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

高级计量经济学

理论经济学博士课程 2023-2024

Lecture 10: Instrumental Variable Estimation

Davidson, R. & MacKinnon, J. (2009). Econometrics Theory and Methods. Oxford University Press.

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 $\boldsymbol{W}_t$，以保证 $E[u_t \mid \boldsymbol{W}_t] = 0$。

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。内生性可以导致 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

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

且 $\boldsymbol{X}$ 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 $E[u_t \mid \Omega_t] = 0$，并且能定义 $n \times k$ 矩阵 $\boldsymbol{W}$ 使其第 t 行 $\boldsymbol{W}_t$ 的要素都包含在 Ω_t 中。

这样定义的 $\boldsymbol{W}$ 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 $\boldsymbol{X}$ 中所有外生或前定变量。

IV 估计量

Instrumental Variables Estimator

工具变量 $\boldsymbol{W}$ 满足矩条件

$$\boldsymbol{W}^\top(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) = \boldsymbol{0}$$

此等式的解 $\hat{\boldsymbol{\beta}}_{\text{IV}}$ 称为 IV 估计量，即

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

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，$\hat{\boldsymbol{\beta}}_{\text{IV}}$ 满足一致性和渐进正态性。基于样本 $\boldsymbol{X}$ 的可识别性是 $\boldsymbol{W}^\top\boldsymbol{X}$ 可逆，而渐进可识别性条件是

$$\operatorname*{plim}_{n\to\infty} \frac{1}{n}\boldsymbol{W}^\top\boldsymbol{X} \text{ 是非奇异确定矩阵}$$

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

$$\operatorname*{plim}_{n\to\infty} \frac{1}{n}\boldsymbol{W}^\top\boldsymbol{u} = \boldsymbol{0}$$

$$E[u_t \mid \boldsymbol{W}_t] = 0 \underset{\not\Leftarrow}{\Rightarrow} \operatorname*{plim}_{n\to\infty} \frac{1}{n}\boldsymbol{W}^\top\boldsymbol{u} = \boldsymbol{0}$$

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

1. 当 $\boldsymbol{X}$ 和 $\boldsymbol{W}$ 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
   $\text{Cov}[\boldsymbol{W}_t, u_t] = 0$

IV 估计量的有效性

当真实参数值是 $\boldsymbol{\beta}_0$ 和 σ_0^2 时，

$$\operatorname{Var}\left[\operatorname*{plim}_{n\to\infty}\sqrt{n}(\hat{\boldsymbol{\beta}}_{\mathrm{IV}}-\boldsymbol{\beta}_0)\right]=\sigma_0^2\operatorname*{plim}_{n\to\infty}\left(\tfrac{1}{n}\boldsymbol{X}^\top\boldsymbol{P}_{\boldsymbol{W}}\boldsymbol{X}\right)^{-1}$$

因此，IV 估计量的渐进有效性取决于如何选择 $\boldsymbol{W}$ 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 $\bar{\boldsymbol{X}}$，使其第 t 行为 $\bar{\boldsymbol{X}}_t=E[\boldsymbol{X}_t\mid\Omega_t]$，且满足

$$\boldsymbol{X}=\bar{\boldsymbol{X}}+\boldsymbol{V},\qquad E[\boldsymbol{V}_t\mid\Omega_t]=0$$

可以将其理解为生成 $\boldsymbol{X}$ 的 DGP

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

令 $\boldsymbol{Z}$ 为包含 $\boldsymbol{X}$ 中外生或前定变量的子矩阵，则 $\bar{\boldsymbol{Z}}=\boldsymbol{Z}$，因此 $\boldsymbol{Z}$ 也是 $\bar{\boldsymbol{X}}$ 的子矩阵。这就解释了为什么 $\boldsymbol{W}$ 应该包含 $\boldsymbol{X}$ 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 $\bar{\boldsymbol{X}}$，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 $\boldsymbol{W}$ 是 $n \times k$ 矩阵，所以工具变量的个数等于 $\boldsymbol{\beta}$ 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 $\boldsymbol{W}$。根据矩条件 $\boldsymbol{W}^\top(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) = \mathbf{0}$，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 $\boldsymbol{J}$，从而使 $\boldsymbol{WJ}$ 为 $n \times k$ 矩阵，并建立矩条件 $\boldsymbol{J}^\top \boldsymbol{W}^\top(\boldsymbol{y} - \boldsymbol{X\beta}) = \boldsymbol{0}$ 来完成。

在选择矩阵 $\boldsymbol{J}$ 时，应当使其满足下列条件：

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

因此，矩条件 $\boldsymbol{J}^\top \boldsymbol{W}^\top(\boldsymbol{y} - \boldsymbol{X\beta}) = \boldsymbol{0}$ 之解 $(\boldsymbol{J}^\top \boldsymbol{W}^\top \boldsymbol{X})^{-1} \boldsymbol{J}^\top \boldsymbol{W}^\top \boldsymbol{y}$ 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 $\boldsymbol{X}=\bar{\boldsymbol{X}}+\boldsymbol{V}$，$E[V_t \mid \Omega_t]=0$时，用 $\boldsymbol{WJ}$ 替代 $\boldsymbol{W}$ 获得的渐进协方差矩阵是

$$\sigma_0^2 \operatorname*{plim}_{n\to\infty}\left(\frac{1}{n}\bar{\boldsymbol{X}}^\top \boldsymbol{P}_{\boldsymbol{WJ}}\bar{\boldsymbol{X}}\right)^{-1}$$

简单 IV 估计量中的最优工具变量是 $\boldsymbol{W}=\bar{\boldsymbol{X}}$。

根据定义，$\bar{\boldsymbol{X}}_t \in \Omega_t$，因此 $\bar{\boldsymbol{X}}_t$ 是 $\boldsymbol{W}_t$ 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 $\bar{\boldsymbol{X}}=\boldsymbol{WJ}$ 的矩阵 $\boldsymbol{J}$。

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

$$\boldsymbol{WJ}=\boldsymbol{P}_{\boldsymbol{W}}\bar{\boldsymbol{X}}=\boldsymbol{W}(\boldsymbol{W}^\top\boldsymbol{W})^{-1}\boldsymbol{W}^\top\bar{\boldsymbol{X}}$$

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

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

广义 IV 估计量

Generalized IV Estimator

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

$$\begin{aligned}\operatorname*{plim}_{n\to\infty} \boldsymbol{J} &= \operatorname*{plim}_{n\to\infty} (\frac{1}{n}\boldsymbol{W}^\top \boldsymbol{W})^{-1}(\frac{1}{n}\boldsymbol{W}^\top \bar{\boldsymbol{X}}) \\ &= \operatorname*{plim}_{n\to\infty} (\frac{1}{n}\boldsymbol{W}^\top \boldsymbol{W})^{-1}(\frac{1}{n}\boldsymbol{W}^\top \boldsymbol{X})\end{aligned}$$

$\operatorname*{plim}_{n\to\infty} \frac{1}{n}\boldsymbol{W}^\top \bar{\boldsymbol{X}} = \operatorname*{plim}_{n\to\infty} \frac{1}{n}\boldsymbol{W}^\top \boldsymbol{X}$ （p.318）

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

当选择 $\boldsymbol{WJ} = \boldsymbol{P_W X}$ 时，矩条件变为

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

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

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

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

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

$$\boldsymbol{y} = \boldsymbol{P}_{\boldsymbol{W}} \boldsymbol{X} \boldsymbol{\beta} + \boldsymbol{v}$$

的 OLS 估计量。其中的解释变量 $\boldsymbol{P}_{\boldsymbol{W}} \boldsymbol{X}$ 是用 $\boldsymbol{W}$ 回归每一个 $\boldsymbol{x}_i$ 所得的预测值所组成的矩阵。

因此，GIV 估计量 $\hat{\boldsymbol{\beta}}_{\text{GIV}}$ 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 $\boldsymbol{x}_i = \boldsymbol{W}\boldsymbol{\beta} + \boldsymbol{w}$ 进行 OLS 估计，并计算 $\hat{\boldsymbol{x}}_i$；
2. 第二阶段（second stage）：令 $\hat{\boldsymbol{X}} = [\hat{\boldsymbol{x}}_1, \ldots, \hat{\boldsymbol{x}}_k]$，并对 $\boldsymbol{y} = \hat{\boldsymbol{X}}\boldsymbol{\beta} + \boldsymbol{v}$ 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 $\hat{\boldsymbol{X}}$ 的取值非常接近 $\boldsymbol{X}$。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 $\boldsymbol{X}$ 对 $\boldsymbol{y}$ 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

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

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

在此基础上，我们假设 $E[u_t u_s \mid \boldsymbol{W}_t, \boldsymbol{W}_t] = \omega_{ts}$，$\omega_{ts}$ 是协方差矩阵 $\boldsymbol{\Omega}$ 的 (t, s) 要素。

广义矩估计法简介

Generalized Method of Moments Estimation

从条件 $E[u_t \mid W_t] = 0$ 可推出

$$E[\boldsymbol{W}_t^\top (y_t - \boldsymbol{X}_t \boldsymbol{\beta})] = \boldsymbol{0}$$

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

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

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

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

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

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

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：

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

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

当 $\boldsymbol{\Omega}$ 未知时，我们需要对 $\boldsymbol{\Sigma} = \operatorname*{plim}_{n\to\infty} \frac{1}{n} \boldsymbol{W}^\top \boldsymbol{\Omega} \boldsymbol{W}$ 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 $\hat{\boldsymbol{\Sigma}}$，称之为 heteroskedasticisy 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	$\boldsymbol{X}^\top(\boldsymbol{\beta})\big(\boldsymbol{y}-\boldsymbol{x}(\boldsymbol{\beta})\big)=\mathbf{0}$	矩条件的唯一解	$\big(\boldsymbol{y}-\boldsymbol{x}(\boldsymbol{\beta})\big)^\top\big(\boldsymbol{y}-\boldsymbol{x}(\boldsymbol{\beta})\big)$
GLS	$\boldsymbol{X}^\top\boldsymbol{\Omega}^{-1}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})=\mathbf{0}$	$(\boldsymbol{X}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{y}$	$(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})^\top\boldsymbol{\Omega}^{-1}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})$
IV (GIV)	$\boldsymbol{J}^\top\boldsymbol{W}^\top(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})=\mathbf{0}$ $\boldsymbol{J}=(\boldsymbol{W}^\top\boldsymbol{W})^{-1}\boldsymbol{W}^\top\bar{X}$	$(\boldsymbol{X}^\top\boldsymbol{P}_W\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{P}_W\boldsymbol{y}$	$(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})^\top\boldsymbol{P}_W(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})$
GMM	$\boldsymbol{J}^\top\boldsymbol{W}^\top(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})=\mathbf{0}$ $\boldsymbol{J}=(\boldsymbol{W}^\top\boldsymbol{\Omega}\boldsymbol{W})^{-1}\boldsymbol{W}^\top\boldsymbol{X}$	$\big(\boldsymbol{X}^\top\boldsymbol{W}(\boldsymbol{W}^\top\boldsymbol{\Omega}\boldsymbol{W})^{-1}\boldsymbol{W}^\top\boldsymbol{X}\big)^{-1}$ $\times\boldsymbol{X}^\top\boldsymbol{W}(\boldsymbol{W}^\top\boldsymbol{\Omega}\boldsymbol{W})^{-1}\boldsymbol{W}^\top\boldsymbol{y}$	$(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})^\top\boldsymbol{W}(\boldsymbol{W}^\top\boldsymbol{\Omega}\boldsymbol{W})^{-1}$ $\times\boldsymbol{W}^\top(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})$

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 Endogeneous 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