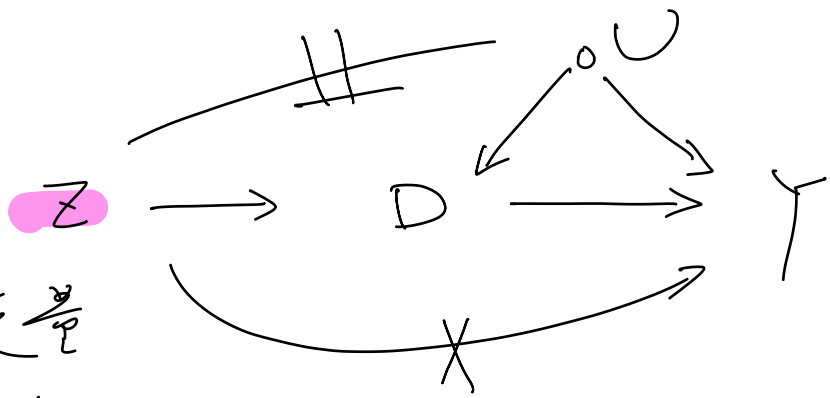
Chapter 23 IV: Econometric perspective

遗传学 计量经济学

S. Wright

路径分析 path analysis



工具变量

IV

经济学的背景

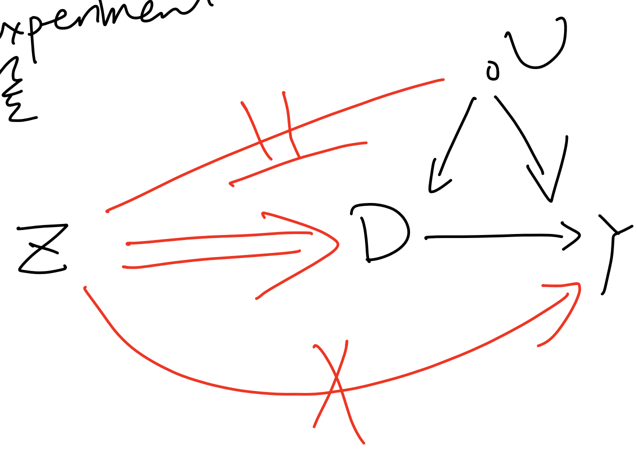
— 大部分都是观察性研究

— D→Y 因果作用

— D 没有随机化

—— 给定协变量，也无法完全消除混杂

natural experiment
自然实验



问题：何处找 Z？

<u>例 23.1</u> 鼓励性实验

<u>例 23.2</u> Angrist (1990)

TSLS
后面讲

Hearst et al (1986 NEJM)

越战

彩票数 $\Longrightarrow$ 服兵役

彩票数 $\longrightarrow$ 死亡率

$$\begin{tikzcd} & U \arrow[dl] \arrow[dr] & \\ \text{彩票数} \arrow[r] & \text{服兵役} \arrow[r] & \text{死亡率} \end{tikzcd}$$

<u>例23.3</u> Angrist & Krueger (1991)

$$\begin{tikzcd} & U \arrow[dl] \arrow[dr] & \\ \text{出生季度} \arrow[r, Rightarrow] & \text{教育年数} \arrow[r] & \text{收入} \end{tikzcd}$$

<u>例23.4</u> Angrist & Evans (1998)

$$\begin{tikzcd} & U \arrow[dl] \arrow[dr] & \\ \text{前两个孩子同性别与否} \arrow[r, Rightarrow] & \text{孩子数目} \arrow[r] & \text{母亲工作} \end{tikzcd}$$

例 23.5　Card (1993)

U

成长地是否离大学近 ⟶ 教育年数 ⟶ 工资

反例：孟母三迁

例 23.6　Chapter 25

Mendelian Randomization

U

SNPs ⟶ HDL/LDL ⟶ 心脏病

回顾 OLS: $\mathrm{lm}(Y \sim D)$

$$\hat{\beta} \xrightarrow{P} E(DD^T)^{-1} E(DY)$$

$\neq$ $Y = \beta D + \varepsilon$ 中的 β

OLS 线性投影

$\hat{\beta}$ 矩估计

OLS

线性模型的数据生成机制

$E(D\varepsilon) \neq 0$

内生性

endogenous

解决内生性: IV

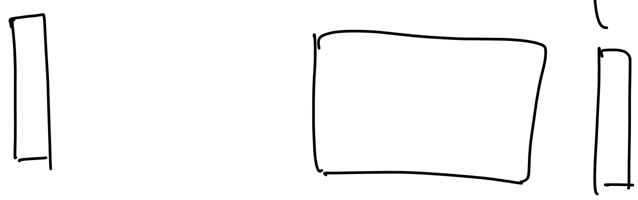
$$\begin{cases} Y = D^T \beta + \varepsilon \\ E(\varepsilon Z) = 0 \end{cases}$$

$$\begin{tikzcd} & & \varepsilon \arrow[dl] \arrow[dr] & \\ Z \arrow[r] & D \arrow[rr] & & Y \end{tikzcd}$$

Z 为 D 的 IV

$$\Rightarrow E\Big(Z\big(\underbrace{Y - D^T\beta}_{=\,\varepsilon}\big)\Big) = 0$$

$$\Rightarrow E(ZY) = E(ZD^T)\beta$$



① $\dim(Z) < \dim(D)$

β 有 ∞ 多解

under identified

② $\dim(Z) = \dim(D)$

$E(ZD^T)$ 满秩

just identified

$$\Rightarrow \beta = E(ZD^T)^{-1} E(ZY)$$

唯一解

$$\hat{\beta}_{IV} = \left(\sum_{i=1}^{n} Z_i D_i^T \right)^{-1} \sum_{i=1}^{n} Z_i Y_i$$

矩方

$$\begin{cases} Y = \alpha + \beta D + \varepsilon & D \rightarrow \text{一维} \\ E(\varepsilon) = 0, \quad \mathrm{cov}(\varepsilon, Z) = 0 \end{cases}$$

$$\Rightarrow \mathrm{cov}(Z, Y) = \beta \, \mathrm{cov}(Z, D)$$

$$\Rightarrow \beta = \frac{\mathrm{cov}(Z, Y)}{\mathrm{cov}(Z, D)}$$

if $\mathrm{cov}(Z, D) \neq 0$

$$= \frac{\mathrm{cov}(Z, Y) / \mathrm{Var}(Z)}{\mathrm{cov}(Z, D) / \mathrm{Var}(Z)}$$

$$= \frac{\text{lm}(Y \sim Z)\$\text{coef}[2]}{\text{lm}(D \sim Z)\$\text{coef}[2]}$$

Z 二值

$$= \frac{E(Y \mid Z=1) - E(Y \mid Z=0)}{E(D \mid Z=1) - E(D \mid Z=0)}$$

D=值 CACE 的公式

③ $\dim(Z) > \dim(D)$

over identified

$$E(ZY) = E(ZD^T)\beta$$

假定有唯一解

样本水平：$\left(\frac{1}{n}\sum_{i=1}^{n} Z_i Y_i\right) = \left(\frac{1}{n}\sum_{i=1}^{n} Z_i D_i^T\right)\hat{\beta}$

一般无解

GMM: generalized method of moments

(Hansen 1982)

古老的计量经济想法：

两步（阶段）最小二乘

Two-stage least squares (TSLS)

① OLS($D \sim Z$)

OLS(D 的外分量 $\sim Z$)

$$\Rightarrow \hat{D}_i \quad (i = 1 \cdots n)$$

② OLS($Y \sim \hat{D}$)

$$\Rightarrow \hat{\beta}_{TSLS} \xrightarrow{P} \beta$$

直觉:

$$\hat{\beta}_{TSLS} \overset{②}{=} \left(\sum_{i=1}^{n} \hat{D}_i \hat{D}_i^T \right)^{-1} \sum_{i=1}^{n} \hat{D}_i \underbrace{Y_i}_{= D_i^T\beta + \varepsilon_i}$$

模型 $= \left(\sum_{i=1}^{n} \hat{D}_i \hat{D}_i^T \right)^{-1} \left(\sum_{i=1}^{n} \hat{D}_i D_i^T \right) \beta$

$$+ \left(\sum_{i=1}^{n} \hat{D}_i \hat{D}_i^T \right)^{-1} \left(\sum_{i=1}^{n} \hat{D}_i \varepsilon_i \right)$$

① : $D_i = \hat{D}_i + \check{D}_i$

$\sum_{i=1}^{n} \hat{D}_i \check{D}_i^T = 0$

$\Rightarrow \sum_{i=1}^{n} \hat{D}_i \hat{D}_i^T = \sum_{i=1}^{n} \hat{D}_i D_i^T$

$$\Rightarrow \hat{\beta}_{TSLS} = \beta + \left(\sum_{i=1}^{n} \hat{D}_i \hat{D}_i^T\right)^{-1} \sum_{i=1}^{n} \hat{D}_i \varepsilon_i$$

想方设法用 $E(Z\varepsilon) = 0$

简单事实 $\hat{D}_i = \hat{\Gamma}^T Z_i$

$$\text{代入}: \hat{\beta}_{TSLS} = \beta + \left(\hat{\Gamma}^T \left(\frac{1}{n}\sum_{i=1}^{n} Z_i Z_i^T\right) \hat{\Gamma}\right)^{-1} \hat{\Gamma}^T \left(\frac{1}{n}\sum_{i=1}^{n} Z_i \varepsilon_i\right)$$

$\xrightarrow{P} \beta$

$\hat{\Gamma} \xrightarrow{P} \Gamma$, $\frac{1}{n}\sum_{i=1}^{n} Z_i Z_i^T \xrightarrow{P} E(ZZ^T)$, $\frac{1}{n}\sum_{i=1}^{n} Z_i \varepsilon_i \xrightarrow{P} E(Z\varepsilon) = 0$

Chapter 7

SPECIFICATION AND ESTIMATION OF SIMULTANEOUS EQUATION MODELS

关于IV的
章节

JERRY A. HAUSMAN*

Massachusetts Institute of Technology

Contents

更多统计

Statistical Science
2014, Vol. 29, No. 3, 323–358
DOI: 10.1214/14-STS480

Instrumental Variables: An Econometrician's Perspective[1]

Guido W. Imbens

Abstract. I review recent work in the statistics literature on instrumental variables methods from an econometrics perspective. I discuss some of the older, economic, applications including supply and demand models and relate them to the recent applications in settings of randomized experiments with noncompliance. I discuss the assumptions underlying instrumental variables methods and in what settings these may be plausible. By providing context to the current applications, a better understanding of the applicability of these methods may arise.

Key words and phrases: Simultaneous equations models, randomized experiments, potential outcomes, noncompliance, selection models.

HW 23.3   TSLS 等价形式

控制函数 control function

① $OLS(D \sim Z) \Rightarrow \check{D}_i$ 残差

② $OLS(Y \sim D + \check{D})$

$\Rightarrow coef(D) = \hat{\beta}_{CF}$

结论: $\hat{\beta}_{CF} = \hat{\beta}_{TSLS}$

重要特例

structural form

$$
\begin{cases}
Y_i = \beta_0 + \beta_1 \underbrace{D_i}_{\text{一维}} + \underbrace{\beta_2^T X_i}_{\text{协变量}} + \varepsilon_i \\
D_i = \gamma_0 + \gamma_1 Z_i + \gamma_2^T X_i + \varepsilon_{2i}
\end{cases}
$$

记 $\text{"}D\text{"} = \begin{pmatrix} 1 \\ D \\ X \end{pmatrix}$, $\text{"}Z\text{"} = \begin{pmatrix} 1 \\ Z \\ X \end{pmatrix}$

reduced form

$$
\begin{cases}
Y_i = \Gamma_0 + \overbrace{\Gamma_1}^{=\beta_1\gamma_1} Z_i + \Gamma_2^T X_i + \varepsilon_{1i} \\
D_i = \underbrace{\gamma_0 + \gamma_1 Z_i + \gamma_2^T X_i}_{\text{外生变量}} + \varepsilon_{2i}
\end{cases}
$$

内生变量 (Y_i, D_i)

外生变量

间接最小二乘 Indirect Least Squares (ILS)

① $OLS(Y \sim Z + X)$

$$\Rightarrow coef(Z) = \hat{\Gamma}_1$$

② $OLS(D \sim Z + X)$

$$\Rightarrow coef(Z) = \hat{\gamma}_1$$

③ $$\hat{\beta}_{1,ILS} = \hat{\Gamma}_1 / \hat{\gamma}_1$$

定理 23.1

习题

$$\hat{\beta}_{1,ILS} = \hat{\beta}_{1,TSLS}$$

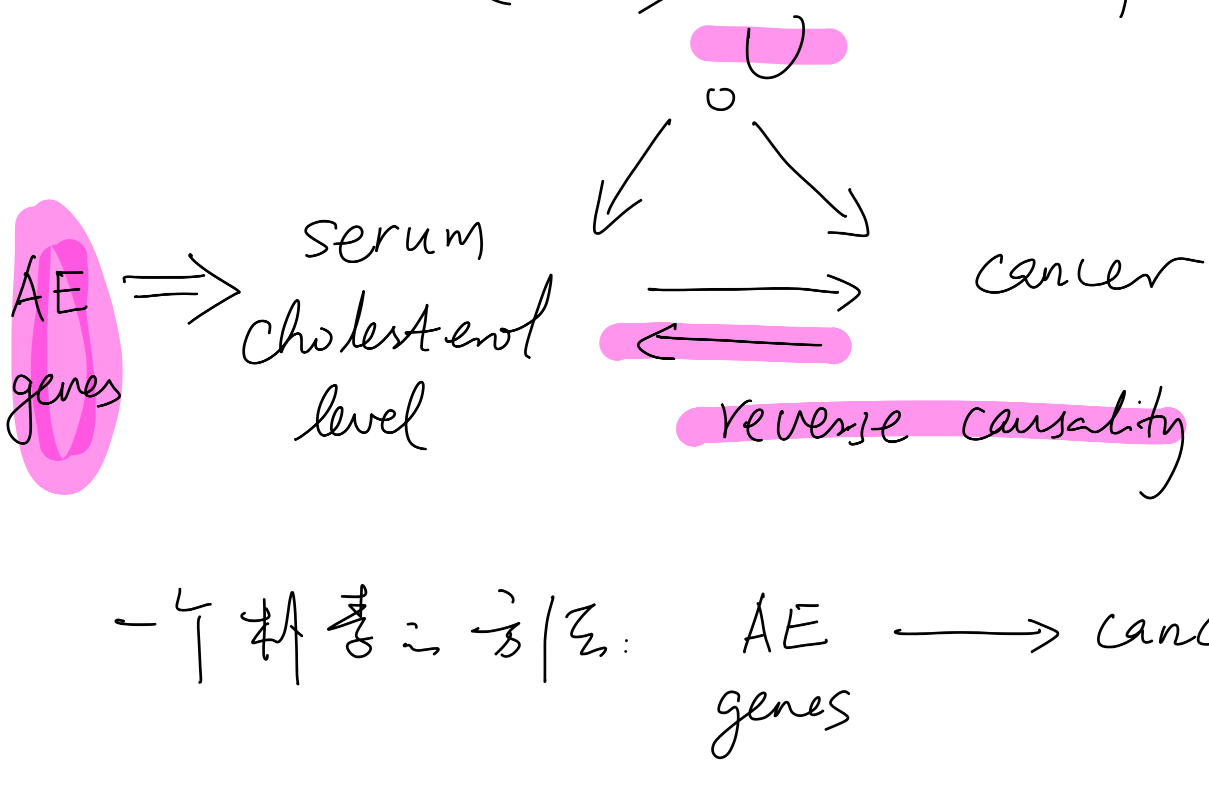
Chapter 25 IV 应用: MR

Mendelian Randomization

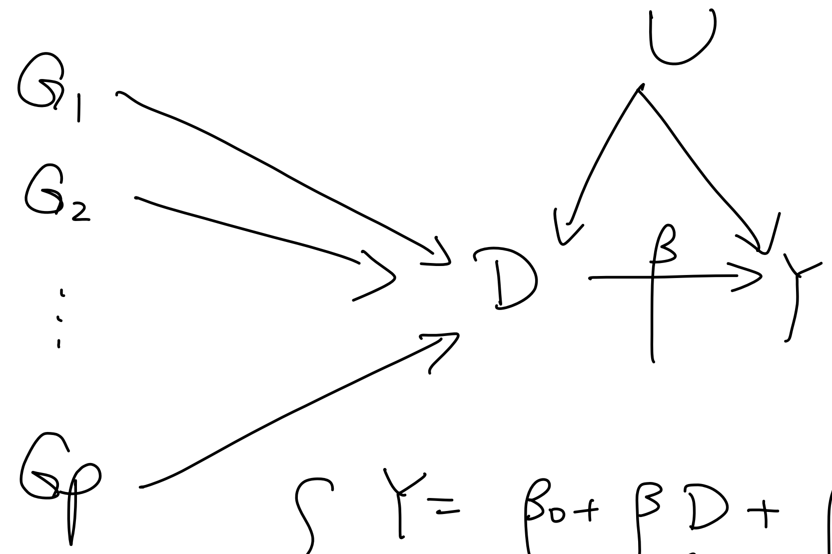
25.4: 批评 MR

先假装这些批评不存在

Katan (1986) 观察性研究



一个朴素之方法: AE genes $\longrightarrow$ cancer



$$\begin{cases} Y = \beta_0 + \beta D + \beta_u U + \varepsilon_Y \\ D = \gamma_0 + \gamma_1 G_1 + \cdots + \gamma_p G_p + \gamma_u U + \varepsilon_D \end{cases}$$

$$\Rightarrow \begin{cases} Y = (\beta_0 + \beta\gamma_0) + \underbrace{\beta\gamma_1}_{=\Gamma_1} G_1 + \cdots + \underbrace{\beta\gamma_p}_{=\Gamma_p} G_p + (\beta_u + \beta_0\gamma_u) U + \varepsilon_Y \\ D = \gamma_0 + \gamma_1 G_1 + \cdots + \gamma_p G_p + \gamma_u U + \varepsilon_D \end{cases}$$

若有个体数据: TSLS

若没有个体数据:

$$\begin{pmatrix} \hat{\Gamma}_1 \\ se_{Y1} \end{pmatrix} \cdots \begin{pmatrix} \hat{\Gamma}_p \\ se_{Yp} \end{pmatrix} : \text{关于Y的 GWAS}$$

$$\begin{pmatrix}\hat{\gamma}_1 \\ se_{D1}\end{pmatrix} \cdots \begin{pmatrix}\hat{\gamma}_p \\ se_{Dp}\end{pmatrix}$$ ：关于D的 GWAS

核心公式： $T_j = \beta\gamma_j \quad (j=1\ldots p)$

β有p个估计： $\hat{T}_j = \hat{\beta}_j \hat{\gamma}_j$

$$\hat{\beta}_j = \hat{T}_j / \hat{\gamma}_j$$

如何组合？—— meta-analysis

荟萃分析

$$\left(\hat{\beta}_j, \quad se_j^2 = \left(se_{\gamma_j}^2 + \hat{\beta}_j^2 se_{D_j}^2\right) / \hat{\gamma}_j^2\right)$$

$$j = 1 \ldots p$$

Fisher 组合：

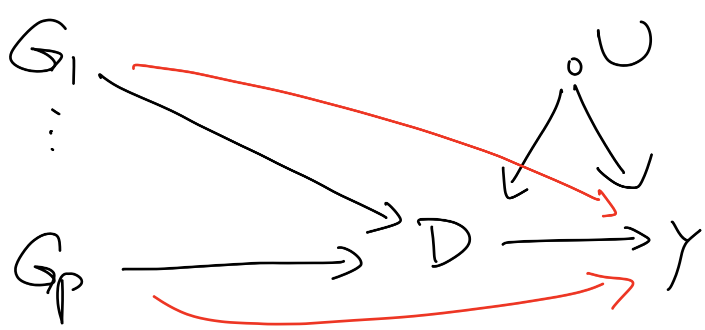
$$\hat{\beta}_{fisher_0} = \frac{\sum_{j=1}^{p} \hat{\beta}_j / se_j^2}{\sum_{j=1}^{p} 1 / se_j^2}$$

$$\hat{\beta}_{fisher_1} = \frac{\sum_{j=1}^{p} \hat{\beta}_j / \text{错误的方差}}{\sum_{j=1}^{p} 1 / \text{错误的方差}}$$

$$= \frac{\sum_{j=1}^{p} \hat{\Gamma}_j \hat{\gamma}_j / se_{Y_j}^2}{\sum_{j=1}^{p} \hat{\gamma}_j^2 / se_{Y_j}^2}$$

(文献中常见

讲义： 比上面模型稍微变了



→ : 偏差

以及何时偏差呢？

工具变量的推广

① OLS → nonparametric

$$\begin{cases} y = \underbrace{g(x)}_{\text{外生}} + \varepsilon \quad \text{相关} \\ E(\varepsilon \mid z) = 0 \end{cases}$$

$$\Rightarrow E(y \mid z) = E\left(g(x) \mid z\right) + 0$$

$$\Rightarrow E(y \mid z) = \int \underbrace{g(x)}_{?} f(x \mid z)\, dx$$

积分方程的 $g(x)$

矩阵求逆 $\longrightarrow$ 算子求逆

② 最近的工作

Jiang, Chen, Ding (Biometrika 2023)

$$
\begin{tikzcd}
 & & U \arrow[dl] \arrow[dr] & \\
Z \arrow[r] & \text{neuron 1} \arrow[rr] & & \text{neuron 2}
\end{tikzcd}
$$

活动点过程 　　　　 活动点过程

point process

JOURNAL ARTICLE CORRECTED PROOF

An instrumental variable method for point processes: generalized Wald estimation based on deconvolution

Get access >

Zhichao Jiang, Shizhe Chen, Peng Ding ✉

Biometrika, asad005, https://doi.org/10.1093/biomet/asad005

Published: 23 January 2023 **Article history** ▾