高计(考完啦)

题目整理

第一次作业

1. 对于简单线性模型：$y_i = \beta_1 + \beta_2 x_{i2} + \epsilon_i$, 估计$\beta_1, \beta_2$;考虑新的样本$(y_{n+1}, x_{n+1})$, ta的预测方差$Var(y_{n+1} - \hat{y_{n+1}}|X, x_{n+1})$的表达式，并求其最小值及取值条件
   1. 最小化的条件是$x_{n+1} = \bar{X}$, 即新样本落在被估计样本的均值时估计方差最小。
2. 证明：对于OLS模型，解释变量个数越多，残差平方和$e'e$越小, R^2越大
   1. 在(k+1)元模型视角下(k)元模型可视为一簇特例：$\{\boldsymbol{\beta} : \beta_{k+1} = 0\}$
   2. 证明：定义$\mathbf{e}_k$是k元模型下的残差，我们要证明的是$\mathbf{e}_k'\mathbf{e}_k > \mathbf{e}_{k+1}'\mathbf{e}_{k+1}$, 我们可以去证明ta的充分条件:$\mathbf{e}_{k+1} \perp (\mathbf{e}_k - \mathbf{e}_{k+1})$（直角三角形）
   3. 为了证明正交，我们要借助消灭矩阵:
      $\mathbf{e}_{k+1}' \cdot (\mathbf{e}_k - \mathbf{e}_{k+1}) = \mathbf{y'M_{k+1}'M_{k+1}y - y'M_{k+1}'M_k y}$
      $= \mathbf{y'M_{k+1}y - y'M_{k+1}y} = 0$
   4. 消灭矩阵的含义：保留一个向量与$\mathbf{X}$各个分量都正交的部分，这解释了3中$\mathbf{M_{k+1}'M_k = M_{k+1}'}$
3. 若DGP(data generating process)为:$y = \beta_1 x_1 + \beta_2 x_2 + \epsilon_i$, 而我们估计的模型为$y = \gamma_1 x_1 + u_i$, 证明:$\gamma_{OLS}$对于$\beta_1$是有偏估计
4. 若DGP(data generating process)为:$y = \beta_1 x_1 + \epsilon_i$, 而我们估计的模型为$y = \gamma_1 x_1 + \gamma_2 x_2 + u_i$, 证明:
   $\gamma_{1,OLS}$对于β_1是无偏估计，但是$Var(\gamma_{1,OLS}|X) > Var(\beta_{1,OLS}|X)$
   1. 一般化的计算(利用分块矩阵，把矩阵整体放入线性方程组中，简化方程结构)
   2. 定义$\mathbf{X} = (\mathbf{X_1} \vdots \mathbf{X_2})$ $\mathbf{b} = (\mathbf{b_1} \vdots \mathbf{b_2})'$ $\boldsymbol{\beta} = (\boldsymbol{\beta_1} \vdots \boldsymbol{\beta_2})'$
   3. OLS一阶条件方程组可以写成：

      $$(\mathbf{X_1'} \vdots \mathbf{X_2'})'(\mathbf{X_1} \vdots \mathbf{X_2})(\mathbf{b_1} \vdots \mathbf{b_2})' = (\mathbf{X_1'} \vdots \mathbf{X_2'})'\mathbf{y}$$

      即：

      $$\begin{cases} \mathbf{X_1'X_1b_1 + X_1'X_2b_2 = X_1'y} \\ \mathbf{X_2'X_1b_1 + X_2'X_2b_2 = X_2'y} \end{cases}$$

4. 第一个方程左乘$\mathbf{X_1(X_1'X_1)^{-1}}$得到:$\mathbf{X_1b_1 + P_1'X_2b_2 = P_1'y}$
5. 带入到第二个方程:对称地，可以得到$\mathbf{b_1}$
   $\mathbf{X_2'(y - X_2b_2) + X_2'X_2b_2 = X_2'y} \quad \mathbf{b_2 = (X_2'M_1X_2)^{-1}X_2'M_1y}$
6. $\mathbf{b_2 = (X_2'M_1X_2)^{-1}X_2'M_1y = (X_2'M_1'M_1X_2)^{-1}X_2'M_1y}$
   $\mathbf{= (X_{2to1}'X_{2to1})^{-1}X_{2to1})'y}$ $\mathbf{X_{2to1}}$表示
   $\mathbf{X_2}$中与$\mathbf{X_1}$各个分量都正交的部分
7. $Var(\mathbf{b_1|X_1,X_2}) - Var(\mathbf{b_1^*|X_1,X_2}) = \sigma^2((\mathbf{X_1'M_2X_1})^{-1} - (\mathbf{X_1'X_1})^{-1})$
8. 定理：已知$\mathbf{B}$，$\mathbf{A}$半正定 : $\mathbf{(B^{-1} - A^{-1})}$半正定 iff $\mathbf{A - B}$半正定
9. 因为$\mathbf{X_1'X_1 - X_1'M_2X_1 = X_1'P_2X_1 = (P_2X_1)'(P_2X_1)} \geqslant 0$ 得证

期中考试

- 在大样本模型下考虑$\sqrt{n}(s^2 - E(\epsilon^2))$的渐近分布

$$
\begin{aligned}
s^2 &= \frac{1}{n}\sum_{i=1}^{n} e_i^2 \\
&= \frac{1}{n}\sum_{i=1}^{n} \epsilon_i^2 + 2(b-\beta)'\bar{g} + (b-\beta)'\mathbf{S}_{xx}(b-\beta)
\end{aligned}
$$

$$
\sqrt{n}(s^2 - E(\epsilon^2)) = \sqrt{n}(\frac{1}{n}\sum_{i=1}^{n} \epsilon_i^2 - E(\epsilon^2)) + 2(b-\beta)'\sqrt{n}\bar{g} + \sqrt{n}(b-\beta)'\mathbf{S}_{xx}(b-
$$

$$
\xrightarrow{d} N(0, Var(\epsilon_i^2))
$$

$$
[\sqrt{n}\bar{g} \xrightarrow{d} N(0,S) \quad (b-\beta) \xrightarrow{p} 0 \quad \sqrt{n}(b-\beta) \xrightarrow{d} N(0, Avar(b))]
$$

4. In the restricted least square, the sum of squared residuals is minimized subject to the constraint implied by the null hypothesis $\mathrm{R}\beta = \mathrm{r}$. Form the Lagrangian as

$$
\mathfrak{L} = \frac{1}{2}(\mathrm{y} - \mathrm{X}\tilde{\beta})'(\mathrm{y} - \mathrm{X}\tilde{\beta}) + \lambda'(\mathrm{R}\tilde{\beta} - \mathrm{r}),
$$

where λ here is the #r-dimensional vector of Lagrange multipliers. Let $\hat{\beta}$ be the restricted least square estimator of β. It is the solution of the constrained minimization problem.

(a) (10 points) Let b be the unconstrained OLS estimator. Show that both $\hat{\beta} - b$ and λ can be expressed as the form $\mathrm{C_i(Rb - r)}$ and solve the corresponding matrices $\mathrm{C_i}$.

(b) (10 points) Let $\hat{\varepsilon} = \mathrm{y} - \mathrm{X}\hat{\beta}$, $\mathrm{e} = \mathrm{y} - \mathrm{X}\mathrm{b}$, and $\mathrm{P} = \mathrm{X(X'X)^{-1}X'}$. Show: $\hat{\varepsilon}'\hat{\varepsilon} - \mathrm{e'e} = \hat{\varepsilon}'\mathrm{P}\hat{\varepsilon}$.

-

(b) 可以移项改为$\hat{\epsilon}'\mathbf{M}\hat{\epsilon} = e'e : \mathbf{M = I - P}$，这个明显更好证明

$$
\hat{\epsilon}'\mathbf{M}\hat{\epsilon} = \mathbf{(M(y - X\hat{\beta}))'(M(y - X\hat{\beta})) = e'e}
$$

因为$\mathbf{M}$矩阵依据OLS方法得到，有着很好的性质$\mathbf{MX=0,My=e}$，从而得证

第二次作业

hayashi教材

Q4R：3.2.1/3.2.2/3.4.1

AE:2/3/5

参考

- 林文夫《计量经济学》
- 洪永淼《高级计量经济学》
- keep random
- 数理统计
- 多元统计分析学习记录目录 - 知乎 (zhihu.com)
- AE答案参考：
  第一章答案
  第三章答案
  第三章答案

有限样本OLS

有限样本假设

1. 线性模型：总体模型可以表示为:$\mathbf{y}=\mathbf{X}\beta+\epsilon$
2. 严格外生性：$E(\epsilon|\mathbf{X})=\mathbf{0}$
    1. $E(\epsilon)=\mathbf{0}$
    2. $E(\mathbf{X}\epsilon)=\mathbf{0}$
    3. $Cov(\mathbf{X,}\epsilon)=\mathbf{0}$
3. 无多重共线性：$rank(\mathbf{X})=K \quad w.p.1$（以概率1）
4. 误差球面方差：$Var(\epsilon|\mathbf{X})=\sigma^2\mathbf{I_n}$
    1. 同方差：$Var(\epsilon_\mathbf{i}|\mathbf{X})=\sigma^2$
    2. 观测值不相关：$Var(\epsilon_\mathbf{i},\epsilon_\mathbf{j}|\mathbf{X})=0$
5. 误差正态分布：$\epsilon|\mathbf{X}\sim N(\mathbf{0},\sigma^2\mathbf{I_n})$
    1. 独立随机样本

OLS代数计算

OLS的目标是：

$$\begin{aligned}\min_{\mathbf{b}} \mathbf{e}'\mathbf{e} &= (\mathbf{y}-\mathbf{Xb})'(\mathbf{y}-\mathbf{Xb}) \\ &= \mathbf{yy}' - \mathbf{2y}'\mathbf{Xb} + \mathbf{b}'\mathbf{X}'\mathbf{Xb}\end{aligned}$$

接下来对b求导可以得到：

$$\frac{\partial \mathbf{e}'\mathbf{e}}{\partial \mathbf{b}} = 2((\mathbf{X}'\mathbf{X})\mathbf{b} - \mathbf{X}'\mathbf{y})$$

另上式为0可以得到：$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{A}_{\mathbf{k}\times\mathbf{n}}\mathbf{y}$, 抽样误差
$b-\beta = \mathbf{A}\epsilon = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\epsilon)$
基于这个表达式得到的OLS渐近偏误可以解释内生性出现时偏误的方向。

有限样本OLS性质

代数、几何性质

1. 定义投影矩阵：$\mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \mathbf{XA}$:
    1. $\mathbf{P}$对称且二阶幂等
        1. 二阶幂等矩阵具有性质：$tr(\mathbf{P}) = rank(\mathbf{P})$
    2. $\mathbf{PX} = \mathbf{X}, \mathbf{Py} = \mathbf{Xb} = \hat{\mathbf{y}}$
    3. 之所以如此命名是因为$\mathbf{y}$在$\mathbf{M}$作用下被投影到了$span(\mathbf{X})$上
2. 定义消灭矩阵：$\mathbf{M} = \mathbf{I_n} - \mathbf{P}$
    1. $\mathbf{M}$对称且二阶幂等
        1. 二阶幂等矩阵具有性质：$tr(\mathbf{P}) = rank(\mathbf{P})$
    2. $\mathbf{MX} = \mathbf{0}, \mathbf{My} = \mathbf{M}\epsilon = \mathbf{e}$
3. $\mathbf{X}'\mathbf{e} = 0, \hat{\mathbf{y}}'\mathbf{e} = 0$
    1. $\mathbf{X}'\mathbf{e} = \mathbf{X}'(\mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}) = \mathbf{X}'\mathbf{My} = \mathbf{0}$

统计性质

点估计的性质

1. 无偏性：$E(\mathbf{b}|\mathbf{X}) = \beta$
2. OLS方差：$\begin{aligned}&Var(\mathbf{b}|\mathbf{X}) = \sigma^{\mathbf{2}}(\mathbf{X}'\mathbf{X})^{-1} \\ &(\mathbf{AA}' = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{XX}'(\mathbf{X}'\mathbf{X})^{-1} = (\mathbf{X}'\mathbf{X})^{-1})\end{aligned}$
3. 高马定理：OLS估计量是BLUE：最优线性无偏估计量

1. 线性无偏估计量：在形如“$\mathbf{Cy}$的对β的无偏估计量”的范围
2. 最优：最小条件协方差阵$Var((\mathbf{C}-\mathbf{A})\mathbf{y}|\mathbf{X})$半正定
3. 证明：考虑$\mathbf{D}=(\mathbf{C}-\mathbf{A})$, 则：
   1. $\mathbf{Cy}=\mathbf{DX}\beta+\mathbf{b}+\mathbf{C}\epsilon$
   2. $$\begin{aligned}&E(\mathbf{Cy}|\mathbf{X})\\ &=E((\mathbf{DX}\beta+\mathbf{b}+\mathbf{C}\epsilon)|\mathbf{X})\\ &=\beta+E((\mathbf{DX}\beta|\mathbf{X})\\ &\to E((\mathbf{DX}\beta|\mathbf{X})=0\\ &\to \mathbf{DX}=0\end{aligned}$$
   3. $$\begin{aligned}&Var(\mathbf{Cy}|\mathbf{X})\\ &=Var((\mathbf{D}+\mathbf{A})\epsilon|\mathbf{X})\\ &=\sigma^2((\mathbf{D}+\mathbf{A})(\mathbf{D}+\mathbf{A})')\\ &=\sigma^2(\mathbf{DD}'+\mathbf{AA}'+\mathbf{AD}'+\mathbf{DA}')\\ &(\mathbf{DA}'=\mathbf{DX}(\mathbf{X}'\mathbf{X})^{-1}=\mathbf{0})\\ &=\sigma^2(\mathbf{DD}'+\mathbf{AA}')\\ &=\sigma^2(\mathbf{D}'\mathbf{D}+(\mathbf{X}'\mathbf{X})^{-1})\\ &\geqslant \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\end{aligned}$$
4. $Cov(\mathbf{b},\mathbf{e}|\mathbf{X})=\mathbf{0}$
   1. $$\begin{aligned}&Cov(\mathbf{b},\mathbf{e}|\mathbf{X})\\ &=E((\mathbf{b}-\beta)\mathbf{e}'|\mathbf{X})\\ &=E(\mathbf{A}\epsilon(\mathbf{M}\epsilon)'|X)\\ &=\sigma^2\mathbf{AM}'\\ &=(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{M}'\\ &=0\end{aligned}$$
5. σ^2的无偏估计$s^2=\frac{\mathbf{e}'\mathbf{e}}{n-k}$
   1. $E(\chi^2_n)=n\sigma^2$
   2. 证明：$$\begin{aligned}&E(\mathbf{e}'\mathbf{e}|\mathbf{X})\\ &=E(\epsilon\mathbf{M}\epsilon)\\ &=\sigma^2tr(\mathbf{M})\\ &=\sigma^2rank(\mathbf{M})\\ &=\sigma^2rank(\mathbf{I_n}-\mathbf{P})\\ &=\sigma^2(n-k)\end{aligned}$$

点估计的条件分布

三大分布具体看 数理统计

1. $b|\mathbf{X}\sim N(\beta,\sigma^2(\mathbf{X}'\mathbf{X})^{-1})$
   1. $\frac{b-\beta}{\sqrt{\sigma^2(\mathbf{X}'\mathbf{X})^{-1}}}\sim N(0,1)$

2. $(n-k)\frac{s^2}{\sigma^2}|\mathbf{X} \sim \chi^2_{n-k}$ （这里有**前提条件**：$\mathbf{M}$是幂等矩阵）
   1. $$\begin{aligned}(n-k)\frac{s^2}{\sigma^2} &= \frac{\epsilon'\mathbf{M}\epsilon}{\sigma^2} \\ &= (\frac{\epsilon}{\sigma})'\mathbf{M}(\frac{\epsilon}{\sigma})\end{aligned}$$
   2. $\frac{\epsilon}{\sigma}|\mathbf{X} \sim N(0, \mathbf{I_n})$, $(\frac{\epsilon}{\sigma})'\mathbf{M}(\frac{\epsilon}{\sigma})|\mathbf{X} \sim \chi^2_{rank(\mathbf{M})}$
   3. $rank(\mathbf{M}) = n-k$

   假设检验：
3. $\frac{b-\beta}{\sqrt{\sigma^2(\mathbf{X}'\mathbf{X})^{-1}}}|\mathbf{X} \sim N(0,1)$
   1. $\frac{b-\beta}{\sqrt{s^2(\mathbf{X}'\mathbf{X})^{-1}}} \sim t_{n-k}$
      1. $\frac{b-\beta}{\sqrt{s^2(\mathbf{X}'\mathbf{X})^{-1}}} = \frac{b-\beta}{\sqrt{\sigma^2(\mathbf{X}'\mathbf{X})^{-1}}} / \frac{s}{\sigma} \sim \frac{N(0,1)}{\sqrt{\chi^2_{n-k}/n-k}}$
4. $\mathbf{Rb} \sim N(\mathbf{R}\beta, \sigma^2\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}')$
   1. $(\mathbf{Rb}-\mathbf{R}\beta)'(\sigma^2\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}')^{-1}(\mathbf{Rb}-\mathbf{R}\beta) \sim \chi^2_{rank(\mathbf{R})}$
   2. $(\mathbf{Rb}-\mathbf{R}\beta)'(s^2\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}')^{-1}(\mathbf{Rb}-\mathbf{R}\beta)/rank(\mathbf{R}) \sim F_{rank(\mathbf{R}),n-k}$
      1. $$\begin{aligned}&(\mathbf{Rb}-\mathbf{R}\beta)'(s^2\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}')^{-1}(\mathbf{Rb}-\mathbf{R}\beta)/rank(\mathbf{R}) \\ &= \frac{(\mathbf{Rb}-\mathbf{R}\beta)'(\sigma^2\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}')^{-1}(\mathbf{Rb}-\mathbf{R}\beta)/rank(\mathbf{R})}{s^2/\sigma^2} \\ &\sim [\chi^2_{rank(\mathbf{R})}/rank(\mathbf{R})]/[\chi^2_{n-k}/(n-k)]\end{aligned}$$

大样本OLS(基于随机'DGP'，而非i.i.d随机样本)

DGP介绍

DGP（数据生成过程），或者称为STSP（随机时间序列过程），数学上可以写作一个随机过程$Z_t = Z(t, \ )$, ta的一条数据路径写作$Z(t,\omega), \omega \in \Omega$, 这是一个关于时间$t$的确定性函数，具体值由$\omega$决定

模型所要研究的DGP

平稳序列 stationary

严平稳序列

定义 5.2 ［严平稳］：一个随机时间序列 $\{Z_t\}$ 是严平稳的，如果对于任意的不同时间点 $t_1, t_2, \cdots, t_m$，以及对于任意的正整数 k 和 m，$\{Z_{t_1}, Z_{t_2}, \cdots, Z_{t_m}\}$ 与 $\{Z_{t_1+k}, Z_{t_2+k}, \cdots, Z_{t_m+k}\}$ 具有相同的联合概率分布，即

$$F_{Z_{t_1}Z_{t_2}\cdots Z_{t_m}}(z_1, \cdots, z_m) = F_{Z_{t_1+k}Z_{t_2+k}\cdots Z_{t_m+k}}(z_1, \cdots, z_m)$$

弱平稳(二阶平稳 协方差平稳)

对于一个随机时间序列$\{Z_t\}$, 若对任意t满足以下条件则称为弱平稳

(1) $E(Z_t) = \mu$。

(2) $\operatorname{var}(Z_t) = \sigma^2 < \infty$。

(3) $\operatorname{cov}(Z_t, Z_{t-j}) = \gamma(j)$ 仅是滞后阶数 j 的函数, 与时间点 t 无关。

遍历序列 ergodic

定义 5.9 [**遍历性 (Ergodicity)**]: 一个严平稳序列 $\{Z_t\}$ 是遍历的, 如果对于任意两个有界函数 $f : \mathbb{R}^{k+1} \to \mathbb{R}$ 和 $g : \mathbb{R}^{l+1} \to \mathbb{R}$, 及任意正整数 k, l, 有

$$\begin{aligned}&\lim_{m\to\infty} |E[f(Z_t, \cdots, Z_{t+k})g(Z_{m+t}, \cdots, Z_{m+t+l})]| \\ &= |E[f(Z_t, \cdots, Z_{t+k})]| \cdot |E[g(Z_{m+t}, \cdots, Z_{m+t+l})]|\end{aligned}$$

从随机过程角度理解:

“遍历性”=“不可约 非周期 正常返”

直观理解1:

对时间序列中任一时期取期望，与对时间序列各时期取值求均值等价；数据路径可以反映每一个时点的样本的数量特征。

直观理解2:

这意味着弱相关性：只要时间间隔足够长，时点间渐近独立

白噪声 white noise

定义 5.5 [**白噪声 (White Noise, WN)**]: 随机时间序列 $\{Z_t\}$ 是白噪声(或序列无关), 如果对所有 t, 有

(1) $E(Z_t) = 0$。

(2) $\operatorname{var}(Z_t) = \sigma^2$。

(3) $\operatorname{cov}(Z_t, Z_{t-j}) = \gamma(j) = 0$, 对所有的滞后阶 $j > 0$。

鞅 martingale

对于一个随机时间序列$\{Z_t\}$, 若对任意t满足以下条件则称为鞅:

$E(Z_t|\sigma_{t-1}) = Z_{t-1}$ $\quad\sigma_{t-1}$指的是前$t-1$期所蕴含的全部信息

$\{\sigma_t\}$构成了一个递增的集合序列，称为σ-域流

鞅差分序列 m.d.s

对于一个鞅$\{Z_t\}$, 做一阶差分可以生成一个新的随机时间序列$\{\epsilon\}$:
$\epsilon_t = Z_t - Z_{t-1}$
根据鞅的性质易得$E(\epsilon_t|\sigma_{t-1}) = 0$:不可预测当期的中性的随机冲击

引理 5.1 如果 $\{\varepsilon_t\}$ 是一个鞅差分序列, 且 $E(\varepsilon_t^2) = \sigma^2 < \infty$, 则 $\{\varepsilon_t\}$ 是一个白噪声。

证明 由重复期望法则, 有

$$E(\varepsilon_t) = E[E(\varepsilon_t|I_{t-1})] = 0$$

且对于任意的 $j > 0$, 有

$$\begin{aligned}\operatorname{cov}(\varepsilon_t, \varepsilon_{t-j}) &= E(\varepsilon_t\varepsilon_{t-j}) - E(\varepsilon_t)E(\varepsilon_{t-j}) \\ &= E[E(\varepsilon_t|I_{t-1})\varepsilon_{t-j}] \\ &= E(0 \cdot \varepsilon_{t-j}) \\ &= 0\end{aligned}$$

这表明, 鞅差分序列是白噪声, 且方差 $\operatorname{var}(\varepsilon_t) = \sigma^2$。

一些实例

AR(1)

MA(1)

ARCH

针对DGP的LLN&CLT

遍历平稳随机序列的WLLN

引理 5.2 [遍历平稳随机样本的弱大数定理 (WLLN for Ergodic Stationary Random Samples)]: 假设 $\{Z_t\}$ 是一个遍历平稳过程, 并且 $E(Z_t) = \mu$, $E|Z_t| < \infty$。则当 $n \to \infty$ 时, 样本均值

$$\bar{Z}_n = n^{-1}\sum_{t=1}^{n} Z_t \xrightarrow{p} \mu$$

遍历平稳鞅差分序列的CLT

定理 5.3 [遍历平稳鞅差分序列的中心极限定理 (**CLT for Ergodic Stationary M.D.S**)]**:** 假设 $\{Z_t\}$ 是一个遍历平稳鞅差分序列，$\text{var}(Z_t) \equiv E(Z_t Z_t') = V$ 是有限、对称与正定的矩阵。则当 $n \to \infty$ 时，有

$$\sqrt{n}\bar{Z}_n = n^{-1/2} \sum_{t=1}^{n} Z_t \xrightarrow{d} N(\mathbf{0}, V)$$

或等价地

$$V^{-1/2}\sqrt{n}\bar{Z}_n \xrightarrow{d} N(\mathbf{0}, I)$$

$$
\begin{aligned}
&Var(\sqrt{n}\bar{Z}_n) = E(\sqrt{n}\bar{Z}_n \cdot \sqrt{n}\bar{Z}_n{}') \\
&= n^{-1} E(\sum_{t=1}^{n} Z_t \cdot \sum_{s=1}^{n} Z_s') \\
&= n^{-1} \sum_{t=1}^{n} \sum_{s=1}^{n} E(Z_t \cdot Z_s') \\
&= n^{-1} \sum_{t=1}^{n} E(Z_t \cdot Z_t') \\
&\{t \neq s \quad e.g.\ t > s : E(Z_t \cdot Z_s') = E(E(Z_t \cdot Z_s')|\sigma_s) = 0\} \\
&= E(Z_t \cdot Z_t') \\
&= V
\end{aligned}
$$

模型假设

假设 5.1 [遍历平稳性 (Ergodic Stationarity)]：$\{Y_t, X_t'\}$ 是一个可观测的遍历平稳过程，其中 Y_t 是一个随机变量，X_t 是一个 $K \times 1$ 随机向量。

假设 5.2 [线性 (Linearity)]:

$$Y_t = X_t'\beta^o + \varepsilon_t$$

其中 β^o 是 $K \times 1$ 未知参数向量，ε_t 是不可观测的扰动项。

假设 5.3 [正确模型设定 (Correct Model Specification)]: $E(\varepsilon_t|X_t) = 0$ 且 $E(\varepsilon_t^2) = \sigma^2 < \infty$.

假设 5.4 [非奇异性 (Nonsingularity)]: $K \times K$ 矩阵

$$Q = E(X_t X_t')$$

是有限、对称和非奇异的。

假设 5.5 [鞅差分 (M.D.S)]: 相对于 $\{X_s\varepsilon_s, s < t\}$ 生成的 σ-域，$\{X_t\varepsilon_t\}$ 为一鞅差分序列，并且 $K \times K$ 矩阵 $V \equiv \mathrm{var}(X_t\varepsilon_t) = E(X_t X_t' \varepsilon_t^2)$ 是有限、对称及正定的。

---

0. 与高计(考完啦) > 有限样本假设相比假设4&5放松了，但是对随机序列$\{y_t, X_t\}$ $\{X_t\epsilon_t\}$增加了一些假设
1. 假设3也称为前定条件，这里有一个隐含假设：存在常数项解释变量。一般情况下假设3表示为$E(\mathbf{X_t}\epsilon_\mathbf{t}) = E(\mathbf{g_t}) = 0$，即$t$期解释变量在这一期之前已确定，与当期随机扰动项“正交”
2. 在接下来的记录中默认$X_{t:k*1}$(按时间)为列向量。从而可以将随机矩阵记作：$\mathbf{X} = (\mathbf{X_1'} \cdots \mathbf{X_n'})', \mathbf{X}' = (\mathbf{X_1} \cdots \mathbf{X_n})$
3. 记$g_t = X_t\epsilon_t$ $\bar{g} = \sum g_t/n = \mathbf{X}'\epsilon/n$
$\mathbf{S} = Var(\mathbf{g_t})$ $\mathbf{S_{xx}} = \sum X_t \cdot X_t'/n$ $(k*k \text{ vec.})$
$\mathbf{S_{xy}} = \sum X_t \cdot y_t/n$ $(k*1 \text{ vec.})$

OLS代数计算

由高计(考完啦) > OLS代数计算结果经过代数变换可以得到：

$$
\begin{aligned}
b &= (\mathbf{X'X})^{-1}\mathbf{X'y} = (\mathbf{X'X/n})^{-1}\mathbf{X'y/n} = \mathbf{S_{xx}^{-1}S_{xy}} \\
b &= \boldsymbol{\beta} + (\mathbf{X'X})^{-1}\mathbf{X'}\boldsymbol{\epsilon} = \boldsymbol{\beta} + (\mathbf{X'X/n})^{-1}\mathbf{X'}\boldsymbol{\epsilon}\mathbf{/n} \\
&= \boldsymbol{\beta} + \mathbf{S_{xx}^{-1}\bar{g}}
\end{aligned}
$$

基于这个式子可以得出$\mathbf{S_{xy}^{-1}} = \mathbf{S_{xx}}\boldsymbol{\beta} + \mathbf{\bar{g}}$
基于这个表达式得到的OLS渐近偏误可以解释内生性出现时偏误的方向。

大样本(渐近)OLS性质

help： 数理统计 > Slutsky引理

示例：$X_n \xrightarrow{\mathscr{D}} X, Y_n \xrightarrow{p} c$

---

b的大样本性质

- $\mathbf{\bar{g}} \xrightarrow{P} \mathbf{0}$
  LLN+模型前定假设立即得到
  一般情况下，渐近偏误依概率收敛到$\Sigma_{\mathbf{XX}^{-1}}E(\mathbf{g_t})$
- $\mathbf{S_{xx}^{-1}} \xrightarrow{P} \boldsymbol{\Sigma}_{\boldsymbol{XX}^{-1}}$
  LLN+模型非奇异假设立即得到
- $\mathbf{b} \xrightarrow{P} \boldsymbol{\beta}$
  $\mathbf{b} - \boldsymbol{\beta} = \mathbf{S_{xx}^{-1}\bar{g}}$: $\mathbf{\bar{g}} \xrightarrow{p} E(\mathbf{g}) = 0$ $\mathbf{S_{xx}^{-1}} \xrightarrow{p} \boldsymbol{\Sigma_{xx}^{-1}}$
  $\mathbf{b} - \boldsymbol{\beta} \xrightarrow{p} 0$ Q.E.D
- $\sqrt{n}\mathbf{S_{xx}^{-1}\bar{g}} \xrightarrow{\mathscr{D}} X \sim N(0, Avar(\mathbf{b}))$ $Avar(\mathbf{b}) = \boldsymbol{\Sigma_{xx}^{-1}}\mathbf{S}\boldsymbol{\Sigma_{xx}^{-1}}$
  ($\boldsymbol{\Sigma_{xx}^{-1}} = E(\mathbf{X'X})$ $\mathbf{S} = E(\mathbf{gg'}) = E(\epsilon^2\mathbf{x_i x_i'})$)
  将左式拆分成两部分 $\sqrt{n}\mathbf{S_{xx}^{-1}\bar{g}} = \mathbf{S_{xx}^{-1}} \cdot \sqrt{n}\mathbf{\bar{g}}$
  $\sqrt{n}\mathbf{\bar{g}}$依据m.d.s版本的CLT可以证明依分布收敛到$N(0,, Var(g_i) = S)$
  又因为$\mathbf{S_{xx}^{-1}} \xrightarrow{P} \boldsymbol{\Sigma_{xx}^{-1}}$，依据Slutsky定理，QED
- 补充命题：对于参数θ_0存在点估计θ_n且$\sqrt{n}(\theta_n - \theta_0) \xrightarrow{\mathscr{D}} N(\mu, \Sigma)$，$\mu, \Sigma$有界，则点估计$\theta_n$是参数$\theta_0$的一致估计
  $(\theta_n - \theta_0) \xrightarrow{\mathscr{D}} N(\mu/n, \Sigma/n)$ RHS r.v.依概率收敛到原点，QED

$E(\epsilon^2)$的估计

至于此处为什么$E(\epsilon^2)$与$Var(\epsilon)$不等价在上面已经提及：

> 假设3也称为前定条件，这里有一个隐含假设：存在常数项解释变量。
> 一般情况下假设3表示为$E(X_t\epsilon_t)=0$，即t期解释变量在这一期之前已确定，与当期随机扰动项“正交”

- $\frac{\mathbf{e}'\mathbf{e}}{n}$ $\frac{\mathbf{e}'\mathbf{e}}{n-k}$ 都是一致估计
  直观理解，后者本来就是无偏估计(更充分)，前者是后者的$\frac{n-k}{n}$倍，倍数收敛到1，从而在渐近视角下(具体证明用Slutsky)二者都是一致估计
  证明：

$$
\begin{aligned}
&\mathbf{e}=\mathbf{Y}-\mathbf{Xb}=\boldsymbol{\epsilon}-\mathbf{X}(\mathbf{b}-\boldsymbol{\beta})\\
&\mathbf{e}'\mathbf{e}=\boldsymbol{\epsilon}'\boldsymbol{\epsilon}+(\mathbf{b}-\boldsymbol{\beta})'\mathbf{X}'\mathbf{X}(\mathbf{b}-\boldsymbol{\beta})-(\mathbf{b}-\boldsymbol{\beta})'\mathbf{X}'\epsilon+\epsilon'\mathbf{X}(\mathbf{b}-\boldsymbol{\beta})\\
&\mathbf{e}'\mathbf{e}/\mathbf{n}=\boldsymbol{\epsilon}'\boldsymbol{\epsilon}/\mathbf{n}+(\mathbf{b}-\boldsymbol{\beta})'(\mathbf{X}'\mathbf{X}/\mathbf{n})(\mathbf{b}-\boldsymbol{\beta})-\mathbf{2\bar{g}}'(\mathbf{b}-\boldsymbol{\beta})
\end{aligned}
$$

  对最后一个式子利用LLN+上面的收敛性结论，QED

假设检验

- $H_0:\boldsymbol{\beta_j}=c$
  如果H_0正确，$\frac{\sqrt{n}(\mathbf{b_j}-c)}{\sqrt{Avar(\mathbf{b_{jj}})}}\sim N(0,1)$ Test: $\frac{\sqrt{n}(\mathbf{b_j}-c)}{\sqrt{\widehat{Avar(\mathbf{b_{jj}})}}}\xrightarrow{\mathscr{D}}N(0,1)$
  不同于有限样本，这里不存在估计效应，即检验统计量分布的扭曲
  高计(考完啦) > 点估计的条件分布
- $H_0:\mathbf{R}\boldsymbol{\beta}=r$
  如果H_0正确，$\frac{\sqrt{n}(\mathbf{Rb}-r)}{\sqrt{RAvar(\mathbf{b})R'}}\sim N(0,1)$
  Test: $n(\mathbf{Rb}-r)'(\mathbf{R}\widehat{Avar}(\mathbf{b})\mathbf{R}')^{-1}(\mathbf{Rb}-r)\xrightarrow{\mathscr{D}}\boldsymbol{\chi}^2_{\#R}$
  把上式视作正态r.vec.的自我内积，从而好理解为什么是$\boldsymbol{\chi}^2$分布：
  $\sqrt{n}'(\mathbf{Rb}-r)'(\mathbf{R}\widehat{Avar}(\mathbf{b})\mathbf{R}')^{-1/2\,\prime}(\mathbf{R}\widehat{Avar}(\mathbf{b})\mathbf{R}')^{-1/2}\sqrt{n}(\mathbf{Rb}-r)$

$\mathbf{S}=E(\mathbf{X_iX_i'}\epsilon_\mathbf{i}^2)=\Sigma_g$的估计

假设6: $\mathbf{X}$的四阶矩存在，则有$\hat{S}=\frac{\mathbf{X'BX}}{n}\xrightarrow{p}S\quad B=diag<e_i^2>$
证明：只考虑存在一个解释变量，被估计量是$E((x\epsilon)^2)$，估计量是$\sum(ex)^2/n$

$$
\begin{aligned}
&e^2=\epsilon^2+[(\mathrm{b}-\beta)x]^2-2(\mathrm{b}-\beta)x\epsilon\\
&(ex)^2=(\epsilon x)^2+[(\mathrm{b}-\beta)x^2]^2-2(\mathrm{b}-\beta)x^3\epsilon\\
&\sum(ex)^2=\sum(\epsilon x)^2+(\mathrm{b}-\beta)^2\sum x^4-2(\mathrm{b}-\beta)\sum x^3\epsilon\\
\\
&\sum(ex)^2/n\\
&=(\sum(\epsilon x)^2/n)+(\mathrm{b}-\beta)^2(\sum x^4/n)-2(\mathrm{b}-\beta)\bar{g}(\sum x^3\epsilon)\\
&(E(x^3\epsilon)=E(x^2x\epsilon)\le E(x^4)E(x\epsilon)^2=E(x^4)E(g^2)<\infty)
\end{aligned}
$$

对最后一个式子利用LLN+上面的收敛性结论，QED

GMM(基于内生性、工具变量主题)

内生性问题介绍

工具变量:回顾中级计量

数据分析方法(中级计量) 看这里就好了

模型假设

其实和CH2很像
这里一共有三类变量:1个被解释变量:$\mathbf{y}$; L个解释变量:$\mathbf{z}$; k个工具变量:$\mathbf{x}$

1. DGP: $\mathbf{y_i}=\mathbf{z_i'}\delta+\epsilon_i$
   这里对于解释变量$\{\mathbf{z_i}\}$没有任何的约束，可以视为是对CH2假设的放宽
2. $\{\mathbf{y_i},\mathbf{z_i},\mathbf{x_i}\}$是遍历平稳的随机序列
   为了使用遍历平稳序列的LLN
3. 工具变量前定:$E(\mathbf{x_i}\epsilon_i')=E(\mathbf{g_i})=E(\mathbf{x_i}\mathbf{y_i}-\mathbf{x_i}\mathbf{z_i'}\delta)=\mathbf{0}$
   这对应着工具变量的外生性
4. $r(E(\mathbf{x_i}\mathbf{z_i'}))=r(\mathbf{\Sigma_{xz}})=L$
   一方面保证了工具变量**至少是恰好识别的**
   另一方面保证了工具变量与解释变量的**相关性**:存在一个(L×L)主子式大于零，即存在L个工具变量与L个解释变量相关。
5. $\{\mathbf{g_i}\}$是一个二阶矩有限的m.d.s.
   为了使用遍历平稳m.d.s.的CLT
6. $E((\mathbf{x_{ik}}\mathbf{z_{il}})^2)$有限：有限四阶矩(相比之下第二章假设的是$X$的有限四阶矩)
   用来证明$\mathbf{S}=E(\mathbf{X_i}\mathbf{X_i'}\epsilon_\mathbf{i}^\mathbf{2})$的一致估计

恰好识别特例：矩估计(MM)

依据假设3给出的矩条件，经代数变形有：$E(\mathbf{x_i y_i}) - E(\mathbf{x_i z_i'})\delta = 0$
由于是恰好估计，依据假设4，矩阵$E(\mathbf{x_i z_i'})$可逆，从而δ存在显式解
$\delta = (E(\mathbf{x_i z_i'}))^{-1}E(\mathbf{x_i y_i}) \ \rightarrow \ \hat{\delta} = (\frac{1}{n}\sum_{i=1}^n \mathbf{x_i z_i'})^{-1}(\frac{1}{n}\sum_{i=1}^n \mathbf{x_i y_i}) = (\mathbf{S_{xz}})^{-1}(\mathbf{S_{xy}})$

一般情形:GMM

如果矩阵$E(\mathbf{x_i z_i'})$不是方阵，那么δ无解，从而不可估计；所以我们只好退而求其次，去寻找一个$\hat{\delta}$使得$\mathbf{S_{xy}} - \mathbf{S_{xz}}\hat{\delta}$ 尽可能接近$\mathbf{0}$，所以我们要最小化
$\bar{\mathbf{g}} = \mathbf{S_{xy}} - \mathbf{S_{xz}}\hat{\delta}$ 的模，然而在统计学中，我们允许不同维度的权重不等，这通过表示二次型的(K×K)矩阵$\hat{\mathbf{W}}$来体现

代数计算

此时我们的优化问题是：

$$\max_{\hat{\delta}} \quad (\mathbf{S_{xy}} - \mathbf{S_{xz}}\hat{\delta})'\hat{\mathbf{W}}(\mathbf{S_{xy}} - \mathbf{S_{xz}}\hat{\delta})$$

目标函数展开为：$\hat{\delta}'\mathbf{S_{xz}}'\hat{\mathbf{W}}\mathbf{S_{xz}}\hat{\delta} - 2\mathbf{S_{xy}'}\hat{\mathbf{W}}\mathbf{S_{xz}}\hat{\delta} + \mathbf{S_{xy}}'\hat{\mathbf{W}}\mathbf{S_{xy}}$
经过求导可以得到：$\hat{\delta} = (\mathbf{S_{xz}}'\hat{\mathbf{W}}\mathbf{S_{xz}})^{-1}\mathbf{S_{xz}}'\hat{\mathbf{W}}\mathbf{S_{xy}}$
此时$\bar{\mathbf{g}} = \mathbf{S_{xy}} - \mathbf{S_{xz}}\hat{\delta} = (\mathbf{I_k} - \mathbf{S_{xz}}(\mathbf{S_{xz}}'\hat{\mathbf{W}}\mathbf{S_{xz}})^{-1}\mathbf{S_{xz}}'\hat{\mathbf{W}})\mathbf{S_{xy}}$
同时我们也可以带入$= \mathbf{S_{xy}} = \mathbf{S_{xz}}\hat{\delta} + \bar{\mathbf{g}}$得到估计误差：
$\hat{\delta} - \delta = (\mathbf{S_{xz}}'\hat{\mathbf{W}}\mathbf{S_{xz}})^{-1}\mathbf{S_{xz}}'\hat{\mathbf{W}}\bar{\mathbf{g}}$

2SLS的GMM表达形式

假设模型:$\mathbf{y_i} = \delta_1 + \delta_2\mathbf{z_{2i}} + \delta_3\mathbf{z_{3i}} + \epsilon_i$

- 第一阶段：$reg \quad \mathbf{z_{2i}} \sim \mathbf{x_i}, \mathbf{z_{3i}}$
  定义矩阵$\mathbf{X} = (\mathbf{1}|\mathbf{X_1}|\mathbf{Z_3})_{n\times 3}$ $\mathbf{P_X} = \mathbf{X(X'X)^{-1}X'}$
  从而得到第一阶段拟合值$\hat{\mathbf{Z}}_2 = \mathbf{P_X Z_2}$
- 第二阶段：$reg \quad \mathbf{y_i} \sim \hat{\mathbf{z}}_{2i}, \mathbf{z_{3i}}$
  定义矩阵$\hat{\mathbf{Z}} = (\mathbf{1}\vdots\hat{\mathbf{Z}}_2\vdots\mathbf{Z_3})_{n\times 3} = \mathbf{P_X}(\mathbf{1}\vdots\mathbf{Z_2}\vdots\mathbf{Z_3})_{n\times 3}$
  $\hat{\delta_{2SLS}} = \mathbf{(\hat{Z}'\hat{Z})^{-1}\hat{Z}'y} = \mathbf{(Z'P_X Z)^{-1}Z'P_X y} = (\mathbf{S'_{xz}S_{xx}^{-1}S_{xz}})^{-1}\mathbf{S'_{xz}S_{xx}^{-1}S_{xy}}$
  由此可以看出$\hat{\mathbf{W}} = \mathbf{S_{xx}^{-1}}$ 第一阶段OLS regressor的协方差阵

渐近统计性质

GMM估计量的性质

1. $\hat{\delta} \xrightarrow{p} \delta$
   实际上就是$\bar{\mathbf{g}} \xrightarrow{p} \mathbf{0}$，这个通过假设3和4很容易得到
2. $\sqrt{n}(\hat{\delta}-\delta) \xrightarrow{d} N(0, Avar(\hat{\delta}))$
   $Avar(\delta) = (\mathbf{\Sigma}_{\mathbf{xz}}'\hat{\mathbf{W}}\mathbf{\Sigma}_{\mathbf{xz}})^{-1}\mathbf{\Sigma}_{\mathbf{xz}}'\hat{\mathbf{W}}\mathbf{S}\hat{\mathbf{W}}\mathbf{\Sigma}_{\mathbf{xz}}(\mathbf{\Sigma}_{\mathbf{xz}}'\hat{\mathbf{W}}\mathbf{\Sigma}_{\mathbf{xz}})^{-1}$
   实际上就是$\sqrt{n}(\bar{\mathbf{g}}) \xrightarrow{d} N(0, \mathbf{S})$，这个通过假设3和5很容易得到
   $\widehat{Avar}(\hat{\delta}) = (\mathbf{S_{xz}}'\hat{\mathbf{W}}\mathbf{S_{xz}})^{-1}\mathbf{S_{xz}}'\hat{\mathbf{W}}\hat{\mathbf{S}}\hat{\mathbf{W}}\mathbf{S_{xz}}(\mathbf{S_{xz}}'\hat{\mathbf{W}}\mathbf{S_{xz}})^{-1}$

GMM残差的性质

$\frac{1}{n}\sum_{i=1}^n e_i^2 \xrightarrow{p} E(\epsilon^2)$ $\ e_i = y_i - \mathbf{z_i}'\hat{\delta}$
$\frac{1}{n}\sum_{i=1}^n e_i^2 = \frac{1}{n}\sum_{i=1}^n \epsilon_i^2 + (\delta-\hat{\delta})'(\frac{1}{n}\sum_{i=1}^n \mathbf{z_i}\mathbf{z_i}')(\delta-\hat{\delta}) - 2(\delta-\hat{\delta})'(\frac{1}{n}\sum_{i=1}^n \mathbf{z_i}\epsilon_{\mathbf{i}}')$
但是这里不能得出$\sqrt{n}(\frac{1}{n}\sum_{i=1}^n e_i^2 - E(\epsilon^2)) \xrightarrow{d} N(0, Avar(\epsilon^2))$

$$\begin{aligned}\sqrt{n}(\frac{1}{n}\sum_{i=1}^n e_i^2 - E(\epsilon^2)) =& \sqrt{n}(\frac{1}{n}\sum_{i=1}^n \epsilon_i^2 - E(\epsilon^2)) \\ &+ (\delta-\hat{\delta})'(\sqrt{n}(\frac{1}{n}\sum_{i=1}^n \mathbf{z_i}\mathbf{z_i}'))(\delta-\hat{\delta}) \\ &- 2\sqrt{n}(\delta-\hat{\delta})'((\frac{1}{n}\sum_{i=1}^n \mathbf{z_i}\epsilon_{\mathbf{i}}'))\end{aligned}$$

这里并没有条件保证$((\frac{1}{n}\sum_{i=1}^n \mathbf{z_i}\epsilon_{\mathbf{i}}'))$的收敛性(相比之下在第二章，我们有前定条件保证$((\frac{1}{n}\sum_{i=1}^n \mathbf{x_i}\epsilon_{\mathbf{i}}'))$依概率收敛到0)，如果ta不依概率收敛到0，那么最后一项不会依概率收敛到0.

基于GMM模型的假设检验

1. $H_0: \delta_j = c$ 如果原假设成立：$\sqrt{n}(\hat{\delta}_j - \delta_j) \xrightarrow{\mathscr{D}} N(0, Avar(\hat{\delta})_{jj})$
   $\frac{\sqrt{n}(\hat{\delta}_j - c)}{\sqrt{\widehat{Avar}(\hat{\delta})}} \xrightarrow{d} N(0,1)$
2. $H_0: \mathbf{R}\delta = \mathbf{r}$ 如果原假设成立：$\sqrt{n}(\mathbf{R}\hat{\delta} - \mathbf{r}) \xrightarrow{d} N(0, \mathbf{R}Avar(\hat{\delta})\mathbf{R}')$
   $n(\mathbf{R}\hat{\delta} - \mathbf{r})'(\mathbf{R}\widehat{Avar}(\hat{\delta})\mathbf{R}')^{-1}(\mathbf{R}\hat{\delta} - \mathbf{r}) \xrightarrow{\mathscr{D}} \chi^2_{\#r}$

$\mathbf{S} = E(\mathbf{X_i X_i'}\epsilon_\mathbf{i}^2) = \Sigma_g$的估计

在假设6之下，类似于CH2，可以得出一致估计$\hat{\mathbf{S}} = \frac{1}{n}\sum_{i=1}^{n} e_i^2(\mathbf{x_i x_i'})$
我们仅证明K=L=1的情景，**过程在AE.2**

2. We wish to prove Proposition 3.4. To avoid inessential complications, consider the case where $K = 1$ and $L = 1$ so that $\mathbf{x}_i$ and $\mathbf{z}_i$ are scalars. So (3.5.5) becomes

$$\hat{\varepsilon}_i^2 = \varepsilon_i^2 - 2(\hat{\delta} - \delta)z_i\,\varepsilon_i + (\hat{\delta} - \delta)^2 z_i^2,$$

and the $\widehat{\mathbf{S}}$ in (3.5.10) becomes

$$\frac{1}{n}\sum_{i=1}^{n} x_i^2\,\varepsilon_i^2 - 2(\hat{\delta} - \delta)\frac{1}{n}\sum_{i=1}^{n} z_i\,x_i^2\,\varepsilon_i + (\hat{\delta} - \delta)^2\frac{1}{n}\sum_{i=1}^{n} z_i^2\,x_i^2. \qquad (*)$$

(a) Show: The first term on the RHS of (∗) converges in probability to S $(\equiv \mathrm{E}(x_i^2\,\varepsilon_i^2))$.

(b) Use the Cauchy-Schwartz inequality

$$\mathrm{E}(|X \cdot Y|) \leq \sqrt{\mathrm{E}(X^2)\cdot \mathrm{E}(Y^2)}$$

to show that $\mathrm{E}(z_i x_i^2 \varepsilon_i)$ exists and is finite. **Hint:** $z_i x_i^2 \varepsilon_i$ is the product of $x_i\varepsilon_i$ and $x_i z_i$. Fact: $\mathrm{E}(x)$ exists and is finite if and only if $\mathrm{E}(|x|) < \infty$. Where do we use Assumption 3.6?

(c) Show that the second term on the RHS of (∗) converges to zero in probability.

AE2 (1) 由于$\{x_i,\varepsilon_i\}$遍历平稳：符合LLN条件，从而对$\frac{1}{n}\sum_{i=1}^{n} x_i^2\varepsilon_i^2$有：$\frac{1}{n}\sum_{i=1}^{n} x_i^2\varepsilon_i^2 \xrightarrow{p} E(x_i^2\varepsilon_i^2)$

(2) $E z_i x_i^2 \varepsilon_i = E(z_i x_i)(x_i\varepsilon_i) \leq \sqrt{E(z_i x_i)^2 E(x_i\varepsilon_i)^2}$，由Assumption 3.5知$E(x_i^2\varepsilon_i^2) < \infty$

由Assumption 3.6知$E(x_i^2 z_i^2) < \infty$ $\therefore E z_i x_i^2 \varepsilon_i < \infty$

(3) 对第二项：$\frac{1}{n}\sum_{i=1}^{n} z_i x_i^2 \varepsilon_i \xrightarrow{p} E z_i x_i^2 \varepsilon_i < \infty$，$\hat{\delta} \xrightarrow{p} \delta$，从而第二项依概率收敛至0

(4) 对第三项：$\frac{1}{n}\sum_{i=1}^{n} z_i^2 x_i^2 \xrightarrow{p} E z_i^2 x_i^2 < \infty$，$\hat{\delta} \xrightarrow{p} \delta$，从而第三项依概率收敛至0

最优权重矩阵的选取

通过上面的计算，我们发现$Avar(\hat{\delta})$与$\hat{\mathbf{W}}$相关，为了最高的估计准度，我们要选取$\hat{\mathbf{W}}$使得$Avar(\hat{\delta})$最小。这里采用先猜后证的方式给出：猜$\hat{\mathbf{W}} = \hat{\mathbf{S}}^{-1}$

$\hat{\delta} = (\mathbf{S_{xz}}'\hat{\mathbf{S}}^{-1}\mathbf{S_{xz}})^{-1}\mathbf{S_{xz}}'\hat{\mathbf{S}}^{-1}\mathbf{S_{xy}}$

$\widehat{Avar}(\hat{\delta}) = (\mathbf{S_{xz}}'\hat{\mathbf{S}}^{-1}\mathbf{S_{xz}})^{-1}\mathbf{S_{xz}}'\hat{\mathbf{S}}^{-1}\hat{\mathbf{S}}\hat{\mathbf{S}}^{-1}\mathbf{S_{xz}}(\mathbf{S_{xz}}'\hat{\mathbf{S}}^{-1}\mathbf{S_{xz}})^{-1} = (\mathbf{S_{xz}}'\hat{\mathbf{S}}^{-1}\mathbf{S_{xz}})^{-1}$

$\widehat{Avar}(\hat{\delta}) \xrightarrow{p} (\mathbf{\Sigma}'_{\mathbf{xz}}\mathbf{S}^{-1}\mathbf{\Sigma_{xz}})^{-1} = Avar(\hat{\delta})$

证明**过程在AE.3**

3. We wish to prove the algebraic result (3.5.11). What needs to be proved is that $\mathbf{A} - \mathbf{B}$ is positive semidefinite where

$$\mathbf{A} \equiv (\mathbf{\Sigma}'_{\mathbf{xz}}\mathbf{W}\mathbf{\Sigma}_{\mathbf{xz}})^{-1}\mathbf{\Sigma}'_{\mathbf{xz}}\mathbf{W}\mathbf{S}\mathbf{W}\mathbf{\Sigma}_{\mathbf{xz}}(\mathbf{\Sigma}'_{\mathbf{xz}}\mathbf{W}\mathbf{\Sigma}_{\mathbf{xz}})^{-1}$$

and

$$\mathbf{B} \equiv (\mathbf{\Sigma}'_{\mathbf{xz}}\mathbf{S}^{-1}\mathbf{\Sigma}_{\mathbf{xz}})^{-1}.$$

A result from matrix algebra states:

> Let $\mathbf{A}$ and $\mathbf{B}$ be two positive definite and symmetric matrices. $\mathbf{A} - \mathbf{B}$ is positive semidefinite if and only if $\mathbf{B}^{-1} - \mathbf{A}^{-1}$ is positive semidefinite.

Therefore, what needs to be proved is that

$$\mathbf{Q} \equiv \mathbf{\Sigma}'_{\mathbf{xz}}\mathbf{S}^{-1}\mathbf{\Sigma}_{\mathbf{xz}} - \mathbf{\Sigma}'_{\mathbf{xz}}\mathbf{W}\mathbf{\Sigma}_{\mathbf{xz}}(\mathbf{\Sigma}'_{\mathbf{xz}}\mathbf{W}\mathbf{S}\mathbf{W}\mathbf{\Sigma}_{\mathbf{xz}})^{-1}\mathbf{\Sigma}'_{\mathbf{xz}}\mathbf{W}\mathbf{\Sigma}_{\mathbf{xz}}$$

is positive semidefinite.

(a) Since $\mathbf{S}$ $(K \times K)$ is positive definite, there exists a nonsingular $K \times K$ matrix $\mathbf{C}$ such that $\mathbf{C}'\mathbf{C} = \mathbf{S}^{-1}$. Verify that $\mathbf{Q}$ can be written as

$$\mathbf{Q} = \mathbf{H}'\mathbf{M_G}\mathbf{H},$$

where

$$\mathbf{H} = \mathbf{C}\mathbf{\Sigma}_{\mathbf{xz}}, \quad \mathbf{M_G} = \mathbf{I}_K - \mathbf{G}(\mathbf{G}'\mathbf{G})^{-1}\mathbf{G}', \quad \mathbf{G} = \mathbf{C}'^{-1}\mathbf{W}\mathbf{\Sigma}_{\mathbf{xz}}.$$

Hint: $\mathbf{C}^{-1}\mathbf{C}'^{-1} = \mathbf{S}$.

(b) Show that $\mathbf{Q}$ is positive semidefinite. **Hint:** First show that $\mathbf{M_G}$ is symmetric and idempotent. As you showed in Exercise 1, a symmetric and idempotent matrix is positive semidefinite.

AE3 (1) $Q=(\Sigma_{xz}'C')(C\Sigma_{xz})-\Sigma_{xz}'W\Sigma_{xz}(\Sigma_{xz}'WC^{-1}C'^{-1}W\Sigma_{xz})^{-1}\Sigma_{xz}'W\Sigma_{xz}$

$=(\Sigma_{xz}'C')(C\Sigma_{xz})-(\Sigma_{xz}'C')C^{-1}W\Sigma_{xz}(\Sigma_{xz}'WC^{-1}C'^{-1}W\Sigma_{xz})^{-1}\Sigma_{xz}'WC^{-1}(C\Sigma_{xz})$

$=H'(I_k-G(G'G)^{-1}G')H = H'M_GH$

(2) 先证 M_G 对称、幂等(idempotent)：

$M_G' = I_k - G(G(G'G)^{-1})' = I_k - G(G'G)^{-1}G' = M_G$

$M_GM_G = I_k - 2G(G'G)^{-1}G' + G(G'G)^{-1}G'G(G'G)^{-1}G' = I_k - G(G'G)^{-1}G' = M_G$

对这类矩阵，任给 x，$x'Ax = x'A'Ax = (Ax)'Ax \geq 0$；$A$ 半正定

同理 $x'H'M_GHx = x'H'M_G'M_GHx = (M_GHx)'M_GHx \geq 0$；$Q$ 半正定

两阶段有效GMM

动机：基于最优权重矩阵$\hat{\mathbf{W}}=\hat{\mathbf{S}}^{-1}$，我们需要估计出$\hat{\mathbf{S}}^{-1}$；这需要$e_i$；而为了得到$e_i$我们需要得到一致估计的$\hat{\delta}$, 从而我们可以先估计出$\hat{\delta}$

1. 先用一些经典的$\hat{\mathbf{W}}$入手得到一致估计$\hat{\delta}$，从而得到$\hat{\mathbf{S}}^{-1}$
   比如取$\hat{\mathbf{W}}=\mathbf{S_{xx}}^{-1}$ $(2SLS)$ $\hat{\mathbf{W}}=\mathbf{I}$ (无差异)
2. 利用第一步得到的$\hat{\mathbf{S}}^{-1}$作为$\hat{\mathbf{W}}$再做一次GMM，使得$Avar(\hat{\delta})$最小

基于GMM的IV外生性检验

动机：通过计算残差向量到原点的"**最小统计距离**"判断，因此，我们选取最优权重矩阵$\hat{\mathbf{W}}=\hat{\mathbf{S}}^{-1}$得到$J-statistic$

$$J(\hat{\delta},\hat{\mathbf{S}})=(\sqrt{n}\bar{\mathbf{g}})'\hat{\mathbf{S}}^{-1}(\sqrt{n}\bar{\mathbf{g}}) \qquad \sqrt{n}\bar{\mathbf{g}}\xrightarrow{d}N(0,(\mathbf{S_{xz}}'\hat{\mathbf{S}}^{-1}\mathbf{S_{xz}})^{-1})$$

如果假设3.1-3.5全部满足：$J-statistic$的渐近分布是χ^2_K(根据$Slutsky$引理)，但小样本分布是χ^2_{K-L}，受制于$\hat{\delta}$中L个等式的约束。这个统计量衡量了$\hat{\delta}(\hat{\mathbf{S}})$对应的残差向量到原点的"最小统计距离"。如果这个数值过大，那很有可能就是外生性条件被违反。

证明**过程在AE.5**

5. (optional) Prove Proposition 3.6 by following the steps below.

(a) Show: $\mathbf{g}_n(\hat{\boldsymbol{\delta}}(\widehat{\mathbf{S}}^{-1})) = \widehat{\mathbf{B}}\,\bar{\mathbf{g}}$ where $\widehat{\mathbf{B}} \equiv \mathbf{I}_K - \mathbf{S}_{\mathbf{xz}}(\mathbf{S}'_{\mathbf{xz}}\widehat{\mathbf{S}}^{-1}\mathbf{S}_{\mathbf{xz}})^{-1}\mathbf{S}'_{\mathbf{xz}}\widehat{\mathbf{S}}^{-1}$.

(b) Since $\widehat{\mathbf{S}}$ is positive definite, there exists a nonsingular $K \times K$ matrix $\mathbf{C}$ such that $\mathbf{C}'\mathbf{C} = \widehat{\mathbf{S}}^{-1}$. Define $\mathbf{A} \equiv \mathbf{C}\mathbf{S}_{\mathbf{xz}}$. Show that $\widehat{\mathbf{B}}'\widehat{\mathbf{S}}^{-1}\widehat{\mathbf{B}} = \mathbf{C}'\mathbf{M}\mathbf{C}$, where $\mathbf{M} \equiv \mathbf{I}_K - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$. What is the rank of $\mathbf{M}$? **Hint:** The rank of an idempotent matrix equals its trace.

(c) Let $\mathbf{v} \equiv \sqrt{n}\,\mathbf{C}\,\bar{\mathbf{g}}$. Show that $\mathbf{v} \to_{\mathrm{d}} N(\mathbf{0}, \mathbf{I}_K)$.

(d) Show that $J(\hat{\boldsymbol{\delta}}(\widehat{\mathbf{S}}^{-1}), \widehat{\mathbf{S}}^{-1}) = \mathbf{v}'\mathbf{M}\mathbf{v}$. What is its asymptotic distribution?

AE5 (1) $g_n(\hat\delta(\hat S^{-1})) = \frac{1}{n}\sum_{i=1}^n x_i(y_i - z_i'\hat\delta(\hat S^{-1})) = S_{xy} - S_{xz}\hat\delta(\hat S^{-1}) = (I - S_{xz}(S_{xz}'\hat S^{-1}S_{xz})^{-1}S_{xz}'\hat S^{-1})S_{xy} = \hat B S_{xy}$

$\hat B S_{xy} = \hat B \cdot \frac{1}{n}\sum_{i=1}^n x_i(z_i'\hat S^{-1} + \varepsilon_i) = \hat B \bar g + \hat B S_{xz}'\hat S^{-1}$

$\hat B S_{xz} S^{-1} = S_{xz}\hat S^{-1} - S_{xz}(S_{xz}'\hat S^{-1}S_{xz})^{-1}S_{xz}'\hat S^{-1}S_{xz}\hat S^{-1} = S_{xz}\hat S^{-1} - S_{xz}\hat S^{-1} = 0$

$\therefore \hat B S_{xy} = \hat B \bar g$

(2) $\hat B'\hat S^{-1}\hat B = (C\hat B)'C\hat B$;

$C\hat B = C - CS_{xz}(S_{xz}'C'CS_{xz})^{-1}S_{xz}'C'C = C - A(A'A)^{-1}A'C = MC$

$\therefore (C\hat B)'C\hat B = C'M'MC$

又因为 $M'M = (I_k - A(A'A)^{-1}A')'(I_k - A(A'A)^{-1}A') = I_k - A(A'A)^{-1}A'$; 上式 $= C'MC$

由于 M 幂等: $\mathrm{rank}(M) = tr(M) = k - tr(A(A'A)^{-1}A') = k - tr(A'A(A'A)^{-1}) = k - l$

(3) 由于 $\sqrt{n}\bar g \xrightarrow{d} N(0, S)$, $\sqrt{n}C\bar g \xrightarrow{d} N(0, CSC')$, $CSC' = C(C'C)^{-1}C' = CC^{-1}C'^{-1}C' = I_k$

(4) $J(\hat\delta(\hat S^{-1}), \hat S^{-1}) = n(\hat B\bar g)'\hat S^{-1}(\hat B\bar g) = (\sqrt{n}\bar g)'\hat B'\hat S^{-1}\hat B(\sqrt{n}\bar g) = (\sqrt{n}\bar g)'C'MC(\sqrt{n}\bar g)$

$= (\sqrt{n}C\bar g)'M(\sqrt{n}C\bar g)$, 由于 $\sqrt{n}C\bar g \xrightarrow{d} N(0, I_k)$, M 为幂零矩阵,

所以 $J \xrightarrow{d} \chi^2_{k-l = \mathrm{rank}(M)}$

期末小抄记录内容

1. ~~三类模型的假设、结论（平行）~~
2. ~~有限样本的几何性质~~
3. ~~GMM的一些特殊性质(有效GMM 两步GMM 2SLS-GMM Hansen J-test)~~
4. ~~一些例题~~