工具变量 Part V

Instrumental Variable (IV)

计量经济：instrument 成为一个动词
表示 IV 的方法

Chapter 21 An experimental perspective

IV 有实验设计的对应

IV ~ 鼓励性实验

⇓

这是一个极其平凡的视角

随机化实验很多时候是不会做的

但是可做鼓励性实验

$$Z_i : \begin{cases} 1 & \text{分配到处理} \\ 0 & \text{分配到对照} \end{cases}$$

treatment assignment

$$D_i: \begin{cases} 1 & \text{选择了处理} \\ 0 & \text{选择了对照} \end{cases}$$

如果 $Z_i \neq D_i$，有不依从 noncompliance

Y_i：结果

一个极为平凡的思考方式：

$$\tau_D = E(D(1) - D(0))$$

$$= E(D \mid Z=1) - E(D \mid Z=0)$$

$$\tau_Y = E(Y(1) - Y(0))$$

$$= E(Y \mid Z=1) - E(Y \mid Z=0)$$

只报告 τ_Y：intention-to-treat 分析

ITT

ITT 毛病：Z→Y 而不是 D→Y 的作用

问题：$(Z_i, D_i, Y_i)_{i=1}^n$ 有不依从性，如何思考和分析数据？

Econometrica, Vol. 62, No. 2 (March, 1994), 467–475

IDENTIFICATION AND ESTIMATION OF LOCAL AVERAGE TREATMENT EFFECTS[1]

BY GUIDO W. IMBENS AND JOSHUA D. ANGRIST

计量经济

Nobel Prize

1. INTRODUCTION

RANDOM ASSIGNMENT OF TREATMENT and concurrent data collection on treatment and control groups is the norm in medical evaluation research. In contrast, the use of random assignment to evaluate social programs remains controversial. Following criticism of parametric evaluation models (e.g., Lalonde (1986)), econometric research has been geared towards establishing conditions that guarantee nonparametric identification of treatment effects in observational studies, i.e. identification without relying on functional form restrictions or distributional assumptions. The focus has been on identification of average treatment effects in a population of interest, or on the average effect for the subpopulation that is treated. The conditions required to nonparametrically identify these parameters can be restrictive, however, and the derived identification results fragile. In particular, results in Chamberlain (1986), Manski (1990), Heckman (1990), and Angrist and Imbens (1991) require that there be some subpopulation for whom the probability of treatment is zero, at least in the limit.

The purpose of this paper is to show that even when there is no subpopulation available for whom the probability of treatment is zero, we can still identify an average treatment effect of interest under mild restrictions satisfied in a wide range of models and circumstances. We call this a *local average treatment effect* (LATE). Examples of problems where the local average treatment effect is identified include latent index models and evaluations based on natural experiments such as those studied by Angrist (1990) and Angrist and Krueger (1991). LATE is the average treatment effect for individuals whose treatment status is influenced by changing an exogenous regressor that satisfies an exclusion restriction.

JASA

Identification of Causal Effects Using Instrumental Variables

Joshua D. ANGRIST, Guido W. IMBENS, and Donald B. RUBIN

We outline a framework for causal inference in settings where assignment to a binary treatment is ignorable, but compliance with the assignment is not perfect so that the receipt of treatment is nonignorable. To address the problems associated with comparing subjects by the ignorable assignment—an "intention-to-treat analysis"—we make use of instrumental variables, which have long been used by economists in the context of regression models with constant treatment effects. We show that the instrumental variables (IV) estimand can be embedded within the Rubin Causal Model (RCM) and that under some simple and easily interpretable assumptions, the IV estimand is the average causal effect for a subgroup of units, the compliers. Without these assumptions, the IV estimand is simply the ratio of intention-to-treat causal estimands with no interpretation as an average causal effect. The advantages of embedding the IV approach in the RCM are that it clarifies the nature of critical assumptions needed for a causal interpretation, and moreover allows us to consider sensitivity of the results to deviations from key assumptions in a straightforward manner. We apply our analysis to estimate the effect of veteran status in the Vietnam era on mortality, using the lottery number that assigned priority for the draft as an instrument, and we use our results to investigate the sensitivity of the conclusions to critical assumptions.

KEY WORDS: Compliers; Intention-to-treat analysis; Local average treatment effect; Noncompliance; Nonignorable treatment assignment; Rubin-Causal-Model; Structural equation models.

数学工具. 测度概率论

Z 随机

D, Y 在 Z 之后

潜在结果 $\underbrace{D_i(1), D_i(0)}_{\text{二值}}, \underbrace{Y_i(1), Y_i(0)}_{\text{可以一般}}$

隐变量

$$U_i = \left(D_i(1), D_i(0)\right) = \begin{cases} a & (1,1) \\ c & (1,0) \\ d & (0,1) \\ n & (0,0) \end{cases}$$

a : always taker

c : complier

d : defier

n : never taker

$$\tau_Y = E\left(Y(1) - Y(0)\right)$$

$$
\begin{aligned}
\text{全概率} = \; & \cancel{E\left(Y(1)-Y(0) \mid U=a\right) \Pr(U=a)}_{=0 \;\; ER} \\
& + E\left(Y(1)-Y(0) \mid U=c\right) \Pr(U=c) \\
& + \cancel{E\left(Y(1)-Y(0) \mid U=d\right) \Pr(U=d)}_{=0 \;\; \text{单调性}} \\
& + \cancel{E\left(Y(1)-Y(0) \mid U=n\right) \Pr(U=n)}_{=0 \;\; ER}
\end{aligned}
$$

单调性 (monotonicity): $D_i(1) \geq D_i(0)$

$$\Rightarrow \Pr(U=d) = 0$$

特例: one-sided noncompliance

$$D_i(0) = 0$$

排除约束 (exclusion restriction) ER

如果 $D_i(1)=D_i(0)$ 那么 $Y_i(1)=Y_i(0)$

或 Z 仅通过 D 改变 Y

或 Z 到 Y 无直接作用

或 $Y_i(1)=Y_i(0)$ 对 $U_i=$ a, n 成立

$$\Rightarrow \tau_Y = \underbrace{E\left(Y(1)-Y(0)\,\middle|\,U=c\right)}_{\text{定义 CACE 或 LATE}} \Pr(U=c)$$

CACE: complier average causal effect

LATE: local average treatment effect

$$\tau_D = E\big(D(1) - D(0)\big) \quad \text{与 } \tau_Y \text{ 平行}$$

$$= \cancel{E\big(D(1)-D(0) \mid U=a\big)}^{=0} \Pr(U=a)$$

$$+ \underbrace{E\big(D(1)-D(0) \mid U=c\big)}_{=1} \Pr(U=c)$$

$$+ E\big(D(1)-D(0) \mid U=d\big) \underbrace{\cancel{\Pr(U=d)}}_{=0 \text{ 单调性}}$$

$$+ \cancel{E\big(D(1)-D(0) \mid U=n\big)}^{=0} \Pr(U=n)$$

$$\Rightarrow \tau_D = \Pr(U=c)$$

两式取子起来: CACE (LATE)

$$= E\big(Y(1) - Y(0) \mid U=c\big)$$

或

$$= E\big(Y(1) - Y(0) \mid D(1)=1, D(0)=0\big)$$

$$= E\big(Y(1) - Y(0) \mid D(1) > D(0)\big)$$

$$= \frac{\tau_Y}{\tau_D}$$

$$\overset{\text{随机化}}{=\!=} \frac{E(Y \mid Z=1) - E(Y \mid Z=0)}{E(D \mid Z=1) - E(D \mid Z=0)}$$

矩估计．方差估计（见讲义）

记号的问题

$$CACE = E\big(\underbrace{Y(Z=1) - Y(Z=0)}_{\text{仍然是 } Z \to Y \text{ 作用}} \mid U=c\big)$$

回答了之前的问题吗？

$U=c$: $Z=D$

$\Rightarrow$ CACE 也是 $D \to Y$ 作用

这种写法是 Frangakis & Rubin 2002 (Biometrics)

Imbens & Angrist: $D(1), D(0)$ $\to$ Z

$Y(d)$ 记号暗含 ER

Angrist, Imbens, Rubin: $D(1), D(0)$

$Y(z, d)$: 2×2 因子试验 factorial experiment

ER: $Y(1, d) = Y(0, d)$

或 $Y(z, d) = Y(d)$

$\Rightarrow$ 结论不变

Identification of Causal Effects Using Instrumental Variables

Joshua D. ANGRIST, Guido W. IMBENS, and Donald B. RUBIN (AIR)

个人经历:

兴趣的AIR来(2007)

We outline a framework for causal inference in settings where assignment to a binary treatment is ignorable, but compliance with the assignment is not perfect so that the receipt of treatment is nonignorable. To address the problems associated with comparing subjects by the ignorable assignment—an "intention-to-treat analysis"—we make use of instrumental variables, which have long been used by economists in the context of regression models with constant treatment effects. We show that the instrumental variables (IV) estimand can be embedded within the Rubin Causal Model (RCM) and that under some simple and easily interpretable assumptions, the IV estimand is the average causal effect for a subgroup of units, the compliers. Without these assumptions, the IV estimand is simply the ratio of intention-to-treat causal estimands with no interpretation as an average causal effect. The advantages of embedding the IV approach in the RCM are that it clarifies the nature of critical assumptions needed for a causal interpretation, and moreover allows us to consider sensitivity of the results to deviations from key assumptions in a straightforward manner. We apply our analysis to estimate the effect of veteran status in the Vietnam era on mortality, using the lottery number that assigned priority for the draft as an instrument, and we use our results to investigate the sensitivity of the conclusions to critical assumptions.

KEY WORDS: Compliers; Intention-to-treat analysis; Local average treatment effect; Noncompliance; Nonignorable treatment assignment; Rubin-Causal-Model; Structural equation models.

主观判断: 接近完美的统计文章

Data Analysis Using Stein's Estimator and Its Generalizations

BRADLEY EFRON and CARL MORRIS*

另一篇的
因果推断
但也接近
完美的
统计文章

In 1961, James and Stein exhibited an estimator of the mean of a multivariate normal distribution having uniformly lower mean squared error than the sample mean. This estimator is reviewed briefly in an empirical Bayes context. Stein's rule and its generalizations are then applied to predict baseball averages, to estimate toxomosis prevalence rates, and to estimate the exact size of Pearson's chi-square test with results from a computer simulation. In each of these examples, the mean square error of these rules is less than half that of the sample mean.

1. INTRODUCTION

Charles Stein [15] showed that it is possible to make a uniform improvement on the maximum likelihood estimator (MLE) in terms of total squared error risk when estimating several parameters from independent normal observations. Later James and Stein [13] presented a

rewards of procedures like Stein's. They have the added advantage of having the true parameter values available for comparison of methods. The examples chosen are the first and only ones considered for this report, and the favorable results typify our previous experience.

To review the James-Stein estimator in the simplest setting, suppose that for given θ_i

$$X_i \mid \theta_i \overset{\text{ind}}{\sim} N(\theta_i, 1), \quad i = 1, \cdots, k \geq 3\,, \qquad (1.1)$$

meaning the $\{X_i\}$ are independent and normally distributed with mean $E_{\theta_i} X_i \equiv \theta_i$ and variance $\text{Var}_{\theta_i}(X_i) = 1$. The example (1.1) typically occurs as a reduction to this canonical form from more complicated situations, as when X_i is a sample mean with known variance that is

CACE / LATE: 诺奖级别的工作

想法其实不是新的

BIOMETRICS 45, 619–622
June 1989

鼓励性实验

Simultaneous-Equation Estimation in a Clinical Trial of the Effect of Smoking on Birth Weight

Thomas Permutt and J. Richard Hebel

Department of Epidemiology and Preventive Medicine,
University of Maryland School of Medicine, 655 W. Baltimore Street,
Baltimore, Maryland 21201, U. S. A.

SUMMARY

Controlled experiments can be used to study the effects on health of behaviors that cannot be perfectly controlled. A simple statistical technique allows causal effects to be distinguished from selection effects. The technique is applied to measure the effect of maternal smoking on birth weight.

One particular nonlinear form is instructive to consider. Suppose the effect of smoking is all or nothing. Then the subjects (who all smoked at randomization) may be divided into four classes:

(1) would have stopped smoking regardless of intervention,
(2) would have kept smoking regardless of intervention,
(3) would have stopped only with intervention,
(4) would have stopped only without intervention.

定义的

This article examines how to estimate the effect of a program in the presence of no-shows—persons who are assigned to the program but do not participate. The article briefly discusses the methodological problems involved, describes two current experimental evaluations that are subject to these problems, presents several estimators that overcome these problems, outlines the conditions necessary for these estimators to be feasible, and describes two extensions of the analysis that illustrate a potentially broad range of further applications.

估计量 $\frac{\hat{\tau}_Y}{\hat{\tau}_D}$

1984

ACCOUNTING FOR NO-SHOWS IN EXPERIMENTAL EVALUATION DESIGNS

Bloom's estimator

HOWARD S. BLOOM
Harvard University

Chapter 22 工具变量不等式

IV inequalities

IV 的不等式

IV 的假定

随机化 $Z \perp\!\!\!\perp (D(1), D(0), Y(1), Y(0))$

单调性 $D(1) \geq D(0)$ 或 $Pr(U = d) = 0$

工具变量有用性 $\tau_D \neq 0$

排除约束 $Y(1) = Y(0)$ 对 $U = a$ 或 n 成立

falsify

无法证伪

如果 $\tau_D > 0$

无法证伪

$$\Rightarrow \tau_c = E(Y(1) - Y(0) \mid U = c)$$

$$= \frac{\tau_Y}{\tau_D}$$

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

更精细的结果：

观察

Z	D	隐变量 $U = (D(1), D(0))$
1	1	c, a
1	0	n
0	1	a
0	0	c, n

知道U

混合分布

$$\Pr(D=0 \mid Z=1) = \Pr(U=n) = \pi_n$$

$$\Pr(D=1 \mid Z=0) = \Pr(U=a) = \pi_a$$

$$\Rightarrow \Pr(U=c) = \pi_c = \underbrace{1 - \Pr(D=0 \mid Z=1)}_{\Pr(D=1 \mid Z=1)} - \Pr(D=1 \mid Z=0)$$

$$= \tau_D$$

$$E(\underset{(1)}{Y} \mid \underbrace{Z=1, D=0}_{\Updownarrow \; Z=1,\, U=n}) = E(\underset{\overset{\| \text{ER}}{Y(0)}}{Y(1)} \mid U=n)$$

$$E(Y \mid Z=0, D=1) = E(\underset{\overset{\| \text{ER}}{Y(1)}}{Y(0)} \mid U=a)$$

关注混合分布

$$E(Y \mid Z=1, D=1) = E(Y(1) \mid \cancel{Z=1}, D(1)=1)$$

$$= E(Y(1) \mid \underbrace{D(1)=1}_{\substack{\text{可能 } D(0)=1 \\ \text{可能 } D(0)=0}})$$

$$\overset{\text{全概率}}{=} \underbrace{\Pr(D(0)=1 \mid D(1)=1)}_{=\frac{\pi_a}{\pi_a+\pi_c}} \underline{E(Y(1) \mid U=a)} \; = E(Y \mid Z=0, D=1)$$

$$+ \underbrace{\Pr(D(0)=0 \mid D(1)=1)}_{=\frac{\pi_c}{\pi_a+\pi_c}} E(Y(1) \mid U=c) \; ?$$

$$\Rightarrow E\big(Y(1) \mid U=c\big) \quad \text{一元一次方程}$$

$$= \frac{E(DY \mid Z=1) - E(DY \mid Z=0)}{\pi_c}$$

$$E(Y \mid Z=0, D=0) = E\big(Y(0) \mid \cancel{Z=0}, D(0)=0\big)$$

$$= E\big(Y(0) \mid D(0)=0\big)$$

全概率

$$= \underbrace{\Pr\big(D(1)=1 \mid D(0)=0\big)}_{=\frac{\pi_c}{\pi_c+\pi_n}} E\big(Y(0) \mid U=c\big) \; ?$$

$$+ \underbrace{\Pr\big(D(1)=0 \mid D(0)=0\big)}_{=\frac{\pi_n}{\pi_c+\pi_n}} \underbrace{E\big(Y(0) \mid U=n\big)}_{=E(Y \mid Z=1, D=0)}$$

$$\Rightarrow E\big(Y(0) \mid U=c\big) \quad \text{一元一次}$$

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

简单验证: CACE

$$= E(Y(1)\mid U=c) - E(Y(0)\mid U=c)$$

两项分别都可识别

代入公式化简

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

一样!

如果 Y 为二值

那么 $0 \le E(Y(z)\mid U=u) \le 1$.

特别地 $\begin{cases} 0 \le E(Y(1)\mid U=c) \le 1 \\ 0 \le E(Y(0)\mid U=c) \le 1 \end{cases}$

代入上面的公式

$$\Rightarrow \begin{cases} E(DY|Z=1) - E(DY|Z=0) \geq 0 \\ E(DY|Z=1) - E(DY|Z=0) \leq E(D|Z=1) - E(D|Z=0) \\ E((1-D)Y|Z=0) - E((1-D)Y|Z=1) \geq 0 \\ E((1-D)Y|Z=0) - E((1-D)Y|Z=1) \leq E(D|Z=1) - E(D|Z=0) \end{cases}$$

$$\Rightarrow E(Q|Z=1) - E(Q|Z=0) \geq 0$$

其中 $Q = \begin{cases} DY \\ D(1-Y) \\ (D-1)Y \\ D+Y-DY \end{cases}$ 可识别的条件

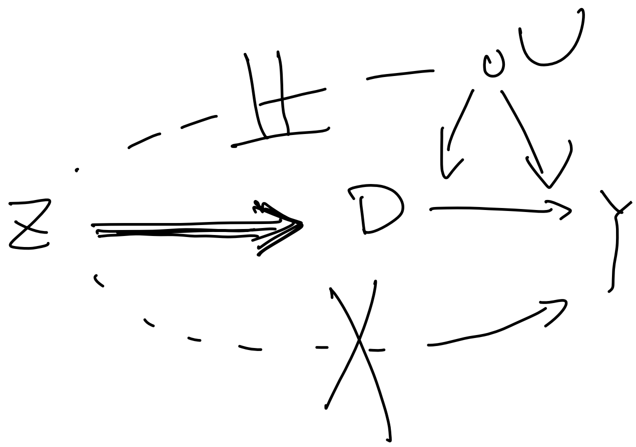
Balke & Pearl (1997 JASA)

Imbens & Rubin (1997 RES)

目的估计 $\begin{cases} f(Y(1)=y \mid U=c) \\ f(Y(0)=y \mid U=c) \end{cases}$

HW 22.4 一般连续 Y 的 IV 不等式

Chapter 23 计量经济的角度



艺术性：
从观察性数据中找到近似的随机的 Z 作为 IV

线性模型:

甚至可以为向量

$$Y = D^T \beta + \varepsilon$$

OLS

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

$$E(\varepsilon D) = 0$$

$$\longrightarrow \beta$$

对统计学家: 废话!

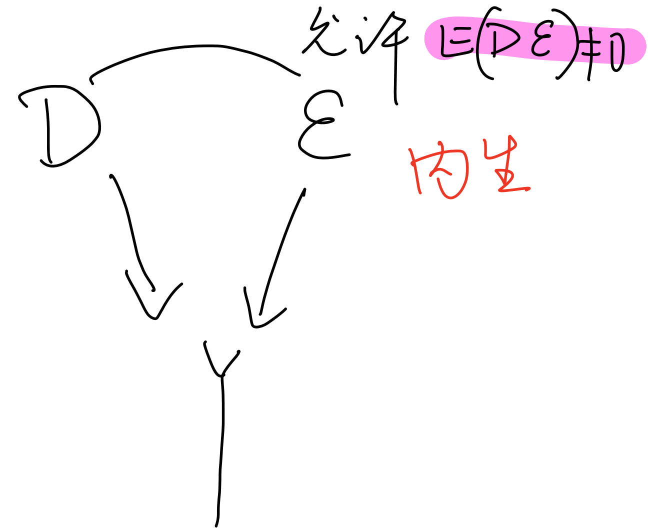
$$(D, Y)$$

定义 $\beta = \arg\min_b E\left(Y - D^T b\right)^2$

定义 $\varepsilon = Y - D^T \beta$

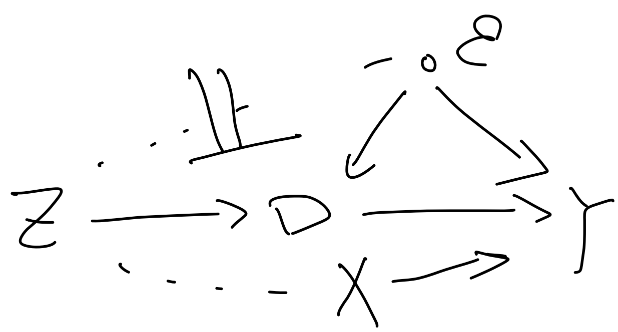
$\Rightarrow E(D\varepsilon) = 0$ 成立!

计量经济:



$\hat{\beta}_{OLS}$ 可能不相合.

Z 满足 $E(Z\varepsilon)=0$



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

exclusion restriction [illegible]

$$\Rightarrow E\left(Z\left(Y - D^T\beta\right)\right) = 0$$

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

解这个线性方程 !!

一个特例: Z, D 二值

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

$$\Rightarrow \mathrm{cov}(Z, Y) = \beta\, \underbrace{\mathrm{cov}(Z, D)}_{\neq 0}$$

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

Z 二值 $= \dfrac{E(Y \mid Z=1) - E(Y \mid Z=0)}{E(D \mid Z=1) - E(D \mid Z=0)}$

IA、AIR

贡献

CACE