应用微观计量经济学

Applied Microeconometrics

江　艇
中国人民大学经济学院

Last updated: May 6, 2019

Time:	周一晚 18:00-21:15
Venue:	明法 0502
Office:	明主 952
E-mail:	`ting.jiang@ruc.edu.cn`
Phone:	8250-0201
Grading:	期末闭卷考试

Lecture 6　工具变量方法

6.1　传统 IV 方法

工作原理

- 关键解释变量内生导致其系数的 OLS 估计不一致。

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

$$b_1^{\text{OLS}} = \frac{\widehat{\text{Cov}(y_i, x_i)}}{\widehat{\text{Var}(x_i)}} \to_{\text{p}} \frac{\text{Cov}(y_i, x_i)}{\text{Var}(x_i)} = \beta_1 + \frac{\text{Cov}(\varepsilon_i, x_i)}{\text{Var}(x_i)} \neq \beta_1$$

if $\text{E}(x_i \varepsilon_i) \neq 0$.

- 工具变量方法

$$b_1^{\text{IV}} \triangleq \frac{\widehat{\text{Cov}(y_i, z_i)}}{\widehat{\text{Cov}(x_i, z_i)}} \to_{\text{p}} \frac{\text{Cov}(y_i, z_i)}{\text{Cov}(x_i, z_i)} = \beta_1 + \frac{\text{Cov}(\varepsilon_i, z_i)}{\text{Cov}(x_i, z_i)} = \beta_1$$

if $\text{Cov}(\varepsilon_i, z_i) = 0$ and $\text{Cov}(x_i, z_i) \neq 0$.

- 直观含义：可以这样思考问题,

$$\frac{dy}{dx} = \beta_1 + \frac{d\varepsilon}{dx}$$

因为 $d\varepsilon/dx \neq 0$，所以无法识别 β_1。在教育回报率的例子中，如果存在某个工具变量 z（先不讨论它具体是什么），它每变动 1 单位，受教育年限和年收入就会分别变动 0.2 年和 500 元，并且**收入变动是由 z 的变动引起的受教育年限变动所间接导致的**，这意味着 0.2 年的受教育年限变动对应着 500 元的收入变动, 那么受教育年限每增加 1 年将对应 $500/0.2 = 2,500$ 元的收入增加，因果效应估计就是 2500。

也就是说，我们同时估计了 dx/dz 和 dy/dz，因果效应估计量就是

$$\beta^{\mathsf{IV}} = \frac{dy/dz}{dx/dz}$$

那么如何一致估计 dy/dz 和 dx/dz 呢？当然是分别用 y 和 x 对 z 进行 OLS 回归，于是

$$\hat{\beta}^{\mathsf{IV}} = \frac{\widehat{\mathsf{Cov}(y,z)}/\widehat{\mathsf{Var}(z)}}{\widehat{\mathsf{Cov}(x,z)}/\widehat{\mathsf{Var}(z)}} = \frac{\widehat{\mathsf{Cov}(y,z)}}{\widehat{\mathsf{Cov}(x,z)}}$$

- 两阶段最小二乘法 (2SLS)
  - 第 1 阶段：x_i 对 z_i 进行 OLS 回归，得到拟合值 $\hat{x}_i$。

$$x_i = \gamma_0 + \gamma_1 z_i + \omega_i = \hat{x}_i + \omega_i$$

  - 第 2 阶段：y_i 对 $\hat{x}_i$ 进行 OLS 回归。

$$y_i = \beta_0 + \beta_1 \hat{x}_i + [\varepsilon_i + \beta_1 (x_i - \hat{x}_i)]$$

$$b_1^{\text{2SLS}} \triangleq \frac{\widehat{\text{Cov}(y_i, \hat{x}_i)}}{\widehat{\text{Var}(\hat{x}_i)}}$$

  - 第 2 阶段的解释变量确实与扰动项不相关。

$$\text{Cov}(\hat{x}_i, \varepsilon_i + \beta_1(x_i - \hat{x}_i)) = \text{Cov}(\hat{x}_i, \varepsilon_i) + \beta_1 \text{Cov}(\hat{x}_i, \omega_i) = 0$$

内生性的来源

- 联立性偏误

$$q_i^d = \alpha_0 + \alpha_1 p_i + u_i$$

$$q_i^s = \beta_0 + \beta_1 p_i + v_i$$

$$q_i^d = q_i^s$$

$$\mathsf{E}(u_i) = \mathsf{E}(v_i) = \mathsf{Cov}(u_i, v_i) = 0$$

$$p_i = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1} + \frac{v_i - u_i}{\alpha_1 - \beta_1}$$

$$q_i = \frac{\alpha_1\beta_0 - \alpha_0\beta_1}{\alpha_1 - \beta_1} + \frac{\alpha_1 v_i - \beta_1 u_i}{\alpha_1 - \beta_1}$$

$$\mathsf{Cov}(p_i, u_i) = -\frac{\mathsf{Var}(u_i)}{\alpha_1 - \beta_1}$$

$$\mathsf{Cov}(p_i, v_i) = \frac{\mathsf{Var}(v_i)}{\alpha_1 - \beta_1}$$

$$
\begin{aligned}
b^{\mathrm{OLS}} &= \frac{\widehat{\mathrm{Cov}(p_i, q_i)}}{\widehat{\mathrm{Var}(p_i)}} \to_{\mathrm{p}} \frac{\mathrm{Cov}(p_i, q_i)}{\mathrm{Var}(p_i)} \\
&= \alpha_1 + \frac{\mathrm{Cov}(p_i, u_i)}{\mathrm{Var}(p_i)} \text{ or } \beta_1 + \frac{\mathrm{Cov}(p_i, v_i)}{\mathrm{Var}(p_i)}
\end{aligned}
$$

- 反向因果

$$
g_i = \alpha_0 + \alpha_1 d_i + u_i
$$
$$
d_i = \beta_0 + \beta_1 g_i + v_i
$$

- 遗漏变量偏误

$$
y = \beta_0 + \beta_1 x + \gamma q + \varepsilon
$$

$$
y = \beta_0 + \beta_1 x + u, \; u \triangleq \gamma q + \varepsilon
$$

$$
b_1^{\mathrm{OLS}} \to_{\mathrm{p}} \beta_1 + \frac{\mathrm{Cov}(x, u)}{\mathrm{Var}(x)} = \beta_1 + \gamma \frac{\mathrm{Cov}(x, q)}{\mathrm{Var}(x)}
$$

- 测量误差

$$y = \beta_0 + \beta_1 x^* + \varepsilon$$

$$x = x^* + e$$

$$\mathsf{E}(e) = 0,\ \mathsf{Cov}(\varepsilon, e) = 0$$

$$\mathsf{Cov}(x^*, e) = 0$$

$$b_1^{\texttt{OLS}} \to_{\mathsf{p}} \frac{\mathsf{Cov}(y, x)}{\mathsf{Var}(x)} = \beta_1 \left(\frac{\sigma_{x^*}^2}{\sigma_{x^*}^2 + \sigma_e^2} \right) = \beta_1 \left(\frac{1}{1 + \frac{\sigma_e^2}{\sigma_{x^*}^2}} \right)$$

当模型中包含协变量时，趋零偏误会增大。

$$b_1^{\texttt{OLS}} \to_{\mathsf{p}} \frac{\mathsf{Cov}(y, \tilde{x})}{\mathsf{Var}(\tilde{x})} = \frac{\mathsf{Cov}(y, \tilde{x}^* + e)}{\mathsf{Var}(\tilde{x})} = \beta_1 \left(\frac{\sigma_{\tilde{x}^*}^2}{\sigma_{\tilde{x}^*}^2 + \sigma_e^2} \right) = \beta_1 \left(\frac{1}{1 + \frac{\sigma_e^2}{\sigma_{\tilde{x}^*}^2}} \right)$$

$$\sigma_{x^*}^2 > \sigma_{\tilde{x}^*}^2$$

类似地，控制固定效应会放大趋零偏误。为了说明的方便，假定采用一阶差分方法消除固定效应，

$$\Delta y_{it} = \beta_1 \Delta x^*_{it} + \Delta \varepsilon_{it}$$

$$x_{it} = x^*_{it} + e_{it}$$

$$b_1^{\mathtt{FE}} \to_p \beta_1 \left(\frac{\sigma^2_{\Delta x^*}}{\sigma^2_{\Delta x^*} + \sigma^2_{\Delta e}} \right)$$

如果 x_{it} 平稳，

$$\sigma^2_{\Delta x^*} = Var(x^*_{it}) - 2Cov(x^*_{it}, x^*_{i,t-1}) + Var(x^*_{i,t-1}) = 2\sigma^2_{x^*}(1 - \rho_{x^*})$$

ρ_e 类似定义，则

$$b_1^{\mathtt{FE}} \to_p \beta_1 \left(\frac{\sigma^2_{x^*}(1 - \rho_{x^*})}{\sigma^2_{x^*}(1 - \rho_{x^*}) + \sigma^2_e(1 - \rho_e)} \right) = \beta \left(\frac{1}{1 + \frac{\sigma^2_e(1-\rho_e)}{\sigma^2_{x^*}(1-\rho_{x^*})}} \right)$$

通常

$$\rho_{x^*} > \rho_e$$

工具变量需要满足的条件

$$
\begin{tikzcd}
x \arrow[r] & y \\
\varepsilon \arrow[ur] &
\end{tikzcd}
\qquad
\begin{tikzcd}
x \arrow[r] & y \\
\varepsilon \arrow[u] \arrow[ur] &
\end{tikzcd}
\qquad
\begin{tikzcd}
z \arrow[r] & x \arrow[r] & y \\
& \varepsilon \arrow[u] \arrow[ur] &
\end{tikzcd}
\qquad
\begin{tikzcd}
z \arrow[r, "\checkmark"] \arrow[rr, bend left, dashed, "\times"] & x \arrow[r] & y \\
& \varepsilon \arrow[u] \arrow[ur] \arrow[ul, dashed, "\times"] &
\end{tikzcd}
$$

- 相关性：工具变量与内生解释变量相关。

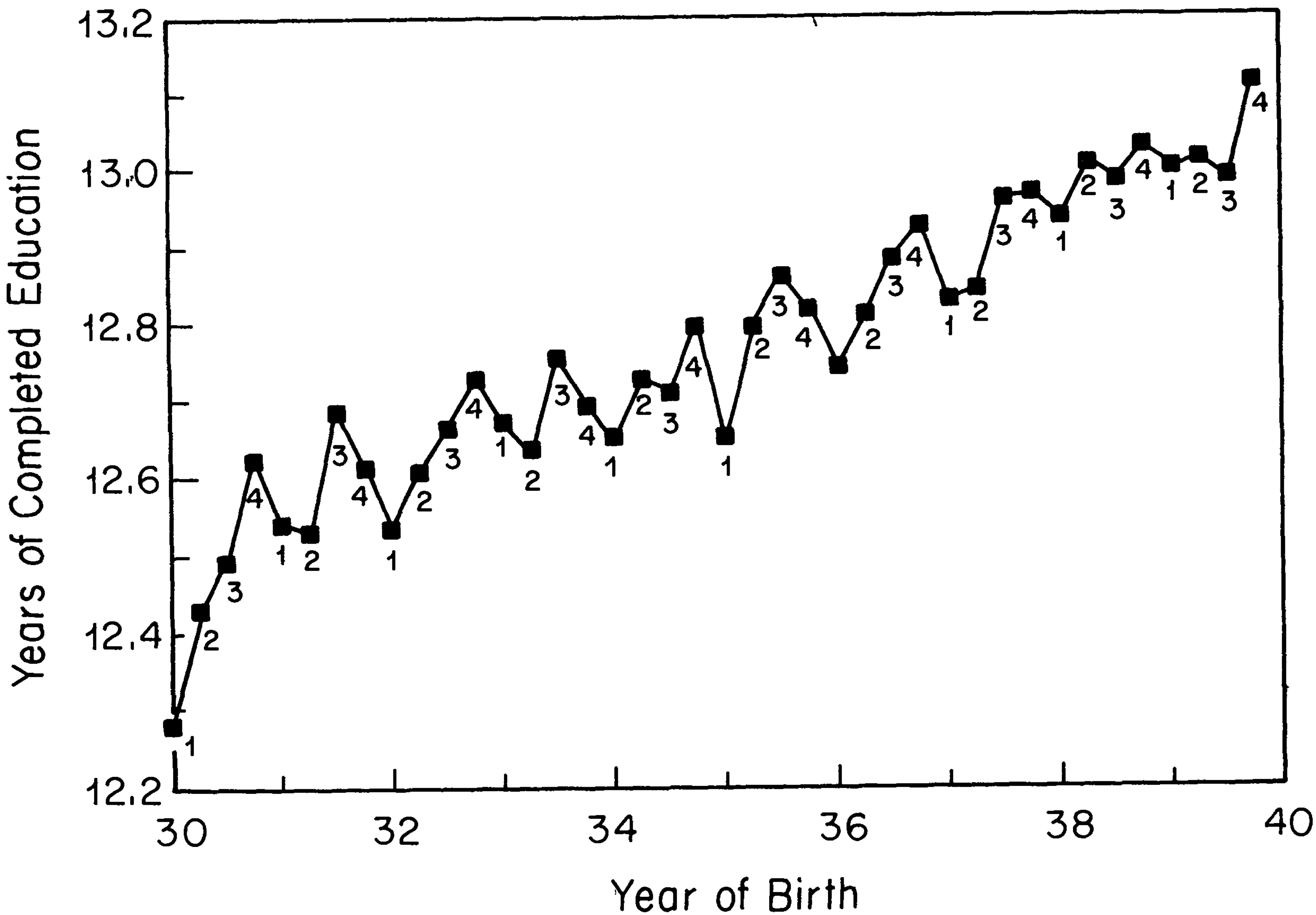
$$\text{Cov}(z, x) \neq 0$$

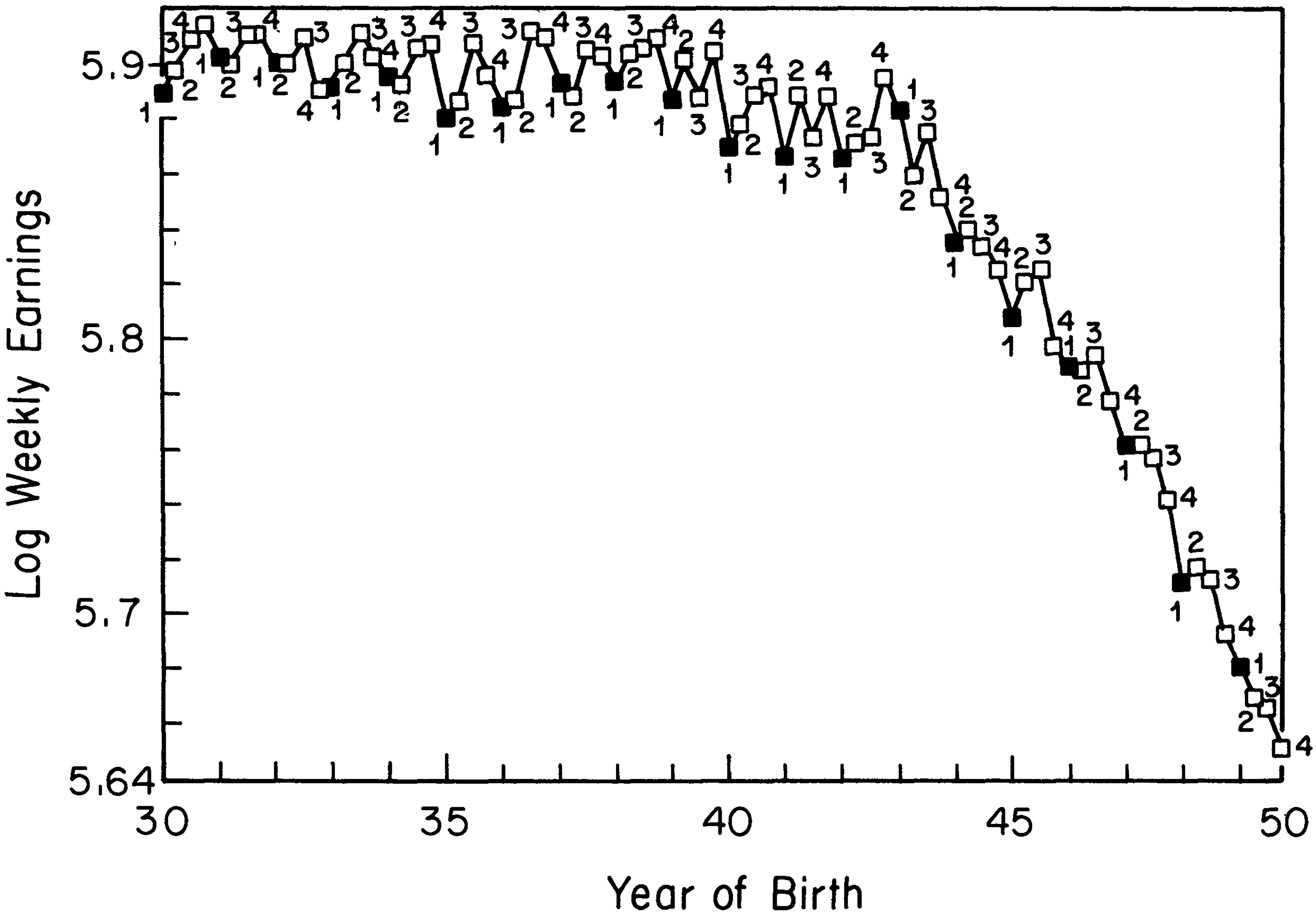
- 排斥性约束：工具变量不出现在结构方程右边。
- 外生性或独立性：工具变量与扰动项不相关。

$$\text{Cov}(z, \varepsilon) = 0$$

- 那么，什么是教育回报率例子中的 z？

示例 1. Angrist and Krueger (1991, *QJE*)




Log Weekly Earnings
5.9
5.8
5.7
5.64
30
35
40
45
50
Year of Birth

- Qin (2015, *unpublished*) 对工具变量方法的批评：工具变量只讲相关性，不讲因果性。这是真的么？

  答：尽管没有明确表述，但是 z 到 x 的因果关系必然隐含在相关性的叙事中。试想，如果是 x 影响了 z，这就意味着 z 和 ε 产生了相关；如果是另外的因素 w 同时影响了 x 和 z，那么也只有当 w 出现在 ε 之中时才会对研究设计造成威胁。因此，工具变量的相关性是否隐含着因果性，实际上可以被归结为对工具变量独立性条件的讨论。反过来说，**一个因果性不强的相关性叙事往往不容易满足独立性条件**。一个典型的例子是用邻近个体的 x 的均值作为本身的 x 的工具变量。

了中国发展高速铁路的基础——“四纵”和“四横”客运专线。该规划方案提出了基于区位的铁路网需要联结的 26 个重要节点城市，而除了这 26 个节点城市之外，每个城市是否通高铁，很大程度上取决于其与四纵四横连接线的地理距离，而每个城市距离规划直线的距离又是外生决定的，因此可以用作该城市实际是否通高铁的工具变量。我们借鉴 Baber（2014）的研究方法，使用该上市公司所在城市与高铁规划线的距离（取对数，IV_ln*Planlane*）作为该城市附近是否有高铁站的工具变量，考察高铁与距离变量的交乘项结果是否显著，表 7 中的回归结果显示：加入了工具变量之后，高铁哑变量与距离的交乘项回归结果依然显著为负，这也进一步印证了我们的结论。

示例 2. Faber (2014, *RES*)

y: 县 1997-2006 经济增长

x: 县 1992-2003 是否通高速

z: 根据最小生成树算法得到的 placebo 高速公路网（最小成本、欧氏距离）

示例 3. Donaldson (2018, *AER*)

y: 县农业收入

x: 是否通铁路

I begin (in Section VC) by estimating equation (16) using OLS. Unbiased OLS estimates require there to be no correlation between the error term (ε_{ot}) and the regressor $(RAIL_{ot})$, conditional on the district and year fixed effects. This requirement would fail if railroads were built in districts and years that were expected to experience real agricultural income growth, or if railroads were built in districts that were on differing unobserved trends from non-railroad districts. For this reason, in Section VD I also estimate three different "placebo" specifications in order to assess the potential magnitude of bias in my OLS results due to nonrandom railroad placement.

Four-Stage Planning Hierarchy.—From 1870–1900, India's Railways Department used one constant system for the evaluation of new railroad projects. Line proposals received from the Indian and provincial governments would appear as "proposed" in the Department's annual *Railway Report*. This invited further discussion, and if the proposed line survived this criticism it would be "reconnoitered." Providing this reconnaissance uncovered no major problems, every meter of the proposed line would then be "surveyed," this time in painstaking and costly detail (usually taking several years to complete).[38] These detailed surveys would provide accurate estimates of expected construction costs, and lines whose surveys revealed modest costs would then be passed on to the Government to be "sanctioned," or given final approval. The railroad planning process was therefore arranged as a four-stage hierarchy of tests that proposed lines would have to pass.

Dependent variable: log real agricultural income	(1)	(2)
Railroad in district	0.164	0.158
	(0.049)	(0.048)
Unbuilt railroad in district, abandoned after proposal stage		0.057
		(0.058)
Unbuilt railroad in district, abandoned after reconnaissance stage		0.013
		(0.099)
Unbuilt railroad in district, abandoned after survey stage		–0.069
		(0.038)

如果拿 placebo 铁路作为 IV 会怎样？

$$
\begin{aligned}
y =& \beta_0 + \beta_1 x + \gamma z + \varepsilon \\
x =& \pi_0 + \pi_1 z + v
\end{aligned}
$$

$$
\hat{\beta}_1 = \frac{\widehat{\text{Cov}(y, z)}}{\widehat{\text{Cov}(x, z)}} \rightarrow_{\text{p}} \beta_1 + \frac{\gamma}{\pi_1}
$$

```
. reghdfe RAIL RAIL_prop RAIL_rec RAIL_sur, absorb(distid year) vce(cluster distid)
(converged in 7 iterations)

HDFE Linear regression                            Number of obs   =      7,086
Absorbing 2 HDFE groups                           F(   3,    191) =       7.53
Statistics robust to heteroskedasticity           Prob > F        =     0.0001
                                                  R-squared       =     0.6651
                                                  Adj R-squared   =     0.6527
                                                  Within R-sq.    =     0.0066
Number of clusters (distid)  =        192         Root MSE        =     0.1670

                                 (Std. Err. adjusted for 192 clusters in distid)
------------------------------------------------------------------------------
             |               Robust
        RAIL |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   RAIL_prop |   .0898069   .0448243     2.00   0.047     .0013927    .1782211
    RAIL_rec |  -.0687267   .0214951    -3.20   0.002    -.1111249   -.0263284
    RAIL_sur |   .0087389   .0306528     0.29   0.776    -.0517227    .0692004
------------------------------------------------------------------------------
```

```
. reghdfe ln_realincome (RAIL = RAIL_prop RAIL_rec RAIL_sur), est(2sls) absorb(distid year) vce(cluster distid)
(converged in 7 iterations)

HDFE IV (2SLS) estimation
-------------------------

Estimates efficient for homoskedasticity only
Statistics robust to heteroskedasticity and clustering on distid

Number of clusters (distid) =      192              Number of obs =      7086
                                                    F(  1,   191) =      1.61
                                                    Prob > F      =    0.2057
Total (centered) SS     =   870.438354              Centered R2   =    0.8250
Total (uncentered) SS   =   870.438354              Uncentered R2 =         .
Residual SS             =  993.8346944              Root MSE      =     .3813

------------------------------------------------------------------------------
             |               Robust
ln_realinc~e |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        RAIL |   .9820156   .7732828     1.27   0.206    -.5432553    2.507287
------------------------------------------------------------------------------
```

```
. reghdfe ln_realincome (RAIL = RAIL_prop), est(2sls) absorb(distid year) vce(cluster distid)
(converged in 7 iterations)

HDFE IV (2SLS) estimation
-------------------------

Estimates efficient for homoskedasticity only
Statistics robust to heteroskedasticity and clustering on distid

Number of clusters (distid) =      192              Number of obs =       7086
                                                    F(  1,   191) =       1.73
                                                    Prob > F      =     0.1905
Total (centered) SS     =   870.438354              Centered R2   =     0.8050
Total (uncentered) SS   =   870.438354              Uncentered R2 =          .
Residual SS             =  1107.169318              Root MSE      =      .4025

------------------------------------------------------------------------------
             |               Robust
ln_realinc~e |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        RAIL |   1.286303   .9790651     1.31   0.190    -.6448661    3.217471
------------------------------------------------------------------------------
```

```
. reghdfe ln_realincome RAIL RAIL_prop, absorb(distid year) vce(cluster distid)
(converged in 7 iterations)

HDFE Linear regression                              Number of obs   =      7,086
Absorbing 2 HDFE groups                             F(   2,    191) =       6.91
Statistics robust to heteroskedasticity             Prob > F        =     0.0013
                                                    R-squared       =     0.8479
                                                    Adj R-squared   =     0.8423
                                                    Within R-sq.    =     0.0076
Number of clusters (distid)  =        192           Root MSE        =     0.3556

                               (Std. Err. adjusted for 192 clusters in distid)
------------------------------------------------------------------------------
             |               Robust
ln_realinc~e |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        RAIL |   .1568211   .0486572     3.22   0.001     .0608467    .2527955
   RAIL_prop |   .0969578   .0559735     1.73   0.085    -.0134478    .2073633
------------------------------------------------------------------------------
```

```
. reghdfe RAIL RAIL_prop, absorb(distid year) vce(cluster distid)
(converged in 7 iterations)

HDFE Linear regression                          Number of obs   =      7,086
Absorbing 2 HDFE groups                         F(   1,    191) =       3.42
Statistics robust to heteroskedasticity         Prob > F        =     0.0661
                                                R-squared       =     0.6648
                                                Adj R-squared   =     0.6525
                                                Within R-sq.    =     0.0060
Number of clusters (distid)  =        192       Root MSE        =     0.1671

                             (Std. Err. adjusted for 192 clusters in distid)
```

RAIL	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
RAIL_prop	.0858427	.0464494	1.85	0.066	-.0057769	.1774623

$$1.286303 \approx 0.1568211 + 0.0969578/0.0858427$$

- 在包含控制变量的工具变量回归中，排斥性约束要求工具变量 z 只通过关键解释变量 x_1 影响 y，那么 z 可不可以通过控制变量 x_2 影响 y 呢？

  答：当然可以。但这个问题应该反过来理解，正是因为担心工具变量有除了 x_1 之外影响 y 的渠道，故而把这些潜在的渠道尽量控制起来，这些被控制起来的渠道就成了控制变量。这正是排斥性约束检验的主要思路。

示例 4. Nunn and Wantchekon (2011, *AER*)

$$\texttt{trust}_{i,e,d,c} = \alpha_c + \beta \texttt{slave export}_e + \mathbf{X}'_{i,e,d,c}\mathbf{\Gamma}_1 + \mathbf{X}'_{e,d,c}\mathbf{\Gamma}_2 + \mathbf{X}'_{d,c}\mathbf{\Gamma}_3 + \mathbf{X}'_e\mathbf{\Gamma}_4 + \varepsilon_{i,e,d,c}$$

其中 i 表示个体，e 表示种族，d 表示地区，c 表示国家

- 被解释变量：非洲民意调查中受访者报告的各种信任指标（亲属、邻居、当地政府、族群内、族群间，离散变量按连续变量处理）
- 关键解释变量：受访者所在种族 (ethnicity) 历史上奴隶出口数量
- 个体特征 ($\mathbf{X}_{i,e,d,c}$)：年龄、性别、是否城市、生活条件、受教育水平、宗教、职业
- 种族-区域特征 ($\mathbf{X}_{e,d,c}$)：区域内同种族人口
- 区域特征 ($\mathbf{X}_{d,c}$)：种族分化程度
- 种族特征 ($\mathbf{X}_e$)

TABLE 1—OLS ESTIMATES OF THE DETERMINANTS OF TRUST IN NEIGHBORS

Dependent variable: Trust of neighbors	Slave exports (thousands) (1)	Exports/ area (2)	Exports/ historical pop (3)	ln (1 + exports) (4)	ln (1 + exports/ area) (5)	ln (1 + exports/ historical pop) (6)
Estimated coefficient	−0.00068	−0.019	−0.531	−0.037	−0.159	−0.743
	[0.00014]	[0.005]	[0.147]	[0.014]	[0.034]	[0.187]
	(0.00015)	(0.005)	(0.147)	(0.014)	(0.034)	(0.187)
	{0.00013}	{0.005}	{0.165}	{0.015}	{0.034}	{0.212}
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes
District controls	Yes	Yes	Yes	Yes	Yes	Yes
Country fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	20,027	20,027	17,644	20,027	20,027	17,644
Number of ethnicities	185	185	157	185	185	157
Number of districts	1,257	1,257	1,214	1,257	1,257	1,214
R^2	0.16	0.16	0.15	0.15	0.16	0.15

Notes: The table reports OLS estimates. The unit of observation is an individual. Below each coefficient three standard errors are reported. The first, reported in square brackets, is standard errors adjusted for clustering within ethnic groups. The second, reported in parentheses, is standard errors adjusted for two-way clustering within ethnic groups and within districts. The third, reported in curly brackets, is T. G. Conley (1999) standard errors adjusted for two-dimensional spatial autocorrelation. The standard errors are constructed assuming a window with weights equal to one for observations less than five degrees apart and zero for observations further apart. The individual controls are for age, age squared, a gender indicator variable, five living conditions fixed effects, ten education fixed effects, 18 religion fixed effects, 25 occupation fixed effects, and an indicator for whether the respondent lives in an urban location. The district controls include ethnic fractionalization of each district and the share of the district's population that is the same ethnicity as the respondent.

- 控制种族层面的殖民统治的影响
  - ▷ 控制决定殖民统治变动性的变量：初始疾病环境、前殖民时期的发展水平（人口密度、是否有城市、聚落形态、司法层级）
  - ▷ 直接控制反映殖民统治影响的变量：是否有铁路、是否有欧洲探险家经过、欧洲传教士数量
- 工具变量：奴隶贸易时期种族到海岸线的距离

as progressing in waves of destruction that originated from the coast. The critical issue is whether an ethnic group's distance from the coast in the past is uncorrelated with factors, other than the slave trade, that may affect how trusting the ethnic group is today—for example, initial prosperity, which may have affected an ethnic group's susceptibility to the slave trade, as well as its subsequent trust. Generally, we would expect distance from the coast to be correlated with overseas trade, and thus with initial prosperity. However, because of Africa's particular history, this is not a concern. In the regions in our sample, there was no overseas trade prior to the transatlantic and Indian Ocean slave trades. This alleviates concerns that initial distance from the coast may have had a direct effect on initial development via preexisting trade.

Despite this fact, there remain a number of other reasons why the exclusion restriction may not be satisfied. First, distance from the coast may be correlated with other forms of European contact, like colonial rule, which followed the slave trade. For this reason, we only report IV estimates after controlling for our full set of ethnicity-level colonial control variables. Second, locations closer to the coast were more likely to rely on fishing as a form of subsistence. Although it is not obvious how this may affect future trust, to be as thorough as possible we control for ethnicities' historical reliance on fishing. Third, for some parts of Africa, proximity to the coast implies greater distance from the ancient trade networks across the Sahara Desert. Because long-term trust may have been affected by a group's involvement in this inland trade, we also control for the average distance to the closest city in the Saharan trade, as well as the average distance to the closest route of the Saharan trade.[22]

- 排斥性约束的证伪检验 (falsification test) 之一：In samples where the first stage is zero, the reduced form should be zero as well. On the other hand, a statistically significant reduced-form estimate with no evidence of a corresponding first stage is cause for worry, because this suggests some channel other than the treatment variable links instruments with outcomes. We can construct “no-first-stage samples” and check whether they generate no evidence of significant reduced-form effects (Angrist and Pischke, 2014).

Table 7—Reduced Form Relationship between the Distance from the Coast and Trust within Africa and Asia

	Trust of local government council			
	Afrobarometer sample		Asiabarometer sample	
	(1)	(2)	(3)	(4)
Distance from the coast	0.00039***	0.00031***	−0.00001	0.00001
	(0.00009)	(0.00008)	(0.00010)	(0.00009)
Country fixed effects	Yes	Yes	Yes	Yes
Individual controls	No	Yes	No	Yes
Number of observations	19,913	19,913	5,409	5,409
Number of clusters	185	185	62	62
R^2	0.16	0.18	0.19	0.22

Notes: The table reports OLS estimates. The unit of observation is an individual. The dependent variable in the Asiabarometer sample is the respondent's answer to the question: "How much do you trust your local government?" The categories for the answers are the same in the Asiabarometer as in the Afrobarometer. Standard errors are clustered at the ethnicity level in the Afrobarometer regressions and at the location (city) level in the Asiabarometer and the WVS samples. The individual controls are for age, age squared, a gender indicator, education fixed effects, and religion fixed effects.

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

TABLE 8—REDUCED FORM RELATIONSHIP BETWEEN THE DISTANCE FROM THE COAST AND TRUST WITHIN AND OUTSIDE OF AFRICA

	Intergroup trust				
	Afrobarometer sample		WVS non-Africa sample		WVS Nigeria
	(1)	(2)	(3)	(4)	(5)
Distance from the coast	0.00039***	0.00037***	−0.00020	−0.00019	0.00054***
	(0.00013)	(0.00012)	(0.00014)	(0.00012)	(0.00010)
Country fixed effects	Yes	Yes	Yes	Yes	n/a
Individual controls	No	Yes	No	Yes	Yes
Number of observations	19,970	19,970	10,308	10,308	974
Number of clusters	185	185	107	107	16
R^2	0.09	0.10	0.09	0.11	0.06

Notes : The table reports OLS estimates. The unit of observation is an individual. The dependent variable in the WVS sample is the respondent's answer to the question: "How much do you trust <nationality> people in general?" The categories for the respondent's answers are: "not at all," "not very much," "neither trust nor distrust," "a little," and "completely." The responses take on the values 0, 1, 1.5, 2, and 3. Standard errors are clustered at the ethnicity level in the Afrobarometer regressions and at the location (city) level in the Asiabarometer and the WVS samples. The individual controls are for age, age squared, a gender indicator, an indicator for living in an urban location, and occupation fixed effects.

*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

- 排斥性约束的证伪检验之二 (Conley et al, 2012, *REStat*).

$$\begin{aligned} y =& \beta_0 + \beta_1 x + \gamma z + \varepsilon \\ x =& \pi_0 + \pi_1 z + v \end{aligned}$$

$$\hat{\beta}_1 = \frac{\widehat{\text{Cov}(y, z)}}{\widehat{\text{Cov}(x, z)}} \rightarrow_{\text{p}} \beta_1 + \frac{\gamma}{\pi_1}$$

若知道关于 γ 取值的信息，则

$$(y - \gamma z) = \beta_0 + \beta_1 x + \varepsilon$$

用 z 作为 x 的 IV 估计 β_1 并构造 $(1-\alpha)$ 置信区间：

$$CI(1-\alpha, \gamma) = [\hat{\beta}_1(\gamma) \pm t_{\alpha/2} \widehat{SE(\hat{\beta}_1(\gamma))}]$$

假设 $\gamma > 0$, $\pi_1 > 0$, $\beta_1 > 0$（其它情形类似），则 γ 越大，该置信区间越接近于零。若使得置信区间下限为零的 γ 过大，则认为结果对于 IV 的疑似内生性稳健。有 `STATA` 命令 `plausexog` 可以实现，但手动也很方便。

禁止回归 (forbidden regression)

- 当结构方程包含内生解释变量的平方项，用工具变量的平方项作为其工具变量。
- 当结构方程包含内生解释变量和控制变量的交互项，用工具变量和控制变量的交互项作为其工具变量。
- 当内生解释变量是离散变量，先回归非线性的第一阶段，然后**以第一阶段的拟合值作为工具变量进行工具变量回归**。
- 当被解释变量是离散变量，尤其对于比较复杂的数据，建议直接采用线性概率模型。

广义矩方法

- 过度识别情形下，样本矩条件方程组无解。

$$\mathrm{E}\left(\mathbf{z}_i(y_i - \mathbf{x}_i'\beta)\right) = 0$$

$$\frac{1}{n}\sum_{i=1}^{n}\mathbf{z}_i(y_i - \mathbf{x}_i'\tilde{\beta}) = 0$$

- 若总体矩成立，则样本矩整体上应该接近于零 (Lars Peter Hansen, 1982)。

$$\min_{\tilde{\beta}} J_n(\tilde{\beta}, \mathbf{W}_n) = n\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{z}_i(y_i - \mathbf{x}_i'\tilde{\beta})\right)'\mathbf{W}_n\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{z}_i(y_i - \mathbf{x}_i'\tilde{\beta})\right)$$

where $\mathbf{W}_n$ is a symmetric and positive definite weighting matrix.

- GMM 估计量有无穷多个。

$$\mathbf{b}_{\text{GMM}} = \left(\mathbf{X}'\mathbf{Z}\mathbf{W}_n\mathbf{Z}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Z}\mathbf{W}_n\mathbf{Z}'\mathbf{y}$$

$$\begin{aligned}\text{Var}(\mathbf{b}_{\text{GMM}}) = \frac{1}{n}&\left(\text{E}\left(\mathbf{x}_i\mathbf{z}_i'\right)\mathbf{W}\text{E}\left(\mathbf{z}_i\mathbf{x}_i'\right)\right)^{-1} \\ &\cdot \text{E}\left(\mathbf{x}_i\mathbf{z}_i'\right)\mathbf{W}\text{E}(\varepsilon_i^2\mathbf{z}_i\mathbf{z}_i')\mathbf{W}\text{E}\left(\mathbf{z}_i\mathbf{x}_i'\right) \\ &\left(\text{E}\left(\mathbf{x}_i\mathbf{z}_i'\right)\mathbf{W}\text{E}\left(\mathbf{z}_i\mathbf{x}_i'\right)\right)^{-1}\end{aligned}$$

where $\mathbf{W} = \text{plim}_{n\to\infty}\mathbf{W}_n$.

- 最优 GMM 估计量是渐进方差最小的 GMM 估计量。

$$\mathbf{W}^* = \left(\mathrm{E}(\varepsilon_i^2 \mathbf{z}_i \mathbf{z}_i')\right)^{-1}$$

$$\mathbf{b}_{\mathrm{EGMM}} = \left(\mathbf{X}'\mathbf{Z}\hat{\mathbf{W}}^*\mathbf{Z}'\mathbf{X}\right)^{-1} \mathbf{X}'\mathbf{Z}\hat{\mathbf{W}}^*\mathbf{Z}'\mathbf{y}$$

$$\mathrm{Var}(\mathbf{b}_{\mathrm{EGMM}}) = \frac{1}{n}\left(\mathrm{E}\left(\mathbf{x}_i\mathbf{z}_i'\right)\left(\mathrm{E}(\varepsilon_i^2\mathbf{z}_i\mathbf{z}_i')\right)^{-1}\mathrm{E}\left(\mathbf{z}_i\mathbf{x}_i'\right)\right)^{-1}$$

where

$$\hat{\mathbf{W}}^* = \left(\frac{1}{n}\sum_{i=1}^{n} e_i^2 \mathbf{z}_i \mathbf{z}_i'\right)^{-1}$$

$e_i = y_i - \mathbf{x}_i'\hat{\beta}$, $\hat{\beta}$ is any consistent estimate.

- 2SLS 估计量相当于权重矩阵为 $\mathbf{W} = \left(\sigma^2 \mathrm{E}\left(\mathbf{z}_i\mathbf{z}_i'\right)\right)^{-1}$ 的 GMM 估计量，因此一般而言，2SLS 估计量不是最优 GMM 估计量（只有在条件同方差假定下两者才是等价的）。
- 因为最优权重矩阵的估计需要用到残差，因此可以分两步走：先使用一个非最优 GMM 估计量（比如 2SLS 估计量）得到残差，据此构造最优权重矩阵，然后进行最优 GMM 估计。

相关性检验是一种正式的统计检验

- 检验外部工具变量在第一阶段回归中的联合显著性。若只有一个内生变量，一般要求 $F_{\text{stat}} > 10$.
- 若有多于一个内生变量，使用 Kleibergen-Paap Wald rk F statistic 检验工具变量整体上的相关性（弱识别检验），或使用 Sanderson-Windmeijer F statistics 检验单个工具变量是否为弱工具变量。

而过度识别检验不足信

- 当然，更不足信的是考察工具变量在结构方程中的显著性。
- 过度识别检验 (testing overidentifying restrictions) 的想法是，目标函数 J_n 的最小值如果大于零，则意味着总体距条件不成立。
- Hansen's J test.

$$J_n \to_{\mathrm{d}} \chi^2(L-K)$$

- 本质上这是一个关于方程设定和矩条件的联合假说检验：原假设被拒绝，既有可能意味着工具变量内生，也有可能意味着工具变量应该进入结构方程，还有可能意味着关于控制变量的矩条件不成立。
- 因此工具变量的排斥性/独立性主要通过理论论证。

- 关键解释变量的内生性检验也是基于类似想法，原假说是关键解释变量外生（即关于该变量的矩条件成立）。可以比较两个 GMM 估计，一个使用了 L 的矩条件，另一个使用了 $L+1$ 个矩条件，其中新增的那个是关于内生变量的矩条件。如果两个 GMM 估计的 J_n 最小值相差很大，则拒绝原假说。
- Hayashi's C test.

$$C \triangleq J_n^{L+1} - J_n^{L} \to_{\mathrm{d}} \chi^2(1)$$

6.2 局部平均处理效应

异质处理效应与 LATE

- 考虑二元内生变量 D 和二元工具变量 Z 情形。Wald estimator:

$$\begin{aligned}\beta =& \frac{E(Y|Z=1)-E(Y|Z=0)}{E(D|Z=1)-E(D|Z=0)}\\ =& \frac{Cov(Y,Z)/Var(Z)}{Cov(D,Z)/Var(Z)}\\ =& \frac{Cov(Y,Z)}{Cov(D,Z)}\end{aligned}$$

- IV 估计量估计的是 ATE 么？

$$Y_i^1 - Y_i^0 = \underbrace{E\left(Y_i^1 - Y_i^0\right)}_{\equiv\beta} + \underbrace{Y_i^1 - Y_i^0 - E\left(Y_i^1 - Y_i^0\right)}_{\equiv U_i}$$

$$\begin{aligned} Y =& Y^0 + \left(Y^1 - Y^0\right) D \\ =& E(Y^0) + \underbrace{Y^0 - E(Y^0)}_{\equiv\varepsilon} + (\beta + U) D \\ =& E(Y^0) + \beta D + (\varepsilon + UD) \end{aligned}$$

  - IV 的识别条件为

$$E(\varepsilon + UD|Z) = 0$$

$$E(UD|Z) = E(U|D=1, Z)P(D=1|Z)$$

  - 若 $E(\varepsilon|Z) = 0, U = 0$, 即处理效应为常数, 则 IV 估计量能识别 ATE。
  - 若 $E(\varepsilon|Z) = 0$, $E(U|D, Z) = E(U|Z) = 0$, 尽管存在异质性处理效应，但人们是否选择进入处理组不依赖于个体处理效应的大小，则 IV 估计量能识别 ATE。
  - 若 $E(U|D=1, Z) \neq 0$, 则 IV 估计量不是 ATE 的一致估计。

- 那么，IV 估计量估计的是什么？

  定义 potential treatment:

  - $D^0 = 0, D^1 = 0$: never takers
  - $D^0 = 0, D^1 = 1$: compliers
  - $D^0 = 1, D^1 = 0$: defiers
  - $D^0 = 1, D^1 = 1$: always-takers

  Imbens and Angrist (1994, *ECMA*) 证明，若

  1. $P(D=1|Z=1) \neq P(D=1|Z=0)$: 存在 compliers
  2. $D_i^1 \geq D_i^0 \ \forall i$: 不存在 defiers
  3. $(Y^0, Y^1, D^0, D^1) \perp Z$

则

$$
\begin{aligned}
& E(Y|Z=1)-E(Y|Z=0) \\
=& E\{DY^1+(1-D)Y^0|Z=1\}-E\{DY^1+(1-D)Y^0|Z=0\} \\
=& E\Big\{D^1Y^1+\big(1-D^1\big)Y^0|Z=1\Big\}-E\Big\{D^0Y^1+\big(1-D^0\big)Y^0|Z=0\Big\} \\
=& E\Big\{D^1Y^1+\big(1-D^1\big)Y^0\Big\}-E\Big\{D^0Y^1+\big(1-D^0\big)Y^0\Big\} \\
=& E\Big\{\big(D^1-D^0\big)\big(Y^1-Y^0\big)\Big\} \\
=& E\big(Y^1-Y^0|D^1-D^0=1\big)P\big(D^1-D^0=1\big) \\
=& E\big(Y^1-Y^0|\text{compliers}\big)P\big(\text{compliers}\big)
\end{aligned}
$$

因为

$$
D^1-D^0=1 \Longleftrightarrow D^1=1, D^0=0 \text{ (compliers)}
$$

$$
\begin{aligned}
& E(D|Z=1) - E(D|Z=0) \\
=& P(D=1|Z=1) - P(D=1|Z=0) \\
=& P(\text{always-takers or compliers}) - P(\text{always-takers}) \\
=& P(\text{compliers})
\end{aligned}
$$

因此

$$
E\big(Y^1 - Y^0|\text{compliers}\big) = \frac{E(Y|Z=1) - E(Y|Z=0)}{E(D|Z=1) - E(D|Z=0)}
$$

LATE 的直观含义

Observable:

	$D=1$	$D=0$
$Z=1$	n_{11}	n_{10}
$Z=0$	n_{01}	n_{00}

Unobservable (假定不存在 defiers):

		$Z=0$	
		$D=0$	$D=1$
$Z=1$	$D=0$	Never-taker (n_n^o, n_n^u)	Defier
	$D=1$	Complier (n_c^t, n_c^c)	Always-taker (n_a^o, n_a^u)

	$D=1$	$D=0$
$Z=1$	Complier (n_c^t) Always-taker (n_a^u)	Never-taker (n_n^o)
$Z=0$	Always-taker (n_a^o)	Complier (n_c^c) Never-taker (n_n^u)

因为 Z 是随机的，因此 $Z=0/1$ subset 中的比例代表总体比例。

$$n_a = \frac{n_{01}}{n_{01}+n_{00}},\ n_n = \frac{n_{10}}{n_{11}+n_{10}}$$

$$n_c = 1 - n_a - n_n = \frac{n_{11}n_{00} - n_{10}n_{01}}{(n_{11}+n_{10})(n_{01}+n_{00})}$$

$$n_a^u = \frac{n_{01}}{n_{01}+n_{00}} \cdot (n_{11}+n_{10}),\ n_n^u = \frac{n_{10}}{n_{11}+n_{10}} \cdot (n_{01}+n_{00})$$

$$n_c^t = n_{11} - n_a^u = \frac{n_{11}n_{00} - n_{10}n_{01}}{n_{01}+n_{00}}$$

$$n_c^c = n_{00} - n_n^u = \frac{n_{11}n_{00} - n_{10}n_{01}}{n_{11}+n_{10}}$$

<table>
<tr><th></th><th colspan="2">$D=0$</th><th colspan="2">$D=1$</th></tr>
<tr><td>Always-taker</td><td></td><td></td><td>$Z=0$
n_a^o</td><td>$Z=1$
n_a^u</td></tr>
<tr><td>Complier</td><td colspan="2">$Z=0$
n_c^c</td><td colspan="2">$Z=1$
n_c^t</td></tr>
<tr><td>Never-taker</td><td>$Z=0$
n_n^u</td><td>$Z=1$
n_n^o</td><td></td><td></td></tr>
</table>

- LATE 是关于 compliers: $n_c^c + n_c^t$.
- ATT 是关于 always-takers 和 a subset of compliers: $n_a^o + n_a^u + n_c^t$.
- 若 n_a^o 较小，则认为不存在 always-takers, 此时 ATT 是关于 n_c^t，又因为 Z random assignment, 因此 LATE 等于 ATT.
- 类似地，如果不存在 never-takers, LATE 等于 ATU.

示例 5. 家庭暴力的柔性处置 (Angrist, 2006, *Journal of Experimental Criminology*).

Assigned and delivered treatments in the MDVE

Assigned treatment	Delivered treatment			
		Coddled		
	Arrest	Advise	Separate	Total
Arrest	98.9 (91)	0.0 (0)	1.1 (1)	29.3 (92)
Advise	17.6 (19)	77.8 (84)	4.6 (5)	34.4 (108)
Separate	22.8 (26)	4.4 (5)	72.8 (83)	36.3 (114)
Total	43.4 (136)	28.3 (89)	28.3 (89)	100.0 (314)

Notes: This table shows percentages and counts for the distribution of assigned and delivered treatments in the Minneapolis Domestic Violence Experiment (MDVE). The first three columns show row percentages. The last column reports column percentages. The number of cases appears in parentheses.

示例 6. Adams et al. (2009, *Journal of Empirical Finance*).

- Y: 公司绩效 (Tobin's Q and ROA)
- D: CEO 是否为公司创始人
- "Eligibility" IV: 创始人死亡比例；创始人数量
- Method: Probit and ILS

5.1.4. Local or global effects?

The use of instrumental variables methods implies that we can only hope to estimate local rather than global effects. In other words, our IV estimator γ^{IV} consistently estimates the average impact of founder–CEOs on the performance of those firms that are affected in their CEO choices by the value of the instruments. There is thus a question of whether our instruments define an interesting sub-population over which this effect is averaged.

Consider, for example, the dead founders variable. All firms that are run by founder–CEOs were affected in their choices by the instrument, because they can only have a founder–CEO if at least one of the founders is alive. Some but not all the firms without a founder–CEO were certainly affected by the instruments, too. Thus, the sub-population of firms affected by the instruments includes all firms in the "treatment group" (firms run by founders) plus some others in the "control group." Consequently, we expect our IV approach to estimate an effect that is somewhere in between the "treatment on the treated effect" and the "average treatment effect."

6.3 边际处理效应

离散处理选择模型

$$Y_i^0 = \mu_0(X_i) + U_i^0$$
$$Y_i^1 = \mu_1(X_i) + U_i^1$$

$$D_i^* = \mu_D(X_i, Z_i) - V_i$$
$$D_i = \begin{cases} 1 & \text{if } D_i^* \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

其中 V_i 是不可观测的 “resistance” or “distaste” for treatment.

$$D_i^* \geq 0 \Leftrightarrow \mu_D(X_i, Z_i) \geq V_i \Leftrightarrow F_V\left(\mu_D(X_i, Z_i)\right) \geq F_V(V_i)$$

定义处理概率（倾向得分）$P(X_i, Z_i) \equiv F_V\left(\mu_D(X_i, Z_i)\right)$，而 $U_{Di} \equiv F_V(V_i)$ 即为 V 分布的分位数。

$$
\begin{aligned}
Y_i &= Y_i^0 + D_i(Y_i^1 - Y_i^0) \\
&= \mu_0(X_i) + D_i\Delta_i + U_i^0
\end{aligned}
$$

$$
\Delta_i \equiv Y_i^1 - Y_i^0 = \mu_1(X_i) - \mu_0(X_i) + U_i^1 - U_i^0
$$

$$
\begin{aligned}
ATE(x) &= E(\Delta_i|X_i = x) = \mu_1(x) - \mu_0(x) \\
ATT(x) &= E(\Delta_i|X_i = x, D_i = 1) \\
&= \mu_1(x) - \mu_0(x) + E(U_i^1 - U_i^0|X_i = x, D_i = 1) \\
ATU(x) &= E(\Delta_i|X_i = x, D_i = 0) \\
&= \mu_1(x) - \mu_0(x) + E(U_i^1 - U_i^0|X_i = x, D_i = 0)
\end{aligned}
$$

如果存在正向自选择 (positive selection),

$$
E(U_i^1 - U_i^0|X_i = x, D_i = 1) > 0 > E(U_i^1 - U_i^0|X_i = x, D_i = 0)
$$

则

$$
ATT(x) > ATE(x) > ATU(x)
$$

二元工具变量

$$Wald(x) = \frac{E(Y_i|Z_i = 1, X_i = x) - E(Y_i|Z_i = 0, X_i = x)}{E(D_i|Z_i = 1, X_i = x) - E(D_i|Z_i = 0, X_i = x)}$$

$$\begin{aligned} LATE(x) &= E(Y_i^1 - Y_i^0|D_i^1 > D_i^0, X_i = x) \\ &= \mu_1(X_i) - \mu_0(X_i) + E(U_i^1 - U_i^0|D_i^1 > D_i^0, X_i = x) \end{aligned}$$

$$IV = \sum_x \omega(x) LATE(x)$$

其中

$$\omega(x) = \frac{p_x Var\left(\hat{D}_i|X_i = x\right)}{Var\left(\hat{D}_i\right)}, \ \hat{D}_i = E(D_i|X_i, Z_i)$$

连续工具变量

$$Wald(z,z',x) = \frac{E(Y_i|Z_i=z,X_i=x)-E(Y_i|Z_i=z',X_i=x)}{E(D_i|Z_i=z,X_i=x)-E(D_i|Z_i=z',X_i=x)}$$

对于任意一组 z 和 z' 都需要满足单调性条件

$$D_i^z \geq D_i^{z'}, \forall i$$

$$\begin{aligned} LATE(z,z',x) &= E(Y_i^1 - Y_i^0 | D_i^z > D_i^{z'},\ X_i = x) \\ &= E(Y_i^1 - Y_i^0 | P(z') < U_D < P(z),\ X_i = x) \end{aligned}$$

边际处理效应

$$MTE(X_i = x, U_{Di} = u_D) = E(Y_i^1 - Y_i^0 | X_i = x,\ U_{Di} = u_D)$$

$U_{Di} = P(z)$ 时的 MTE 就是当 $P(z') \to P(z)$ 时 $LATE(z, z', x)$ 的极限。

$$MTE(X_i = x,\ U_{Di} = p) = \frac{\partial E(Y_i | X_i = x,\ P(Z_i) = p)}{\partial p}$$

直观含义：给定倾向得分 $p = p_0$，$U_D < p_0$ 的个体接受处理，$U_D = p_0$ 的个体无差异。当 p 从 p_0 增加 dp，之前无差异的个体将接受处理，他们的边际处理效应为 $MTE(U_D = p_0)$，由此带来的 Y 的增加即为

$$dY = dp \times MTE(U_D = p_0)$$

因此

$$\frac{dY}{dp} = MTE(U_D = p_0)$$

为了表现处理效应的异质性，通常令 X 取均值，画出 MTE 和 U_D 的二元关系，如果 MTE 曲线向下倾斜，则意味着存在正向自选择，反之则意味着存在反向自选择。

边际处理效应的估计

$$\begin{aligned} MTE(x, u_D) &= E(Y_i^1 - Y_i^0 | X_i = x, U_{Di} = u_D) \\ &= x(\beta_1 - \beta_0) + E(U_i^1 - U_i^0 | U_{Di} = u_D) \end{aligned}$$

$$E(Y_i | X_i = z,\ P(Z) = p) = X_i \beta_0 + X_i(\beta_1 - \beta_0)p + K(p)$$

$$MTE(X_i = x,\ U_{Di} = p) = \frac{\partial E(Y_i | X_i = z,\ P(Z) = p)}{\partial p} = x(\beta_1 - \beta_0) + \frac{\partial K(p)}{\partial p}$$

- 半参数估计：

  1. X 和 Xp 分别对 $K(p)$ 回归，得到残差 e_X 和 e_{Xp}.
  2. Y 对 e_X 和 e_{Xp} 回归，得到残差 e_Y.
  3. e_Y 对 $K(p)$ 回归，MTE 曲线即为 $\hat{K}(p)$ 的一阶导。

- 参数估计：

$$Y = X\beta_0 + X(\beta_1 - \beta_0)p + \sum_{k=2}^{K} \alpha_k p^k + v$$

- `STATA` 命令 `margte` 可以实现。

$$ATE = \frac{1}{n}\sum_{i=1}^{n} X_i(\beta_1 - \beta_0) + \frac{1}{100}\sum_{u=1}^{100}(U^1 - U^0 | U_D = u/100)$$

$$ATT = \frac{1}{n}\sum_{i=1}^{n} \frac{p_i}{\bar{p}} X_i(\beta_1 - \beta_0) + \sum_{u=1}^{100} \frac{P(p > u/100)}{100\bar{p}}(U^1 - U^0 | U_D = u/100)$$

$$ATU = \frac{1}{n}\sum_{i=1}^{n} \frac{1 - p_i}{1 - \bar{p}} X_i(\beta_1 - \beta_0) + \sum_{u=1}^{100} \frac{P(p \leq u/100)}{100(1 - \bar{p})}(U^1 - U^0 | U_D = u/100)$$