方面存在局限, 这将导致相应估计量的方差变大 (见蒙特卡洛模拟结果)。为捕捉变量之间的非线性关系, 本文采用可加逻辑回归模型来逼近最优工具变量。在实际操作中, 利用大量潜在工具变量虽然有利于提升内生处理效应估计的精确度, 但是可能存在工具变量个数超过样本量的问题, 进而使得传统工具变量方法失效。为此, 本文借鉴 Belloni 等 (2012) 的方法, 运用变量选择方法来解决高维工具变量问题。

(二) 低维控制变量-可加逻辑回归工具变量估计量 (ALIVE)

假设给定的控制变量是有效的, 考虑使用可加逻辑回归模型捕捉非线性关系。具体而言, 采用可加逻辑回归模型, 来建立处理变量 D_i 与工具变量 (X_i, Z_i) 之间的函数关系, 即

$$\log \left(\frac{P(D_i = 1 | X_i, Z_i)}{P(D_i = 0 | X_i, Z_i)} \right) = \sum_{j=1}^p f_j(x_{ij}) + \sum_{k=1}^q f_k(z_{ik}) \quad (2)$$

其中, $f_j(\cdot)$ 为第 j 个平滑函数, $X = (X_1, X_2, \dots, X_n)^T = (x_{ij})$, $i = 1, 2, \dots, n$, $j = 1, 2, \dots, p$ 和 $Z = (Z_1, Z_2, \dots, Z_n)^T = (z_{ik})$, $i = 1, 2, \dots, n$, $k = 1, 2, \dots, q$, 且允许维度 $p+q$ 超过 n 。在稀疏性假设下, 仅少数 IV 对 $E(D_i | Z_i, X_i)$ 有解释力, 记 $\mathcal{A} = \{t: f_t(\cdot) \neq 0, t = 1, \dots, p+q\}$ 为具有解释力的 IV 指标集。因此, 当 $t \notin \mathcal{A}$ 时, 有 $f_t(\cdot) = 0$ 。为确保模型的可识别性, 进一步假设所有函数 $f_t(\cdot)$ 已通过中心化处理, 即 $E(f_t(T_{it})) = 0$ 对所有 t 均成立, 其中当 $1 \leq t \leq p$ 时, $T_{it} = x_{it}$; 当 $p+1 \leq t \leq p+q$ 时, $T_{it} = z_{it}$ (Huang 等, 2010)。

为得到式 (2) 中非参成分的估计, 本文采用 B 样条基函数 (B-spline) 来展开未知函数 $f_j(\cdot)$ 和

$f_k(\cdot)$ 。引入一个多项式样条空间 $\mathcal{S}_n$ ，在此空间中，一组标准化基函数由 $\{\phi_{\varpi}, \varpi=1, 2, \dots, m_n\}$ 构成， m_n 为自由度 $L \geq 1$ 和节点数 t 的和， d 是样条阶数，满足 $t \geq d$ 。令第 k 个工具变量为 $\mathbf{Z}^k \in \mathbb{R}^n$ ，控制变量作为工具变量可作类似展开。中心化处理后，工具变量 $\mathbf{Z}^k$ 所对应的 B 样条基函数定义为 $\psi_{\varpi} = \phi_{\varpi}(z_{ik}) - \sum_{i=1}^n \phi_{\varpi}(z_{ik})/n$ 。每一个 $f_{nk} \in \mathcal{S}_n$ ，均可表述为这些中心化 B 样条基函数的线性组合 $f_{nk}(z_{ik}) = \sum_{\varpi=1}^{m_n} \gamma_{k\varpi} \psi_{\varpi}(z_{ik})$ ， $1 \leq k \leq q$ 。在式 (2) 中的单变量函数满足适当平滑性的条件下，通过对系数 $\{\gamma_{k1}, \gamma_{k2}, \dots, \gamma_{km_n}\}$ 的恰当选择 (Stone, 1985)，能够利用 B 样条基函数实现对 $f_k(z_{ik})$ 的有效逼近。式 (2) 可以重新表述为

$$\log(P(D_i=1|\mathbf{X}_i, \mathbf{Z}_i)) \approx \sum_{j=1}^p \sum_{\varpi=1}^{m_n} \gamma_{j\varpi} \psi_{\varpi}(x_{ij}) + \sum_{k=1}^q \sum_{\varpi=1}^{m_n} \gamma_{k\varpi} \psi_{\varpi}(z_{ik}) \quad (3)$$

如果 $q \times m_n > n$ ，在给定数据 $(D_i, \mathbf{X}_i, \mathbf{Z}_i)$ 下，对高维非参可加模型式 (2) 利用 B 样条展开，进而选择重要的工具变量集。令 $\gamma_t = (\gamma_{t1}, \gamma_{t2}, \dots, \gamma_{tm_n})^T$ 为第 t 个工具变量的 m_n 维系数， $\gamma = (\gamma_1^T, \gamma_2^T, \dots, \gamma_{p+q}^T)^T$ 为工具变量 $(\mathbf{X}_i, \mathbf{Z}_i)$ 对应 $m_n(p+q)$ 维系数向量。令 $\mathbf{U}_{it} = (\psi_1(T_{it}), \psi_2(T_{it}), \dots, \psi_{m_n}(T_{it}))^T$ ，如果 $1 \leq t \leq p$ ， $T_{it} = x_{it}$ ，如果 $p+1 \leq t \leq p+q$ ， $T_{it} = z_{it}$ ； $\mathbf{U}_i = (\mathbf{U}_{i1}^T, \mathbf{U}_{i2}^T, \dots, \mathbf{U}_{ip+q}^T)^T$ 为对应于第 i 个样本下 $p+q$ 个工具变量的基函数。令真实基参数为 $\gamma^0 = (\gamma_1^{0T}, \gamma_2^{0T}, \dots, \gamma_{p+q}^{0T})^T$ ，记 $s = |\{t \in \{1, 2, \dots, p+q\} : \gamma_t^0 \neq 0\}|$ ，要求 s 是固定的。为筛选出重要的工具变量，使用如下带适应性组 Lasso 惩罚项的目标函数：

$$L_{n1}(D_i, \mathbf{X}_i, \mathbf{Z}_i; \gamma, \lambda_{n1}) = \sum_{i=1}^n \left(-D_i \gamma^T \mathbf{U}_i + \log(1 + \exp(\gamma^T \mathbf{U}_i)) \right) + \lambda_{n1} \sum_{t=1}^{p+q} \omega_t \|\gamma_t\|_2 \quad (4)$$

其中， $\|\cdot\|_2$ 是向量的 L_2 范数， λ_{n1} 是控制模型复杂度的调节参数， ω_t 是第 t 组的权重系数。组 Lasso 估计量可被用来计算得到权重 ω_t ，即

$$\omega_t = \begin{cases} \|\tilde{\gamma}_t\|_2^{-1}, & \text{若 } \|\tilde{\gamma}_t\|_2 > 0 \\ \infty, & \text{若 } \|\tilde{\gamma}_t\|_2 = 0 \end{cases}$$

其中， $\tilde{\gamma} = (\tilde{\gamma}_1^T, \tilde{\gamma}_2^T, \dots, \tilde{\gamma}_{p+q}^T)^T$ 是最小化如下带惩罚项的目标函数所得的估计结果

$$\tilde{\gamma} = \arg \min_{\gamma} \sum_{i=1}^n \left(-D_i \gamma^T \mathbf{U}_i + \log(1 + \exp(\gamma^T \mathbf{U}_i)) \right) + \lambda_{n2} \sum_{t=1}^{p+q} \|\gamma_t\|_2。$$

最小化式 (4) 可得适应性组 Lasso 估计量，即

$$\hat{\gamma} = (\hat{\gamma}_1^T, \hat{\gamma}_2^T, \dots, \hat{\gamma}_{p+q}^T)^T = \arg \min_{\gamma} L_{n1}(D_i, \mathbf{X}_i, \mathbf{Z}_i; \gamma, \lambda_{n1}) \quad (5)$$

根据式 (5) 的估计结果，筛选出重要的控制变量，并将相应指标集记为 $\hat{\mathcal{I}}_1 = \{j : \|\hat{\gamma}_j\|_2 \neq 0, j=1, \dots, p\}$ ，标记得到的工具变量指标集为 $\hat{\mathcal{I}}_2 = \{k : \|\hat{\gamma}_k\|_2 \neq 0, k=p+1, \dots, p+q\}$ ，工具变量估计指标集记为 $\hat{\mathcal{A}} = \hat{\mathcal{I}}_1 \cup \hat{\mathcal{I}}_2$ 。因此，式 (2) 中单变量平滑函数估计量分别为 $\hat{f}_{nj}(x_{ij}) = \sum_{\varpi=1}^{m_n} \hat{\gamma}_{j\varpi} \psi_{\varpi}(x_{ij})$ 和 $\hat{f}_{nk}(z_{ik}) = \sum_{\varpi=1}^{m_n} \hat{\gamma}_{k\varpi} \psi_{\varpi}(z_{ik})$ ，其中 $j \in \hat{\mathcal{I}}_1$ ， $k \in \hat{\mathcal{I}}_2$ ；二元处理变量 D_i 的最优工具变量估计量为

$$\hat{D}_i = \frac{\exp\left(\sum_{j \in \hat{\mathcal{A}}} \hat{f}_{nj}(x_{ij}) + \sum_{k \in \hat{\mathcal{A}}} \hat{f}_{nk}(z_{ik})\right)}{1 + \exp\left(\sum_{j \in \hat{\mathcal{A}}} \hat{f}_{nj}(x_{ij}) + \sum_{k \in \hat{\mathcal{A}}} \hat{f}_{nk}(z_{ik})\right)} \quad (6)$$

记 D_i 的最优工具变量为 $D_i^* = P(D_i=1|\mathbf{X}_i, \mathbf{Z}_i)$ ，有 $\mathbf{D}^* = (D_1^*, D_2^*, \dots, D_n^*)^T$ 。

令 $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)^T$ ， $\hat{\mathbf{D}} = (\hat{D}_1, \hat{D}_2, \dots, \hat{D}_n)^T$ ，投影矩阵为 $\mathbf{P}_X = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$ ，幂等矩阵 $\mathbf{M} = \mathbf{I}_n - \mathbf{P}_X$ ， $\mathbf{I}_n$ 为 $n \times n$ 的单位矩阵。式 (1) 在给定工具变量 $\mathbf{Z}_i$ 和 $\mathbf{X}_i$ 下取条件期望，得到 $E(Y_i|\mathbf{Z}_i, \mathbf{X}_i) = \alpha_0 D_i^* + \beta_0^T \mathbf{X}_i$ ；令 $\nu_i = Y_i - E(Y_i|\mathbf{Z}_i, \mathbf{X}_i)$ ，进而有 $Y_i = \alpha_0 D_i^* + \beta_0^T \mathbf{X}_i + \nu_i$ ，其中 α_0 是式 (1)

中感兴趣的处理效应。由于 D^* 未能观测, 本文利用 $\hat{D}$ 作为最优工具变量 D^* 的估计, 提出二元处理效应估计量 (Additive Logistic Instrumental Variable Estimator, ALIVE) $\hat{\alpha}_{\text{ALIVE}}$ 如下:

$$\hat{\alpha}_{\text{ALIVE}} = (\hat{D}^T M \hat{D})^{-1} (\hat{D}^T M Y) \quad (7)$$

(三) 高维控制变量-双选可加逻辑回归工具变量估计量 (DS-ALIVE)

在本小节, 考虑式 (1) 中控制变量是高维的情形, 即 X_i 的维度 $p > n$; 且式 (1) 关于控制变量是稀疏的, 即 $\|\beta_0\|_0 \leq \omega$ 。在给定大量潜在控制变量下, 可利用变量选择方法筛选出重要的控制变量。不同于 Lin 等 (2015) 使用的单选方法, 本文采用双选法来处理高维控制变量问题, 这能够有效克服单选方法在应对高维数据时遭遇的局限, 即当系数不满足 beta-min 条件 (Belloni 等, 2012) 时产生的遗漏变量偏误问题。

本部分提出如下 DS-ALIVE 算法。第一步, 给定数据 (D_i, X_i, Z_i) , 沿用 ALIVE 方法, 根据式 (5) ~ (6) 得到二元内生处理变量 D 的最优工具变量估计量 $\hat{D}$ 和控制变量估计指标集 $\hat{I}_1$ 。第二步, 借助数据 (Y_i, X_i) , 采用如下带 L_1 惩罚项的目标函数, 来筛选出与 Y_i 相关的控制变量 $L_{n3}(Y_i, X_i; \beta, \lambda_{n3}) = \sum_{i=1}^n (Y_i - \beta^T X_i)^2 + \lambda_{n3} \|\beta\|_1$ 。其中, λ_{n3} 是正则化参数, $\|\cdot\|_1$ 是向量的 L_1 范数。对上述目标函数求解最小化问题, 令估计量 $\hat{\beta} = \arg \min_{\beta} L_{n3}(Y_i, X_i; \beta, \lambda_{n3})$ 。记由数据 (Y_i, X_i) 筛选得到的控制变量指标集为 $\hat{I}_2 = \{j: |\hat{\beta}_j| \neq 0, j=1, 2, \dots, p\}$ 。至此, 完成双选法的两步控制变量筛选步骤, 记控制变量并集为 $\hat{I} = \hat{I}_1 \cup \hat{I}_2$ 。第三步, 由内生处理变量的预测值 $\hat{D}$ 及双选法所得控制变量的并集 $\hat{I}$, 定义内生处理效应 α_0 的 DS-ALIVE 为 $\hat{\alpha}_{\text{DA}} = (\hat{D}^T M_{\hat{I}} \hat{D})^{-1} (\hat{D}^T M_{\hat{I}} Y)$ 。其中, $X_{\hat{I}} = \{X_j, j \in \hat{I}\}$, $P_{\hat{I}} = X_{\hat{I}} (X_{\hat{I}}^T X_{\hat{I}})^{-1} X_{\hat{I}}^T$, $M_{\hat{I}} = I_n - P_{\hat{I}}$, A_Q 表示矩阵 A 在子标集 Q 中的所有列组合。在处理逻辑回归模型与线性模型的估计任务时, 本文通过最小化拓展贝叶斯信息准则 (EBIC) (Chen 和 Chen, 2008) 来确定参数 λ_n 和 m_n ^①。

三、理论性质

当有效的控制变量给定时, 不需要利用双选方法筛选控制变量, 此时 DS-ALIVE 退化为 ALIVE, 因此, 本节旨在研究高维控制变量下的 $\hat{\alpha}_{\text{DA}}$ 的理论性质。

(一) 假设条件

本文提出如下假设条件。(i) 对于一个半正定矩阵 M , 定义其最小 m 稀疏特征值为 $\varphi_{\min}(m)[M] = \min_{1 \leq |\delta|_0 \leq m} \delta^T M \delta / \|\delta\|^2$, 最大 m 稀疏特征值为 $\varphi_{\max}(m)[M] = \max_{1 \leq |\delta|_0 \leq m} \delta^T M \delta / \|\delta\|^2$ 。存在一个序列 a_n , 当 $a_n \rightarrow \infty$ 时, 对于任意小的正数 $\Delta_n > 0$, $0 \leq k_1 \leq k_2 < \infty$, 存在以下不等式以不小于 $1 - \Delta_n$ 的概率成立: $k_1 \leq \varphi_{\min}(a_n s) [\sum_{i=1}^n X_i X_i^T / n] \leq \varphi_{\max}(a_n s) [\sum_{i=1}^n X_i X_i^T / n] \leq k_2$ 。(ii) $E(v_i^2) < \infty$, $E(\varepsilon_i^2) < \infty$, 其中 $v_i = D_i - D_i^*$, $i=1, 2, \dots, n$ 。(iii) $\max_{1 \leq j \leq \omega} |\sum_{i=1}^n x_{ij} D_i^* / n| < \infty$, $\max_{1 \leq j \leq p} (\sum_{i=1}^n x_{ij}^2 v_i^2 / n)^{1/2} < \infty$ 和 $E(D_i^{*2}) < \infty$ 。(iv) $(\log(p))^2 / n \rightarrow 0$ 。

假设 (i) 是 Belloni 等 (2014) 提出的稀疏特征值假设, 假设 (ii) 规定误差项满足二阶矩的存在性, 假设 (iii) 要求一些重要项的矩条件存在, 假设 (iv) 是关于维度阶数的限制。

(二) 渐近性质

下面定义符号, 对于两个序列 $\{a_n, n=1, 2, \dots\}$, $\{b_n, n=1, 2, \dots\}$, 如果存在常数 $0 < c_1 < c_2 < \infty$ 使得

①因篇幅所限, 拓展贝叶斯信息准则内容以附录 1 展示, 见《统计研究》网站所列附件。下同。

$c_1 \leq a_n/b_n \leq c_2$ 对于足够大的 n 成立, 记作 $a_n \asymp b_n$ 。

引理3.1^①: 令初始值为组Lasso估计量 $\hat{\gamma}$, 其中 $\lambda_{n2} \asymp c_3 \sqrt{m_n} \sqrt{(\eta_n + \log(pm_n))/n}$, c_3 为正值常数, η_n 为发散序列; $m_n \asymp n^{1/(2d+1)}$ 。若假设 (A) ~ (D) 成立^②, $\lambda_{n1} \leq O(n^{1/2-(2d-1/2)/(2d+1)})$ 成立, 则适应性组Lasso估计量可实现模型选择, 结论如下:

$$P(\hat{\mathcal{A}} = \mathcal{A}) \rightarrow 1, n \rightarrow \infty \quad (8)$$

$$\sum_{t \in \mathcal{A}} \|\hat{\gamma}_t - \gamma_t^0\|_2^2 = O_p \left(n^{-\frac{2d-1}{2d+1}} \log(n) \right) \quad (9)$$

$$\sum_{t \in \mathcal{A}} \|\hat{f}_t - f_t\|_2^2 = O_p \left(n^{-\frac{2d}{2d+1}} \log(n) \right) \quad (10)$$

$$\|(\hat{\mathbf{D}} - \mathbf{D}^*)/\sqrt{n}\|_2 = O_p \left(n^{-\frac{2d-2}{2d+1}} \log(n) \right) \quad (11)$$

引理3.1中式(8)~(10)拓展Yang和Maiti (2022) 中定理2的结论, 此结果类似于Huang等 (2010) 中引论4在广义线性模型中连接函数取恒等时的特例结果; 式(11)给出二元内生处理变量最优工具变量估计量的收敛速率。

定理3.1^③: 在引理3.1和假设 (A) 至假设 (D) 下, 二元内生处理效应 α_0 的 $\hat{\alpha}_{DA}$ 是 $\sqrt{n}$ 相合的和渐近正态的, 有 $\sigma_n^{-1} \sqrt{n}(\hat{\alpha}_{DA} - \alpha_0) \rightarrow N(0, 1)$, 当 $n \rightarrow \infty$ 。在结构模型误差项分别为条件异方差与条件同方差时, 渐近方差 σ_n^2 的表达式分别为

$$\sigma_n^2 = \left(E(\mathbf{D}^{*T} \mathbf{M}_{\mathcal{I}} \mathbf{D}^*) \right)^{-1} E(\mathbf{D}^{*T} \mathbf{M}_{\mathcal{I}} \mathbf{D}^* v_i^2) \left(E(\mathbf{D}^{*T} \mathbf{M}_{\mathcal{I}} \mathbf{D}^*) \right)^{-1};$$

$$\sigma_n^2 = \left(E(\mathbf{D}^{*T} \mathbf{M}_{\mathcal{I}} \mathbf{D}^*) \right)^{-1} \sigma_v^2.$$

定理3.1给出二元内生处理效应的DS-ALIVE在大样本下的相合性和渐近分布。特别地, 在结构模型误差项满足条件同方差的情况下, 即 $E(v_i^2) = \sigma_v^2$, 估计量的渐近方差为 $\hat{\sigma}_n^2 = (\hat{\mathbf{D}}^T \mathbf{M}_{\mathcal{I}} \hat{\mathbf{D}}/n)^{-1} (Y - \hat{\alpha}_{DA} \mathbf{D} - X_{\mathcal{I}} \hat{\beta}_{fit})^T (Y - \hat{\alpha}_{DA} \mathbf{D} - X_{\mathcal{I}} \hat{\beta}_{fit})/n$, 其中 $\hat{\beta}_{fit} = (X_{\mathcal{I}}^T X_{\mathcal{I}})^{-1} X_{\mathcal{I}}^T Y$ 为变量选择后再拟合的估计量。在统计推断中, 二元内生处理效应 α_0 的渐近置信区间为 $(\hat{\alpha}_{DA} - \rho_{\tau/2} \hat{\sigma}_n / \sqrt{n}, \hat{\alpha}_{DA} + \rho_{\tau/2} \hat{\sigma}_n / \sqrt{n})$, 其中 $\rho_{\tau/2}$ 为标准正态分布的上 $\tau/2$ 分位数。

四、蒙特卡洛模拟

本节用蒙特卡洛模拟来评估二元内生处理效应DS-ALIVE的有限样本性质。考虑如下结构模型 $Y_i = \alpha_0 D_i + \beta_0^T X_i + \varepsilon_i$, 其中 Y_i 为目标变量, D_i 为二元内生处理变量, $\alpha_0 = 0.75$ 是 D_i 的真实系数, X_i 为一个 p 维的协变量向量, $\beta_0 = (3, 0.15, 0.18, 1.5, 2, 0, \dots, 0)^T \in \mathbb{R}^p$ 为 X_i 的真实参数向量。基于下面两个模型, 生成处理变量 D_i 。模型1为 $\text{logit}(P(D_i = 1 | X_i, Z_i)) = 2.5x_{i2} + 2.5x_{i3} + 5z_{i1} + v_i$, 模型2为 $\text{logit}(P(D_i = 1 | X_i, Z_i)) = 2.5x_{i2} + 1.5x_{i2}^2 + 2.5x_{i3} + 5\sin(\pi z_{i1}) + v_i$ 。控制变量 X_i 来自多元正态分布 $N(0, \Sigma_x)$, 其中 $\Sigma_x = (\rho_{mn})_{p \times p}$, $\rho_{mn} = 0.5^{|m-n|}$, $m, n = 1, 2, \dots, p$, $p = 200$; 产生工具变量 Z_i 来自多元正态分布 $N(0, \Sigma_z)$, 其中 $\Sigma_z = (\rho_{mn})_{q \times q}$, $\rho_{mn} = 0.5^{|m-n|}$, $m, n = 1, 2, \dots, q$, $q = 20$ 。采用二元正态分布

①因篇幅所限, 引理3.1的理论证明以附录2展示。

②因篇幅所限, 假设 (A) ~ (D) 以附录3展示。

③因篇幅所限, 定理3.1的理论证明以附录4展示。

生成随机扰动项 (ε_i, v_i) ，这些扰动项的均值为0、方差均为1且协方差为0.9，这一特性保证处理变量 D_i 的内生性。选取样本量 $n \in \{100, 300\}$ 。借助 $R=1000$ 次模拟，考察各个估计量的有限样本表现，计算估计量的偏误（简称偏误），即 $R^{-1} \sum_{r=1}^R (\hat{\alpha}_r - \alpha_0)$ ，均方误差（简称均方误），即 $R^{-1} \sum_{r=1}^R (\hat{\alpha}_r - \alpha_0)^2$ ，其中 $\hat{\alpha}_r$ 表示 α_0 的第 r 次模拟估计值。

为获取各工具变量的B样条展开，利用R中splines包所提供的 $\text{bs}(z, \text{degree}, \text{knots})$ 函数。在此过程中，逐步增加自由度，从1至7，并且设定节点knots为NULL或各工具变量的25%、50%、75%分位数。运用BIC或EBIC信息准则来确定节点和自由度。估计可加逻辑回归模型时，使用了R程序包grpreg中 $\text{grpreg}(z, D, \text{group}, \text{penalty} = "grLasso")$ 函数，利用glmnet程序包中的glmnet函数实现对控制变量的筛选。在模拟实验中，对处理效应 α_0 的7种不同估计方法进行对比。这些方法具体为，最小二乘估计法（OLS），两阶段最小二乘法（2SLS），Belloni等（2012）提出的适用于高维线性工具变量模型的post-LASSO方法，Fan和Zhong（2018）提出的基于高维可加模型的工具变量方法（NAIVE），Zhong等（2022）提出的基于高维逻辑回归模型的工具变量估计方法（LIVE）以及双选逻辑回归工具变量估计量^①（DS-LIVE），ALIVE和DS-ALIVE。

本节首先探讨模拟情形一，该情形假设 x_2, x_3, z_1 是三个已知且有效的工具变量，旨在通过模拟验证可加逻辑回归模型在预测 D_i 的最优工具变量方面的功效，并评估双选法在解决遗漏变量偏误问题上的关键作用。7个估计量的偏误及均方误数据^②可知，OLS估计结果表现较差，所以其不能有效解决内生性问题。在模型1和模型2下，发现相较于OLS，2SLS在偏误方面的表现有所提升，这是由于2SLS使用工具变量来克服处理变量的内生性问题。然而，2SLS的均方误相对较大，理由是2SLS对二元变量 D_i 的预测值常超出合理范围（0至1的区间），并且仅考虑工具变量的线性作用。在模型1和模型2中，相较于本文提出的DS-ALIVE，NAIVE的偏误和均方误均较大，究其原因使用了单选法筛选变量。单选法下被遗漏的变量使得 x_2, x_3, z_1 不再满足IV的外生性约束，进而导致LIVE和ALIVE的偏误较大。特别地，在模型1中，DS-LIVE的表现达到最优；然而，在模型2中，DS-LIVE用线性模型预测 D_i 时损失一些工具变量的信息，其偏误和均方误均不可避免地增大。相比之下，DS-ALIVE的表现上述两个模型中均达到最优水平。值得注意的是，在模型1中，DS-ALIVE在偏误和均方误方面的表现与DS-LIVE非常接近，这得益于DS-ALIVE选择与线性模型相适应的自由度和节点。

接下来，考察模拟情形二，即控制变量与工具变量的有效性事先均不明确，同时变量个数高于样本量，这是高维数据情形。7种方法的估计结果和评价指标^③显示，在模型1与模型2中，OLS方法展现出较大的偏误与均方误；而post-LASSO方法尽管偏误相对较小，但其均方误最大；NAIVE方法在偏误与均方误方面均表现不佳。由于单选法忽略关键控制变量 x_2, x_3 ，而这些变量与筛选出的工具变量 x_2, x_3, z_1 相关，导致这些工具变量不再是外生的，因此LIVE和ALIVE在偏误方面表现不好。在模型1下，DS-ALIVE和DS-LIVE的有限样本表现几乎一样。然而，当线性模型假设不成立时，DS-ALIVE相比较于DS-LIVE，在偏误与均方误方面均更小。随着样本数量的增加，DS-ALIVE的有限样本表现得到进一步改善。这为定理3.1中关于DS-ALIVE大样本性质的论断提供了证据支持。

① 本文将基于双选法选择控制变量和基于逻辑回归模型估计最优工具变量的方法得到的内生处理效应估计简称为双选逻辑回归工具变量估计量。

② 因篇幅所限，模拟情形一中各种估计量的偏误和均方误以附表1展示。

③ 因篇幅所限，模拟情形二中各种估计量的偏误和均方误以附表2展示。

进一步地,随着潜在工具变量维度 p 与控制变量维度 q 逐步增加,DS-ALIVE在偏误和均方误方面仍然表现最优^①。

五、应用研究

本节利用DS-ALIVE估计身体健康状态对个人收入的影响。在探讨健康对收入的效应时,众多学者已尝试多种方法来应对模型中可能存在的内生性问题。考虑到仅有非农就业者的收入数据,魏众(2004)采用Heckman两步法来纠正由非随机样本选择所引起的偏误。刘国恩等(2004)进一步采用面板固定效应模型,试图消除不可观测的家庭背景差异可能带来的内生性问题。而高梦滔和姚洋(2005)则在研究中选取前定变量的滞后项作为健康状态的工具变量。然而,上述文献均未考虑高维工具变量和控制变量的筛选问题。

借助中国家庭追踪调查(China Family Panel Survey, CFPS)2020年数据,本部分将探讨身体健康状况对个人收入的效应,样本数量为29155^②。为增强分析结果的稳健性,将进一步考察健康水平对字词测试表现的作用。在CFPS数据集中,被调查者对自身健康状况的评价被划分为5个等级,依次为“非常健康”“很健康”“比较健康”“一般”和“不健康”。为区分个体在初始健康水平上的差异,本研究基于这一自我评价结果,将所有观测对象分为两个组别。具体的划分标准如下,若被调查者的自我评价为“非常健康”或“很健康”,则将其归为健康组;若其自我评价为“比较健康”“一般”或“不健康”,则将其归为不健康组。此外,数据集中个人收入和字词测试得分均为连续型变量。上述变量的属性与特征,与本文的研究设定一致。

考虑以下的实证计量模型, $\log(\text{income}_i) = \alpha_0 D_i + \beta_0^T X_i + \varepsilon_i$ 。其中, $\log(\text{income}_i)$ 代表个人收入 income_i 的对数值, D_i 表示个体 i 的身体健康水平。具体来说,若被调查者位于健康组,则设定 D_i 为0,反之,若其位于不健康组,则 D_i 设定为1。此外, X_i 代表一系列其他控制变量,涵盖性别、受教育程度、年龄、体重、是否居住在城市、所处地理位置、是否患有慢性病以及方圆5公里内是否有高污染企业等因素,且考虑这些变量的最高3次方项。

由于健康状况可能会影响收入水平,而收入也可能反作用于健康水平,且自评健康状况不够客观,因此,在上述计量实证模型中,健康状况被视作内生处理变量。为解决此问题,本研究引入2010年CFPS基线调查数据中的“至最近医疗设施距离”这一变量,假设该变量不随时间变化。寻找到的潜在工具变量包括居住环境因素(如至最近医疗设施的距离)、生活行为特征(如是否吸烟、是否饮酒、是否参与体育锻炼)以及生活环境质量(如烹饪用水是否为自来水、烹饪燃料是否为清洁能源、家庭卫生间是否具备冲水功能、家庭垃圾是否由专人收集处理)等。

本文已经找到影响身体健康状况的许多潜在工具变量,但仍需进一步探究这些变量与健康水平之间的相关性。另外,至最近医疗设施距离等变量可能对健康水平产生非线性的影响(李华等,2011)。鉴于此,本文采用如下的可加逻辑回归模型来描述健康状况变量与潜在工具变量之间的关系, $\log\left\{P(D_i=1|Z_i)/(1-P(D_i=1|Z_i))\right\} = \mu + \sum f_k(z_{ik})$ 。其中, $f_k(z_{ik})$ 为工具变量 z_{ik} 的一个未知函数。对任意的 k 值,允许相应的 $f_k(\cdot)$ 在基函数展开时选择不同的自由度和节点,以适应不同的平滑需求。

在上述模型估计中,首先将从家到最近医疗点距离这一变量的平滑函数进行B样条展开,运用BIC信息准则,选择相应函数以节点数1进行展开。然后用相应的基函数展开值乘以基参数估计值作

①因篇幅所限,DS-ALIVE估计量随工具变量和控制变量维度逐步增加后的有限样本表现以附表3与附表4展示。

②因篇幅所限,变量的描述性统计信息以附表5展示。

为单变量函数的预测值。家到最近医疗点距离与单变量函数估计值之间的关系是非线性的^①。探究健康与收入及字词测试得分关系的研究中, 双选方法发现慢性病和居住地高污染企业是重要控制变量。

回归结果显示^②, OLS在估计时忽略模型可能遭受的内生性问题, 因此, 其估计结果存在偏差。在探究健康状况对收入及字词测试表现的影响时, 本部分运用DS-ALIVE、ALIVE、DS-LIVE、LIVE、NAIVE、post-LASSO、OLS多种方法。具体而言, 在考察健康状况对收入的作用, 及其对字词测试表现的影响时, DS-ALIVE所得的健康状况系数估计量的 t 值绝对值均是最大的。这一统计结果表明, 相较于其他方法, DS-ALIVE为身体健康状态对个人收入和字词测试得分的非零影响提供更强的统计显著性证据。产生这一现象的原因是, 相较于DS-LIVE, DS-ALIVE使用可加逻辑回归模型, 通过识别健康水平与至最近医疗设施距离之间的非线性关系, 成功捕捉到更多关于最优工具变量的信息。而与NAIVE、LIVE及ALIVE采用单选法可能导致关键控制变量遗漏不同, DS-ALIVE凭借双选方法有效避免遗漏变量偏误的问题。然而, post-LASSO方法不仅忽视了至最近医疗设施距离与健康水平之间的非线性关系, 而且在预测最优工具变量时, 所得值常偏离 $[0,1]$ 的合理范围。因此, 本文所提出的DS-ALIVE, 为健康状况对收入产生显著影响的观点提供更坚实的证据支持。

六、结论

本文提出使用可加逻辑回归模型近似二元内生处理变量的最优工具变量, 用适应性组Lasso筛选工具变量后得到二元内生处理效应的ALIVE; 进一步在高维控制变量情况下, 采用双选方法筛选控制变量得到二元内生处理效应的DS-ALIVE。本文证明DS-ALIVE满足相合性和渐近正态性, 并采用数值模拟方法验证ALIVE和DS-ALIVE两种方法具有良好的有限样本性质。此外, 借助中国家庭追踪调查数据, 实证研究运用DS-ALIVE为证实身体不健康状态对个人收入具有负面影响的说法提供了有力的证据。

在众多经济社会现象中, 政策变量往往呈现出有序的特征。例如, 医生为病人进行治疗时, 所采用的药物干预剂量就是有序的。基于此, 未来研究可进一步深入探讨如何运用高维工具变量与控制变量来估计内生有序处理效应问题, 同时研究处理效应估计量的渐近理论性质。

参考文献

- [1] 刘国恩, William H D, 傅正泓. 中国的健康人力资本与收入增长[J]. 经济学(季刊), 2004, 4: 101-118.
- [2] 李华, 俞卫, 赵大海. 农村居民健康影响因素分析——Grossman人力资本扩展模型及其应用[A]. “中国医疗改革与老年健康、福利跨学科研究”学术研讨会, 2011: 19-30.
- [3] 苗旺, 刘春辰, 耿直. 因果推断的统计方法[J]. 中国科学: 数学, 2018, 48(12): 1753-1778.
- [4] 高梦滔, 姚洋. 健康风险冲击对农户收入的影响[J]. 经济研究, 2005, 12: 15-25.
- [5] 魏众. 健康对非农就业及其工资决定的影响[J]. 经济研究, 2004, 2: 64-74.
- [6] Abrevaya J, Hsu Y C, Lieli R P. Estimating Conditional Average Treatment Effects[J]. Journal of Business and Economic Statistics, 2015, 33: 485-505.
- [7] Amemiya T. The Nonlinear Two Stage Least Squares Estimator[J]. Journal of Econometrics, 1974, 2(2): 105-110.
- [8] Angrist J D, Keueger A B. Does Compulsory School Attendance Affect Schooling and Earnings?[J]. The Quarterly Journal of Economics, 1991, 106(4): 979-1014.
- [9] Angrist J D, Imbens G W, Rubin D B. Identification of Causal Effects Using Instrumental Variables[J]. Journal of the American Statistical Association, 1996, 91(434): 444-455.
- [10] Belloni A, Chen D, Chernozhukov V, et al. Sparse Models and Methods for Optimal Instruments with an Application to Eminent Domain[J].

①因篇幅所限, 家到最近医疗点距离与单变量函数的估计值之间的关系图以附图1展示。

②因篇幅所限, 身体健康状态对收入或字词测试得分的影响结果以附表6展示。

- Econometrica, 2012, 80(6): 2369–2429.
- [11] Belloni A, Chernozhukov V, Hansen C. Inference on Treatment Effects after Selection among High-dimensional Controls[J]. The Review of Economic Studies, 2014, 81(2): 608–650.
- [12] Belloni A, Hansen C, Newey W. High-dimensional Linear Models with Many Endogenous Variables[J]. Journal of Econometrics, 2022, 228(1): 4–26.
- [13] Bennett A, Kallus N, Schnabel T. Deep Generalized Method of Moments for Instrumental Variable Analysis[J]. In Advances in Neural Information Processing Systems, 2019, 3559–3569.
- [14] Chen J, Chen Z. Extended Bayesian Information Criterion for Model Selection with Large Model Space[J]. Biometrika, 2008, 94: 759–771.
- [15] Chernozhukov V, Hansen C, Spindler M. Post Selection and Post Regularization Inference in Linear Models with Many Controls and Instruments[J]. American Economic Review: Papers & Proceedings, 2015, 105(5): 486–490.
- [16] Cheng T, Dong C, Gao J, et al. GMM Estimation for High-dimensional Panel Data Models[J]. Journal of Econometrics, 2024, 244(1): 105853.
- [17] Donohue III J J, Levitt S D. The Impact of Legalized Abortion on Crime[J]. Quarterly Journal of Economics, 2001, 116: 379–420.
- [18] Fan Q, Zhong W. Nonparametric Additive Instrumental Variable Estimator: A Group Shrinkage Estimation Perspective[J]. Journal of Business and Economic Statistics, 2018, 36: 388–399.
- [19] Frölich M. Nonparametric IV Estimation of Local Average Treatment Effects with Covariates[J]. Journal of Econometrics, 2007, 139(1): 35–75.
- [20] Hastie T, Tibshirani R, Friedman J H, et al. The Elements of Statistical Learning: Data Mining, Inference and Prediction[M]. New York: Springer, 2009: 1–758.
- [21] Hartford J, Lewis G, Leyton-Brown K, et al. Deep IV: A Flexible Approach for Counterfactual Prediction[C]. International Conference on Machine Learning. PMLR, 2017: 1414–1423.
- [22] Huang J, Horowitz J, Wei F. Variable Selection in Nonparametric Additive Models[J]. The Annals of Statistics, 2010, 38(4): 2282–2313.
- [23] Kloeck T L, Mennens. Simultaneous Equations Estimation Based on Principal Components of Predetermined Variables[J]. Econometrica, 1960, 28(1): 46–61.
- [24] Kueck J, Luo Y, Spindler M, et al. Estimation and Inference of Treatment Effects with L2-boosting in High-dimensional Settings[J]. Journal of Econometrics, 2023, 234(2): 714–731.
- [25] Lin W, Feng R, Li H. Regularization Methods for High-dimensional Instrumental Variables Regression with an Application to Genetical Genomics[J]. Journal of the American Statistical Association, 2015, 110: 270–288.
- [26] Stone C. Additive Regression and Other Nonparametric Models[J]. The Annals of Statistics, 1985, 13(2): 689–705.
- [27] Yang K, Maiti T. Ultra-high Dimensional Generalized Additive Model: Unified Theory and Methods[J]. Scandinavian Journal of Statistics, 2022, 49(3): 917–942.
- [28] Zhong W, Zhou W, Fan Q, et al. Dummy Endogenous Treatment Effect Estimation Using High Dimensional Instrumental Variable[J]. The Canadian Journal of Statistics, 2022, 50(3): 795–819.

作者简介

高扬, 安徽财经大学经济学院讲师。研究方向为计量经济、经济统计。

周玮 (通讯作者), 西南财经大学数据科学与商业智能联合实验室、统计与数据科学学院讲师。研究方向为因果推断、高维统计。电子邮箱: zhouwei23@swufe.edu.cn。

钟威, 厦门大学王亚南经济研究院、经济学院统计学与数据科学系教授、博士生导师。研究方向为高维统计、计量经济。

(责任编辑: 葛雪凝)