

高级计量经济学

Lecture 10: Instrumental Variable Estimation

黄嘉平

工学博士 经济学博士
深圳大学中国经济特区研究中心 讲师

办公室	粤海校区汇文楼1510
E-mail	huangjp@szu.edu.cn
Website	https://huangjp.com

工具变量估计

内生性

Endogeneity

至今为止我们讨论过的估计方法（OLS, MM, GLS）都需要假设解释变量（或者信息集）是外生的或者前定的。

例如在 MM 估计中，我们需要从信息集 Ω_t 中选取变量 W_t ，以保证 $E[u_t | W_t] = 0$ 。

在实践中很难保证所有解释变量都和误差项不相关。如果某个解释变量和误差项相关，它就是内生变量。内生性会导致 OLS 估计量有偏且不一致。

内生性可以分为以下几种：

- Errors in variables (measurement error)
- Simultaneity（联立方程、或称双向因果）
- Omitted variables（遗漏变量）

工具变量

Instrumental Variables

考虑下面的线性回归模型

$$y = X\beta + u, \quad E[uu^\top] = \sigma^2 I$$

且 X 中至少有一个内生变量。

假设针对任意观测值 t ，我们都能找到信息集 Ω_t 使其满足 $E[u_t | \Omega_t] = 0$ ，并且能定义 $n \times k$ 矩阵 W 使其第 t 行 W_t 的要素都包含在 Ω_t 中。

这样定义的 W 中的变量被称为工具变量 (instrumental variables, or instruments)。

工具变量应该是外生的或者前定的，且包含 X 中所有外生或前定变量。

IV 估计量

Instrumental Variables Estimator

工具变量 W 满足矩条件

$$W^T(y - X\beta) = 0$$

此等式的解 $\hat{\beta}_{IV}$ 称为 IV 估计量，即

$$\hat{\beta}_{IV} = (W^T X)^{-1} W^T y$$

如果忽略模型上的假设，IV 估计量和 MM 估计量的表达式相同

从 MM 估计量的性质可知，在前定性和可识别性条件下， $\hat{\beta}_{IV}$ 满足一致性和渐进正态性。基于样本 X 的可识别性是 $W^T X$ 可逆，而渐进可识别性条件是

$$\text{plim}_{n \rightarrow \infty} \frac{1}{n} W^T X \text{ 是非奇异确定矩阵}$$

事实上， $\hat{\beta}_{IV}$ 的一致性不需要前定性条件 $E[u_t | W_t] = 0$ ，而只需要

$$\text{plim}_{n \rightarrow \infty} \frac{1}{n} W^T u = 0 \quad E[u_t | W_t] = 0 \Rightarrow \text{plim}_{n \rightarrow \infty} \frac{1}{n} W^T u = 0$$

1. 当 X 和 W 都只包含一个变量时，可识别性条件意味着

$$\text{Cov}[w_t, x_t] \neq 0$$

2. 前定性条件可以推出

$$\text{Cov}[W_t, u_t] = 0$$

这个条件被称为工具变量的渐进不相关（asymptotic uncorrelated）条件。

IV 估计量的有效性

当真实参数值是 β_0 和 σ_0^2 时,

$$\text{Var}\left[\text{plim}_{n \rightarrow \infty} \sqrt{n}(\hat{\beta}_{\text{IV}} - \beta_0)\right] = \sigma_0^2 \text{plim}_{n \rightarrow \infty} \left(\frac{1}{n} X^\top P_W X\right)^{-1}$$

因此, IV 估计量的渐进有效性取决于如何选择 W 中的变量。我们称 IV 估计量满足渐进有效性的工具变量为最优工具变量 (optimal instruments)。

理论上, 我们可以定义矩阵 \bar{X} , 使其第 t 行为 $\bar{X}_t = E[X_t | \Omega_t]$, 且满足

$$X = \bar{X} + V, \quad E[V_t | \Omega_t] = 0 \quad \text{可以将其理解为生成 } X \text{ 的 DGP}$$

由此假设可以证明 $\text{plim}_{n \rightarrow \infty} \frac{1}{n} X^\top P_W X = \text{plim}_{n \rightarrow \infty} \frac{1}{n} \bar{X}^\top P_W \bar{X}$ 。当 $W = \bar{X}$ 时, 右侧的概率极限等于 $\text{plim}_{n \rightarrow \infty} \frac{1}{n} \bar{X}^\top \bar{X}$, 在所有可选择的 W 中最有效, 因此 \bar{X} 是最优工具变量 (详见 p.318)。

令 Z 为包含 X 中外生或前定变量的子矩阵, 则 $\bar{Z} = Z$, 因此 Z 也是 \bar{X} 的子矩阵。这就解释了为什么 W 应该包含 X 中的所有外生或前定变量。

和 MM 估计量的渐进有效性类似, 我们无法观测 \bar{X} , 而只能想办法找到它的一致估计量。

IV 估计中的识别

Identification in IV Estimation

至此，我们假设了工具变量矩阵 W 是 $n \times k$ 矩阵，所以工具变量的个数等于 β 中参数的个数。

在实践中，有时我们可以从信息集中找出 ℓ 个工具变量，从而构建 $n \times \ell$ 矩阵 W 。根据矩条件 $W^\top(y - X\beta) = \mathbf{0}$ ，可知其中共包含 ℓ 个等式，因此：

- 当 $\ell > k$ 时，我们称模型为过度识别（overidentified），此时矩条件的个数大于参数的个数，满足条件的估计量往往不存在；
- 当 $\ell = k$ 时，我们称模型为恰好识别（just/exactly identified），此时矩条件的个数等于参数的个数，因此存在唯一解；
- 当 $\ell < k$ 时，我们称模型为识别不足（underidentified），此时矩条件不存在唯一解。

当 $\ell > k$ 时，最有效的 IV 估计量称为广义 IV 估计量（generalized IV estimator, or GIVE）。 $\ell = k$ 时的 IV 估计量可称为简单 IV 估计量（simple IV estimator）。

广义 IV 估计量

Generalized IV Estimator

当 $\ell > k$ 时，我们可以从 ℓ 个工具变量中选取 k 种线性结合，从而构筑 k 个矩条件。这可以通过定义 $\ell \times k$ 矩阵 J ，从而使 WJ 为 $n \times k$ 矩阵，并建立矩条件 $J^\top W^\top (y - X\beta) = \mathbf{0}$ 来完成。

在选择矩阵 J 时，应当使其满足下列条件：

1. $\text{rank}(WJ) = k$ ，这是为了保证可识别性
2. J 至少应该是渐进确定的 (asymptotic deterministic)
3. J 应当使 IV 估计量满足渐进有效性

因此，矩条件 $J^\top W^\top (y - X\beta) = \mathbf{0}$ 之解 $(J^\top W^\top X)^{-1} J^\top W^\top y$ 代表一种估计量的集合，而是其中最有效的是 GIVE。

广义 IV 估计量

Generalized IV Estimator

从简单 IV 估计量的性质可知，当 $X = \bar{X} + V$ ， $E[V_t | \Omega_t] = 0$ 时，用 WJ 替代 W 获得的渐进协方差矩阵是

$$\sigma_0^2 \text{plim}_{n \rightarrow \infty} \left(\frac{1}{n} \bar{X}^\top P_{WJ} \bar{X} \right)^{-1}$$

简单 IV 估计量中的最优工具变量是 $W = \bar{X}$ 。

根据定义， $\bar{X}_t \in \Omega_t$ ，因此 \bar{X}_t 是 W_t 中变量的确定函数，但不一定是线性函数。一般情况下不存在满足 $\bar{X} = WJ$ 的矩阵 J 。

WJ 是 W 的列空间 $\mathcal{S}(W)$ 中的点集，如果无法在 $\mathcal{S}(W)$ 中寻找最优解，作为次优解，我们可以选择 \bar{X} 在 $\mathcal{S}(W)$ 上的正交投影。即令

$$WJ = P_W \bar{X} = W(W^\top W)^{-1} W^\top \bar{X}$$

此时 $J = (W^\top W)^{-1} W^\top \bar{X}$ 。

可以证明当 $WJ = P_W \bar{X}$ 时，IV 估计量满足渐进有效性（比较对象是所有可能的 WJ ，详见 p.320）。

广义 IV 估计量

Generalized IV Estimator

和前面一样， \bar{X} 未知，因此无法直接计算 $P_W \bar{X}$ 。但是从 $J = (W^\top W)^{-1} W^\top \bar{X}$ 可得，

$$\begin{aligned} \text{plim}_{n \rightarrow \infty} J &= \text{plim}_{n \rightarrow \infty} \left(\frac{1}{n} W^\top W \right)^{-1} \left(\frac{1}{n} W^\top \bar{X} \right) \\ &= \text{plim}_{n \rightarrow \infty} \left(\frac{1}{n} W^\top W \right)^{-1} \left(\frac{1}{n} W^\top X \right) \end{aligned} \quad \text{plim}_{n \rightarrow \infty} \frac{1}{n} W^\top \bar{X} = \text{plim}_{n \rightarrow \infty} \frac{1}{n} W^\top X \quad (\text{p.318})$$

因此，我们可以用 $P_W X$ 替代 $P_W \bar{X}$ 而不改变估计量的渐进性质。

当选择 $WJ = P_W X$ 时，矩条件变为

$$X^\top P_W (y - X\beta) = 0$$

GIV 估计量为 $\hat{\beta}_{\text{GIV}} = (X^\top P_W X)^{-1} X^\top P_W y$ 。 (在 $\ell = k$ 时, $\hat{\beta}_{\text{GIV}} = \hat{\beta}_{\text{IV}}$)

GIV 估计量也可以作为最优化问题 $\min_{\beta} (y - X\beta)^\top P_W (y - X\beta)$ 的解导出。

两阶段最小二乘估计

Two Stage Least Squares Estimation

GIV 估计量可以写成

$$\hat{\beta}_{\text{GIV}} = (X^{\top} P_W X)^{-1} X^{\top} P_W y = (X^{\top} P_W^{\top} P_W X)^{-1} X^{\top} P_W^{\top} y$$

从最后一项可以看出, $\hat{\beta}_{\text{GIV}}$ 是回归模型

$$y = P_W X \beta + v$$

的 OLS 估计量。其中的解释变量 $P_W X$ 是用 W 回归每一个 x_i 所得的预测值所组成的矩阵。

因此, GIV 估计量 $\hat{\beta}_{\text{GIV}}$ 可以通过下面的两阶段最小二乘回归 (2SLS) 获得:

1. 第一阶段 (first stage) : 对 $x_i = W\beta + w$ 进行 OLS 估计, 并计算 \hat{x}_i ;
2. 第二阶段 (second stage) : 令 $\hat{X} = [\hat{x}_1, \dots, \hat{x}_k]$, 并对 $y = \hat{X}\beta + v$ 进行 OLS 估计。

在第二阶段回归中获得的 OLS 估计量就是原模型的 GIV 估计量。

在计算能力缺乏的时代, 2SLS 不失为一种计算 GIV 估计量的好方法。但是通过 2SLS 无法求出原模型中 σ^2 的一致估计量。在实际应用中不需要特意使用 2SLS, 因为现在的计量软件都可以直接进行 IV 估计, 并正确计算回归标准误差。2SLS 的优点是可以帮助我们理解 IV 估计量的一些性质。

IV 估计量的小样本性质

即使 IV (GIV) 估计量能满足一致性、渐进有效性等大样本性质，在有限样本下，它几乎永远是有偏的。

导致 IV 估计量有限样本偏差的原因包括：

- 工具变量的数量 ℓ 过多，使第一阶段回归的拟合效果非常好 (R^2 接近于 1)，导致 \hat{X} 的取值非常接近 X 。此时第二阶段回归的结果就非常接近于原模型的 OLS 估计。这种情况下 IV 估计和 OLS 估计的偏差相似。
- 第一阶段中存在解释能力很低的模型 (R^2 很小或 F 统计值不显著)，称之为存在弱工具变量 (weak instruments)。此时 IV 估计量的有限样本分布和其渐进分布可能差别很大，导致有限样本偏差。

因此，选择工具变量时应使其满足：

1. 工具变量与误差项不相关（外生性、前定性、或渐近不相关性）；
2. 工具变量与内生变量相关（第一阶段回归存在 OLS 解）。

通常我们把这两项总结为：工具变量只通过 X 对 y 产生影响。

广义矩估计法简介

Generalized Method of Moments Estimation

广义矩估计 (GMM) 和最大似然估计 (ML) 是参数估计的两大方法体系。我们至今为止学过的 OLS、GLS、IV 估计都是 GMM 估计的特例。

我们考虑线性回归模型

$$y = X\beta + u, \quad E[uu^\top] = \Omega$$

X 中至少有一个内生变量。假设存在工具变量 W ，满足 $E[u_t | W_t] = 0$ ，且 $\ell \geq k$ 。

在此基础上，我们假设 $E[u_t u_s | W_t, W_t] = \omega_{ts}$ ， ω_{ts} 是协方差矩阵 Ω 的 (t, s) 要素。

广义矩估计法简介

Generalized Method of Moments Estimation

从条件 $E[u_t | W_t] = 0$ 可推出

$$E[W_t^\top (y_t - X_t \beta)] = 0$$

此等式被称为理论矩条件 (theoretical moment condition) , 与其相对应的样本矩条件 (sample moment condition) 是我们非常熟悉的

$$J^\top W^\top (y - X\beta) = 0$$

类似于 IV 估计, 我们可以通过选择合适的 J 找到满足一致性和渐进有效性的估计量。这里可以选择

$$J = (W^\top \Omega W)^{-1} W^\top X$$

此时, 有效 GMM 估计量 (efficient GMM estimator) 是

$$\hat{\beta}_{\text{GMM}} = (X^\top W (W^\top \Omega W)^{-1} W^\top X)^{-1} X^\top W (W^\top \Omega W)^{-1} W^\top y$$

广义矩估计法简介

Generalized Method of Moments Estimation

有效 GMM 估计量 (efficient GMM estimator) 也可以通过下面的最小化问题导出:

$$\min_{\beta} (\mathbf{y} - \mathbf{X}\beta)^{\top} \mathbf{W} (\mathbf{W}^{\top} \mathbf{\Omega} \mathbf{W})^{-1} \mathbf{W}^{\top} (\mathbf{y} - \mathbf{X}\beta)$$

一阶条件是 $\mathbf{X}^{\top} \mathbf{W} (\mathbf{W}^{\top} \mathbf{\Omega} \mathbf{W})^{-1} \mathbf{W}^{\top} (\mathbf{y} - \mathbf{X}\beta) = \mathbf{0}$, 和矩条件一致。

当 $\mathbf{\Omega}$ 未知时, 我们需要对 $\mathbf{\Sigma} = \text{plim}_{n \rightarrow \infty} \frac{1}{n} \mathbf{W}^{\top} \mathbf{\Omega} \mathbf{W}$ 进行一致估计。当我们允许异方差性和自相关性时, 我们可以获得类似于 HCCME 的估计量 $\hat{\mathbf{\Sigma}}$, 称之为 heteroskedasticity and autocorrelation consistent (HAC) estimator。

MM 估计量总结

	矩条件	估计量	最小化目标函数
NLS	$X^{\top}(\beta)(y - x(\beta)) = \mathbf{0}$	矩条件的唯一解	$(y - x(\beta))^{\top}(y - x(\beta))$
GLS	$X^{\top}\Omega^{-1}(y - X\beta) = \mathbf{0}$	$(X^{\top}\Omega^{-1}X)^{-1}X^{\top}\Omega^{-1}y$	$(y - X\beta)^{\top}\Omega^{-1}(y - X\beta)$
IV (GIV)	$J^{\top}W^{\top}(y - X\beta) = \mathbf{0}$ $J = (W^{\top}W)^{-1}W^{\top}\bar{X}$	$(X^{\top}P_WX)^{-1}X^{\top}P_Wy$	$(y - X\beta)^{\top}P_W(y - X\beta)$
GMM	$J^{\top}W^{\top}(y - X\beta) = \mathbf{0}$ $J = (W^{\top}\Omega W)^{-1}W^{\top}X$	$(X^{\top}W(W^{\top}\Omega W)^{-1}W^{\top}X)^{-1}$ $\times X^{\top}W(W^{\top}\Omega W)^{-1}W^{\top}y$	$(y - X\beta)^{\top}W(W^{\top}\Omega W)^{-1}$ $\times W^{\top}(y - X\beta)$

课外阅读

- Angrist, J. D. and Kruger, A. B. (1991).
Does Compulsory School Attendance Affect Schooling and Earnings?
The Quarterly Journal of Economics, 106:4, 979-1014.
<https://www.jstor.org/stable/2937954>
- Bound, J., Jaeger, D. A., and Baker, R. M. (1995).
Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variable is Weak.
Journal of the American Statistical Association, 90:430, 443-450.
<https://www.jstor.org/stable/2291055>
- Angrist, J. D., Imbens, G. W., and Kruger, A. B. (1999).
Jackknife Instrumental Variables Estimation.
Journal of Applied Econometrics, 14:1, 57-67.
<https://www.jstor.org/stable/223249>