

工具变量

工具变量

IV工具变量基础知识

遗漏变量

遗漏变量的后果

遗漏变量的案例

解决方法

测量误差

存在误差的后果

测量误差案例

解决方案

反向因果关系

反向因果后果

案例举例

动态面板模型

出现的问题

解决方案

样本选择性偏误（自选择问题）

出现的原因

造成的后果

解决方案

两阶段最小二乘估计

2SLS的一致性

2SLS的渐近正态性

几点观察

工具变量的选择与有效性检验

工具变量选择

工具变量的选择标准

弱工具变量问题

工具变量的选择方法

其他选择工具变量的思路

广义矩估计

GMM估计思想

恰好识别：工具变量估计

过度识别：以差分GMM估计量为例

其他GMM估计量

过度识别的工具变量之有效性检验

工具变量stata案例

工具变量法与因果推断

异质性工具变量

工具变量识别依从者平均处理效应推导

违反一些假设的情形

违反排除性假设，同时保持独立性和单调性

违反单调性假设

工具变量的实证应用

实例应用: Radio's Impact on Public Spending

研究主题及背景

投票模型

(一) 选民的效用

(二) 县救济金的分配情况

实证分析

(一) 基准回归及分样本回归

(二) 工具变量回归

(三) 其他稳健性检验

(四) 无线电广播对投票率的影响——进一步的机制分析

结论

参考文献:

当内生性问题存在时, 工具变量是解决问题最常用的办法。本文的内容重点, 是在内生性影响了因果推断时, 如何有效利用工具变量法来顺利解决。报告的前两章分别介绍了工具变量法的基础知识、和两阶段最小二乘法的重要性。第三章简述了工具变量法的选择和假设检验。第四章是报告的重点, 介绍了因果推断框架下的工具变量法。第五章是应用工具变量法的一篇实证案例。

IV工具变量基础知识

在因果推断中, 内生性会导致结果出现偏误。遗漏变量、测量误差、反向因果等情况会导致内生性的产生。

遗漏变量

遗漏变量的后果

假设真实的模型为: $Y = \beta_1 X_1 + \beta_2 X_2 + \epsilon_i$

遗漏变量后的模型为: $Y = \widetilde{\beta}_1 X_1 + \epsilon_i$

如果 $cov(X_1, X_2) \neq 0$, 则 $\widetilde{\beta}_1 \neq \beta_1$, 反正 $\widetilde{\beta}_1 = \beta_1$ 。这表明遗漏变量和模型其他变量没有相关性时, 模型估计系数不一致, 反正, 一致。

如果推广到多元回归矩阵形式, 则:

$$\widetilde{\beta}_{OLS} = (X_1^T M_2 X_1)^{-1} X_1^T M_2 Y, \text{ 其中 } M_2 = I - P_x, P_x = \widetilde{\beta}_{OLS} = X_2 (X_2^T X_2)^{-1} X_2^T$$

其中 P_x 为投影矩阵, 即遗漏变量在其他变量上的投影。

遗漏变量的案例

当研究个人能力与教育年限的关系时，个人能力变量是非常抽象的，很多文献一般都用教育年限作为个人能力的代理变量，但会遗漏IQ情商等其他变量。

解决方法

- (1) 寻找只通过影响遗漏变量进而影响因变量的自变量（多余性）；
- (2) 使用面板数据

测量误差

存在误差的后果

我们还是假设真实的模型为： $Y = \beta_1 X_1 + \beta_2 X_2 + \epsilon_i$

次数变量 X_1 存在测量误差，变为 \widetilde{X}_1 ，且 $\widetilde{X}_1 = X_1 + u_i$ ，则很明显， $cov(\widetilde{X}_1, \epsilon_i) \neq 0$

可得： $\tilde{\beta} = [(X + \epsilon)^T(X + \epsilon)]^{-1}(X + \epsilon)^T Y$ ，很明显， $\tilde{\beta} \neq \beta$ ，估计不一致。

测量误差案例

仍然是研究个人能力与工资回报关系，个人能力如果只用受教育年限作为代理变量的话，就忽视了IQ方面的测量误差。

解决方案

寻找工具变量进行估计吧！

反向因果关系

反向因果后果

还是假设模型为： $Y = \beta_1 X_1 + \beta_2 X_2 + \epsilon_i$ ，但 $X_1 = \beta_3 Y + u_i$ 。

联立上述两式，可以推导出 $cov(X_1, \epsilon_i) \neq 0$ ，说明系数估计不一致

案例举例

文化多样性、文化相似性与经济增长的关系，其中文化多样性与经济增长具有很强的反向因果关系。具体论文可以参考论文【赵当如. 方言多样性、相似性与区域经济开放[D].广西大学,2021.】

动态面板模型

出现的问题

动态面板设定如下：

$$y_{it} = \alpha y_{i,t-1} + X\beta + u_i + \epsilon_{it}$$

将上述方程滞后一期得：

$$y_{i,t-1} = \alpha y_{i,t-2} + X\beta + u_i + \epsilon_{i,t-1}$$

$$\text{由于 } cov(y_{i,t-1}, u_i + \epsilon_{it}) = cov(y_{i,t-1}, u_i) + cov(y_{i,t-1}, \epsilon_{it})$$

由于 u_i 是 $y_{i,t-1}$ 的组成部分，因此： $cov(u_i, y_{i,t-1}) \neq 0$ ，故 $cov(y_{i,t-1}, u_i + \epsilon_{it}) \neq 0$

因此，动态面板模型存在内生性问题。

解决方案

系统GMM和差分GMM进行估计，后文会提及。

样本选择性偏误（自选择问题）

出现的原因

由于可以获得数据的限制，导致一部分个体观测者无法观测，只能获得一部分特征个体数据。

造成的后果

设定模型为： $Y = \beta X + \epsilon_i$

变量 X 的观测由于一定特征的受限只能选择一部分样本，假设该阻碍样本收集的特征为 D ，且： $D_i = \begin{cases} 1 \\ 0 \end{cases}$

假设我们只能收集到 $D_i = 1$ 的数据，这时估计的模型相当于：

$$Y = \beta X + \epsilon_i \quad \text{when } D_i = 1$$

只对全样本中的子样本做回归，因此估计系数 β 不一致

解决方案

运用heckman两阶段模型估计。

两阶段最小二乘估计

考虑模型：

$$Y_i = X'_{1i}\beta_1 + X'_{2i}\beta_2 + \varepsilon_i$$

当 $E(\varepsilon_t | X_t) \neq 0$ 时，OLS估计量 $\hat{\beta}$ 不再是 β 的一致估计。那在这种情况下，应该怎样得到 β 的一致估计量呢？

最有效的多元工具变量的方法时两阶段最小二乘法（2SLS），由Theil(1953)最早提出。该方法可以用来估计 β ，即 X_t 对 Y_t 的边际效应。

在模型 (2.1) 中， $X_i = \begin{bmatrix} X_{1i} \\ X_{2i} \end{bmatrix}$ 是 $k * 1$ 维向量， $Z_i = \begin{bmatrix} Z_{1i} \\ Z_{2i} \end{bmatrix}$ 是 $l * 1$ 维向量。

这里 X_{1i} 内生， X_{2i} 外生，即 $E(X_{1t}\varepsilon_i) \neq 0$ 且 $E(X_{2t}\varepsilon_i) = 0$ 。工具变量 Z_i 需要满足总体矩条件： $E(Z'_i X_i) \neq 0$ 且 $E(Z'_i \varepsilon_i) = 0$ 。

当 $l = k$ 时，为恰好识别状态， $\hat{\beta}_{IV} = (Z'X)^{-1}Z'Y$ 。

当 $l > k$ 时，IV过度识别，需要使用两阶段最小二乘法，其基本步骤如下。

第一阶段： X_i 对 Z_i 进行 OLS 回归，得到拟合值 \hat{X}_i 。

第一步实质上是求垂直投影，

$$P_z = Z(Z'Z)^{-1}Z'$$

$$\hat{X} = P_z X = Z(Z'Z)^{-1}Z'X$$

第二阶段： Y_i 对拟合值 \hat{X}_i 进行回归，得到的 OLS 估计量称为 2SLS 估计量，记为 $\hat{\beta}_{2SLS}$ 。则

$$\begin{aligned} \hat{\beta}_{2SLS} &= (\hat{X}'\hat{X})^{-1}\hat{X}'Y \\ &= (X'P_z'P_zX)^{-1}X'P_z'Y \\ &= (X'P_zX)^{-1}X'P_zY \end{aligned} \tag{2.2}$$

把 P_z 的表达式代入，就是 $\hat{\beta}_{2SLS} = (X'Z(Z'Z)^{-1}Z'X)^{-1}X'Z(Z'Z)^{-1}Z'Y$

由于 $Y = \hat{X}\beta + u$ ，所以

$$\begin{aligned} \hat{\beta}_{2SLS} - \beta &= (\hat{X}'X)^{-1}\hat{X}'u \\ &= (X'Z(Z'Z)^{-1}Z'X)^{-1}X'Z(Z'Z)^{-1}Z'u \\ &= \left(\frac{X'Z}{n} \left(\frac{Z'Z}{n} \right)^{-1} \frac{Z'X}{n} \right)^{-1} \frac{X'Z}{n} \left(\frac{Z'Z}{n} \right)^{-1} \frac{Z'u}{n} \end{aligned} \tag{2.4}$$

2SLS的一致性

由弱大数定律，

$$\begin{aligned}
\frac{X'Z}{n} &\xrightarrow{p} E[X_i Z'_i] \\
\frac{Z'X}{n} &\xrightarrow{p} E[Z_i X'_i] \\
\frac{Z'Z}{n} &\xrightarrow{p} E[Z_i Z'_i] \\
\frac{Z'u}{n} &\xrightarrow{p} E[Z_i u_i] = 0
\end{aligned} \tag{2.5}$$

由连续性引理, $\hat{\beta}_{2SLS} - \beta$ 依概率收敛到 $(E[X_i Z'_i] E[Z_i Z'_i] E[Z_i X'_i])^{-1} E[X_i Z'_i] E[Z_i Z'_i]^{-1} * 0 = 0$

所以, 2SLS的估计量具有渐进一致性。

2SLS的渐近正态性

渐进正态性:

$$\sqrt{n}(\hat{\beta}_{2SLS} - \beta) = (\frac{X'Z}{n}(\frac{Z'Z}{n})^{-1}\frac{Z'X}{n})^{-1}\frac{X'Z}{n}(\frac{Z'Z}{n})^{-1}\frac{Z'u}{n} \tag{2.6}$$

由中心极限定理, $\frac{Z'u}{\sqrt{n}}$ 依分布收敛到 $N(0, E(u_i^2 Z_i Z'_i)) = N(0, V)$, 这里记 $V = E(u_i^2 Z_i Z'_i)$,

记 $A = (E[X_i Z'_i] E[Z_i Z'_i] E[Z_i X'_i])^{-1} E[X_i Z'_i] E[Z_i Z'_i]^{-1}$,

由 (2.5) 并根据Slutsky定理, $\sqrt{n}(\hat{\beta}_{2SLS} - \beta)$ 依分布收敛到
 $(E[X_i Z'_i] E[Z_i Z'_i] E[Z_i X'_i])^{-1} E[X_i Z'_i] E[Z_i Z'_i]^{-1} * N(0, V) \sim N(0, A V A')$ $\sim N(0, \Omega)$,

这里 $\Omega = avar(\sqrt{n}\hat{\beta}_{2SLS}) = A V A'$,

所以它是渐进正态的。

几点观察

第一, 当同方差假定成立时, $\hat{\beta}_{2SLS}$ 是最有效的。

第二, 当模型恰好识别时, $\hat{\beta}_{2SLS}$ 退化成 $\hat{\beta}_{IV}$ 。

第三, 当同方差假定不满足时, $\hat{\beta}_{2SLS}$ 不是最有效的。此时最有效 (即方差最小) 的估计量是“efficient GMM”估计量。

工具变量的选择与有效性检验

工具变量选择

工具变量的选择标准

在模型估计过程中被作为工具使用，以替代模型中与误差项相关的随机解释变量的变量，称为工具变量。作为工具变量，必须满足下述四个条件：

- (1) 与所替的随机解释变量高度相关；
- (2) 与随机误差项不相关；
- (3) 与模型中其他解释变量不相关；
- (4) 同一模型中需要引入多个工具变量时，这些工具变量之间不相关。

弱工具变量问题

由于工具变量只包含了原内生解释变量的一部分信息，因此工具变量的引入在减小估计偏误的同时，却增大了估计方差。

如果工具变量与原内生解释变量的相关性很弱，则工具变量仅包含原解释变量较少的信息。因此利用这些少量信息进行的工具变量估计就不准确。即使在样本容量很大的情况下，参数的估计值也很难收敛到真实值。

工具变量的选择方法

较为理想的工具变量：

对于内生的解释变量，可选择相邻较近的滞后变量，而无需太早的信息。因为随着工具变量个数的增加，工具变量集与扰动项相关的可能性也在增加，外生性有可能被违背。

其他选择工具变量的思路

(1) 将集聚数据作为工具变量

将集聚数据作为工具变量会克服内生性问题，但由于无法保证集聚数据具有完全外生性，所以可能会引入噪音甚至导致遗漏偏误。因此使用该类工具变量的研究呈减少趋势。

(2) 将社会空间作为工具变量

社会空间包括地理空间、市场空间等，其在特定的分析层面上具有随机独立性，但却与人类行为及其社会结果密切相关。

(3) 将实验作为工具变量

实验作为一种外来干预，虽然会对被考察的解释变量产生冲击，但是却可置身模型之外，从而为实证研究提供工具变量，其中包括政策干预、改革创新等社会实验。

广义矩估计

GMM估计思想

在球形扰动的假定下，以及IV估计量恰好识别时，2SLS是最有效的估计方法。但如果扰动项存在异方差或自相关，以及当IV估计量是过度识别时，则存在更有效的方法，即广义矩估计。

矩估计方法：从总体具有的某些特征（总体矩）出发，认为如果样本是从某总体中抽出的，则样本也应具有类似的特征（样本矩）。通过计算样本的相关特征，寻找总体参数的估计值。

GMM (Generalized Method of Moments) 估计又称广义矩估计，是基于模型实际参数满足一定矩条件而形成的一种参数估计方法。

恰好识别：工具变量估计

这里以动态面板数据模型中个体效应导致的内生性为例，说明恰好识别的工具变量估计。

在以下动态面板数据模型中：

$$y_{it} = \rho y_{i,t-1} + \varepsilon_{it}$$

$$\varepsilon_{it} = \mu_i + u_{it}$$

个体效应 μ_i 使得 ε_{it} 与 $\varepsilon_{i,t-1}$ （从而与 $y_{i,t-1}$ ）相关，从而导致了动态面板数据模型固有的内生性。

因此想到进行一阶差分以消除内生性：

$$\Delta y_{it} = \rho \Delta y_{i,t-1} + \Delta u_{it}$$

$$\Delta y_{i,t-1} = y_{i,t-1} - y_{i,t-2}$$

$$\Delta u_{it} = u_{it} - u_{i,t-1}$$

但差分产生了新的内生性： $u_{i,t-1}$ 使得 Δu_{it} 与 $\Delta y_{i,t-1}$ 相关。因此动态面板模型具有固有的内生性。

由于动态面板模型固有的内生性，LSDV估计和GLS估计的结果都是有偏且非一致的。这时，只能引入IV，再对模型进行估计。

在Anderson和 Hsiao (1981) 一文中，作者提出将 $y_{i,t-2}$ 作为工具变量，因为扰动项 Δu_{it} 只包含第t期和第t-1期的信息，因此 $y_{i,t-2}$ 对扰动项来说是前定变量，是外生的（和扰动项不相关）；并且 $y_{i,t-2}$ 是 $\Delta y_{i,t-1}$ 的一部分，因此与 $\Delta y_{i,t-1}$ 是相关的（和内生解释变量相关）。因此 $y_{i,t-2}$ 满足了工具变量需具备的两个条件。基于总体矩为0，我们有理由相信参数应该使样本矩也恰好为0。通过这一方法可恰好识别出 ρ 的估计值 $\hat{\rho}$ 。

过度识别：以差分GMM估计量为例

但为了减小方差，就必须增加工具变量的信息。因此考虑可以选择多于内生解释变量个数的工具变量，成为工具变量集（这里的 Y^{t-2} 是一个工具变量向量）：

$$Y^{t-2} = (y_{i1}, y_{i2}, \dots, y_{i,t-2})$$

$$E(Y^{t-2} \Delta u_{it}) = 0$$

基于总体矩为0，我们有理由相信参数应该使样本矩也恰好为0。但此时需要用t-2个方程解1个未知数 $\hat{\rho}$ ，因此很有可能不存在这样一个 $\hat{\rho}$ 使得这t-2个方程同时成立。

退而求其次，我们只能选择一个能最小化样本矩与0之间的距离的 $\hat{\rho}$ 。而向量取平方即其与0之间的欧氏距离，因此样本矩与0之间的距离即样本矩的加权平方和： $g(\hat{\rho})'Wg(\hat{\rho})$ 。

但考虑到这t-2个工具变量包含的信息多少有所不同，为了得到更准确的参数估计结果，应进行赋权：信息含量多（少）的变量应赋予更高（低）的权重。而信息含量的多少可由变量的方差决定，因此方差大（小）的变量信息含量就少（高），应赋予更小（大）的权重。因此加权矩阵W的最优选择是样本方差-协方差矩阵的逆矩阵 $var(g(\hat{\rho}))^{-1}$ 。

样本与0的距离最小化等价于找到一个 $\hat{\rho}$ 以使得 $g(\hat{\rho})'Wg(\hat{\rho})$ 最小化。这个 $\hat{\rho}$ 即为差分GMM估计量。

其他GMM估计量

除了以上提到的差分GMM估计量之外，还可以选择：

(1) 水平GMM (Arellano & Bover, 1995)

$$y_{it} = \rho y_{i,t-1} + \varepsilon_{it}$$

$$\varepsilon_{it} = \mu_i + u_{it}$$

$$\Delta Y_t = (\Delta y_{i,t-1}, \Delta y_{i,t-2}, \Delta y_{i,t-3}, \dots, \Delta y_{i,1})$$

(2) 系统GMM估计 (Blundell & Bond, 1998)

系统GMM的估计思想综合了差分GMM和水平GMM的矩条件。

(3) 前向正交离差GMM (Arellano & Bover, 1995)

前向正交离差GMM的估计思想是：取离差消除个体效应，与差分GMM的差分功能类似。但在取离差时所用均值不是1~T期的组内均值，而是t+1~T期的均值。 ε_{it} 的离差只包含t期之后的信息，故称前向正交离差。由于t-1期之前的信息外生于 ε_{it} 的离差，因此可使用t-1期之前的变量作为工具变量。

过度识别的工具变量之有效性检验

对于一个工具变量集，其中所选择的工具变量是否都满足外生性条件呢？就需要检验工具变量集内的工具变量是否均与扰动项正交。如若正交，样本矩 $g(\hat{\rho})$ 的期望值应该为0。（样本矩 $g(\hat{\rho})$ ：工具变量集与 $\Delta\varepsilon_{it}$ 的样本协方差）。

Hansen's J 检验：

原假设：过度识别的IV是有效的，即 $E(g(\hat{\rho})) = 0$ 。

J检验统计量：

$$J = NTg(\hat{\rho})'Wg(\hat{\rho}) \sim \chi^2_\alpha(m - k)$$

其中，m为矩条件的个数（也即工具变量个数），k为待估参数的个数。

Hansen's J 检验的原理如下：因为总体矩为0 ($E(g(\hat{\rho})) = 0$)，因此GMM的目标函数 $g(\hat{\rho})'Wg(\hat{\rho})$ 作为样本矩的加权平方和应该很接近0。因此若检验统计量 $NTg(\hat{\rho})'Wg(\hat{\rho})$ 显著地异于0，则工具变量集内至少存在一个工具变量不满足外生性假定。

工具变量stata案例

```
*注意：数据这里无法附上，如需要者可以找本组索取
*****检验是否遗漏重要解释变量*****
use "D:\stata13\lixiti\li434.dta",clear
label variable q "人均香烟消费"
label variable y "人均收入"
label variable p "香烟平均销售价格"
label variable tax "香烟平均消费税"
label variable taxs "香烟平均的特别消费税"
gen lnq=ln(q)
gen lny=ln(y)
gen lnp=ln(p)
reg lnq lny lnp //p147(4.4.13)式
predict lnqhat
gen lnq2=(lnqhat)^2
ereturn list
scalar kr=e(df_m)
scalar n=e(N)
scalar R2r=e(r2)
reg lnq lny lnp lnq2 //p147式
ereturn list
scalar ku=e(df_m)
scalar R2u=e(r2)
scalar list n R2r R2u kr ku
scalar f_test=((R2u-R2r)/(ku-kr))/((1-R2u)/(n-ku-1)) //P98(3.7.14)式
dis f_test
* F统计量的临界概率
dis Ftail((ku-kr),(n-ku-1),f_test)
* F统计量的临界值
dis invFtail((ku-kr),(n-ku-1),0.05)

*或是
qui reg lnq lny lnp lnq2
test lnq2
```

*另一方法

```
qui reg lnq lny lnp
linktest //从第二个表格中，可以看到hatsq项的p值为0.054，不拒绝了hatsq系数为零的假设。
*即说明被解释变量lnq的拟合值的平方项不具有解释能力，所以可以得出结论原模型不遗漏了重要的解释变量
*注意到hatsq项的t值为1.98, 1.98^2=3.92, 为上面的f_test值p=0.054
```

```
reg q y p //p148(4.4.14)式
predict qhat
gen qq=(qhat)^2
ereturn list
scalar kr=e(df_m)
scalar n=e(N)
scalar R2r=e(r2)
```

```

reg q y p qq //p148式
ereturn list
scalar ku=e(df_m)
scalar R2u=e(r2)
scalar list n R2r R2u kr ku
scalar f_test=((R2u-R2r)/(ku-kr))/((1-R2u)/(n-ku-1)) //P98(3.7.14)式
dis f_test
* F统计量的临界概率
dis Ftail((ku-kr),(n-ku-1),f_test)
* F统计量的临界值
dis invFtail((ku-kr),(n-ku-1),0.05)

```

*或是

```

qui reg q y p qq
test qq

```

*另一方法

```

qui reg q y p
linktest ///*注意到hatsq项的t值为2.8, 2.8^2=7.84, 为上面的f_test值[]p=0.008

```

*拒绝了hatsq系数为零的假设, 即说明被解释变量q的拟合值的平方项具有解释能力, 所以可以得出结论原模型可能遗漏了重要的解释变量。

*还可以检验被解释变量q的拟合值的平方项, 被解释变量q的拟合值的立方项, 被解释变量q的拟合值 4方项具有解释能力?

```

gen q3=(qhat)^3
gen q4=(qhat)^4
reg q y p qq q3 q4
test qq q3 q4 //检验的p值为0.05194>显著性水平为0.05, 不拒绝原假设, 即认为原模型被解释变量q的拟合值的平方项, 被解释变量q的拟合值的立方项, 被解释变量q的拟合值 4方项
*的系数同时为0, 不同时具有解释能力。

```

*或是

```

qui reg q y p
estat ovtest //在这个命令语句中, estat ovtest是进行Ramsey检验的命令语句?
                // 如果设定rhs, 则在检验过程中使用解释变量, 如果不设定rhs, 则在检验中使用被解释
变量的拟合值。
                //第二部分则是Ramsey检验的结果, 不难发现Ramsey检验的原假设是模型不存在遗漏变
量,
*检验的p值为0.05194>显著性水平为0.05, 不拒绝原假设,
*即认为原模型不存在遗漏变量: qq q3 q4。

```

*****工具变量法以及内生性检验*****

```
use "D:\stata13\lixiti\li434.dta",clear
```

```

gen lnq=ln(q)
gen lny=ln(y)
gen lnp=ln(p)

```

```
reg lnq lny lnp
ivregress 2sls lnq lny (lnp=tax) //用香烟消费税Tax为工具变量，对模型进行IV估计；（采用搜索学习工具变量法回归命令ivregress的一系列知识）
ivregress 2sls lnq lny (lnp=tax taxes) //用香烟消费税Tax和额外的特别消费税 Taxes作为2个工具变量，对模型进行2SLS估计；
```

*过度识别检验

```
quietly ivregress 2sls lnq lny (lnp=tax taxes) //用香烟消费税Tax和额外的特别消费税 Taxes作为2个工具变量，对模型进行2SLS估计；
predict u,residual //记残差变量名为u
reg u tax taxes lny //u关于tax , taxes, lny的回归，辅助回归
dis e(N)*e(r2) //计算n*R平方
dis chi2tail(1,e(N)*e(r2)) // * 卡方统计量的临界概率？
display invchi2tail(1,0.05) //自由度为1，显著性水平取0.05时的卡方临界值
```

*过度识别检验另一方法

*检验工具变量是否与干扰项相关，即工具变量是否为外生变量。目前仅限于在过度识别的情况下，进行过度识别检验。

*2SLS根据Sargan统计量进行过度识别检验，GMM使用Hansen J Test进行过度识别检验。

- * 命令均为： estat overid 检验工具变量的外生性
- * H0：所有工具变量都是外生的。
- * H1：至少有一个工具变量不是外生的，与扰动项相关。

```
quietly ivregress 2sls lnq lny (lnp=tax taxes)
estat overid //P= 0.5621,接受原假设H0：所有工具变量都是外生的，即tax , taxes工具变量都是外生
```

*豪斯曼检验在检验一个模型是否存在内生性时的具体操作下面进行介绍

```
reg lnp lny tax taxes //P141倒数第二个方程，将lnp关于lny tax taxes进行OLS法估计得
predict v,residual //记残差变量名为v
reg lnq lny lnp v //P141倒数第一式：检验结果显然表明在10%的显著性水平下，模型以p=0.087的概率拒绝原假设，说明解释变量p为内生解释变量。
```

*豪斯曼检另一种方法

```
quietly reg lnq lny lnp
estimates store ols //这两个命令在对模型进行回归之后，存储OLS的估计结果为估计的有效估计量
qui ivregress 2sls lnq lny (lnp=tax taxes) //假设怀疑p为内生解释变量，找到的工具变量tax , taxes进行2sls回归估计
estimates store iv //此命令存储2SLS估计的结果为估计的一致估计量。
hausman iv ols,constant sigmamore //此命令是根据以上的存储结果进行豪斯曼检验，然后根据得到的结果图进行判断，
*从结果图可以看到豪斯曼检验的原假设是所有解释变量都是外生的，Stata检验结果显然表明在10%的显著性水平下。
*模型以p=0.0866的概率拒绝原假设，说明解释变量p为内生解释变量
```

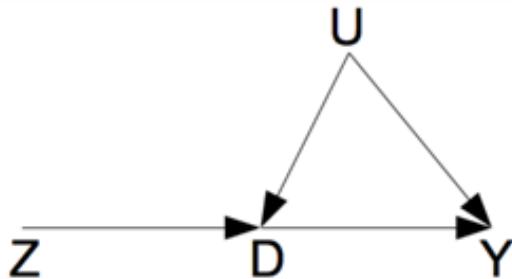
工具变量法与因果推断

异质性工具变量

传统工具变量法假设干预对所有个体的影响是相同的，而现实中干预的影响往往是异质的，同样的政策干预对有些个体影响大，对有些个体影响小，同样是接受大学教育，有些人因此收入有很大提高，有些人收入并没有因教育面变化多少，因而，同质性假设往往不符合现实。如果干预对个体的影响具有异质性，那么工具变量只是识别一部分群体的平均因果效应，不能将工具变量估计量解释为总体的平均因果效应。《Identification of Causal Effects Using Instrumental Variables》这篇文章将工具变量引入了因果推断的框架，并且是在异质性工具变量情形下讨论的，文章还提出了依从者平均处理效应的概念。

例子：美国评价越南战争服兵役对个人健康的影响。提出问题：是否老兵由于长期服役影响了健康？

研究思路：在现实中，一个人是否参军往往受多种因素影响，一般很难穷尽所有影响参军和健康结局的共同因素。为了将因果路径长期服役→健康造成 ($D \rightarrow Y$) 的相关性分离出来，这里对当兵做一个随机化，通过造一个二元工具变量出来，获得抽签号之后，如果完全遵守征兵制度，小于一定阈值（令 $Z = 1$ ）的人就会服兵役（令 $D = 1$ ）；大于阈值（令 $Z = 0$ ）的则可免于兵役（令 $D = 0$ ）。抽签号的产生如同彩票一样是完全随机的，但又直接影响到是否服兵役这一重要的社会变量。利用各人获得的抽签号作为工具变量。这样，去当兵的人就是随机化的了。在这里抽签就是一个工具变量。（这里就相当于控制了混杂因素，因为进入处理组的人数和未进入处理组的人数是随机的）



为什么认为 Z 就是一个较好的工具变量呢？：

首先，一个人抽中的数字的大小不会对健康有直接影响，只会通过影响服役 (D) 来影响健康，这满足排除性假设；其次，抽中数字的大小与潜在结果是无关的（因为抽签是随机分配的），这满足独立性假设；最后抽中数字小的人——尽管存在不依从的现象——参加兵役的可能性会更高，即相关性的假设成立。但遗憾的是，即使上述条件均成立，工具变量依然不足以识别我们通常感兴趣的平均处理效应。因为人群中存在四种人，列出如下：

- a) $D_i(0) = 0, D_i(1) = 0$ 。即无论抽签大小，均不服兵役。----从不参与者(never takers)
- b) $D_i(0) = 1, D_i(1) = 1$ 。即无论抽签大小，均服兵役。----总是参与者(always takers)
- c) $D_i(0) = 0, D_i(1) = 1$ 。即如果抽签抽中，（即抽签的数字小， $Z = 1$ ）就去服兵役，反之不服兵役。----依从者(compliers).
- d) $D_i(0) = 1, D_i(1) = 0$ 。即如果抽签没抽中，（即抽签的数字小， $Z = 1$ ）就去服兵役，反之不服兵役。----叛逆者(defiers)

因此，工具变量估计量估计的是依从者的平均处理效应。这意味着工具变量只能估计出受工具变量影响的这一部分群体的平均处理效应，那些对该工具变量没有反应的个体的平均处理效应是没有办法利用工具变量法估计出来的。由于工具变量只能估计出依从者的平均处理效应，而依从者的处理效应并不一定能推广到其他个体，从而不一定具有外部有效性。利用工具变量法进行因果效应估计时，必须要清楚这一点，不能将工具变量估计量直接推广到

其他群体。

实际上，传统工具变量法假设干预对所有个体的影响是相同的，而现实中干预的影响往往是异质的，抽签抽到同样的号码，不同的个体所做的决定也不一样，因而，同质性假设往往不符合现实。如果干预对个体的影响具有异质性，工具变量只是识别一部分个体的平均处理效应，不能将工具变量估计量解释为总体的平均处理效应或简单地解释为干预组的平均处理效应。不同的工具变量往往估计的是不同群体的平均处理效应，因而，不同的工具识别不同的因果效应参数。

可以将工具变量 Z_i 对结果变量 Y_i 的影响看作是一个因果链，工具变量 Z_i 直接影响原因变量 D_i ，原因变量 D_i 最终影响到结果变量 Y 。对于每一个人*i*来说，让 Y 成为观察到的健康结果， D_i 表示否在军队服役， Z_i 表示是否抽中服兵役。在这个模型中， β_1 代表 D 对 Y 的因果效应。涉及 D_i^* 的潜在指数表述源于这样一个概念，即遵从是一种通过比较服役和不服役的预期效用而决定的选择。因为 $Z_i = 1$ 不一定带来 $D_i = 1$ ，即抽签抽中了不一定实际服兵役，因为他们可能逃避兵役，或者可能因为医疗或家庭原因而推迟服兵役。没有摇到号的也可能自愿应征。

接下来对符号表示进行说明， $Z_i = 1$ 表示抽签号码低（可能会被召去参军），而 $Z_i = 0$ 表示抽签号码高（不会被召去参军）。让 $D_i(Z)$ 表示在随机分配的征兵抽签 Z 的情况下，*i*是否会服役。在理想的情况下， $D_i(Z)$ 将等于所有的 Z_i 。但实际上，由于各种原因， $D_i(Z)$ 和 Z_i 可能不同，因为个人可能自愿服兵役，可能逃避兵役，或者可能因为医疗或家庭原因而推迟服兵役。 $Y_i(Z, D)$ 表示个体*i*在干预状态 $D_i = d$ 和工具变量 $Z_i = z$ 下的潜在结果，这里原因变量 D_i 和工具变量 Z_i 均为二元变量，我们将 $D_i(Z)$ 称为原因变量， $Y_i(Z, D)$ 称为“潜在结果”。

$$Y_i = \beta_0 + \beta_1 \cdot D_i + \varepsilon_i$$

$$D_i^* = \alpha_0 + \alpha_1 \cdot Z_i + v_i$$

其中 $D_i = \begin{cases} 1 & \text{if } D_i^* > 0 \\ 0 & \text{if } D_i^* \leq 0 \end{cases}$, $E[Z_i \cdot \varepsilon_i] = 0$, $E[Z_i \cdot v_i] = 0$, $\text{cov}(D_i, Z_i) \neq 0$ 那么依从者平均处理效应可以写成：

$$\begin{aligned} \hat{\beta}_1^{\text{IV}} &= \widehat{\text{cov}}(Y_i, Z_i) / \widehat{\text{cov}}(D_i, Z_i) \\ &= \frac{\sum_{i=1}^N Y_i Z_i / \sum_{i=1}^N Z_i - \sum_{i=1}^N Y_i (1 - Z_i) / \sum_{i=1}^N (1 - Z_i)}{\sum_{i=1}^N D_i Z_i / \sum_{i=1}^N Z_i - \sum_{i=1}^N D_i (1 - Z_i) / \sum_{i=1}^N (1 - Z_i)} \end{aligned} \quad (1)$$

工具变量识别依从者平均处理效应推导

接下来给出工具变量识别依从者平均处理效应的过程。定义 Z 对 D 的处理效应为： $D_i(1) - D_i(0)$,

Z 对 Y 的处理效应： $Y_i(1, D_i(1)) - Y_i(0, D_i(0))$.

当个体*i*已经知道自己是否获得了参军资格，那么他是否服役对其健康的影响就是： $Y_i(Z, 1) - Y_i(Z, 0)$.

给定个体*i*的服役状态，他是否获得参军资格对其健康的影响就是： $Y_i(1, D) - Y_i(0, D)$.

为了识别依从者平均处理效应，接下来给出5个假设

假设1：独立性假设：

- a. 若 $Z_i = Z'_i$, 那么 $D_i(Z) = D_i(Z')$;
- b. 若 $Z_i = Z'_i$ 且 $D_i = D'_i$, 那么 $Y_i(Z, D) = Y_i(Z', D')$.

独立性假设可以引申为： Z_i 独立于 $\{Y_i(0, D_i(0)), Y_i(1, D_i(1)), D_i(0), D_i(1)\}$ 。每个*i*的潜在结果与其他个体的处理状态是独立的，这一假设允许将 $Y_i(Z, D)$ 和 $D_i(Z)$ 分别写成 $Y_i(Z_i, D_i)$ 和 $D_i(Z_i)$ 。

假设2：随机分配，即抽签 Z_i 是随机的： $\Pr(Z = c) = \Pr(Z = c')$.

从独立性假设和随机分配假设可以推导出 Z 对 Y 的平均处理效应: $E[Y_i | Z_i = 1] - E[Y_i | Z_i = 0]$,

以及 Z 对 D 的平均处理效应: $E[D_i | Z_i = 1] - E[D_i | Z_i = 0]$.

因此, 可以推导出 D 对 Y 的平均处理效应, 首先, D 对 Y 的平均处理效应可以写成 $\frac{\text{Cov}(Y_i, Z_i)}{\text{Cov}(D_i, Z_i)}$,

其中

$$\begin{aligned}\text{Cov}(Y_i, Z_i) &= E[Y_i Z_i] - E[Y_i]E[Z_i] \\ &= E[Z_i E[Y_i | Z_i]] - pE[E[Y_i | Z_i]] \\ &= pE[Y_i | Z_i = 1] - p(pE[Y_i | Z_i = 1] \\ &\quad + (1-p)E[Y_i | Z_i = 0])) \\ &= p(1-p)\{E[Y_i | Z_i = 1] - E[Y_i | Z_i = 0]\}\end{aligned}\tag{2}$$

同样的, $\text{Cov}(D_i, Z_i) = p(1-p)\{E[D_i | Z_i = 1] - E[D_i | Z_i = 0]\}$.

因此, D 对 Y 的平均处理效应可以表示为:

$$\frac{E[Y_i | Z_i = 1] - E[Y_i | Z_i = 0]}{E[D_i | Z_i = 1] - E[D_i | Z_i = 0]}\tag{3}$$

假设3: 排除性假设: $Y(Z, D) = Y(Z', D)$ for all Z, Z' and for all D . 这意味着: $Y_i(0, d) = Y_i(1, d), d = 0, 1$. 说明一旦原因变量 D_i 取值给定, 工具变量 Z_i 不会改变潜在结果。根据假设3, 我们现在可以将潜在结果 $Y(Z, D)$ 定义为 D 的函数:

$$Y(D) = Y(Z, D) = Y(Z', D) \quad \forall Z, Z' \text{ and } \forall D$$

然后根据假设1, 我们可以进一步写成 $Y_i(D_i)$ 。

假设4: Z 对 D 的平均处理效应 $E[D_i(1) - D_i(0)]$ 不等于零。

假设5: 单调性: $D_i(1) \geq D_i(0)$ 。

独立性假设和排除性假设足以在单位水平上建立 Z 对 Y, D 的处理效应与 D 对 Y 的因果效应之间的基本关系:

$$\begin{aligned}Y_i(1, D_i(1)) - Y_i(0, D_i(0)) \\ &= Y_i(D_i(1)) - Y_i(D_i(0)) \\ &= [Y_i(1) \cdot D_i(1) + Y_i(0) \cdot (1 - D_i(1))] \\ &\quad - [Y_i(1) \cdot D_i(0) + Y_i(0) \cdot (1 - D_i(0))] \\ &= (Y_i(1) - Y_i(0)) \cdot (D_i(1) - D_i(0))\end{aligned}\tag{4}$$

对于每个人*i*来说, Z 对 Y 的处理效应是 D 对 Y 的因果效应和 Z 对 D 的处理效应的乘积。因此, 我们可以把 Z 对 Y 的平均处理效应写成两个子群体的平均处理效应的加权和, 这两个子群体的 $D_i(0)$ 不等于 $D_i(1)$ 。

$$\begin{aligned}E[Y_i(1, D_i(1)) - Y_i(0, D_i(0))] \\ &= E[(Y_i(1) - Y_i(0))(D_i(1) - D_i(0))] \\ &= E[(Y_i(1) - Y_i(0)) | D_i(1) - D_i(0) = 1] \\ &\quad \cdot P[D_i(1) - D_i(0) = 1] \\ &\quad - E[(Y_i(1) - Y_i(0)) | D_i(1) - D_i(0) = -1] \\ &\quad \cdot P[D_i(1) - D_i(0) = -1].\end{aligned}\tag{5}$$

上述方程缺乏单调性假设。单调性假设要求 $D_i(1) - D_i(0)$ 要么等于0, 要么等于1, 因此 Z 对 Y 的平均因果效应等于 $D_i(0) = 0$ 和 $D_i(1) = 1$ 的人的平均处理效应的乘积及其在总体中所占比例:

$$\begin{aligned}E[Y_i(D_i(1), 1) - Y_i(D_i(0), 0)] \\ &= E[(Y_i(1) - Y_i(0)) | D_i(1) - D_i(0) = 1] \\ &\quad \cdot P[D_i(1) - D_i(0) = 1]\end{aligned}\tag{6}$$

由于单调性假设意味着 $E[D_i(1) - D_i(0)] = P[D_i(1) - D_i(0) = 1]$ 且 $E[D_i(1) - D_i(0)]$ 不为零。在上述5个假设下，工具变量估计为：

$$\begin{aligned} & \frac{E[Y_i(D_i(1), 1) - Y_i(D_i(0), 0)]}{E[D_i(1) - D_i(0)]} \\ &= E[(Y_i(1) - Y_i(0)) | D_i(1) - D_i(0) = 1] \end{aligned} \quad (7)$$

我们称之为局部平均处理效应 (LATE)。

违反一些假设的情形

违反排除性假设，同时保持独立性和单调性

Z 对 Y 的处理效应为： $H_i = Y_i(1, d) - Y_i(0, d)$.

其中， $d = 0$ 表示受试者总是不接受服役， $d = 1$ 表示受试者 i 总是接受服役。在排除性假设下，对于所有非依从者， $H_i = 0$ 。在违反排除性假设的环境下工具变量估计相对于不违反情形下的局部平均处理效应 (LATE) 的偏差，等于 Z 对 Y 的平均直接影响乘以成为非依从者的几率。

违反单调性假设

由于我们保留了排除性假设，对于 $D_i(1) \neq D_i(0)$ 的 i ， D 对 Y 的因果关系仍然是唯一定义的，并等于 D_i 。由于违反单调性而产生的偏差由两个因素组成：

$$\begin{aligned} & E[Y_i(1, D_i(1)) - Y_i(0, D_i(0))] / E[D_i(1) - D_i(0)] \\ & - E[Y_i(1) - Y_i(0) | i \text{ is a complier}] \\ &= -\lambda \cdot \{E[Y_i(1) - Y_i(0) | i \text{ is a defier}] \\ & - E[Y_i(1) - Y_i(0) | i \text{ is a complier}]\} \end{aligned} \quad (8)$$

其中 $\lambda = \frac{P(i \text{ is a defier})}{P(i \text{ is a complier}) - P(i \text{ is a defier})}$.

由于违反单调性而产生的偏差由两个因素组成。第一个因素，其中 λ 与叛逆者的比例有关，在单调性假设下，违约者存在的概率是0，因此 λ 也等于零。在违反单调性的假设下，我们也可以发现违约者的比例越小，违反单调性假设的偏差就越小。然而，由于这个因素的分母是 Z 对 D 的平均处理效应，只要 Z 对 D 的平均处理效应很小，即使违约者很少，偏差也会很大。第二个因素是 D 对 Y 的处理效应的变化越小，违反单调性假设的偏差就越小。

工具变量的实证应用

20世纪70年代越战期间，为了保证征兵的透明性和公正性，美国采用了一种特殊的征兵机制，为每一天赋予一个随机的数字，一年365天共1-365个数字随机地分配到每一天，从而使每个人的生日均对应一个随机化的数字，不妨称之为征兵随机数。征兵随机数产生之后，美国国防部会给出一个门槛，比如1970年为195，所有生日对应的随机数字低于195的获得参军资格，而生日对应随机数字高于195的不需要参军，当然获得参军资格的人也可能由于身体或升学原因而免除参军义务，没有参军资格的人也有部分志愿参军，但数量非常少，可以忽略。

由于征兵随机数的产生机制使参军资格完全独立于个体特征，但参军资格却对是否参军具有重要影响，因而，参军资格成为一个天然的工具。用 Z_i 来表示参军资格， $Z_i = 1$ 表示获得参军资格(即个人生日对应的随机数低于门槛)， $Z_i = 0$ 表示没有参军资格。用 D_i 表示最终是否参军， $D_i = 1$ 表示个体*i*在越战期间参军了， $D_i = 0$ 表示没有参军。 Z_i 和 D_i 之间密切相关，由于征兵随机数生成的随机性，可以认为 Z_i 独立于个体的潜在健康情况，即是否获得参军资格并不直接影响个人健康，因而 Z_i 是一个合适的工具变量。本例中的潜在结果 $Y_i(Z, D)$ 是一个指标变量，如果在抽签分配 Z 和兵役指标 D 的情况下，如果*i*会在1974年和1983年之间死亡，则 Y_i 等于1。

对于IV估计的有效因果解释，我们需要满足5个假设：

假设1（独立性假设）：任何有被抽签选中风险的人的退伍军人身份不受其他有被选中风险的人的退伍军人身份的影响，同样，任何这样的人的平民死亡率不受其他人的退伍军人身份的影响。

假设2（随机分配）：抽签的分配是随机的；

假设3（排除性假设）：一旦考虑到退伍军人身份（给定 D_i ），死亡风险（ Y_i ）不受征兵身份（ Z_i ）的影响；

假设4（Z对D的非零平均处理效应）：抽中低号码数量增多（ $Z_i = 1$ ）可以增加平均服役概率（ $D_i = 1$ ）；

假设5（单调性假设）：人群中没有这种看似非常奇怪的人：他们如果没被抽中就坚持要去服兵役，但是被抽中的就不去服兵役。

下表列出了1950年和1951年出生的白人男子按出生年份和征兵情况分列的数据和对服兵役对平民死亡率影响的一些估计。第2列是服兵役资格，第3列显示1974--1983年间宾夕法尼亚州和加利福尼亚州的死亡人数。第5列和第6列分别显示了每千人中平民死亡和自杀的平均人数，计算方法是死亡人数除以使用1970年人口普查估计的风险人口。第7列显示退伍军人身份的频率，根据1984年的收入和项目参与调查（SIPP）估计。在第5-7列中，第三行的条目给出了抽签号码低和高的人（有无征兵资格）之间的死亡、自杀和退伍军人身份的概率差异。第5列和第6列中的第四行给出了这些差异与按征兵资格划分的退伍军人概率差异的比率。这些是标准的IV估计。

表1 1950年和1951年出生的白人男子的死亡率数据

Year	Draft eligibility ^a	Number of deaths ^b	Number of suicides ^c	Probability of death ^d	Probability of suicide	Probability of military service ^e
1950	Yes	2,601	436	.0204 (.0004)	.0034 (.0002)	.3527 (.0325)
	No	2,169	352	.0195 (.0004)	.0032 (.0002)	.1934 (.0233)
<i>Difference (Yes minus No)</i>				.0009 (.0006)	.0002 (.0002)	.1593 (.0401)
<i>IV estimates^f</i>				.0056 (.0040)	.0013 (.0013)	
1951	Yes	1,494	279	.0170 (.0004)	.0032 (.0002)	.2831 (.0390)
	No	2,823	480	.0168 (.0003)	.0029 (.0001)	.1468 (.0180)
<i>Difference (Yes minus No)</i>				.0002 (.0005)	.0003 (.0002)	.1362 (.0429)
<i>IV estimates</i>				.0015 (.0037)	.0022 (.0016)	

输出结果表明，从对于1950年出生的男性，在抽中服兵役($Z_i = 1$)的男性中，有35.3%的人曾在军队服役。没有抽中服兵役($Z_i = 0$)的男性中，只有19.3%的人服兵役。抽签服兵役的随机分配表明，抽签服兵役具有因果关系，使服役的概率平均增加了15.9%。同样，在抽中服兵役($Z_i = 1$)的男性中，2.04%的人在1974年至1983年间去世，而没有抽中服兵役($Z_i = 0$)的男性中，这一比例为1.95%，0.09%的差异可以解释为对抽签服兵役这一工具变量对死亡率的平均因果效应的估计（ Z 对 Y ）。兵役对15.9%被征召入伍的平民死亡率的平均因果效应是0.56%（ D 对 Y ）。

实例应用：Radio's Impact on Public Spending

研究主题及背景

政治的传播不可避免地要通过媒体驱动：选民在进行投票前需要评估候选人是否会为他们做出贡献，而大众媒体则可以将这些信息传递给选民，帮助选民进行决策。此外，大众媒体的成本结构以及收入结构会导致其受众的分布有所不同，即知情选民和不知情选民是由特定的地域分布的，而知情的选民更倾向于参与投票。

本文研究的主题是无线电广播对公共支出的影响作用。这篇文章的故事背景发生在20世纪30年代，在这一时期，针对大萧条带来的经济危机与社会矛盾，罗斯福推行了新政，其中有一个项目就是发放失业救济金，而这个项目恰好开始于无线电广播发展的扩张期。本文分别基于美国2500个县的横截面数据以及1920-1940年美国县级面板数据这两组数据研究无线电广播对政府救济金支出的影响，从两个方面具体分析其影响机制，其一是无线电广播的普及直接影响救济金支出，其二是无线电广播的普及通过影响投票率进而影响政府的救济金支出。

投票模型

文章首先通过一个理论模型（即投票模型）阐述无线电广播对财政支出的作用机制。模型是在州长选举过程中展开的，一共涉及三个主体：现任州长、竞选者（现任州长的竞争者）、选民。每个选民都可以自主选择是否参与投票，参与投票的选民依据自身的效用，将选票投给现任州长或者竞选者中的一方，当一方获得的投票数超过总投票数的 $1/2$ 时，该方获选。以下是一些符号说明，其中下标*i,c,s*分别代表选民、县、州。

表1 符号说明

符号	含义
n_c	c 县的人口数
z_c	c 县的人均救济金
u_c	c 县的选民从救济金中所获得的效用
ξ_i	选民 <i>i</i> 是否得知现任州长将拨款 z_c 实施救济，二元变量， $\xi_i = 1$ 表示选民了解州长的政策， $\xi_i = 0$ 表示选民不了解现任州长的救济政策， c 县所有选民独立同分布
σ_c	选民 <i>i</i> 知道州长将实施救济项目的概率，即 $\sigma_c = \sigma_c(r_c) = P(\xi_i = 1) = \bar{\sigma}_c(r_c)^{c_{\sigma r}}$ ， σ_c 与无线电广播的普及率 r_c 有关，文中假设函数形式为 $\sigma_c = \bar{\sigma}_c(r_c)^{c_{\sigma r}}$ ，其中 $\bar{\sigma}_c$ 和 $c_{\sigma r}$ 为正常数
τ_i	二元变量，表示选民 <i>i</i> 是否参与投票， $\tau_i = 1$ 表示参与投票， $\tau_i = 0$ 表示不参与投票， c 县所有选民独立同分布
t_c	选民参与投票的概率，即 $t_c = P(\tau_i = 1)$
v_i	二元变量，表示选民 <i>i</i> 是否投票给现任州长， $v_i = 1$ 表示投给现任州长， $v_i = 0$ 表示投给竞争者
$-\eta$	现任州长的大众声望
$-\beta_i$	选民 <i>i</i> 在个人意识形态上对州长的偏爱（ β_i ：在个选民 <i>i</i> 人意识形态上对竞选者的偏爱）
I_s	s 州的救济金预算

（一）选民的效用

州长可以为选民带来效用，包括州长实施的政策为其带来的效用以及州长本人为其带来的效用（州长的大众声望）。

选民从救济金中获得的效用为： $u_c(z_c) = k + \frac{a_c}{1-1/\alpha}(z_c)^{1-1/\alpha}$ ，其中 $a_c > 0$ ，反映了选民对救济金的依赖程度（需求程度），其取值越大表示选民越依赖救济金。此外， $0 < \alpha < 1$ 。

于是选民的总效用表示为： $u_c(z_c) - \beta_i - \eta$ 。

那么，选民如何在现任州长和竞争者之间进行抉择呢？选民根据其效用进行投票，在门槛值 \bar{u}_i 下，若选民的效用不低于门槛效用 \bar{u}_i ，则选民将投票给现任州长，反之，则投票给竞选者。选民的投票准则用数学语言表示为：

- (1) 若 $\xi_i u_c(z_c) - \beta_i - \eta \geq \bar{u}_i$ ，则投票给现任州长；
- (2) 若 $\xi_i u_c(z_c) - \beta_i - \eta < \bar{u}_i$ ，则投票给竞争对手。

此外，本文假设 $\beta_i + \bar{u}_i$ 服从均值为 ϕ_c ，密度函数为 f_c 的均匀分布。

注：在投票过程中。存在这样一群选民，他们关注自身的效用最大化，但是意识形态薄弱（即政治立场不坚定），这类选民被称为摇摆选民或者边际选民。参与投票的选民不是支持现任州长，就是支持现任州长的竞争对手。由于选民为现任州长投票的前提条件为 $\xi_i u_c(z_c) - \beta_i - \eta \geq \bar{u}_i$ ，即 $\xi_i u_c(z_c) - \eta \geq \beta_i + \bar{u}_i$ 。由于前面我们已经解释了 β_i 的含义，即反映选民在个人意识形态上对挑战者的偏爱，又由于 $\beta_i + \bar{u}_i \sim f_c$ ，因此我们可以将 f_c 理解为边际选民密度。

那么很容易可以计算出选民*i*将选票投给现任州长的概率（同时也代表投票给州长的选民的比例），即：

$$P(v_i = 1) = P(\xi_i u_c(z_c) - \beta_i - \eta \geq \bar{u}_i) = 1/2 + f_c(\sigma_c u_c(z_c) - \eta - \phi_c).$$

(二) 县救济金的分配情况

前面分析了选民的效用以及选民根据自身效用进行的投票决策，接下来我们考虑州长的选择。当州长获得的投票数超过总投票数的1/2时，州长胜选，即州长胜选的条件用数学式可表示为：

$$\sum_i \tau_i v_i > 1/2 \sum_i \tau_i. \quad (9)$$

因此，州长胜选的概率为：

$$\begin{aligned} P\left(\sum_i \tau_i v_i > 1/2 \sum_i \tau_i\right) &= P\left(\sum_c t_c n_c P(v_i = 1) > 1/2 \sum_c t_c n_c\right) \\ &= P\left(\sum_c t_c n_c f_c (\sigma_c u_c(z_c) - \eta - \phi_c) > 0\right) \\ &= H\left(\frac{1}{\sum_c t_c n_c f_c} \sum_c t_c n_c f_c (\sigma_c u_c(z_c) - \phi_c)\right). \end{aligned} \quad (10)$$

其中， H 为 η 的分布函数。

州长的目标是在救济金的预算约束下最大化自身的胜选概率，即：

$$\begin{cases} \max & H\left(\frac{1}{\sum_c t_c n_c f_c} \sum_c t_c n_c f_c (\sigma_c u_c(z_c) - \phi_c)\right) \\ s.t. & \sum_c n_c z_c = I_s. \end{cases} \quad (11)$$

则FOC为： $t_c n_c f_c u_c'(z_c) = \lambda_s$.

代入效用函数的具体形式，并取对数，有：

$$\ln z_c = \alpha c_{\sigma r} \ln r_c + \alpha \ln \bar{\sigma}_c + \alpha \ln t_c + \alpha \ln f_c + \alpha \ln a_c - \alpha \ln \lambda_s. \quad (1)$$

实证分析

我们注意到在理论模型(1)中 $\bar{\sigma}_c$ 和 a_c 这两个参数是比较抽象的（前者反映了选民是否了解政策，后者则反映了选民对救济金的依赖或者需求程度），因此在实证分析中，我们要通过一些具体的变量对这两个参数进行估计。假设：

$$\ln \bar{\sigma}_c = c_\sigma' x_{\sigma c} + \varepsilon_\sigma, \quad (2)$$

$$\ln a_c = c_a' x_{ac} + \varepsilon_a, \quad (3)$$

则有：

$$\ln z_c = c_1 \ln r_c + c_2 \ln t_c + \beta_1' x_{c1} + \mu_s + \varepsilon_{c1}, \quad (4)$$

其中，

$$c_1 = \alpha c_{\sigma r}, \quad c_2 = \alpha, \quad \beta_1 = \begin{bmatrix} \alpha \\ \alpha c_\sigma \\ \alpha c_a \end{bmatrix}, \quad x_{c1} = \begin{bmatrix} \ln f_c \\ x_{\sigma c} \\ x_{ac} \end{bmatrix}. \quad (12)$$

除了拥有无线电广播设备的家庭比例 (r_c)外，本文所选择的影响选民是否了解政策的变量还有文盲率和入学率；文章选择的影响选民对救济金的需求程度的变量有：失业率、人均银行存款、人均农作物价值、女性比例、黑人比例、移民比例、平均租金等；此外，文章还引入了一些控制变量（具体见表2）。值得指出的是，本文中边际选民密度 f_c 是用 c 县所在的州1922-193年间举行州长选举时民主党得票率的标准差衡量的。在样本的选择上，凡是在1933年到1937年进行过选举的州都被选为样本。

表2 变量的选取

Theoretical variables	Empirical variables	Theoretical variables	Empirical variables
$z_c =$	FERA spending/capita		
$\sigma_c =$	+ share hhlds with radios		+ median rent
	- share illiterate		- share 21+
	+ school enrollment		+ share 65+
$t_c =$	+ voter turnout		+ share female
$f_c =$	+ marginal voter density		+ share black
$a_c =$	+ unempl. 1930		+ share immigrants
	+ unempl. 1937	controls	+ share partisans
	- bank deposits/capita		share urban
	- %Δbank deposits/capita		rural dummy
	- median dwell. value		gas sales/capita
	- farm value/capita		population
	- retail wage	instruments	pop. density
	- crop value/capita		ground conductivity
			share woodland

(一) 基准回归及分样本回归

根据模型(4)，本文关注的核心是 c_1 和 c_2 两个系数，两者分别反映了救济金支出相对于拥有无线广播设备的家庭比例（无线电设备拥有率）和选民投票率的弹性。根据表3第一列的基准回归结果，在0.01的显著水平下，无线广播设备拥有率的系数和投票率的系数均显著为正。无线电广播通过向选民传递相关的政策信息，使选民更好地监督和评价州长的行为，为了获得更多的选票，州长将更多的救济金分配给拥有收音机的人口比例更大的地区，即无线广播设备普及率越高的地区得到的救济金越多。此外，投票率越高的地区吸引政府救济金的能力也更强。值得注意的是，文盲率对政府救济金支出的影响作用在0.05的显著性水平下显著为负，这意味着文盲率越高的地区吸引救济金的能力越差，这是因为文盲率高的地区居民的信息识别水平较低，政治参与度也比较低（投票率低，低政治影响力）。第二列是忽略部分控制变量从而扩大样本量之后的回归结果，结果表明无线广播设备持有率和投票率对政府救济金支出都有显著的正向作用。

表3 基准回归及分样本回归结果

	A	B	I	II	III	IV
σ_c	+ c_1 : log share hhlds with radios		0.138** (2.6)	0.145** (3.0)	0.264** (3.8)	0.201** (3.6)
	- share illiterate		-1.111* (-2.1)	-1.216* (-2.2)	-2.133 (-1.9)	-1.577** (-2.6)
	+ school enrollment		0.856* (2.3)	0.789* (2.3)	.877 (1.6)	0.847 (1.9)
t_c	+ c_2 : log voter turnout		0.165** (2.9)	0.189** (3.3)	0.389** (2.9)	0.120* (2.4)
f_c	+ marginal voter density		0.034 (0.1)	0.137 (0.4)	0.077 (0.1)	-0.185 (-0.4)
a_c	+ unempl. 1930		7.837** (3.9)	6.493** (3.2)	8.449** (3.7)	0.088 (0.03)
	+ unempl. 1937		9.750*** (10.6)	9.153*** (10.6)	10.165*** (6.5)	8.248*** (5.6)
			⋮	⋮	⋮	⋮
R^2			0.63	0.63	0.59	0.69
Number of observations			2492	2679	1749	984
p -value, $H_0: c_1$ subsample =						
$c_{1,full\ sample - subsample}$					0.00	0.08
p -value, $H_0: c_2$ subsample =						
$c_{2,full\ sample - subsample}$					0.04	0.67

考虑到样本的异质性，文章还进行了分样本回归。首先，选举可以分为竞争性选举和非竞争性选举，竞争性表现为获胜方的优势低于30%。在进行非竞争性选举的州中，一个政党具有绝对的统治地位和呼吸权，于是这个政党可能不会为了赢得选举而高度关注救济金预算。因此我们假设选举竞争激烈的地区政治效应更大，即拥有收音机的家庭比例和选民投票率对救济支出的影响应该更大。第三列是利用竞争性选举的州组成的子样本进行回归得到的结果，无线电系数和投票率系数同样显著为正，并且F检验的p值表明两个系数在竞争州与非竞争州是显著不同的。此外，考虑到在边际意义上无线电广播对农村选民的信息改善作用大于城市选民，因为城市居民拥有更多的媒体选择，具有更好的途径获取信息，因此，我们假设农村地区广播的影响更大。表3最后一列是农村家庭子样本的回归结果，结果表明无线电广播对政府救济金支出的效应在农村地区明显更大。

(二) 工具变量回归

无线广播设备持有比例较高的地区通常具有如下特征：（1）对失业救济金的需求更低（低失业率、高工资、高存款等）；（2）政治影响力更高（高投票率、低文盲率等）。那么，与救济金需求相关的和与政治影响力相关的未观察到的因素将会导致 r_c 与扰动项相关，造成估计结果的不一致。为了检验OLS估计结果是否存在遗漏变量偏误，文章引入了两个无线广播设备持有率 r_c 的工具变量，大地电导率以及林地的占地比例。根据表4的前三列可知，IV估计的无线电系数 c_1 比OLS估计的要大，但是仅在农村子样本中显著。由于2SLS中第一阶段的F统计量均大于10，因此我们有理由认为不存在弱工具变量。使用工具变量法的前提是工具变量的有效性，为此，进行过度识别检验，考察是否所有工具变量均为外生，即与扰动项不相关，检验结果表明我们应接受过度识别的原假设。此外，使用工具变量法的前提是存在内生解释变量，为此必须进行Hausman检验，其原假设为“所有解释变量均为外生”，即不存在内生变量，结果表明所有解释变量均为外生，因此不存在内生解释变量，应使用OLS。

表4 工具变量回归

	I	II	III	IV
sample	full sample	competitive states	rural counties	full sample
instrumented variable	radio	radio	radio	turnout
c_1 : log share hhlds with radios	0.238 (1.0)	0.617 (1.9)	0.717* (2.3)	0.143* (2.7)
c_2 : log voter turnout	0.162** (2.9)	0.331** (2.8)	0.060 (0.8)	0.120 (1.9)
:	:	:	:	:
state effects	yes	yes	yes	yes
R^2	0.63	0.58	0.66	0.63
Number of observations	2490	1748	981	2470
F -stat, instruments, 1st stage	36	22	22	148
Overid restrictions, χ^2_{df} (p -value)	2.42 ₁ (0.12)	1.4 ₁ (.23)	0.43 ₁ (.51)	1.71 ₃ (0.64)
Hausman test for endogeneity, p -value	0.67	0.26	0.07	0.35

(三) 其他稳健性检验

无线电广播设备持有率对公共救济支出的影响源于其自身的作用还是仅仅因为它代表了一些与公众对救济金的需求（或者对新消费品的需求）相关的因素？若如此，那么汽车的拥有应该与无线电广播设备的拥有具有相似的表现。然而，汽车拥有量的相关数据缺失严重，而人均汽油销售额与汽车持有量具有很强的正相关性。此外，由于人均汽油销售额与无线电广播普及率一样，都与工资、就业等正相关，并且人均汽油销售额与拥有无线电广播设备的家庭比例高度相关，于是本文用人均汽油销售额代替 r_c 引入模型(4)。回归结果表明其对公共救济金支出并不存在显著影响。

(四) 无线电广播对投票率的影响——进一步的机制分析

从直觉上来看，无线电广播为选民提供了州长所要实施的政策对的相关信息。当了解到更多的政策信息时，选民参与投票的积极性将会提升，从而提高了投票率。拥有无线电广播设备是否提高了投票率呢？本文通过以下面板数据模型检验上述机制：

$$t_{ct} = b_1 r_{ct} + \beta_2 X_{ct2} + \mu_c + \mu_t + \varepsilon_{ct2} \quad (5)$$

结果表明无线电广播的普及率确实对投票率有正向的促进作用。

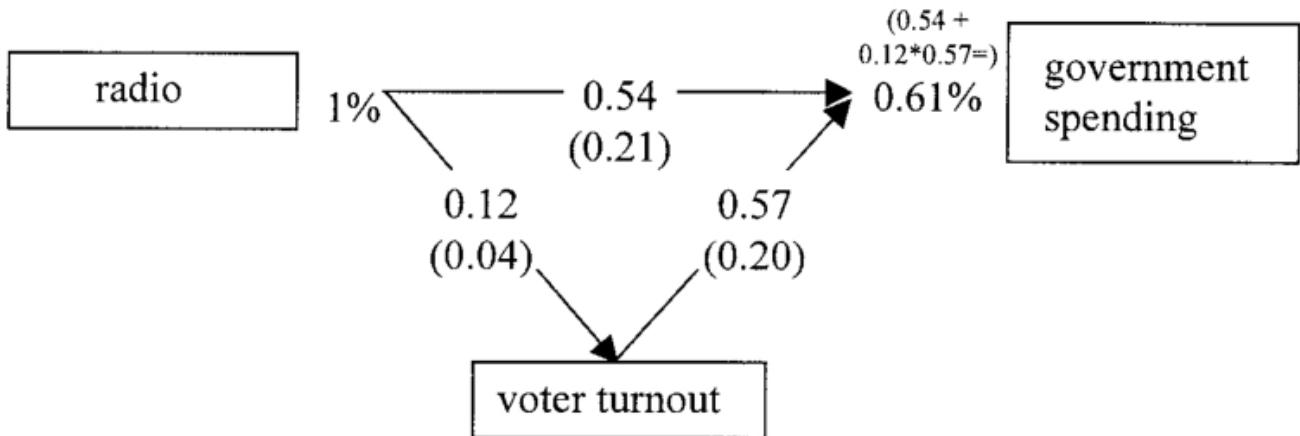


FIGURE II
Total Estimated Effect of Radio on FERA Spending

图1 作用机制

结论

- (1) 无线电广播通过向选民传递相关的政策信息，使选民更好地监督和评价政客的行为，并且更有可能参与投票（新政时期的美国，州长将更多的救济资金分配给那些拥有收音机的人口比例更大的地区）；
- (2) 无线电广播提高了美国农村吸引政府救济金的相对能力（相对于城市），因为无线电广播对政府救济金支出的效应在农村地区明显更大；
- (3) 文盲率越高的地区吸引救济金的能力越差，这同样反映了信息传导在财富再分配过程中的重要作用。

参考文献：

- Anderson, T.W. and Hsiao, C. (1981). Estimation of dynamic models with error components. *American Statistics Association*, 76, 375.
- Arellano, M. and Bover, O. (1995) Another Look at the Instrumental Variable Estimation of Error-Components Models. *Journal of Econometrics*, 68, 29-51.
- Blundell R. and Bond, S. (1998) Initial Conditions and Moment Restrictions in Dynamic Panel Data Models. *Journal of Econometrics*, 87, 115-143.
- M.伍德里奇. 计量经济学导论:现代观点[M]. 中国人民大学出版社, 2003.
- 陈强. 高级计量经济学及Stata应用[M]. 高等教育出版社, 2010.
- Angrist, J. D., & Krueger, A. B. (2001). Instrumental variables and the search for identification: From supply and demand to natural experiments. *Journal of Economic perspectives*, 15(4), 69-85.

Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American statistical Association*, 91(434), 444-455.

Abadie, A. (2003). Semiparametric instrumental variable estimation of treatment response models. *Journal of econometrics*, 113(2), 231-263.

Deaton, A. (2010). Instruments, randomization, and learning about development. *Journal of economic literature*, 48(2), 424-55.

Imbens, G. W. (2010). Better LATE than nothing: Some comments on Deaton (2009) and Heckman and Urzua (2009). *Journal of Economic literature*, 48(2), 399-423.

Clingingsmith, D., Khwaja, A. I., & Kremer, M. (2009). Estimating the impact of the Hajj: religion and tolerance in Islam's global gathering. *The Quarterly Journal of Economics*, 124(3), 1133-1170.

Strömberg, D. (2004). Radio's impact on public spending. *The Quarterly Journal of Economics*, 119(1), 189-221.