

# 超高维数据下半参数工具变量模型的双惩罚正则估计

赵培信<sup>1\*</sup>, 唐新蓉<sup>2</sup>

<sup>1</sup>重庆工商大学数学与统计学院, 重庆

<sup>2</sup>重庆工商大学资产管理处, 重庆

收稿日期: 2022年7月8日; 录用日期: 2022年8月2日; 发布日期: 2022年8月11日

## 摘要

在超高维数据下, 考虑一类含内生性协变量的半参数工具变量模型的估计。结合惩罚估计技术以及SIS特征筛选方法, 提出了一种识别有效工具变量的双惩罚正则估计方法, 并得到了模型参数分量和非参数分量的工具变量调整估计。在一些正则条件下, 证明了参数分量和非参数分量的估计均是相合的。最后通过一些数值仿真模拟研究了所提方法的有限样本性质。

## 关键词

超高维数据, 工具变量, 惩罚估计, 半参数模型, 内生性变量

# Double Penalized Regularization Estimation for Semiparametric Instrumental Variable Models with Ultrahigh Dimensional Data

Peixin Zhao<sup>1\*</sup>, Xinrong Tang<sup>2</sup>

<sup>1</sup>College of Mathematics and Statistics, Chongqing Technology and Business University, Chongqing

<sup>2</sup>Department of Assets Management, Chongqing Technology and Business University, Chongqing

Received: Jul. 8<sup>th</sup>, 2022; accepted: Aug. 2<sup>nd</sup>, 2022; published: Aug. 11<sup>th</sup>, 2022

## Abstract

The estimation of a semiparametric instrumental variable model with endogenous covariates is considered under ultra-high dimensional data. Combined with penalty estimation technology and

\*通讯作者。

**SIS feature screening method, a double penalized regularization estimation method for identifying optimal instrumental variables is proposed, and an optimal instrumental variable adjusted estimator of model parameter and nonparametric components are obtained. Under some regular conditions, it is proved that the estimators of parametric and nonparametric components are consistent. Finally, the finite sample properties of the proposed method are studied by some numerical simulations.**

## Keywords

**Ultra-High Dimensional Data, Instrumental Variable, Penalty Estimation, Semiparametric Model, Endogenous Variables**

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

设  $\{X_i, Z_i, U_i, Y_i\}$ ,  $i = 1, \dots, n$  表示一组容量为  $n$  的样本, 其中  $X_i$ ,  $Z_i$  和  $U_i$  是协变量,  $Y_i$  是响应变量。本文考虑如下带变系数的半参数模型:

$$Y_i = X_i^T \beta + Z_i^T \theta(U_i) + \varepsilon_i, \quad i = 1, \dots, n \quad (1)$$

其中  $\beta = (\beta_1, \dots, \beta_p)^T$  是一个  $p$  维的未知参数向量,  $\theta(U_i) = (\theta_1(U_i), \dots, \theta_k(U_i))^T$  是一个  $k$  维的未知非参数函数系数向量,  $\varepsilon_i$  表示模型误差。模型(1)同时含有参数分量和非参数函数分量, 因此在对实际问题统计建模过程中具有较好的灵活性和广泛的适应性。该模型近来引起了许多学者的广泛关注, 并且对模型中的参数和非参数分量提出了大量的估计方法(参见文[1] [2] [3])。但是以上文献是在假定所有协变量均为外生性协变量的前提下对模型的估计问题进行讨论, 而对高维数据进行统建模中假定所有协变量均为外生性协变量是不切实际的。如果某些协变量具有内生性, 上述文献提出的估计方法给出的估计则不是相合的, 而是会产生一定的内生性偏差。

对于具有内生性变量的模型, 工具变量调整技术则可以提供一个消除内生性偏差的方法, 进而得到模型的相合估计。由于半参数模型具有较好的灵活性和广泛的适应性, 近年来对含内生性变量的半参数模型的统计推断也越来越受到广泛的关注。例如, Cai 和 Xiong [4]对一类含内生协变量的变系数部分线性模型的估计问题, 提出了一种基于工具变量调整的三阶段估计方法。Yuan 等[5]对一类含内生协变量的变系数部分线性模型的变量选择问题, 提出了一种基于工具变量调整的变量选择方法。Yang 等[6]对一类含内生性变量的部分线性单指标模型的经验似然统计推断问题, 提出了一种工具变量调整的经验似然估计方法。

注意到上述文献所提出工具变量调整方法均是在工具变量的维数为固定的低维数据下进行的, 对超高维数据下的半参数工具变量模型的估计问题, 上述给出的基于工具变量调整的估计方法将不能直接使用。为此, 本文讨论当某些协变量为内生性协变量并且工具变量为超高维数据的情况下, 研究模(1)的估计问题。具体地, 假定模型(1)中的  $Z_i$  和  $U_i$  为外生性协变量, 而  $X_i$  为内生性协变量, 并满足:

$$X_i = \Gamma \xi_i + e_i, \quad i = 1, 2, \dots, n \quad (2)$$

其中  $\xi_i$  是一个由工具变量构成的  $q$  维向量,  $\Gamma$  是一个  $p \times q$  维的未知参数矩阵。  $e_i$  是模型误差。注意到  $Z_i$

和  $U_i$  为外生性协变量, 而  $X_i$  为内生性协变量, 那么本文可假定  $Z_i$  和  $U_i$  直接影响响应变量  $Y_i$ , 这表明  $Z_i$  和  $U_i$  均与  $X_i$  相互独立。另外, 在接下来的估计过程中假定协变量  $X_i$  的维数  $p$  是固定的, 而工具变量  $\xi_i$  为超高维工具变量, 即  $\xi_i$  的维数  $q$  远远大于样本容量  $n$ 。进一步, 本文假设未知参数矩阵  $\Gamma$  是稀疏的, 即  $\Gamma$  中只有一小部分参数是非零的, 这也意味着只有小部分工具变量是有效工具变量。

本文的目标是识别并估计  $\Gamma$  中的非零参数, 进而识别有效工具变量。并利用识别出的有效工具变量以及工具变量调整技术对模型(1)中的参数和非参数分量给出一个估计过程。具体地, 首先结合内生协变量  $X_i$  和工具变量  $\xi_i$  的相关关系构建一个辅助的回归模型, 基于构建的辅助回归模型, 并利用 SIS (Sure Independence Screening) 特征筛选方法提出了一种有效工具变量的识别方法。值得一提的是, 注意到  $X_i$  是一个多维向量以及  $\Gamma$  是一个超高维矩阵, 因此经典的 SIS 方法(参见文[7])将不能再直接应用。本文提出了一种基于辅助回归模型的 SIS 特征筛选方法, 并给出了一种基于双惩罚的迭代算法, 改进并推广了已有的 SIS 变量筛选过程。其次, 利用识别出的有效工具变量并采用工具变量调整技术, 对模型(1)中的参数和非参数分量给出了一种基于工具变量调整的估计方法, 并在一些正则条件下证明了所得估计是渐近相合估计量。

## 2. 估计过程及迭代算法

注意到  $X_i$  为内生性协变量,  $\xi_i$  为对应的工具变量, 所以有  $E(\varepsilon_i | X_i) \neq 0$  和  $E(\varepsilon_i | \xi_i) = 0$ 。由第 1 节中的讨论可知外生性协变量  $Z_i$  和  $U_i$  均与内生性协变量  $X_i$  相互独立。那么结合  $X_i = \Gamma \xi_i + e_i$ , 可以证明工具变量  $\xi_i$  也独立于外生性协变量  $Z_i$  和  $U_i$ 。记  $\xi_{is}$  为工具变量  $\xi_i = (\xi_{i1}, \dots, \xi_{iq})^T$  的第  $s$  个分量, 那经计算可得:

$$Cov(Y_i, \xi_{is}) = \sum_{j=1}^p Cov(X_{ij}, \xi_{is}) \beta_j + \sum_{j=1}^k Cov(Z_{ij}, \xi_{is}) \theta_j(U_i) + Cov(\varepsilon_i, \xi_{is}) = \sum_{j=1}^p Cov(X_{ij}, \xi_{is}) \beta_j \quad (3)$$

另外, Bound 等[8]以及 Didelez 等[9]指出, 当  $\xi_{is}$  是有效工具变量时,  $\xi_{is}$  应与内生性变量  $X_i$  高度相关。即如果对所有  $j \in \{1, \dots, p\}$  有  $Cov(X_{ij}, \xi_{is}) = 0$ , 那么  $\xi_{is}$  是一个无效的工具变量。另外注意到  $\beta \neq 0$ , 因此(3)式表明如果  $Cov(Y_i, \xi_{is}) \neq 0$  显著成立, 那么  $\xi_{is}$  则是一个有效工具变量。基于该结论并利用 SIS 特征筛选方法, 本文可以根据响应变量  $Y_i$  和工具变量  $\xi_{is}$ ,  $s \in \{1, \dots, q\}$  之间的边际相关性选择有效工具变量。具体地, 首先定义响应变量  $Y_i$  和工具变量  $\xi_{is}$ ,  $s \in \{1, \dots, q\}$  的一个辅助回归模型如下:

$$Y_i = \omega_1 \xi_{i1} + \omega_2 \xi_{i2} + \dots + \omega_q \xi_{iq} + v_i, \quad i = 1, \dots, n \quad (4)$$

其中  $v_i$  是模型误差满足  $E(v_i | \xi_{i1}, \dots, \xi_{iq}) = 0$ 。对于任给定的  $s \in \{1, \dots, q\}$ , 结合模型(4)并利用类似张等[10]的估计方法, 定义关于分量的边际回归模型为  $Y_i = \omega_s \xi_{is} + v_i$ , 并且  $\omega_s$  的边际最小二乘估计可通过最小化以下目标函数给出。

$$\sum_{i=1}^n Q(Y_i, \omega_s \xi_{is}) \equiv \sum_{i=1}^n (Y_i - \omega_s \xi_{is})^2 \quad (5)$$

通过最小化(5)式, 可以得到边际最小二乘估计量为

$$\hat{\omega}_s^M = \left( \sum_{i=1}^n \xi_{is}^2 \right)^{-1} \sum_{i=1}^n \xi_{is} Y_i \quad (6)$$

结合(4)和(6)式, 则可识别出有效工具变量的子集如下:

$$\hat{M}_{v_n} = \{1 \leq s \leq q : |\hat{\omega}_s^M| \geq v_n\} \quad (7)$$

其中  $v_n$  是预先给定的一个门限值。另外, 利用类似张等[10]的讨论可知, 基于边际回归所识别出的有效工具变量可能会存在一定的选择性偏差。比如, 某一有效工具变量与内生性变量  $X_i$  的某些分量存在高度

相关, 但该工具变量可能与响应变量  $Y_i$  相关性不显著, 进而导致无法识别出该有效工具变量。为了克服这个问题, 本文提出了一个新的基于双惩罚的迭代算法。在如下迭代算法中, 我们还利用识别出的有效工具变量提出了内生性变量的调整过程, 并给出了基于有效工具变量调整的模型参数分量  $\beta$  和非参数分量  $\theta(u)$  的估计过程。具体算法如下: 在此算法中, 本文首先预先确定一个稀疏参数大小  $d = O(n/\ln n)$ 。然后, 算法如下。

### 第一阶段: 有效工具变量识别算法

第 1 步: 对所有的  $s \in \{1, \dots, q\}$ , 由(7)式识别出一个包含  $S_1$  个有效工具变量的子集  $\mathcal{A}_1$ , 其中  $S_1 = \lfloor 2d/3 \rfloor$  以及  $d = n/\ln n$ 。

第 2 步: 记  $\xi_{\mathcal{A}_1}$  为集合  $\mathcal{A}_1$  对应的有效工具变量, 并且模型(2)中的对应系数矩阵则为  $\Gamma_{\mathcal{A}_1}$ 。然后基于模型(2), 通过最小化如下惩罚目标函数来选择  $\mathcal{A}_1$  的子集  $\mathcal{M}_1$ :

$$\frac{1}{n} \sum_{i=1}^n \|X_i - \Gamma_{\mathcal{A}_1} Z_{i, \mathcal{A}_1}\|_2^2 + p_\lambda(\Gamma_{\mathcal{A}_1})$$

其中  $p_\lambda(\cdot)$  是一个惩罚函数。在实际应用中, Lasso 惩罚, SCAD 惩罚以及 MCP 惩罚等惩罚函数均可以使用。令  $\hat{\Gamma}_{\mathcal{A}_1}$  为  $\Gamma_{\mathcal{A}_1}$  的惩罚估计矩阵, 有  $\mathcal{M}_1$  是  $\hat{\Gamma}_{\mathcal{A}_1}$  的非零列对应的指标集。

第 3 步: 对任给定的  $s \in \mathcal{M}_1^c = \{1, \dots, q\} \setminus \mathcal{M}_1$ , 基于模型(4), 最小化如下惩罚目标函数:

$$\sum_{i=1}^n \left[ Y_i - (\xi_{i, \mathcal{M}_1}^\top \omega_{\mathcal{M}_1} + \xi_{is} \omega_s) \right]^2 + p_\lambda(\omega_s)$$

对任意  $s \in \mathcal{M}_1^c$ , 令  $\omega_s$  的估计量为  $\hat{\omega}_s^{(2)}$ , 那么通过排序  $\left\{ \left| \hat{\omega}_s^{(2)} \right| : s \in \mathcal{M}_1^c \right\}$  选择一个大小为  $k_2 = \min(5, d - |\mathcal{M}_1|)$  的指数子集  $\mathcal{A}_2$ 。进一步, 用第 2 步的算法从  $\mathcal{M}_1 \cup \mathcal{A}_2$  中选择一个子集  $\mathcal{M}_2$ 。

第 4 步: 对第 3 步进行迭代直到有  $|\mathcal{M}_l| \geq d$  或  $\mathcal{M}_l = \mathcal{M}_{l-1}$ , 所选的最终子集表示为  $\mathcal{M}_{final}$ 。注意  $\mathcal{M}_{final}$  满足  $|\mathcal{M}_{final}| \leq d < n$ , 因此, 提出的迭代算法可实现降低维数的效果。

### 第二阶段: 基于有效工具变量的内生性协变量的工具变量调整算法

首先用  $\xi^*$  表示  $\mathcal{M}_{final}$  集合中对相应的工具变量。那么  $\xi^*$  则是识别出的有效工具变量组成的一个向量。另外, 从模型(2)中可知  $E\{X\xi^{\top}\} = \Gamma E\{\xi\xi^{\top}\}$ , 那么  $\Gamma$  中有效工具变量对应的参数矩阵的矩估计量可表示为:

$$\hat{\Gamma}^* = \left( \sum_{i=1}^n X_i \xi_i^{*\top} \right) \left( \sum_{i=1}^n \xi_i^* \xi_i^{*\top} \right)^{-1}$$

因此, 基于有效工具变量调整的协变量可定义为  $X^* = \hat{\Gamma}^* \xi^*$ 。

### 第三阶段: 模型参数和非参数分量的估计

记  $B(u) = (B_1(u), \dots, B_L(u))^\top$  为表示阶数为  $m$  的 B 样条基数, 其中  $L = \kappa + m$ ,  $\kappa$  为内部结点的个数。那么  $\theta_s(u)$  可以渐近表示为  $\theta_s(u) \approx B(u)^\top \gamma_s$ , 其中  $\gamma_s = (\gamma_{s1}, \dots, \gamma_{sL})^\top$  是基函数系数向量。记  $W_i = I_k \otimes B(U_i) \cdot Z_i$  以及  $\gamma = (\gamma_1^\top, \dots, \gamma_k^\top)^\top$ , 那么通过最小化如下目标函数来给出  $\beta$  和  $\gamma$  的估计。

$$M_n(\beta, \gamma) = \frac{1}{n} \sum_{i=1}^n (Y_i - X_i^{*\top} \beta - W_i^\top \gamma)^2 \quad (8)$$

记  $\hat{\beta}$  和  $\hat{\gamma}$  为最小化(8)式的解, 那么  $\hat{\beta}$  则是基于工具变量调整的参数分量  $\beta$  估计量, 并且非参数分量  $\theta_s(u)$  的估计可以定义为  $\hat{\theta}_s(u) \approx B(u)^\top \hat{\gamma}_s$ 。

另外在上述算法过程中对非参数函数系进行样条展开时, 内部结点取为等距结点, 并且内部结点个数  $\kappa$  通过最小化如下交叉验证得分函数进行得到。

$$CV(\kappa) = \frac{1}{n} \sum_{i=1}^n \left\{ Y_i - X_i^{*T} \hat{\beta}_{[i]} - Z_i^T \hat{\theta}_{[i]}(U_i) \right\}^2$$

其中  $\hat{\beta}_{[i]}$  和  $\hat{\theta}_{[i]}(u)$  表示删除第  $i$  个样本个体后, 基于(8)式得到的  $\beta$  和  $\theta(u)$  估计量。

### 3. 渐近性质

令  $S(Y_i, \omega_s, \xi_{is}) = \partial Q(Y_i, \omega_s, \xi_{is}) / \partial \omega_s$ , 其中  $Q(\cdot, \cdot)$  由式(5)定义, 那么由(6)式定义的估计量  $\hat{\omega}_s^M$  也是估计方程  $\sum_{i=1}^n S(Y_i, \omega_s, \xi_{is}) = 0$  的解。另外对每一个  $s \in \{1, \dots, q\}$ , 定义  $\omega_s^M$  为  $E\{S(Y_i, \omega_s, \xi_{is})\} = 0$  的解。那么结合(5)式并简单计算可得

$$\omega_s^M = E(\xi_s Y) / E(\xi_s^2) \tag{9}$$

进一步, 记  $\mathcal{M}_* = \{1 \leq s \leq q : \omega_s \neq 0\}$  表示所有有效工具变量构成的指标集, 假设  $|\mathcal{M}_*| = m < q$ , 即有效工具变量只是所有工具变量一个子集。此外令  $\mathcal{B} = \{\omega_s : |\omega_s| \leq B\}$  其中  $B$  为一个给定的常数。为了给出有效工具变量识别以及所得估计量的渐近性质, 首先给出如下的一些正则性条件。

C1: 记  $I(\omega_s^M) = E\{S(Y_i, \omega_s^M, \xi_{is})\}$ , 那么假定  $I(\omega_s^M)_B = \sup_{\omega_s^M \in \mathcal{B}} I(\omega_s^M)$  存在上界  $c_0$ , 其中  $c_0$  表示某一正常数。

C2: 对  $\mathcal{B}$  中任意的  $\omega_s$  和  $\omega'_s$ , 函数  $Q(\cdot, \cdot)$  满足以下利普席茨条件:

$$|Q(Y_i, \omega_s, \xi_{is}) - Q(Y_i, \omega'_s, \xi_{is})| \leq c_1 |\xi_{is}| |\omega_s - \omega'_s|$$

其中  $c_1$  表示某一正常数。

C3: 对所有的  $\omega_s \in \mathcal{B}$ , 函数  $Q(Y_i, \omega_s, \xi_{is})$  关于  $\omega_s$  是凸函数且满足:

$$\left| E\{Q(Y_i, \omega_s, \xi_{is}) - Q(Y_i, \omega_s^M, \xi_{is})\} \right| \geq c_2 |\omega_s - \omega_s^M|^2$$

其中  $c_2$  表示某一正常数。

C4: 对任意的  $s \in \{1, \dots, q\}$ , 概率趋于 1,  $|\xi_{is}| \leq c_3$  和  $|Y_i| \leq c_3$  以概率趋于 1 成立, 其中  $c_3$  表示某一正常数。

C5: 存在一个正常数  $c_4$ , 使得  $\sup_{s \in \mathcal{M}_*} |\omega_s^M| \leq c_4$  成立。此外, 对任给定的  $s \in \mathcal{M}_*$ , 存在一个正常数  $c_5$  有:

$$\left| E\{Q(Y_i, \omega_s^M, \xi_{is}) - Q(Y_i, 0)\} \right| \geq c_5 n^{-k}$$

其中  $k \in (0, 1/2)$ 。

在这些正则性条件下, 如下定理 1 给出了边际估计量  $\hat{\omega}_s^M (s=1, \dots, q)$  的一致收敛性。

**定理 1.** 假定正则性条件 C1~C5 成立。那对任给定的  $c_6 > 0$ , 存在一个正常数  $c_7$  有:

$$P\left\{ \max_{1 \leq s \leq q} |\hat{\omega}_s^M - \omega_s^M| \geq c_6 n^{-k} \right\} \leq q \exp(-c_7 n^{1-2k})$$

其中  $k$  在条件 C5 中定义。

定理 1 表明本文提出的有效工具变量识别方法允许工具变量的维数为  $q = o(\exp(n^{1-2k}))$ , 即在一些正则性条件下, 提出的有效工具变量识别方法可以处理超高维的工具变量。另外, 如下定理 2 表明本文提出的有效工具变量识别方法可以渐近以概率 1 识别出所有的有效工具变量。

**定理 2.** 令  $v_n = c_8 n^{-2k}$ , 其中  $c_8$  是一个正常数并且满足  $0 < c_8 \leq b_1/2$ ,  $b_1 = c_5/(c_1 c_3)$ , 那么在正则性条件 C1~C5 下, 可得

$$P\{\mathcal{M}_* \subset \hat{\mathcal{M}}_{v_n}\} \geq 1 - m \exp(-c_7 n^{1-2k}),$$

其中  $m = |\mathcal{M}_*|$  表示  $\mathcal{M}_*$  中的指标个数。

显然, 如果  $\max_{s \in \mathcal{M}_*} |\omega_s^M| = o(n^{-k})$  成立, 那么从定理 1 有, 对于任给定的  $b_2 > 0$ ,  $\max_{s \in \mathcal{M}_*} |\hat{\omega}_s^M| \leq b_2 n^{-k}$  以概率趋于 1 成立。因此通过定理 2 中  $v_n$  的选择, 可以保证有效工具变量的识别具有一致相合性, 即  $P\{\mathcal{M}_* = \hat{\mathcal{M}}_{v_n}\} = 1 - o(1)$ 。

接下来研究模型参数分量估计  $\hat{\beta}$  和非参数函数系数估计  $\hat{\theta}(u)$  的渐进性质。记  $\Gamma^*$  表示所识别出的有效工具变量  $\xi^*$  对应的系数矩阵。再结合定理 2 可知, 内生性协变量  $X_i$  以概率趋于 1 可写为  $X_i = \Gamma^* Z_i^* + e_i$ ,  $i = 1, \dots, n$ 。为了给出  $\hat{\beta}$  和  $\hat{\theta}(u)$  的渐进性质, 首先给出一些关于模型(1)和(2)的正则性条件如下:

C6: 非参数函数系数  $\theta(u)$  在区间  $[0, 1]$  内  $r$  阶连续可微, 其中  $r \geq 2$ 。

C7: 记  $c_1, \dots, c_k$  是  $[0, 1]$  的内部结点。那么存在常数  $c$ , 使得  $\max\{|\Delta_{i+1} - \Delta_i|\} = o(\kappa^{-1})$  以及  $\max\{\Delta_i\} / \min\{\Delta_i\} \leq c$ , 其中  $\Delta_i = c_i - c_{i-1}$ ,  $c_0 = 0$ ,  $c_{k+1} = 1$  和  $\kappa$  表示内部结点的个数。

C8: 有效工具变量对应的系数矩阵  $\Gamma^*$  是一个行满秩的矩阵。另外记  $\tau_1$  和  $\tau_2$  分别成为矩阵  $E(\xi^* \xi^{*T})$  的最小和最大的特征值, 并且  $\rho_1$  和  $\rho_2$  分别是矩阵  $\Gamma^* \Gamma^{*T}$  的最小和最大的特征值, 那么存在常数  $0 < \rho_1^* < \rho_2^* < \infty$  和  $0 < \tau_1^* < \tau_2^* < \infty$ , 使得  $\rho_1^* < \rho_1 < \rho_2 < \rho_2^*$  和  $\tau_1^* < \tau_1 < \tau_2 < \tau_2^*$  成立。

在正则性条件 C1~C8 下, 如下定理 3 表明估计量  $\hat{\beta}$  是一致相合的。

**定理 3.** 假设正则性条件 C1~C8 成立, 并且内部结点的数量  $\kappa$  满足  $\kappa = O(n^{1/(2r+1)})$ 。那么有

$$\|\hat{\beta} - \beta_0\| = O_p(\sqrt{\kappa/n}),$$

其中  $\beta_0$  是  $\beta$  的参数真值,  $r$  由条件 C6 定义。

另外如下定理 4 表明非参数函数系数的估计量  $\hat{\theta}(u)$  也是一致相合的, 并达到了最优的非参数收敛速度。

**定理 4.** 假设正则性条件 C1~C8 成立, 并且内部结点的数量  $\kappa$  满足  $\kappa = O(n^{1/(2r+1)})$ 。那么有

$$\|\hat{\theta}_s(u) - \theta_s(u)\| = O_p(n^{-r/(2r+1)}),$$

其中  $r$  由条件 C6 定义。

## 4. 数值模拟研究

在本节中, 我们通过一些蒙特卡洛数值模拟实验来研究本文估计方法的限样本性质。为实施模拟, 数据从如下模型产生:

$$\begin{cases} Y_i = X_i \beta + Z_i \theta(U_i) + \varepsilon_i \\ X_i = \Gamma \xi_i + \alpha \varepsilon_i \end{cases}$$

其中  $\beta = 2$  和  $\theta(u) = 0.5 \sin(u) + 0.5 \cos(u)$ 。我们取  $\Gamma = (2, 1.5, 1, 0.5, 0, \dots, 0)$  是一个  $1 \times q$  维矩阵, 并且工具变量由  $\xi_s \sim N(1, 1.5)$ ,  $s = 1, \dots, q$  产生。显然, 从工具变量的生成机制可以看出  $\xi_s$ ,  $s = 1, \dots, 4$  是有效工具变量, 且其他工具变量是无效工具变量。内生性变量  $X_i$  根据模型产生, 其中模型误差取为  $\varepsilon \sim N(0, 0.5)$ , 并且  $\alpha$  可分别取为  $\alpha = 0.2$  和  $\alpha = 0.8$  两种情况用来表示协变量内生性的不同水平。另外, 外生性变量  $U_i$  来自  $U_i \sim U(0, 1)$ ,  $Z_i$  来自  $Z_i \sim N(0, 1)$ , 响应变量  $Y_i$  根据模型生成。

在我们的迭代算法中有许多惩罚性函数可供使用, 例如 SCAD 惩罚(参见 Tibshirani [11])、Lasso 惩罚(参见 Fan 和 Li [12])和 MCP 惩罚(参见 Zhang [13])等。注意到 Zhao 和 Xue [14]指出变量选择的模拟结果对惩罚函数的选择不是太敏感, 因此在接下来的模拟中, 我们仅给出基于 SCAD 惩罚的数值模拟结果。

另外样本大小选择分别为  $n = 200$ 、 $400$  以及  $600$  三种情况, 工具变量的维数也分别为  $q = 500$ 、 $1000$  和  $1500$  三种情况, 并且对于每一种情况, 数值模拟实验进行重复  $1000$  次模拟运行。

接下来首先模拟有效工具变量识别的有限样本性质。为此我们选取如下指标进行模拟: 包含所有有效工具变量的最小模型大小(MMS)、把真实的有效工具变量正确识别为有效工具变量的个数(TP), 把无效工具变量错误识别为有效工具变量的个数(FP)以及错误选择率(FSR), 其中 FSR 定义为  $FSR = FP / (TP + FP)$ 。显然, FSR 代表了错误识别成有效工具变量的无效工具变量个数在所有识别成有效工具变量个数中所占的比例。在我们的模拟中, 所有这些评价指标都是基于  $1000$  次重复模拟实验的平均值。基于  $1000$  次模拟运行, 模拟结果如表 1 所示。从表 1 中, 我们可以得出以下观察结果:

1) 对模拟过程中所设定的每一种情况, 所提出的有效工具变量识别方法均可以给出相对较小的 MMS。这表明所提出的有效工具变量识别方法可以有效地控制有效工具变量的边际信号, 对于工具变量维数的变化是比较稳定的。

2) 对于任意给定的样本量  $n$  和工具变量维数  $q$ , 在内生性协变量的两种内生性水平下得到的 TP、FP 和 FSR 等指标的表现是非常相似的。这表明当内生性水平较低时, 所提出的有效工具变量识别方法仍可以识别出有效的工具变量。

3) 对于任意给定的协变量内生性水平, 模拟的 FP 和 FSR 指标均随着样本量  $n$  的增加而减小, 而 TP 指标随着样本量的增加趋于真实有效工具变量的个数。这表明所提出的有效工具变量识别方法可以渐近相合地识别出有效工具变量。

**Table 1.** The simulation results of the optimal instrumental variable identification  
**表 1.** 有效工具变量识别的模拟结果

$n$	$q$	$\alpha = 0.2$				$\alpha = 0.8$			
		MMS	TP	FP	FSR	MMS	TP	FP	FSR
200	500	14.897	2.398	0.041	0.016	14.874	2.396	0.045	0.018
	1000	15.791	2.397	0.042	0.017	15.725	2.392	0.047	0.019
	1500	16.649	2.392	0.048	0.019	16.648	2.391	0.048	0.020
400	500	8.457	3.636	0.013	0.004	8.465	3.619	0.015	0.004
	1000	10.358	3.625	0.017	0.005	10.393	3.616	0.018	0.005
	1500	13.347	3.629	0.021	0.006	13.422	3.615	0.023	0.006
600	500	5.745	3.994	0.002	0.001	5.763	3.982	0.003	0.001
	1000	7.899	3.989	0.005	0.001	7.918	3.979	0.005	0.002
	1500	9.584	3.987	0.007	0.002	9.688	3.972	0.006	0.002

接下来, 我们模拟参数分量  $\beta$  和非参数函数系数  $\theta(u)$  的估计效果。为了进行比较, 我们把本文所提出的基于工具变量调整的估计(IVE)方法与忽略协变量的内生性的朴素估计(NE)方法进行比较。这里的朴素估计方法是指忽略协变量  $X_i$  的内生性, 并通过直接最小化如下目标函数来得到  $\beta$  和  $\theta(u)$  的估计量:

$$\sum_{i=1}^n (Y_i - X_i^T \beta - W_i^T \gamma)^2,$$

其中  $W_i = B(U_i) \cdot Z_i$ ,  $\theta(u)$  由  $\theta(u) \approx B(u)^T \gamma$  近似逼近。

对于参数分量  $\beta$ , 本文仅给出当工具变量维数为  $q=1000$  时的模拟结果。在工具变量维数为其他情况下的模拟结果是类相似的, 因此不再一一展示。基于 1000 次重复模拟实验, 关于参数分量  $\beta$  估计量的 1000 个绝对偏差值  $|\hat{\beta} - \beta|$  的箱线图见图 1 和图 2, 其中图 1 是在内生性水平为  $\alpha = 0.02$  下的模拟结果, 图 2 是在内生性水平为  $\alpha = 0.08$  下的模拟结果。由图 1 和图 2 可以得到如下结论:

1) 随着样本量的增加, 基于 IVE 估计方法给出的绝对偏差值则明显减小, 而基于 NE 方法给出的绝对偏差值即使样本量增加的情况下仍然较大。这就表明所提出的 IVE 方法优于 NE 方法, 而且忽略变量内生性的 NE 方法将会产生一定的内生性偏差。

2) 对于任意给定的样本量  $n$ , 基于 IVE 估计方法所给出的模拟结果在不同的内生性水平下是非常相似的。这也表明本文提出的工具变量调整机制可以减弱协变量内生性对估计精度带来的影响。

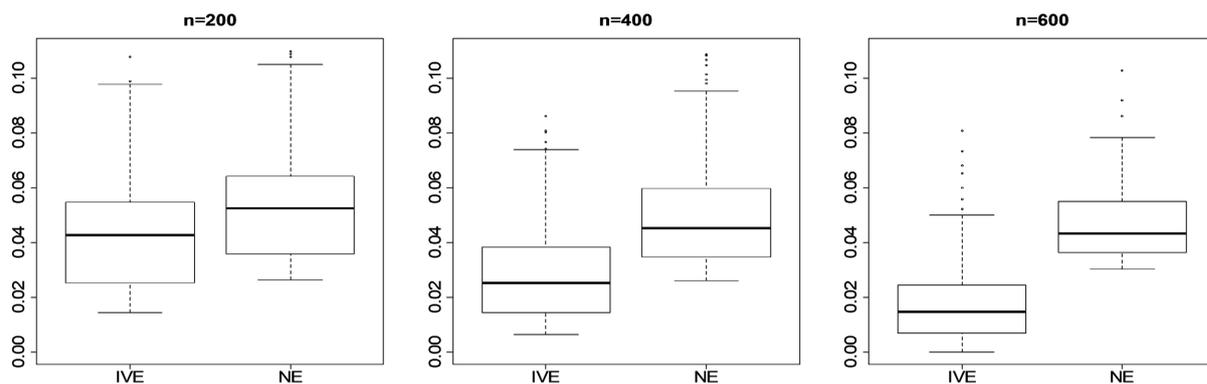


Figure 1. Under the endogenous level  $\alpha = 0.2$ , the box-plots of absolute bias values for the estimator  $\hat{\beta}$

图 1. 在内生性水平  $\alpha = 0.2$  下, 估计量  $\hat{\beta}$  的绝对偏差箱线图

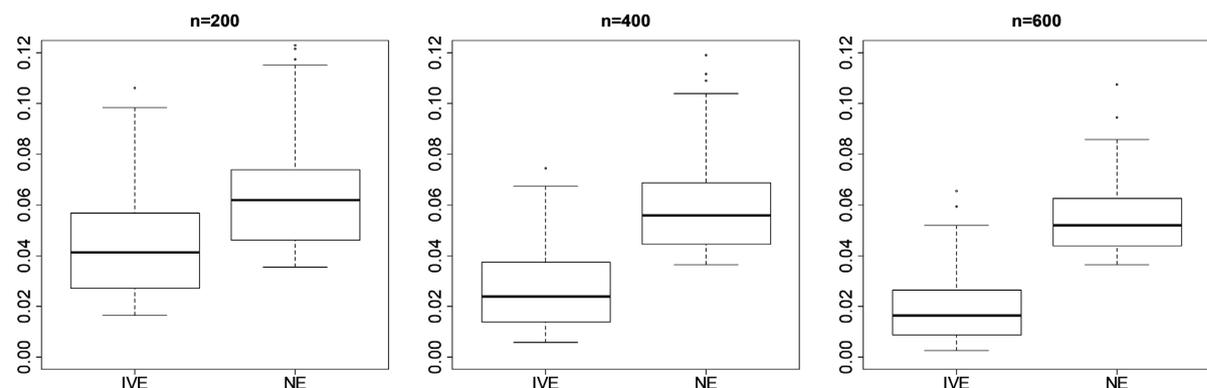
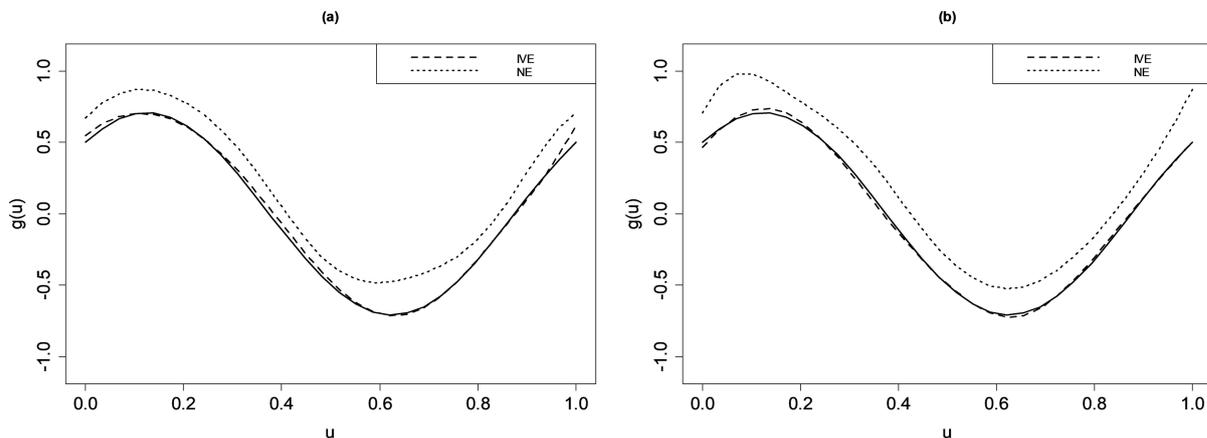


Figure 2. Under the endogenous level  $\alpha = 0.8$ , the box-plots of absolute bias values for the estimator  $\hat{\beta}$

图 2. 在内生性水平  $\alpha = 0.8$  下, 估计量  $\hat{\beta}$  的绝对偏差箱线图

另外, 图 3 给出了在样本量  $n = 400$  的情况下, 非参数函数系数  $\theta(u)$  的模拟结果, 其中短虚线表示基于 IVE 方法得到的估计曲线, 点虚线表示基于 NE 方法得到的估计曲线, 实线表示  $\theta(u)$  的真实曲线。从图 3 可以看出, NE 方法给出的估计曲线有相对较大的估计偏差, 而本文提出的 IVE 方法则可以减弱估计偏差, 给出的估计曲线非常接近真实曲线。另外还可以看出, 在不同的内生性水平下, 基于 IVE 方法给出的估计曲线是非常相似的。综上, 上述模拟结果表明, 协变量的内生性将会影响参数分量和非参数分量估计的一致相合性, 而本文所提出的 IVE 方法可有效地削弱协变量的内生性对模型参数分量及非参数分量估计的影响。



**Figure 3.** The estimator of  $\theta(u)$  under the the endogenous level  $\alpha = 0.2$  (a) and  $\alpha = 0.8$  (b), respectively

**图 3.**  $\theta(u)$  的估计曲线, 其中(a)是在内生性水平为  $\alpha = 0.2$  时的模拟结果, (b)是在  $\alpha = 0.8$  时的模拟结果

### 5. 定理的证明

在本节中, 我们将给出定理 1~4 的详细证明过程。令  $E_n \{m(Y, \xi_s)\} = n^{-1} \sum_{i=1}^n m(Y_i, \xi_{is})$  表示数学期望对应的经验测度。另外, 定义  $\mathcal{B}_s(N) = \{\omega_s \in \mathcal{B} : |\omega_s - \omega_s^M| \leq N\}$  其中  $N > 0$  为某一正常数, 并进一步定义  $F_s(N) = \sup_{\omega_s \in \mathcal{B}_s(N)} \left| (E_n - E) \left\{ Q(Y_i, \omega_s \xi_{is}) - Q(Y_i, \omega_s^M \xi_{is}) \right\} \right|$ 。首先给出一些定理证明过程中需要的引理。

**引理 1.** 假设正则条件 C1~C5 成立, 那么有

$$\min_{s \in \mathcal{M}_*} |\omega_s^M| \geq b_1 n^{-k},$$

其中  $k \in (0, 1/2)$ ,  $b_1 = c_5 / (c_1 c_3)$  和  $\omega_s^M$  由(9)式定义。

**证明:** 根据条件 C2 可知

$$\left| Q(Y_i, \omega_s^M \xi_{is}) - Q(Y_i, 0) \right| \leq c_1 |\omega_s^M| |\xi_{is}| \tag{10}$$

对(10)式两边同时取期望可得

$$\left| E \left\{ Q(Y_i, \omega_s^M \xi_{is}) - Q(Y_i, 0) \right\} \right| \leq E \left\{ \left| Q(Y_i, \omega_s^M \xi_{is}) - Q(Y_i, 0) \right| \right\} \leq c_1 |\omega_s^M| E \{ |\xi_{is}| \} \leq c_1 c_3 |\omega_s^M| \tag{11}$$

另外由条件 C5 可知

$$\left| E \left\{ Q(Y_i, \omega_s^M \xi_{is}) - Q(Y_i, 0) \right\} \right| \geq c_5 n^{-k} \tag{12}$$

从(11)和(12)式可知, 对任意  $s \in \mathcal{M}_*$ , 则有  $|\omega_s^M| \geq b_1 n^{-k}$ 。即引理 1 的证明完毕。

**引理 2.** 假设正则条件 C1~C5 成立, 那么对任给定的  $\zeta > 0$  和  $s \in \{1, \dots, q\}$ , 则有:

$$P \{ F_s(N) \geq 4Nb_3 n^{-1/2} (1 + \zeta) \} \leq \exp(-2\zeta^2)$$

其中  $b_3 = c_1 c_3$ 。

**证明:** 结合条件 C2, 并利用对称化定理(参见 Vaart 和 Wellner [15])可得

$$\begin{aligned} E \{ F_s(N) \} &\leq 2E \left\{ \sup_{\omega_s \in \mathcal{B}_s(N)} \left| E_n \delta_i \left[ Q(Y_i, \omega_s \xi_{is}) - Q(Y_i, \omega_s^M \xi_{is}) \right] \right| \right\} \\ &\leq 4c_1 E \left\{ \sup_{\omega_s \in \mathcal{B}_s(N)} \left| E_n \left[ \delta_i \xi_{is} (\omega_s - \omega_s^M) \right] \right| \right\} \end{aligned} \tag{13}$$

其中  $\delta_1, \dots, \delta_n$  表示分别以概率为 1/2 取值为  $\pm 1$  的独立且同分布随机变量序列。另外再通过利用 Jensen 不等式可以得到

$$\begin{aligned} E \left\{ \sup_{\omega_s \in \mathcal{B}_s(N)} \left| E_n \left[ \delta \xi_{is} (\omega_s - \omega_s^M) \right] \right| \right\} &\leq \sup_{\omega_s \in \mathcal{B}_s(N)} |\omega_s - \omega_s^M| \times E \left[ |E_n \delta \xi_{is}| \right] \\ &\leq N \left( E |E_n \delta \xi_{is}|^2 \right)^{1/2} \\ &= \frac{N}{\sqrt{n}} \left[ E \{ \delta^2 \xi_{is}^2 \} \right]^{1/2} \\ &\leq N c_3 n^{-1/2} \end{aligned} \quad (14)$$

结合(13)和(14)式则可以得到  $E \{ F_s(N) \} \leq 4N b_1 n^{-1/2}$ 。另外再由条件 C2 可知在集合  $\mathcal{B}_s(N)$  上有

$$\left| Q(Y_i, \omega_s \xi_{is}) - Q(Y_i, \omega_s^M \xi_{is}) \right| \leq b_1 N$$

因此, 令  $L = 2N b_1$  并利用集中定理(参见 Massart [16])可得

$$\begin{aligned} \Pr \{ F_s(N) \geq 4N b_1 n^{-1/2} (1 + \zeta) \} &\leq \Pr \left[ F_s(N) \geq E \{ F_s(N) \} + 4N b_1 n^{-1/2} \zeta \right] \\ &\leq \exp \left[ -n (4N b_1 n^{-1/2} \zeta)^2 / (2L^2) \right] \\ &= \exp(-2\zeta^2) \end{aligned}$$

即引理 2 证明完毕。

**定理 1 证明.** 令  $\omega_{\lambda s} = \lambda \hat{\omega}_s^M + (1 - \lambda) \omega_s^M$ , 其中  $\lambda = \left( 1 + \left| \hat{\omega}_s^M - \omega_s^M \right| / N \right)^{-1} \leq 1$ , 且  $N$  满足条件(C1)和(C2)。则一个简单的计算有  $|\omega_{\lambda s} - \omega_s^M| = \lambda |\hat{\omega}_s^M - \omega_s^M| \leq N$ , 其表明  $\omega_{\lambda s} \in \mathcal{B}_s(N)$ 。因为  $\hat{\omega}_s^M$  是  $E_n \{ Q(Y_i, \omega_s \xi_{is}) \}$  的最小取值, 然后调用凸性, 我们得到

$$\begin{aligned} E_n \{ Q(Y_i, \omega_{\lambda s} \xi_{is}) \} &\leq \lambda E_n \{ Q(Y_i, \hat{\omega}_s^M \xi_{is}) \} + (1 - \lambda) E_n \{ Q(Y_i, \omega_s^M \xi_{is}) \} \\ &\leq \lambda E_n \{ Q(Y_i, \omega_s^M \xi_{is}) \} + (1 - \lambda) E_n \{ Q(Y_i, \omega_s^M \xi_{is}) \} \\ &= E_n \{ Q(Y_i, \omega_s^M \xi_{is}) \} \end{aligned} \quad (15)$$

另外, 注意到  $\omega_s^M$  是  $E \{ S(Y_i, \omega_s \xi_{is}) \} = 0$  的解, 因此有  $\omega_s^M$  是  $E \{ Q(Y_i, \omega_s \xi_{is}) \}$  的最小值点。因此有

$$E \{ Q(Y_i, \omega_s^M \xi_{is}) \} \leq E \{ Q(Y_i, \omega_{\lambda s} \xi_{is}) \} \quad (16)$$

进而结合(15)和(16)式, 通过简单计算可以得到

$$E \{ Q(Y_i, \omega_{\lambda s} \xi_{is}) \} - E \{ Q(Y_i, \omega_s^M \xi_{is}) \} \leq (E - E_n) \{ Q(Y_i, \omega_{\lambda s} \xi_{is}) - Q(Y_i, \omega_s^M \xi_{is}) \} \leq F_s(N) \quad (17)$$

其中  $F_s(N) = \sup_{\omega_s \in \mathcal{B}_s(N)} \left| (E - E_n) \{ Q(Y_i, \omega_{\lambda s} \xi_{is}) - Q(Y_i, \omega_s^M \xi_{is}) \} \right|$ 。另外再由条件 C5 可得

$$\left| E \{ Q(Y_i, \omega_{\lambda s} \xi_{is}) \} - E \{ Q(Y_i, \omega_s^M \xi_{is}) \} \right| \geq c_2 |\omega_{\lambda s} - \omega_s^M|^2 \quad (18)$$

因此, 结合(17)和(18)式可以得到

$$|\omega_{\lambda s} - \omega_s^M| \leq (F_s(N)/c_2)^{1/2} \quad (19)$$

对任给定的  $\delta$  并结合(19)式可得  $P \left( |\omega_{\lambda s} - \omega_s^M| \geq \delta \right) \leq P \left( F_s(N) \geq c_2 \delta^2 \right)$ 。取  $\delta = N/2$ , 那么有

$$P \left( |\omega_{\lambda s} - \omega_s^M| \geq N/2 \right) \leq P \left( F_s(N) \geq c_2 N^2 / 4 \right) \quad (20)$$

结合  $\omega_{\lambda_s}$  的定义, 简单计算可得  $\omega_{\lambda_s} - \omega_s^M = \lambda(\hat{\omega}_s^M - \omega_s^M)$ 。因此可以得到

$$\begin{aligned} P\left(|\omega_{\lambda_s} - \omega_s^M| \geq N/2\right) &= P\left(|\hat{\omega}_s^M - \omega_s^M| \geq \lambda^{-1}N/2\right) \\ &= P\left(|\hat{\omega}_s^M - \omega_s^M| \geq \left[1 + |\hat{\omega}_s^M - \omega_s^M|/N\right]N/2\right) \\ &= P\left(|\hat{\omega}_s^M - \omega_s^M| \geq N\right) \end{aligned} \tag{21}$$

取  $N = 16(c_1c_3/c_2)n^{-1/2}(1 + \zeta)$ , 则由(20)和(21)可得

$$P\left(|\hat{\omega}_s^M - \omega_s^M| \geq 16(c_1c_3/c_2)n^{-1/2}(1 + \zeta)\right) \leq P\left(F_s(N) \geq 4Nc_1c_3n^{-1/2}(1 + \zeta)\right) \tag{22}$$

因此结合(22)式并利用引理 2 可得

$$P\left\{\sqrt{n}|\hat{\omega}_s^M - \omega_s^M| \geq 16(c_1c_3/c_2)(1 + \zeta)\right\} \leq \exp(-2\zeta^2)$$

取  $1 + \zeta = (c_2c_6)/(c_1c_3)n^{1/2-k}/16$ , 那么则有

$$P\left(|\hat{\omega}_s^M - \omega_s^M| \geq c_6n^{-k}\right) \leq \exp\left(-2\left[(c_2c_6)/(c_1c_3)n^{1/2-k}/16-1\right]^2\right) \tag{23}$$

注意到  $\left[(c_2c_6)/(c_1c_3)n^{1/2-k}/16-1\right]^2 = O(n^{1-2k})$ , 因此由(23)式可得

$$P\left(|\hat{\omega}_s^M - \omega_s^M| \geq c_6n^{-k}\right) \leq \exp(-c_7n^{1-2k}) \tag{24}$$

其中  $c_7$  为一个正常数。另外简单计算可得

$$P\left(\max_{1 \leq s \leq q} |\hat{\omega}_s^M - \omega_s^M| \geq c_6n^{-k}\right) \leq \sum_{s=1}^q P\left(|\hat{\omega}_s^M - \omega_s^M| \geq c_6n^{-k}\right) \tag{25}$$

进而结合(24)和(25)式可得

$$P\left(\max_{1 \leq s \leq q} |\hat{\omega}_s^M - \omega_s^M| \geq c_6n^{-k}\right) \leq q \exp(-c_7n^{1-2k})$$

这就完成了定理 1 的证明。

**定理 2 证明.** 我们首先定义  $\mathcal{A}_n$  如下

$$\mathcal{A}_n = \left\{ \max_{s \in \mathcal{M}_*} |\hat{\omega}_s^M - \omega_s^M| \leq b_1n^{-k}/2 \right\}$$

其中  $b_1$  在引理 1 中定义。由引理 1 可知对任意  $s \in \mathcal{M}_*$  有  $|\omega_s^M|^2 \geq b_1n^{-k}$ , 因此在  $\mathcal{A}_n$  上利用三角不等式可以得到

$$|\hat{\omega}_s^M| \geq |\omega_s^M| - |\hat{\omega}_s^M - \omega_s^M| \geq |\omega_s^M| - \max_{s \in \mathcal{M}_*} |\hat{\omega}_s^M - \omega_s^M| \geq b_1n^{-k}/2$$

取  $v_n = c_8n^{-k}$ , 其中  $0 < c_8 < b_1/2$ , 那么可得

$$P\left(\mathcal{M}_* \subset \hat{\mathcal{M}}_{v_n}\right) \geq 1 - P\left(\mathcal{A}_n^c\right) = 1 - P\left(\max_{s \in \mathcal{M}_*} |\hat{\omega}_s^M - \omega_s^M| \geq b_1n^{-k}/2\right) \tag{26}$$

另外由定理 1 可得

$$P\left(\max_{s \in \mathcal{M}_*} |\hat{\omega}_s^M - \omega_s^M| \geq b_1n^{-k}/2\right) \leq m \exp(-c_7n^{1-2k}) \tag{27}$$

其中  $m$  是  $\mathcal{M}_*$  中的元素数。那么结合(26)和(27)式则完成了定理 2 的证明。

**定理 3 证明.** 由第 3 节中的讨论可知  $X_i = \Gamma^* \xi_i^* + e_i$  以概率趋于 1 成立。另外注意  $\hat{\Gamma}^*$  是  $\Gamma^*$  的矩估计量, 因此可得  $\hat{\Gamma}^* = \Gamma^* + O_p\left(\sqrt{1/n}\right)$ 。再结合  $E(e_i) = 0$ , 并通过简单计算可得

$$\begin{aligned}
\frac{1}{n} \sum_{i=1}^n (X_i - X_i^*)^\top \beta &= \frac{1}{n} \sum_{i=1}^n (\Gamma^* \xi_i^* + e_i - \hat{\Gamma}^* \xi_i^*)^\top \beta \\
&= \frac{1}{n} \sum_{i=1}^n \xi_i^{*\top} (\Gamma^* - \hat{\Gamma}^*)^\top \beta + \frac{1}{\sqrt{n}} \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n e_i^\top \beta \right) \\
&= O_p(\|\Gamma^* - \hat{\Gamma}^*\|) + O_p(n^{-1/2}) \\
&= O_p(\sqrt{1/n})
\end{aligned} \tag{28}$$

记  $\beta = \beta_0 + \delta_n \eta_1$ ,  $\gamma = \gamma_0 + \delta_n \eta_2$  和  $R(U_i) = Z_i^\top \theta(U_i) - W_i^\top \gamma_0$ , 其中  $\delta_n = \sqrt{k/n}$ ,  $\beta_0$  和  $\gamma_0$  是参数真值, 且  $\eta_1$  和  $\eta_2$  表示常数向量。那么由(28)式可得到

$$\begin{aligned}
M_n(\beta, \gamma) &= \frac{1}{n} \sum_{i=1}^n (Y_i - X_i^{*\top} \beta - W_i^\top \gamma)^2 \\
&= \frac{1}{n} \sum_{i=1}^n [e_i^\top \beta_0 + \varepsilon_i + R(U_i) - \delta_n \xi_i^{*\top} \Gamma^{*\top} \eta_1 - \delta_n W_i^\top \eta_2]^2 + O_p(\|\Gamma^* - \hat{\Gamma}^*\|^2) \\
&= \frac{1}{n} \sum_{i=1}^n [e_i^\top \beta_0 + \varepsilon_i + R(U_i) - \delta_n \psi_i^\top \eta]^2 + O_p(1/n)
\end{aligned} \tag{29}$$

其中  $\psi_i = (\xi_i^{*\top} \Gamma^{*\top}, W_i^\top)^\top$  和  $\eta = (\eta_1^\top, \eta_2^\top)^\top$ 。类似地, 我们可得到

$$M_n(\beta_0, \gamma_0) = \frac{1}{n} \sum_{i=1}^n (Y_i - X_i^{*\top} \beta_0 - W_i^\top \gamma_0)^2 = \frac{1}{n} \sum_{i=1}^n [e_i^\top \beta_0 + \varepsilon_i + R(U_i)]^2 + O_p(1/n) \tag{30}$$

进一步记  $\Delta_n(\beta, \gamma) = M_n(\beta, \gamma) - M_n(\beta_0, \gamma_0)$ , 那么由(29)和(30)可得

$$\begin{aligned}
\Delta_n(\beta, \gamma) &= \frac{1}{n} \sum_{i=1}^n \delta_n^2 \psi_i^\top \eta \eta^\top \psi_i - \frac{2}{n} \sum_{i=1}^n \delta_n (e_i^\top \beta_0 + \varepsilon_i) \psi_i^\top \eta \\
&\quad - \frac{2}{n} \sum_{i=1}^n \delta_n R(U_i) \psi_i^\top \eta + o_p(\delta_n^2) \\
&\equiv J_1 + J_2 + J_3 + O_p(\delta_n^2)
\end{aligned} \tag{31}$$

结合  $E(e_i^\top \beta_0 + \varepsilon_i | \xi_i^*, W_i) = 0$ , 简单计算可得

$$J_2 = -\frac{2\delta_n}{\sqrt{n}} \left[ \frac{1}{\sqrt{n}} \sum_{i=1}^n (e_i^\top \beta_0 + \varepsilon_i) \psi_i^\top \eta \right] = O_p\left(\frac{\delta_n}{\sqrt{n}}\right) \|\eta\| = o_p(\delta_n^2) \|\eta\| \tag{32}$$

另外由 Schumaker [17]可知  $\|R(U_i)\| = O_p(k^{-r}) = O_p(\delta_n)$ 。因此简单计算可得

$$J_3 = O_p(\delta_n \|R(U_i)\|) \|\eta\| = O_p(\delta_n^2) \|\eta\| \tag{33}$$

注意到  $J_1 = O_p(\delta_n^2) \|\eta\|^2$ , 那么对于一个充分大的常数  $C$ , 在  $\|\eta\| = C$  条件下  $J_2$  和  $J_3$  完全可由  $J_1$  控制。另外注意  $J_1$  是一个正数, 因此结合(31)~(33)式可得, 对任给的  $\nu > 0$ , 存在一个大常数  $C$  使得

$$P\left\{\inf_{\|\eta\|=C} M_n(\beta, \gamma) > M_n(\beta_0, \gamma_0)\right\} \geq 1 - \nu \tag{34}$$

另外由  $M_n(\cdot, \cdot)$  的凸性可知, (34)式表明存在一个局部最小值点  $\hat{\beta}$  和  $\hat{\gamma}$ , 使得  $\hat{\beta} - \beta_0 = O_p(\delta_n)$  和  $\hat{\gamma} - \gamma_0 = O_p(\delta_n)$  以不小于  $1 - \nu$  的概率成立。再由  $\nu$  的任意性, 则完成定理 3 的证明。

**定理 4 证明.** 利用类似 Zhao 和 Xue [14]的讨论可得

$$\|\hat{\theta}_s(u) - \theta_s(u)\|^2 \leq 2(\hat{\gamma}_s - \gamma_{s0})^\top H(\hat{\gamma}_s - \gamma_{s0}) + \int_0^1 R_s(u)^2 du \tag{35}$$

其中  $R_s(u) = \theta_s(u) - B^T(u)\gamma_{s0}$  和  $H = \int_0^1 B(u)B^T(u)du$ 。另外, 由定理 3 的证明过程可知  $\|\hat{\gamma}_s - \gamma_{s0}\| = O_p(\sqrt{\kappa/n})$ 。结合条件 C6 和  $\kappa = O(n^{1/(2r+1)})$ , 则可以得到  $O_p(\sqrt{\kappa/n}) = O_p(n^{-r/(2r+1)})$ 。因此再结合  $\|H\| = O(1)$  可得

$$(\hat{\gamma}_s - \gamma_{s0})^T H (\hat{\gamma}_s - \gamma_{s0}) = O_p\left(n^{\frac{-2r}{2r+1}}\right) \quad (36)$$

另外注意到  $\|R_s(u)\| = O(\kappa^{-r}) = O\left(n^{\frac{-r}{2r+1}}\right)$ , 则可以得到

$$\int_0^1 R_s(u)^2 du = O_p\left(n^{\frac{-2r}{2r+1}}\right) \quad (37)$$

结合(35)~(37)式则完成了定理 4 的证明。

## 6. 结论

本文在工具变量为超高维的情况下, 对一类含内生性协变量的半参数模型的估计问题进行了研究。结合构造辅助回归模型和惩罚估计技术, 提出了一种基于双惩罚的正则估计方法。所提出的估计方法不仅可以给出模型参数分量和非参数分量的相合估计, 还可以识别出有效工具变量。模拟研究表明, 即使工具变量为超高维数据的情况下, 本文提出的方法仍可以相合的识别出所有有效工具变量。另外, 模拟结果还表明本文提出的基于工具变量调整的估计方法对模型参数和非参数函数系数的估计也是行之有效的。

注意到本文考虑的半参数模型具有较好的灵活性和广泛的适应性, 经典的线性模型、部分线性模型以及部分线性可加模型均是本文所研究模型的特殊情形。因此, 本文提出的双惩罚正则估计方法可以方便地推广到超高维数据下部分线性模型以及部分线性可加模型等半参数模型的统计推断研究上。另外, 在本文的研究中, 工具变量  $\xi_i$  假定为超高维数据, 而内生性协变量  $X_i$  则假定为维数固定的低维数据。因此, 接下来一个值得进一步研究的课题是在工具变量  $\xi_i$  和内生性协变量  $X_i$  均为超高维数据的情况下, 研究超高维半参数工具变量模型的统计推断问题。

## 基金项目

本文研究成果得到国家社会科学基金项目《高维内生协变量的半参数建模及其在环境治理绩效测度中的应用》(编号: 18BTJ035)资助。

## 参考文献

- [1] Fan, J. and Huang, T. (2005) Profile Likelihood Inferences on Semiparametric Varying Coefficient Partially Linear Models. *Bernoulli*, **11**, 1031-1057. <https://doi.org/10.3150/bj/1137421639>
- [2] Kai, B., Li, R. and Zou, H. (2011) New Efficient Estimation and Variable Selection Methods for Semiparametric Varying-Coefficient Partially Linear Models. *The Annals of Statistics*, **39**, 305-332. <https://doi.org/10.1214/10-AOS842>
- [3] Li, D., Chen, J. and Lin, Z. (2011) Statistical Inference in Partially Time-Varying Coefficient Models. *Journal of Statistical Planning and Inference*, **141**, 995-1013. <https://doi.org/10.1016/j.jspi.2010.09.004>
- [4] Cai, Z. and Xiong, H. (2012) Partially Varying Coefficient Instrumental Variables Models. *Statistica Neerlandica*, **66**, 85-110. <https://doi.org/10.1111/j.1467-9574.2011.00497.x>
- [5] Yang, Y., Chen, L. and Zhao, P. (2017) Empirical Likelihood Inference in Partially Linear Single Index Models with Endogenous Covariates. *Communications in Statistics, Theory and Methods*, **46**, 3297-3307. <https://doi.org/10.1080/03610926.2015.1060341>
- [6] Yuan, J., Zhao, P. and Zhang, W. (2016) Semiparametric Variable Selection for Partially Varying Coefficient Models with Endogenous Variables. *Computational Statistics*, **31**, 693-707. <https://doi.org/10.1007/s00180-015-0601-y>

- 
- [7] Fan, J. and Lv, J. (2008) Sure Independence Screening for Ultrahigh Dimensional Feature Space. *Journal of the Royal Statistical Society: Series B*, **70**, 849-911. <https://doi.org/10.1111/j.1467-9868.2008.00674.x>
- [8] Bound, J., Jaeger, D. and Baker, R. (1995) Problems with Instrumental Variables Estimation When the Correlation between the Instruments and the Endogenous Explanatory Variable Is Weak. *Journal of the American Statistical Association*, **90**, 443-450. <https://doi.org/10.2307/2291055>
- [9] Didelez, V., Meng, S. and Sheehan, N. (2010) Assumptions of IV Methods for Observational Epidemiology. *Statistical Science*, **25**, 22-40. <https://doi.org/10.1214/09-STS316>
- [10] Zhang, S., Zhao, P., Li, G. and Xu, W. (2019) Nonparametric Independence Screening for Ultra-High Dimensional Generalized Varying Coefficient Models with Longitudinal Data. *Journal of Multivariate Analysis*, **171**, 37-52. <https://doi.org/10.1016/j.jmva.2018.11.002>
- [11] Tibshirani, R. (1996) Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society: Series B*, **58**, 267-288. <https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>
- [12] Fan, J. and Li, R. (2001) Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties. *Journal of the American Statistical Association*, **96**, 1348-1360. <https://doi.org/10.1198/016214501753382273>
- [13] Zhang, C. (2010) Nearly Unbiased Variable Selection under Minimax Concave Penalty. *The Annals of Statistics*, **38**, 894-942. <https://doi.org/10.1214/09-AOS729>
- [14] Zhao, P. and Xue, L. (2010) Variable Selection for Semiparametric Varying Coefficient Partially Linear Errors-in-Variables Models. *Journal of Multivariate Analysis*, **101**, 1872-1883. <https://doi.org/10.1016/j.jmva.2010.03.005>
- [15] Vaart, A. and Wellner, J. (1996) *Weak Convergence and Empirical Processes*. Springer, New York.
- [16] Massart, P. (2000) About the Constants in Talagrand's Concentration Inequalities for Empirical Processes. *The Annals of Probability*, **28**, 863-884. <https://doi.org/10.1214/aop/1019160263>
- [17] Schumaker, L. (1981) *Spline Function*. Wiley, New York.