高级计量经济学II课堂笔记*

刘旭浚[†]　赵凡青[‡]

2024 年 4 月 12 日：第2版

2020 年 9 月 6 日：第1版

*笔记主要内容由赵凡青于2020年9月份的第一版整理完成，第2版内容由刘旭浚进行的补充与调整（作者排序按照拼音顺序）。本笔记主要为上海财经大学经济学院李聪老师所开设的高级计量经济学II的课堂内容笔记，仅供学习参考使用。

[†]上海财经大学金融学院2023级博士研究生；

[‡]上海财经大学金融学院2019级博士研究生。

写在前面

特殊时期为了便于自己复习，顺便学习下LaTeX的用法，特此整理高计II笔记。有错误还请及时告知我。

赵凡青，2020年9月

偶然发现的很棒的计量笔记，顺便为了整理自己学习计量经济学以来的笔记和思考，决定在此基础上对赵凡青学长的笔记进行一定程度上的更新与整理。本次更新纯属个人行为，如有与之前笔记的龃龉之处，或有所纰漏，都与赵学长无关，也欢迎与我讨论笔记中的不足之处。

刘旭浚，2024年4月

目录

第一章　计量I回顾

1.1　最小二乘

1.1.1　最小二乘估计

对于经典的线性回归模型

$$y = X\beta + u \tag{1.1}$$

其中，y和u为$n\times 1$维的向量，$X = (x_1, x_2, \ldots, x_k)$为$n\times k$维的矩阵、$x_i$为$n\times 1$的向量，$\beta$为$k\times 1$维的向量。[1]在古典假设成立的情况下，参数估计量为BLUE（Best Linear Unbiased Estimator）.

为求解所估计的$\hat{\beta}$，根据OLS定义，最小化残差平方和：

$$\min_{\beta}\ \underbrace{(y-X\beta)'}_{1\times n}\underbrace{(y-X\beta)}_{n\times 1}$$

为求解β，有

$$\begin{aligned}
\text{FOC}:\ & \frac{\partial}{\partial\beta}(y'y-\beta'X'y-y'X\beta+\beta'X'X\beta)\\
&= -X'y-X'y+(X'X+X'X)\beta\\
&= -2X'y+2X'X\beta = 0\\
\Longrightarrow \hat{\beta} &= (X'X)^{-1}X'y\\
&= (X'X)^{-1}X'u+\beta
\end{aligned} \tag{1.2}$$

由于$\hat{\beta}$为u的线性组合，在古典假设中，我们假定$u\sim N(0,\sigma^2 I_n)$，因此为求得估计量$\hat{\beta}$的分布，我们仅需求得$\hat{\beta}$的均值与方差即可。

[1]通常，我们设定x_1为单位向量为保证模型包括了常数项。

$\hat{\beta}$的均值.

$$\begin{aligned}
\mathrm{E}[\hat{\beta}] &= \mathrm{E}[(\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}'y] \\
&= \mathrm{E}[(\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}'(\mathrm{X}\beta+u)] \\
&= \beta+\mathrm{E}[(\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}'u] \\
(\because \mathrm{E}[\mathrm{X}'u]=0) &= \beta \qquad (1.3)
\end{aligned}$$

由于$\mathrm{E}[\hat{\beta}]=\beta$，故OLS估计$\hat{\beta}$为无偏估计量。

$\hat{\beta}$的方差.

$$\begin{aligned}
\mathrm{V}ar(\hat{\beta}) &= \mathrm{E}[(\hat{\beta}-\beta)'(\hat{\beta}-\beta)] \\
&= \mathrm{E}[(\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}'uu'\mathrm{X}(\mathrm{X}'\mathrm{X})] \\
&= (\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}'\mathrm{E}[uu']\mathrm{X}(\mathrm{X}'\mathrm{X}) \\
&= (\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}'\sigma^2\mathrm{I}\mathrm{X}(\mathrm{X}'\mathrm{X}) \\
&= \sigma^2(\mathrm{X}'\mathrm{X})^{-1} \qquad (1.4)
\end{aligned}$$

当σ^2已知时，$\hat{\beta}\sim\mathrm{N}(\beta,\sigma^2(\mathrm{X}'\mathrm{X})^{-1})$。但实际中，我们通常并不知道随机扰动项方差的真值（true value），因此，我们需要估计σ^2。

$$\hat{\sigma}^2=\frac{\hat{u}'\hat{u}}{n-k} \qquad (1.5)$$

其中

$$\begin{aligned}
\hat{u} &= \hat{y}-\mathrm{X}\hat{\beta} \\
&= \underbrace{[\mathrm{I}_n-\overbrace{\mathrm{X}(\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}'}^{\text{P：投影矩阵（Projection matrix）}}]}_{\text{M：消灭矩阵（Annihilator matrix）}}y \qquad (1.6)
\end{aligned}$$

对于矩阵P和矩阵M而言，他们具有如下性质：（1）$\mathrm{PX}=\mathrm{X},\mathrm{P}u=0,\mathrm{MX}=0$；（2）P和M都是对称矩阵和

幂等矩阵。[2] 回到Equation 1.5，我们需要求得$E[\hat{u}'\hat{u}]$来得到无偏估计量$\hat{\sigma}^2$，故有

$$
\begin{aligned}
E[\hat{u}'\hat{u}] &= u'Mu \quad (\because \hat{u} = My = M(X\beta + u) = Mu) \\
&= tr(E[u'Mu]) \\
&= E[tr(u'Mu)] \text{（线性算子可交换运算顺序）} \\
&= E[tr(Muu')] \mathrm{tr(AB)=tr(BA)} \\
&= tr(E[Muu']) \\
&= tr(E[ME[uu'|X]]) \\
&= \sigma^2 E[tr(M)] \\
&= \sigma^2 E[tr(I) - tr(X(X'X)^{-1}X')] \text{（tr(A+B)=tr(A)+tr(B)）} \\
&= \sigma^2 E[tr(I) - tr((X'X)^{-1}X'X)] = \sigma^2(n-k)
\end{aligned} \tag{1.7}
$$

故，可得σ^2的无偏估计如下：

$$E[\hat{\sigma}^2] = E[\frac{\hat{u}'\hat{u}}{n-k}] = \sigma^2 \tag{1.8}$$

所以估计量$\hat{\beta}$的分布为：

$$\hat{\beta} \sim N(\beta, \sigma^2(X'X)^{-1}) \tag{1.9}$$

将向量$\hat{\beta}$中的元素标准化后，有

$$z_j = \frac{\hat{\beta}_j - \beta_j}{\sqrt{\sigma^2(X'X)^{-1}_{jj}}} \sim N(0,1) \tag{1.10}$$

1.1.2 参数检验

*t*检验：单变量

但通常情况下，σ^2是未知的，故我们需要使用σ^2的无偏估计来进行代换。

$$
\begin{aligned}
\frac{\hat{\beta}_j - \beta_j}{\sqrt{\hat{\sigma}^2(X'X)^{-1}_{jj}}} &= \frac{(\hat{\beta}_j - \beta_j)/\sqrt{\sigma^2(X'X)^{-1}_{jj}}}{\sqrt{[(n-k)\hat{\sigma}^2/\sigma^2]/(n-k)}} \\
&= \frac{N(0,1)}{\sqrt{\chi^2_{(n-k)}/(n-k)}}
\end{aligned} \tag{1.11}
$$

[2] 矩阵M是将向量变换到与x向量所张成平面垂直的平面上，所以$MX = 0$；矩阵P是将向量变换到与x所张成平面平行的向量（该平面的投影projection）,所以$PX = X$；P和M是一组正交分解，所以$P + M = I_n$，$Py + My = y$向量加法（平行四边形法则)，P称为projection matrix，M称为orthogonal projection matrix，M和P都是symetric idempotent matrix（对称、幂等矩阵）。

其中，分母中的$(n-k)s^2/\sigma^2 \sim \chi^2_{(n-k)}$，由Theorem 1.1.1可知，

$$(n-k)\hat{\sigma}^2/\sigma^2 = \frac{\hat{u}'\hat{u}}{\sigma^2} = \frac{u'\mathrm{M}u}{\sigma^2} = (\frac{u}{\sigma})'\mathrm{M}(\frac{u}{\sigma}) \sim \chi^2(tr(\mathrm{M})) = \chi^2(n-k)$$

其中$u/\sigma \sim \mathrm{N}(0, 1_n)$.

Theorem 1.1.1. **标准正态向量中幂等二次型的分布**

if $\underset{m\times 1}{z} \sim \mathrm{N}(0, \mathrm{I}_m)$*, matrix A is a* $m \times m$ *symetric idempotent matrix, then*

$$z'\mathrm{A}z \sim \chi^2(df)$$

其中，χ^2分布的自由度$df = rank(\mathrm{A}) = tr(\mathrm{A})$（幂等矩阵的秩等于它的迹）。

但为了确定Equation 1.11的分布，我们还需证明分子分母随机变量的独立性。在证明其独立性之前，我们先介绍线性函数与二次型独立性的定理。

Theorem 1.1.2. **线性函数与二次型的独立性**

一个标准正态向量x的线性函数$\mathrm{L}x$与对称幂等二次型$x'\mathrm{A}x$独立，如果$\mathrm{LA} = 0$。

证明.首先，因为A为幂等矩阵，故$x'\mathrm{A}x = x'\mathrm{A}'\mathrm{A}x = (\mathrm{A}x)'\mathrm{A}x$。由于$x \sim \mathrm{N}(0, \mathrm{I}_n)$，故$\mathrm{L}x$与$x'\mathrm{A}x$均服从正态分布。对于正态分布而言，若$Cov(\mathrm{L}x, (\mathrm{A}x)') = 0$，则说明二者独立。

$$\begin{aligned} \mathrm{C}ov(\mathrm{L}x, (\mathrm{A}x)') &= \mathrm{E}[\mathrm{L}xx'\mathrm{A}'] - \mathrm{E}[\mathrm{L}x]\mathrm{E}[\mathrm{A}x] \\ &= \mathrm{LI}_n\mathrm{A}' = \mathrm{LA} \end{aligned}$$

故，当$\mathrm{LA} = 0$时，$Cov(\mathrm{L}x, (\mathrm{A}x)') = 0$，有$\mathrm{L}x$与$x'\mathrm{A}x$独立（这个证明稍微有点奇怪）。

Equation 1.11的分子分母分别为

$$\begin{aligned} \text{分子：} \quad & \frac{\hat{\beta}_j - \beta_j}{\sqrt{\sigma^2 (\mathrm{X}'\mathrm{X})^{-1}_{jj}}} = \frac{(\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}'}{\sqrt{(\mathrm{X}'\mathrm{X})^{-1}_{jj}}} \cdot \frac{u}{\sigma} \\ \text{分母：} \quad & \frac{(n-k)\hat{\sigma}^2}{\sigma^2} = (\frac{u}{\sigma})'\mathrm{M}(\frac{u}{\sigma}) \end{aligned}$$

又因为

$$\frac{(\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}'}{\sqrt{(\mathrm{X}'\mathrm{X})^{-1}_{jj}}}\mathrm{M} = \frac{(\mathrm{X}'\mathrm{X})^{-1}(\mathrm{XM})'}{\sqrt{(\mathrm{X}'\mathrm{X})^{-1}_{jj}}} = 0$$

故，Equation 1.11的分子分母独立，有Equation 1.11中的

$$\frac{\hat{\beta}_j - \beta_j}{\sqrt{\hat{\sigma}^2 (\mathrm{X}'\mathrm{X})^{-1}_{jj}}} \sim t(n-k)$$

***F*检验：多变量联合**

已知方差：Wald test. 为检验所估计参数（$\hat{\beta}$）是否满足J个线性约束是否，本小节将介绍F检验。首先，假设

$$\begin{cases} \mathrm{H}_0 : \mathrm{R}\hat{\beta} = q \\ \mathrm{H}_1 : \mathrm{R}\hat{\beta} \neq q \end{cases} \Longrightarrow \begin{matrix} \mathrm{H}_0 : d = \mathrm{R}\hat{\beta} - q = 0 \\ \mathrm{H}_1 : d = \mathrm{R}\hat{\beta} - q \neq 0 \end{matrix}$$

由于$\hat{\beta}$服从正态分布，d为$\hat{\beta}$的线性组合、服从于正态分布。故，对于d的分布，我们仅需要知道其均值与方差。

- **均值**：$\mathrm{E}(d) = \mathrm{RE}(\hat{\beta}) - q \overset{\mathrm{H}_0}{=} 0$
- **方差**：$\mathrm{V}ar(d) = \mathrm{RV}ar(\hat{\beta})\mathrm{R}' = \sigma^2 \mathrm{R}(\mathrm{X}'\mathrm{X})^{-1}\mathrm{R}'$

故有，

$$d = \mathrm{R}\hat{\beta} - q \sim \mathrm{N}(0, \Sigma) \tag{1.12}$$

其中，$\Sigma^2 = \sigma^2 \mathrm{R}(\mathrm{X}'\mathrm{X})^{-1}\mathrm{R}'$。故，在方差$\sigma^2$已知的情况下，可直接使用标准化$d$（服从于标准正态分布）来进行假设检验。如前所述，在通常的实证中，我们并不知道随机扰动项的真实方差σ^2，因而，我们将使用Wald检验：

$$w = d'\Sigma^{-1}d = (\mathrm{R}\hat{\beta} - q)'[\sigma^2 \mathrm{R}(\mathrm{X}'\mathrm{X})^{-1}\mathrm{R}']^{-1}(\mathrm{R}\hat{\beta} - q)$$

其中，J为约束条件的个数。根据Theorem 1.1.3，可知$w \sim \chi^2(\mathrm{J})$。

Theorem 1.1.3. *If a $n \times 1$ vector $x \sim \mathrm{N}(\mu, \Sigma)$, then*

$$\begin{aligned} \Sigma^{-1/2}(x - \mu) &\sim \mathrm{N}(0, \mathrm{I}_n) \\ (x-\mu)'\Sigma^{-1}(x-\mu) &\sim \chi^2(n) \end{aligned}$$

未知方差：F test. 类似于对$\hat{\beta}$的检验，由于Σ中σ^2未知，因此，我们将使用$s^2 = \hat{\sigma}^2$来替代σ^2。

$$\begin{aligned} \mathrm{F} &= (\mathrm{R}\hat{\beta} - q)'[\hat{\sigma}^2 \mathrm{R}(\mathrm{X}'\mathrm{X})^{-1}\mathrm{R}']^{-1}(\mathrm{R}\hat{\beta} - q)/\mathrm{J} \\ &= \frac{(\mathrm{R}\hat{\beta} - q)'[\sigma^2 \mathrm{R}(\mathrm{X}'\mathrm{X})^{-1}\mathrm{R}']^{-1}(\mathrm{R}\hat{\beta} - q)/\mathrm{J}}{\frac{s^2(n-k)}{\sigma^2}/(n-k)} \sim \mathrm{F}(\mathrm{J}, n-k) \end{aligned} \tag{1.13}$$

$F=\frac{\chi_1^2/n_1}{\chi_2^2/n_2}$要求分子分母中的随机变量独立(与$t$统计量一样)。而对于正态随机变量而言，随机变量不相关与独立等价，故，仅需证明Equation 1.13中分子分母的随机变量不相关、即可证明其服从于F分布。为了证明不相关，考虑分子

$$\frac{R\hat{\beta}-q}{\sigma}\overset{H_0}{=}\frac{R(\hat{\beta}-\beta)}{\sigma}=\frac{R(X'X)^{-1}X'u}{\sigma}=\frac{Au}{\sigma}$$

而分母，

$$\frac{s^2(n-k)}{\sigma^2}=\frac{\frac{\hat{u}'\hat{u}}{n-k}(n-k)}{\sigma^2}=\frac{\hat{u}'\hat{u}}{\sigma^2}=\frac{u'M'Mu}{\sigma^2}=(\frac{Mu}{\sigma})'(\frac{Mu}{\sigma})$$

因此只需要考虑$\frac{Au}{\sigma}$和$\frac{Mu}{\sigma}$的相关性，有

$$Cov(\frac{Au}{\sigma},\frac{Mu}{\sigma})=E(\frac{Au}{\sigma})(\frac{Mu}{\sigma})'=E(\frac{1}{\sigma^2}Auu'M')=E(AM)=0$$

其中，$AM=R(X'X)^{-1}X'M=0$（$MX=0$）。所以相关性为0，分子分母独立，Equation 1.13服从$F(J,n-k)$。

上述假设检验的思想在于检验实际的估计是否满足所施加的约束，但另一方面，施加了约束的模型的拟合优度通常会比没有约束的模型更低，因而，从拟合有度损失的角度，我们同样可以推导出F检验的等价形式。

等价形式. F检验可等价表示为（H_0：$R\beta=q$），

$$F=\frac{(u^{*\prime}u^*-u'u)/J}{u'u/(n-k)}=\frac{(SSR^*-SSR)/J}{SSR/(n-k)}$$

其中，带有（不带有）上标*表示参数由带有（不带有）约束的模型估计得到。

下面我们将推导上述F统计量的等价形式。首先，对于带有约束的模型而言，在最小化残差平方和的过程变为，

$$\min_{\beta}\quad (y-X\beta)'(y-X\beta)$$
$$s.t.\quad R\beta-q=0$$

有拉格朗日方程ℓ为，

$$\ell(\beta,\lambda)=(y-X\beta)'(y-X\beta)+2\underset{1\times J}{\lambda'}(R\beta-q)$$

故，一阶最优条件（FOC）为，

$$\frac{\partial\ell}{\partial\beta}=\begin{pmatrix}\frac{\partial\ell}{\partial\beta_1}\\\frac{\partial\ell}{\partial\beta_2}\\\vdots\\\frac{\partial\ell}{\partial\beta_k}\end{pmatrix}\quad\text{or}\quad\frac{\partial\ell}{\partial\beta'}=(\frac{\partial\ell}{\partial\beta_1},\frac{\partial\ell}{\partial\beta_2},\dots,\frac{\partial\ell}{\partial\beta_k})$$

保证$\frac{\partial \ell}{\partial \beta}$的维度和微分前一致，即与$\beta$（$k\times 1$）或者$\beta'$（$1\times k$）保持一致。对于带有约束的模型而言，

$$\begin{cases} \frac{\partial \ell}{\partial \beta}|_{\beta=\hat{\beta}^*} = -2\mathrm{X}'(y-\mathrm{X}\hat{\beta}^*)+2\mathrm{R}'\hat{\lambda}^* = 0 & (a) \\ \frac{\partial \ell}{\partial \lambda}|_{\lambda=\hat{\lambda}^*} = 2(\mathrm{R}\hat{\beta}-q) = 0 & (b) \end{cases} \tag{1.14}$$

进一步地，有

$$\begin{aligned} (a)\ in\ \text{Equation 1.14} &\Longrightarrow \mathrm{R}'\hat{\lambda}^* = \mathrm{X}'y-\mathrm{X}'\mathrm{X}\hat{\beta}^* \\ \text{上式两边同时左乘}\mathrm{R}(\mathrm{X}'\mathrm{X})^{-1} &\Longrightarrow \mathrm{R}(\mathrm{X}'\mathrm{X})^{-1}\mathrm{R}'\hat{\lambda}^* \\ &= \mathrm{R}(\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}'y-\mathrm{R}(\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}'\mathrm{X}\hat{\beta}^* \\ (y=\mathrm{X}\hat{\beta}+e) &= \mathrm{R}\hat{\beta}+\mathrm{R}(\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}'u-\mathrm{R}\hat{\beta}^* \\ &= \mathrm{R}(\hat{\beta}-\hat{\beta}^*) = \mathrm{R}\hat{\beta}-q \end{aligned}$$

解得$\hat{\lambda}^*$如下

$$\begin{aligned} \hat{\lambda}^* &= [\mathrm{R}(\mathrm{X}'\mathrm{X})^{-1}\mathrm{R}']^{-1}(\mathrm{R}\hat{\beta}-q) \\ \mathrm{R}'\hat{\lambda}^* &= \mathrm{R}'[\mathrm{R}(\mathrm{X}'\mathrm{X})^{-1}\mathrm{R}']^{-1}(\mathrm{R}\hat{\beta}-q) \\ \mathrm{R}'[\mathrm{R}(\mathrm{X}'\mathrm{X})^{-1}\mathrm{R}']^{-1}(\mathrm{R}\hat{\beta}-q) &= \mathrm{X}'y-\mathrm{X}'\mathrm{X}\hat{\beta}^* \\ &= \mathrm{X}'(\mathrm{X}\hat{\beta}+u)-\mathrm{X}'\mathrm{X}\hat{\beta}^* \\ &= \mathrm{X}'\mathrm{X}(\hat{\beta}-\hat{\beta}^*) \\ \Longrightarrow \hat{\beta}-\hat{\beta}^* &= (\mathrm{X}'\mathrm{X})^{-1}\mathrm{R}'[\mathrm{R}(\mathrm{X}'\mathrm{X})^{-1}\mathrm{R}']^{-1}(\mathrm{R}\hat{\beta}-q) \end{aligned}$$

又$\mathrm{SSR}^* = u^{*\prime}u^*$和$\mathrm{SSR} = u'u$，根据$\hat{\beta}-\hat{\beta}^*$，可得

$$\begin{aligned} u^* = y-\mathrm{X}\hat{\beta}^* &= u+\mathrm{X}(\hat{\beta}-\hat{\beta}^*) \\ u^{*\prime}u^* &= u'u+(\hat{\beta}-\hat{\beta}^*)'\mathrm{X}'\mathrm{X}(\hat{\beta}-\hat{\beta}^*) \\ (\text{代入}\hat{\beta}-\hat{\beta}^*) &= u'u+(\mathrm{R}\hat{\beta}-q)'[\mathrm{R}(\mathrm{X}'\mathrm{X})^{-1}\mathrm{R}']^{-1}(\mathrm{R}\hat{\beta}-q) \end{aligned}$$

所以，

$$\begin{aligned} &\frac{(\mathrm{SSR}^*-\mathrm{SSR})/\mathrm{J}}{\mathrm{SSR}/(n-k)} \\ =& \frac{(\mathrm{R}\hat{\beta}-q)'[\mathrm{R}(\mathrm{X}'\mathrm{X})^{-1}\mathrm{R}']^{-1}(\mathrm{R}\hat{\beta}-q)/\mathrm{J}}{\hat{u}'\hat{u}/(n-\mathrm{K})} \\ =& \frac{(\mathrm{R}\hat{\beta}-q)'[\mathrm{R}(\mathrm{X}'\mathrm{X})^{-1}\mathrm{R}']^{-1}(\mathrm{R}\hat{\beta}-q)/\mathrm{J}}{s^2} \end{aligned} \tag{1.15}$$

化简后Equation 1.13和Equation 1.15一致。

1.2 古典假设的放松

1.2.1 随机解释变量

在之前的所有讨论中，包括了$u \sim \mathrm{N}(0, \sigma^2)$、X非随机等假设，但由于数据的获取的方式往往具有随机性(比如调差问卷)，所以这些假设过于实际问题过于严格。因此，当数据是随机的时候，需要用到以下工具。

Theorem 1.2.1. *Law of Large Number (LLN)*

Let z_i be an i.i.d and $\mathrm{M} \times p$ matrix of observations, with $\mathrm{E}(z_i) = \mu$, assume $\mathrm{E}|z_i|^2$ is finite, then

$$\frac{1}{n}\sum_{i=1}^{n} z_i \xrightarrow{p} \mathrm{E}(z_i) = \mu$$

Theorem 1.2.2. *Central Limit Theory (CLT)*

Let z_i be an i.i.d and $\mathrm{M} \times 1$ vector of observations，with $\mathrm{E}(z_i) = \mu$ and $\mathrm{V}ar(z_i) = \Omega$, then

$$\mathrm{Z}_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(z_i - \mu) \xrightarrow{d} \mathrm{N}(0, \Omega)$$

其中，p和d分别代表依概率与依分布收敛。[3][4]LLN说明样本的均值依概率收敛到总体的均值，而CLT说明样本均值的抽样分布依分布收敛到正态分布。

Definition 1.2.1. *Converge in Mean Square Error（or say, Converge in Mean Square,$\xrightarrow{\mathrm{L}^2}$)*

A sequence of random variable x_n converges to a constant θ in mean square error (MSE), that is $x_n \xrightarrow{\mathrm{MSE}} \theta$, if

$$\lim_{n \to \infty} \mathrm{E}(x_n - \theta)^2 = 0$$

[3]依概率收敛的定义：$\forall \varepsilon > 0$, $\lim_{n\to\infty} Prob.(|x_n - x| > \varepsilon) = 0$，则称$x_n$依概率收敛于$x$，记作$x_n \xrightarrow{p} x$。

[4]依分布收敛的定义：若在$\mathrm{F}(x)$的所有连续点上，均有$\lim_{n\to\infty} |\mathrm{F}_n(x_n) - \mathrm{F}(x)| = 0$，则称$x_n$依分布收敛于$x$，记作$x_n \xrightarrow{d} x$。

对于MSE收敛而言，其同时要求随机变量的方差为零与随机变量无偏，

$$
\begin{aligned}
& \mathrm{E}(x_n-\theta)^2 \\
= & \mathrm{E}(x_n-\mathrm{E}(x_n)+\mathrm{E}(x_n)-\theta)^2 \\
= & \mathrm{E}[(x_n-\mathrm{E}x_n)^2+(\mathrm{E}x_n-\theta)^2+2(x_n-\mathrm{E}x_n)(\mathrm{E}x_n-\theta)] \\
= & \mathrm{E}(x_n-\mathrm{E}x_n)^2+(\mathrm{E}x_n-\theta)^2 \\
= & \mathrm{V}ar(x_n)+\mathrm{B}ias(x_n)^2
\end{aligned}
$$

故，极大收敛之间的关系大致如下：$\xrightarrow{\text{MSE}} \Longrightarrow \xrightarrow{p}(\text{consistency}) \Longrightarrow \xrightarrow{d}$。

在引入了渐近分布相关的性质后，我们开始推导在u不一定服从正态分布时，$\hat{\beta}$的渐近分布（渐近正态）。首先，从Equation 1.1中，我们可以得到所估参数为

$$
\begin{aligned}
\hat{\beta}^{\text{OLS}}-\beta &= (\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}'u \\
\sqrt{n}(\hat{\beta}^{\text{OLS}}-\beta) &= (\frac{1}{n}\mathrm{X}'\mathrm{X})^{-1}\frac{1}{\sqrt{n}}\mathrm{X}'u \\
&= (\frac{1}{n}\sum_i^n x_i x_i')^{-1}\frac{1}{\sqrt{n}}\sum_i^n x_i u_i \\
\text{LLN}: & \quad (\frac{1}{n}\sum_i^n x_i x_i')^{-1} \xrightarrow{p} \mathrm{E}[x_i x_i']^{-1} \\
\text{CLT}: & \quad \frac{1}{\sqrt{n}}\sum_i^n x_i u_i \overset{d}{\sim} \mathrm{N}(0,\sigma^2\mathrm{E}[x_i x_i']) \\
\sqrt{n}(\hat{\beta}_{\text{OLS}}-\beta) &\overset{d}{\sim} [\mathrm{E}(x_i x_i')]^{-1}\mathrm{N}(0,\sigma^2[\mathrm{E}(x_i x_i')])=\mathrm{N}(0,\sigma^2[\mathrm{E}(x_i x_i')]^{-1}) \\
\text{又}\sigma^2[\mathrm{E}(x_i x_i')]^{-1} &= \sigma^2[\frac{1}{n}\sum_i^n x_i x_i']^{-1} \\
&= n\sigma^2(\mathrm{X}'\mathrm{X})^{-1} \\
\hat{\beta}^{\text{OLS}} &\overset{d}{\sim} \mathrm{N}(\beta,\sigma^2(\mathrm{E}[x_i' x_i])^{-1})
\end{aligned}
$$

对于某一个系数$\hat{\beta}_j$而言，有

$$
\frac{\hat{\beta}_j-\beta_j}{\sqrt{\sigma^2(\mathrm{X}'\mathrm{X})^{-1}_{jj}}} \overset{d}{\sim} \mathrm{N}(0,1)
$$

分子分母趋于0的速度相同，所以商不为0，根据LLT和CLT，当$n\to\infty$，上式服从标准正态分布，而不再是小样本下的t分布。

1.2.2 球形扰动项

在经典假设下，随机扰动项的方差结构为

$$\mathrm{V}ar(u) = \mathrm{E}[u'u] = \sigma^2 \mathrm{I}_n = \begin{pmatrix} \sigma^2 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma^2 \end{pmatrix}_{n\times n}$$

放宽假设，如果$\sigma_1^2 \neq \sigma_2^2 \neq \ldots \neq \sigma_n^2 \neq \sigma^2$且对角线以外的元素不为零，会导致$\mathrm{V}ar(\hat{\beta}^{\mathrm{OLS}}) = \hat{\sigma}^2(\mathrm{X}'\mathrm{X})^{-1}$不再成立，即$\hat{\beta}^{\mathrm{OLS}}$不再有效，但仍满足无偏性$\mathrm{E}[\hat{\beta}^{\mathrm{OLS}}] = \beta$。

(1) **Heteroskadasticity (HET)** 需要估计n个参数

$$\begin{pmatrix} \sigma_1^2 & \dots & 0 \\ \vdots & \sigma_i^2 & \vdots \\ 0 & \dots & \sigma_n^2 \end{pmatrix}_{n\times n}$$

(2) **Serial correlation (AUTO)** 由某一个参数数量较少的表达式估计协方差矩阵

$$\begin{pmatrix} \sigma_1^2 & \sigma_{12}^2 & \dots & \sigma_{1n}^2 \\ \sigma_{21}^2 & \sigma_{22}^2 & \dots & \sigma_{2n}^2 \\ \vdots & \sigma_{ij}^2 & \ddots & \vdots \\ \sigma_{n1}^2 & \dots & \dots & \sigma_{nn}^2 \end{pmatrix}_{n\times n}$$

在不满足经典假设的情况下$\hat{\beta}^{\mathrm{OLS}}$不是最优的估计量(not best),所以可以使用GLS（Generalized Least Square）。对$y = \mathrm{X}\beta + u$两边同时左乘一个矩阵R，得到$\mathrm{R}y = \mathrm{RX}\beta + \mathrm{R}u$，使得新的随机误差项R$u$满足古典假设中的球形扰动项假设，但使用GLS的前提是估计$\mathrm{V}ar(u)$的结构。由于$\mathrm{V}ar(u)$一共有$n\times n$个参数，因此需要增加其他假设，减少参数的个数。

一个自然的想法是：协方差矩阵Σ是少量参数θ的函数，即，我们可以将方差σ_i^2考虑为与x_i的某些特征有关，通常我们乘这些特征为z_i（可能为x_i的函数）。

习惯上，我们设定$\sigma_i^2 = \sigma^2 \exp(z_i'\alpha)$，$\alpha$是$k\times 1$向量$k \ll n$，或者记为$\sigma_i^2 = \sigma_i^2(\theta), \theta(\sigma^2, \alpha)$。

假设$y_i \sim \mathrm{N}(x_i'\beta, \sigma_i^2(\theta, z_i))$，由极大似然估计MLE

$$\begin{aligned} \ln \mathscr{L}(\beta, \theta | x, y, z) &= \sum_i^n f(y_i | x_i, z_i, \beta_i, \theta) \\ &= -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\sum_{i=1}^n \ln \sigma_i^2(\theta) - \frac{1}{2}\sum_{i=1}^n \frac{(y_i - x_i'\beta)^2}{\sigma_i^2(\theta)} \end{aligned}$$

如果$\sigma_i^2(\theta) = \sigma^2$，则MLE等价于OLS。现在我们有MLE等价于以$\frac{1}{\sigma_i^2(\theta)}$为权重的GLS。

组间异方差（Groupwise HET）． 实际中，我们通常会遇到的异方差问题即组间异方差，该类异方差假设不同组别之间随机扰动项的方差结构不同，但组内是同方差的。对于*i.i.d*的数据而言（组别出现的顺序可变）：

$$\begin{pmatrix} \sigma_{\mathrm{B}}^2 & & & \\ & \sigma_{\mathrm{S}}^2 & & \\ & & \sigma_{\mathrm{B}}^2 & \\ & & & \ddots \end{pmatrix} \text{or} \begin{pmatrix} \sigma_{\mathrm{B}}^2 & & & \\ & \sigma_{\mathrm{B}}^2 & & \\ & & \sigma_{\mathrm{S}}^2 & \\ & & & \ddots \end{pmatrix}$$

若方差σ^2已知，则有β的估计为

$$\hat{\beta}^{\mathrm{GLS}} = [\sum_{g=1}^{\mathrm{G}} \frac{1}{\hat{\sigma}_g^2} \mathrm{X}_g' \mathrm{X}_g]^{-1} [\sum_{g=1}^{\mathrm{G}} (\frac{1}{\hat{\sigma}_g^2} \mathrm{X}_g' y_g)]$$

$$\mathrm{X} = \begin{pmatrix} \mathrm{X}_1 \\ \vdots \\ \mathrm{X}_{\mathrm{G}} \end{pmatrix}, \hat{\sigma}_g^2 = \frac{u_g' u_g}{n_g}$$

在$\sigma_i^2 = \sigma^2 \exp(z_i'\alpha)$中，Z是分组dummy Matrix（$x_i$属于哪一组，则Z的哪一列就等于1，否则等于0）

$$\mathrm{Z} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ & \vdots & & \end{pmatrix}_{\mathrm{G}\times\mathrm{G}}$$

假设R是变换矩阵（Transformation Matrix）

$$\begin{aligned} \mathrm{R}y &= \mathrm{RX}\beta + \mathrm{R}u \\ \Longrightarrow y^* &= \mathrm{X}^*\beta + u^* \\ \Longrightarrow \hat{\beta}^{\mathrm{WLS}} &= (\mathrm{X}^{*\prime}\mathrm{X}^*)^{-1}\mathrm{X}^{*\prime}y^* \\ &= (\mathrm{X}'\mathrm{R}'\mathrm{RX})^{-1}\mathrm{X}'\mathrm{R}'\mathrm{R}y, \quad \mathrm{R}'\mathrm{R} = \frac{1}{\hat{\sigma}_g^2} \end{aligned}$$

异方差稳健标准误（Heteroskadasticity Robust Standard Error）． 对于经典的OLS估计而言，有

$$\begin{aligned} \hat{\beta}^{\mathrm{OLS}} &= \beta + (\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}'u \\ \mathrm{V}ar(\hat{\beta}^{\mathrm{OLS}}) &= (\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}'\mathrm{E}uu'\mathrm{X}(\mathrm{X}'\mathrm{X})^{-1} \\ &= (\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}'\Sigma'\mathrm{X}(\mathrm{X}'\mathrm{X})^{-1} \end{aligned}$$

由于异方差问题，$\Sigma \neq \sigma^2 \mathrm{I}_n$，所以$\mathrm{V}ar(\hat{\beta}^{\mathrm{OLS}}) \neq \sigma^2(\mathrm{X}'\mathrm{X})^{-1}$。为得到$\mathrm{V}ar(\hat{\beta}^{\mathrm{OLS}})$，我们需要首先估计Σ的结构。但Σ仍旧有$n$个待估参数（对角线上的元素），所以直接估计$(\mathrm{X}'\Sigma\mathrm{X})_{k\times k}$Lfi一般有$k \ll n$。

$$
\begin{aligned}
\mathrm{X}'\Sigma\mathrm{X} &= \sum_i \sum_j \sigma_{ij} x_i x_j' \\
(\because \sigma_{ij} = 0) &= \sum_i \sigma_{ii} x_i x_i'
\end{aligned}
$$

又，

$$
\mathrm{E}(u_i^2 x_i x_i') = \mathrm{E}[\mathrm{E}(u_i^2 | x_i) x_i x_i'] = \mathrm{E}[\sigma_i^2 x_i x_i']
$$

当$n \to \infty$，有

$$
\frac{1}{n}\sum_{i=1}^n \hat{u}_i{}^2 x_i x_i' \to \frac{1}{n}\sum_{i=1}^n \sigma_i^2 x_i x_i'
$$

故，$\hat{\beta}^{\mathrm{OLS}}$方差的估计为

$$
\widehat{\mathrm{V}ar}(\hat{\beta}^{\mathrm{OLS}}) = (\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}' \begin{pmatrix} \hat{u_1^2} & & \\ & \ddots & \\ & & \hat{u_n^2} \end{pmatrix} \mathrm{X}(\mathrm{X}'\mathrm{X})^{-1}
$$

该文件标准误就是在Stata等软件中常用的，虽然看上去只是把Σ矩阵进行了替换,但由于$\hat{u}^2 \neq \hat{\sigma}^2$，所以本质上是替换了$\mathrm{X}'\Sigma\mathrm{X}$矩阵。

自相关. 关于自相关，考虑随机扰动项服从AR(1)，即

$$
\begin{aligned}
u_t &= \rho u_{t-1} + \varepsilon_t, \quad |\rho| < 1 \\
\Longrightarrow \mathrm{V}ar(u_t) &= \sigma_u^2 = \rho\sigma_u^2 + \sigma_\varepsilon^2 \\
&= \frac{\sigma_\varepsilon^2}{1-\rho}
\end{aligned}
$$

其中，ε_t为白噪声（white noise）。

由于u是一个AR(1)，故有，

$$
\mathrm{E}(u_t u_{t-1}) = \rho\sigma_u^2, \quad \mathrm{E}(u_t u_{t-2}) = \rho^2\sigma_u^2, \quad \dots
$$

$$\mathrm{E}(uu') = \frac{\sigma_\varepsilon^2}{1-\rho}\begin{pmatrix} 1 & \rho & \rho^2 & \dots \\ \rho & 1 & \rho & \vdots \\ \rho^2 & \rho & \ddots & \vdots \\ \vdots & \dots & \dots & 1 \end{pmatrix}$$

故，已知随机扰动项协方差矩阵Σ的结构，就可以找到变换矩阵用GLS估计，令变换矩阵

$$\mathrm{R} = \begin{pmatrix} \sqrt{1-\rho^2} & 0 & \dots & 0 \\ -\rho & 1 & \ddots & \vdots \\ 0 & -\rho & 1 & \vdots \\ \ddots & \dots & -\rho & 1 \end{pmatrix}，\text{则}u^* = \mathrm{R}u = \begin{pmatrix} \sqrt{1-\rho^2}u_1 \\ u_2-\rho u_1 \\ u_3-\rho u_2 \\ \vdots \\ u_n-\rho u_{n-1} \end{pmatrix}$$

其中，估计ρ时，首先得到$\hat{u}$序列，然后用$\hat{u}$来得到$\hat{\rho}$。

因此方程$y = \mathrm{X}\beta + u$转化为$\mathrm{R}y = \mathrm{R}x + \mathrm{R}u$, $\mathrm{V}ar(u)$是一个对角线以外元素不为0的方阵，而$\mathrm{V}ar(\mathrm{R}u) = \mathrm{R}\mathrm{V}ar(u)\mathrm{R}'$是一个对角阵，且对角线元素相等（满足球形扰动项假设）。

对于自相关问题，回到$\mathrm{V}ar(\hat{\beta})$的结构上，我们知道$\mathrm{V}ar(\hat{\beta}) = (\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}'\Sigma'\mathrm{X}(\mathrm{X}'\mathrm{X})^{-1}$，其中

$$\Sigma = \hat{u}\hat{u}' = \begin{pmatrix} \hat{u}_1{}^2 & \hat{u}_1\hat{u}_2 & \dots & \hat{u}_1\hat{u}_n \\ \hat{u}_2\hat{u}_1 & \hat{u}_2{}^2 & & \vdots \\ \vdots & & \ddots & \vdots \\ \dots & \dots & \dots & \hat{u}_n{}^2 \end{pmatrix}$$

在异方差问题中，我们直接将估计出的$diag = (\hat{u_1^2}, \hat{u_2^2}, \dots, \hat{u_n^2})$代入为$\hat{\Sigma}$，即可获得修正后的标准方差估计，但对于自相关问题而言，由于$(\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}'\hat{u}\hat{u}'\mathrm{X}(\mathrm{X}'\mathrm{X})^{-1} = 0$（因为$\mathrm{X}'\hat{u} = 0$）,所以不能简单地认为$\hat{\Sigma} = \hat{u}\hat{u}'$。

Newey-West Adjusted SE. Newey和West（1987）通过为协方差赋权的方式来对$\hat{u}\hat{u}'$进行调整，即为越靠近对角线的元素赋予更高的权重。

$$\begin{aligned} \text{Unadjusted: } \mathrm{X}'\hat{\Sigma}\mathrm{X} &= \frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}}\hat{u_t^2}x_t x_t' + \frac{1}{\mathrm{T}}\sum_{l=1}^{\mathrm{L}}\sum_{t=l+1}^{\mathrm{T}}\hat{u_t^2}\hat{u_{t-l}^2}(x_t x_{t-l}' + x_{t-l}x_t') \\ \text{NW-adjusted: } \mathrm{X}'\hat{\Sigma}'\mathrm{X} &= \frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}}\hat{u_t^2}x_t x_t' + \frac{1}{\mathrm{T}}\sum_{l=1}^{\mathrm{L}}\sum_{t=l+1}^{\mathrm{T}}w_l\hat{u_t^2}\hat{u_{t-l}^2}(x_t x_{t-l}' + x_{t-l}x_t') \end{aligned}$$

权重为

$$w_l = 1 - \frac{l}{\mathrm{L}+1}$$

其中，T为样本在时序上的维度，L为研究中所考虑的自相关阶数，通常取$T^{\frac{1}{4}}$。当$L = 0$时，Newey-West调整的标准误实际为 Huber–White标准误（HET consistent）。

1.2.3 内生变量

在之前的假设中，大样本放松了随机扰动项服从正态分布的假设，而关于异方差的讨论放松了球形扰动项的假设，本小节将进一步讨论如何放松u和x不相关的假设。首先，u和X不相关意味着：

$$E[u|X] = 0$$

当X和u相关时，此时$\hat{\beta} - \beta = (X'X)^{-1}X'u$是有偏的：

$$\begin{aligned}
\hat{\beta} - \beta &= (\frac{X'X}{n})^{-1}(\frac{X'u}{n}) \\
&= (\frac{1}{n}\sum x_i x_i')^{-1}(\frac{1}{n}x_i u_i) \\
E[\hat{\beta} - \beta] &= (E[x_i x_i'])^{-1}(E x_i u_i) \neq 0 \; when \; n \to \infty \\
&= 0 \; if \; E([x_i u_i]) = E([x_i \underbrace{E[u_i|x_i]}_{=0}]) = 0
\end{aligned}$$

因此，当$E[u|X] \neq 0$时，但$E[u_i|X_i] = 0$，$\hat{\beta}$虽然不是无偏的但是是一致的，即$\hat{\beta} - \beta \xrightarrow{p} 0$。

Unbiasness vs Consistency

- 考虑$E(y_t) = \mu$，估计量$\hat{\mu}_1 = \frac{1}{n+1}\sum_{t=1}^{n} y_t$，$E\hat{\mu}_1 = \frac{n}{n+1}\mu \neq 0$，但是$p\lim_{n\to\infty} \frac{n}{n+1}\frac{1}{n}\sum_{t=1}^{n} y_t = \mu$。
  - 有偏但是一致。
- 考虑$\hat{\mu}_2 = 0.01y_1 + \frac{0.99}{n-1}\sum_{t=2}^{n} y_t$，$E\hat{\mu}_2 = \mu$，但是$p\lim_{n\to\infty}\hat{\mu}_2 = 0.01y_1 + 0.99\mu \neq \mu$。
  - 是无偏但是不一致。

一致性要求$\forall i, (E[x_i u_i]) \to 0 \; when \; n \to \infty$，但是当存在内生变量时，$E[xu]$就不是零向量，因而会影响所有$\beta_j$的估计。

1.3 内生性问题

内生性问题出现的原因主要有三种：测量误差，互为因果，遗漏变量。接下来内容将逐个介绍三种问题。

1.3.1 测量误差（Measurement Error）

首先，在实际的数据收集过程中，由于各种原因，所收集的数据将会收到一系列的干扰，从而使得所收集的数据包含了某些噪声信息。对于真实模型：

$$y_t^0 = \beta_1 + \beta_2 x_t^0 + u_t^0 \tag{1.16}$$

其中，由于测量误差，实际数据（x_t与y_t）为

$$x_t = x_t^0 + v_{1t} \qquad y_t = y_t^0 + v_{2t}$$

故，实际模型为

$$\begin{aligned} y_t - v_{2t} &= \beta_1 + \beta_2(x_t - v_{1t}) + u_t^0 \\ \Longrightarrow y_t &= \beta_1 + \beta_2 x_t + [\underbrace{u_t^0 + v_{2t} - \beta_2 v_{1t}}_{=u_t}] \end{aligned}$$

此时，$\mathrm{V}ar(u_t) \geq \sigma^2(\mathrm{X}'\mathrm{X})^{-1}$，使得$\hat{\mathrm{V}ar}(\hat{\beta}^{\mathrm{OLS}})$变大，使得对$\hat{\beta}$的统计推断并不准确，即估计并不有效、存在方差更小的估计方法。由于测量误差，有x_t与u_t相关：

$$\mathrm{C}ov(x_t, u_t) = \mathrm{E}(x_t u_t) = \mathrm{E}[(x_t^0 + v_{1t})(u_t^0 + v_{2t} - \beta_2 v_{1t})] = -\beta \mathrm{V}ar(v_{1t})$$

举例而言，考虑收入和消费的关系：

$$\Uparrow\downarrow y_t = \beta_1 + \beta_2 x_t \Uparrow + u_t \downarrow \tag{1.17}$$

x_t增加会导致y_t增加$\Uparrow$，但由于内生性，所以x_t和u_t的负相关性导致u_t对y_t产生反方向的影响，使得低估了β_2（但不会使得符号相反）。

1.3.2 互为因果（Simultaneous Causality）

互为因果为最常见的内生性问题，一般直接的方法是使用工具变量来解决互为因果的问题，本文先介绍互为因果的问题，然后在介绍基于工具变量的回归方法。考虑经典的供求均衡问题，对于均衡时的成交数量而言，有

$$\begin{cases} q_t = \gamma_d p_t + x_t^d \beta_d + u_t^d \\ q_t = \gamma_s p_t + x_t^s \beta_s + u_t^s \end{cases} \tag{1.18}$$

但同时，对于均衡时的成交价格p_t, q_t与u_t^d, u_t^s而言，有

$$\begin{pmatrix} q_t \\ p_t \end{pmatrix} = \begin{pmatrix} 1 & -\gamma_d \\ 1 & -\gamma_s \end{pmatrix}^{-1} + \left[\begin{pmatrix} x_t^d & \beta_d \\ x_t^s & \beta_s \end{pmatrix} \begin{pmatrix} u_t^d \\ u_t^s \end{pmatrix} \right] \tag{1.19}$$

Equation 1.18和Equation 1.19表明价格与数量互相联系，并不能够直接形成所谓的因果推断，因此以下内容将引入工具变量来尝试解决内生性问题。

1.4 工具变量

IV. 工具变量是指与内生变量相关，但与被解释变量无关的变量（满足相关性与无关性两个条件），后续内容会介绍如何检验工具变量的相关性与无关性假设。因而，一个直观的解决内生性变量的方法为，将内生变量中与被解释变量无关的部分提取出来，使得该部分与残差项无关：

$$\begin{aligned} y &= \mathrm{X}_1\gamma_1 + \mathrm{X}_2\gamma_2 + u \\ &\Rightarrow \mathrm{X}_1\gamma_1 + \mathrm{X}_{\mathrm{IV}}\gamma_2 + u \end{aligned}$$

其中X_1是外生变量，X_2是内生变量，而$\rho(\mathrm{X}_{\mathrm{IV}}, u) = 0$。

2SLS. 实现提取内生变量中与被解释变量无关部分，并将其作为新的解释变量的方法叫做两阶段最小二乘（2-stage Least Square）。

$$\begin{aligned} \text{First Stage: } \hat{\mathrm{X}}_2 &= (\mathrm{X}_1, \mathrm{X}_{\mathrm{IV}}) \begin{pmatrix} \hat{\Pi}_1 \\ \hat{\Pi}_2 \end{pmatrix} = \mathrm{W} \begin{pmatrix} \hat{\Pi}_1 \\ \hat{\Pi}_2 \end{pmatrix} \\ \text{Second Stage: } y &= \mathrm{X}_1\gamma_1 + \hat{\mathrm{X}}_2\gamma_2 + u \end{aligned}$$

其中要求$\hat{\Pi}_2$显著（sigificant），通常来讲，$\mathrm{F} \leq 10$代表此时的工具变量为弱工具变量（weak IV）。又因为有

$$\begin{aligned} \hat{\mathrm{X}}_2 &= \mathrm{P_W X_2} \\ \mathrm{X}_1 &= \hat{\mathrm{X}}_1 = \mathrm{P}_w \mathrm{X}_1 \\ \Longrightarrow \hat{\mathrm{X}} &= (\hat{\mathrm{X}}_1, \hat{\mathrm{X}}_2) = \mathrm{P}_w \mathrm{X} \end{aligned}$$

所以，第二阶段的估计为：

$$
\begin{aligned}
y &= \mathrm{X}_1\gamma_1 + \hat{\mathrm{X}}_2\gamma_2 + u \\
&= (\hat{\mathrm{X}}_1, \hat{\mathrm{X}}_2)\begin{pmatrix}\gamma_1 \\ \gamma_2\end{pmatrix} + u \\
&= \hat{\mathrm{X}}\delta + u
\end{aligned}
$$

故，最小二乘估计量为

$$
\begin{aligned}
\hat{\delta}^{2\mathrm{SLS}} &= (\hat{\mathrm{X}}'\hat{\mathrm{X}})^{-1}\hat{\mathrm{X}}'y \\
&= (\mathrm{X}'\mathrm{P}_\mathrm{W}'\mathrm{P}_\mathrm{W}\mathrm{X})^{-1}\mathrm{X}'\mathrm{P}_\mathrm{W}'y \\
&= (\mathrm{X}'\mathrm{P}_\mathrm{W}\mathrm{X})^{-1}\mathrm{X}'\mathrm{P}_\mathrm{W}y
\end{aligned}
$$

工具变量法的检验.

1. 判断是否存在内生性，是否需要工具变量

2. 工具变量与内生变量是否有足够强的关系，F-test，F>10 in first stage

3.X_{IV}不能直接影响y，即不能直接影响u（若X_{IV}影响y，且未出现在$(\mathrm{X}_1, \mathrm{X}_2)$里，则必定与$u$相关）。

Overidentifying Restrictions Test (Sargan Test). 用来判断工具变量是否是外生的。由于检验存在局限性，所以当直觉和检验出现矛盾的时候一般还是更依赖于直觉。对于工具变量的外生性而言，检验假设为

$$
\begin{aligned}
&\mathrm{H}_0 : \mathrm{E}[\mathrm{W}'u] = 0 \quad \mathrm{W} = (\mathrm{X}_1, \mathrm{X}_{\mathrm{IV}}) \\
&\mathrm{H}_1 : \mathrm{E}[\mathrm{W}'u] \neq 0
\end{aligned}
$$

考虑回归模型

$$
\begin{aligned}
y &= \mathrm{X}_1\beta_1 + \mathrm{X}_2\beta_2 + u \\
\mathrm{W} &= (\mathrm{X}_1, \mathrm{X}_{\mathrm{IV}})
\end{aligned}
$$

因此，我们仅需要估计出$\hat{u}$，然后计算$\mathrm{E}[\mathrm{W}'\hat{u}]$，即可检验原假设，利用2SLS估计：

$$
y = \mathrm{X}_1\hat{\beta}_1^{2\mathrm{SLS}} + \mathrm{X}_2\hat{\beta}_2^{2\mathrm{SLS}} + \hat{u}
$$

将$\hat{u}$回归到W上，即$\hat{u} = \mathrm{W}\hat{b} + e$，可以计算出$\mathrm{R}^2$，当$\mathrm{R}^2$越高时，说明W和$\hat{u}$之间的相关性越强。但是由

于R^2的分布不能直接查表判断，所以要构造包含R^2的标准分布，考虑LR检验，计算$n\mathrm{R}^2$：

$$\text{LR test: } n\mathrm{R}^2 = n\frac{\mathrm{SSE}}{\mathrm{SST}} = \frac{\hat{b}'\mathrm{W}'\mathrm{W}\hat{b}}{\frac{1}{n}\hat{u}'\hat{u}} \tag{1.20}$$

根据之前的笔记$\hat{\sigma}^2 = \frac{\hat{u}'\hat{u}}{n-k}$, $y = \mathrm{X}\beta + u$, $\mathrm{M}y = \mathrm{M}u = \hat{u}$, $\mathrm{P}y = \hat{y}$, $y = \hat{y} + \hat{u}$，所以$\mathrm{W}\hat{b} = \mathrm{P_W}\hat{u}$，Equation 1.20可以改写为：

$$\frac{\hat{b}'\mathrm{W}'\mathrm{W}\hat{b}}{\frac{1}{n}\hat{u}'\hat{u}} = \frac{\hat{u}'\mathrm{P}_\mathrm{W}'\mathrm{P_W}\hat{u}}{\frac{1}{n}\hat{u}'\hat{u}} \sim \chi^2(q)$$

接下来将计算χ^2分布的自由度q，首先，

$$\begin{aligned}
\hat{u} &= y - \mathrm{X}\hat{\beta}_{2\mathrm{SLS}} \\
&= [\mathrm{I} - \mathrm{X}(\mathrm{X}'\mathrm{P_W}\mathrm{X})^{-1}\mathrm{X}'\mathrm{P_W}](\mathrm{X}\beta + u) \\
&= [\mathrm{I} - \mathrm{X}(\mathrm{X}'\mathrm{P_W}\mathrm{X})^{-1}\mathrm{X}'\mathrm{P_W}]u
\end{aligned}$$

其中，$\mathrm{W}_{n\times l} = (\mathrm{X}_1, \mathrm{X}_{\mathrm{IV}})$，$\mathrm{X}_{n\times k} = (\mathrm{X}_1, \mathrm{X}_2)$，$\hat{\beta}^{2\mathrm{SLS}} = (\mathrm{X}'\mathrm{P_W}\mathrm{X})^{-1}\mathrm{X}'\mathrm{P_W}y$，故有

$$\begin{aligned}
\hat{u}'\mathrm{P_W}\hat{u} &= u'[\mathrm{I} - \mathrm{P_W}\mathrm{X}(\mathrm{X}'\mathrm{P_W}\mathrm{X})^{-1}\mathrm{X}']\mathrm{P_W}[\mathrm{I} - \mathrm{X}(\mathrm{X}'\mathrm{P_W}\mathrm{X})^{-1}\mathrm{X}'\mathrm{P_W}]u \\
&= u'[\mathrm{P_W} - \mathrm{P_W}\mathrm{X}(\mathrm{X}'\mathrm{P_W}\mathrm{X})^{-1}\mathrm{P_W}]u \sim \chi^2(q)
\end{aligned}$$

而q取决于矩阵$[\mathrm{P_W} - \mathrm{P_W}\mathrm{X}(\mathrm{X}'\mathrm{P_W}\mathrm{X})^{-1}\mathrm{P_W}]$的迹（trace），

$$\begin{aligned}
tr(\mathrm{P_W} - \mathrm{P_W}\mathrm{X}(\mathrm{X}'\mathrm{P_W}\mathrm{X})^{-1}\mathrm{P_W}) &= tr(\mathrm{P_W}) - tr(\mathrm{P_W}\mathrm{X}(\mathrm{X}'\mathrm{P_W}\mathrm{X})^{-1}\mathrm{X}'\mathrm{P_W}) \\
&= tr(\mathrm{W}(\mathrm{W}'\mathrm{W})^{-1}\mathrm{W}') - tr(\mathrm{P_W}\mathrm{X}(\mathrm{X}'\mathrm{P_W}\mathrm{X})^{-1}\mathrm{X}'\mathrm{P_W}) \\
&= tr(\mathrm{W}'\mathrm{W})^{-1}\mathrm{W}'\mathrm{W}) - tr((\mathrm{X}'\mathrm{P_W}\mathrm{X})^{-1}\mathrm{X}'\mathrm{P_W}\mathrm{P_W}\mathrm{X}) \\
&= tr(\mathrm{I}_l) - tr(\mathrm{I}_k) = l - k
\end{aligned}$$

$$\hat{u}'\mathrm{P_W}\hat{u} \sim \chi^2(l-k) \tag{1.21}$$

其中l是W含有变量的个数，k是X含有变量的个数。χ^2分布存在要求自由度$l-k$大于0，也即工具变量的个数大于内生变量的个数，所以又叫Overidentifying Restriction Test，必须要在过度识别的情况下才能检验（如果恰好识别则χ^2的自由度等于0，则不能检验）。

Hausman Test. Hausman Test旨在比较两个估计量的准度与效率的问题，尝试找到两个不同的估计量，一个估计量在H_0和H_1下都是一致的，另一个在H_1下不一致，但在H_0下是一致且有效的。

考虑假设：

$\mathrm{H_0}$：OLS is consistent, 2SLS is consistent ,but *not* efficient

$\mathrm{H_1}$：OLS is *not* consistent, 2SLS is consistent

因此想法将$\hat{\beta}^{\mathrm{OLS}}-\hat{\beta}^{\mathrm{2SLS}}$变成一个标准的分布，记$\hat{\beta}^{\mathrm{IV}}=\hat{\beta}^{\mathrm{2SLS}}$ 已知

$$\hat{\beta}^{\mathrm{OLS}}=(\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}'y$$
$$\hat{\beta}^{\mathrm{IV}}=(\mathrm{X}'\mathrm{P_W X})^{-1}\mathrm{X}'\mathrm{P_W}y$$

$$\begin{aligned}\hat{\beta}^{\mathrm{IV}}-\hat{\beta}^{\mathrm{OLS}}&=(\mathrm{X}'\mathrm{P_W X})^{-1}\mathrm{X}'\mathrm{P_W}y-(\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}'y\\&=(\mathrm{X}'\mathrm{P_W X})^{-1}\mathrm{X}'\mathrm{P_W}(\mathrm{X}\hat{\beta}_{\mathrm{OLS}}+\hat{u}_{\mathrm{OLS}})-\hat{\beta}_{\mathrm{OLS}}\\&=(\mathrm{X}'\mathrm{P_W X})^{-1}\mathrm{X}'\mathrm{P_W M_X}y\\&=(\mathrm{X}'\mathrm{P_W X})^{-1}\mathrm{X}'\mathrm{P_W M_X}u\end{aligned}\tag{1.22}$$

转换成含u的表达式后，思路是找χ^2

$$test=(\hat{\beta}_{\mathrm{IV}}-\hat{\beta}_{\mathrm{OLS}})'[\mathrm{V}ar(\hat{\beta}_{\mathrm{IV}}-\hat{\beta}_{\mathrm{OLS}})]^{-1}(\hat{\beta}_{\mathrm{IV}}-\hat{\beta}_{\mathrm{OLS}})\tag{1.23}$$
$$=\chi^2(?)\tag{1.24}$$

由于$\mathrm{V}ar(\hat{\beta}_{\mathrm{IV}}-\hat{\beta}_{\mathrm{OLS}})$不可逆,所以$\chi^2$分布的自由度不是k，考虑

$$\begin{aligned}\hat{\beta}_{\mathrm{IV}}-\hat{\beta}_{\mathrm{OLS}}&=(\mathrm{X}'\mathrm{P_W X})^{-1}\begin{pmatrix}\mathrm{X}_1'\\\mathrm{X}_2'\end{pmatrix}\mathrm{P_W M_X}u\\&=(\mathrm{X}'\mathrm{P_W X})^{-1}\begin{pmatrix}\mathrm{X}_1'\mathrm{P_W M_X}u\\\mathrm{X}_2'\mathrm{P_W M_X}u\end{pmatrix}\qquad(\mathrm{X}_1'\mathrm{P_W}=\mathrm{X}_1,\mathrm{X}_1\mathrm{M_X}=0)\\&=(\mathrm{X}'\mathrm{P_W X})^{-1}\begin{pmatrix}0\\\mathrm{X}_2'\mathrm{P_W M_X}u\end{pmatrix}\end{aligned}\tag{1.25}$$

$$\begin{aligned}\mathrm{V}ar(\hat{\beta}_{\mathrm{IV}}-\hat{\beta}_{\mathrm{OLS}})&=\mathrm{E}\left[(\mathrm{X}'\mathrm{P_W X})^{-1}\begin{pmatrix}0\\\mathrm{X}_2'\mathrm{P_W M_X}u\end{pmatrix}(0'\,u'\mathrm{M_X P_W X_2})(\mathrm{X}'\mathrm{P_W X})\right]\\&=\mathrm{E}\left[(\mathrm{X}'\mathrm{P_W X})^{-1}\begin{pmatrix}0&0\\0&\mathrm{X}_2'\mathrm{P_W M_X}uu'\mathrm{M_X P_W X_2}\end{pmatrix}(\mathrm{X}'\mathrm{P_W X})^{-1}\right]\end{aligned}\tag{1.26}$$

如果，$\mathrm{E}(uu') = \sigma^2\mathrm{I}$,并且将中间矩阵的左上角部分记为A,得到

$$\mathrm{V}ar(\hat{\beta}_{\mathrm{IV}} - \hat{\beta}_{\mathrm{OLS}}) = \mathrm{E}\left[(\mathrm{X}'\mathrm{P_W}\mathrm{X})^{-1}\begin{pmatrix}\mathrm{A} & 0 \\ 0 & \mathrm{X}_2'\mathrm{P_W}\mathrm{M_X}\sigma^2\mathrm{M_X}\mathrm{P_W}\mathrm{X}_2\end{pmatrix}(\mathrm{X}'\mathrm{P_W}\mathrm{X})^{-1}\right] \tag{1.27}$$

$$\mathrm{V}ar(\hat{\beta}_{\mathrm{IV}} - \hat{\beta}_{\mathrm{OLS}})^{-1} = \mathrm{E}\left[(\mathrm{X}'\mathrm{P_W}\mathrm{X})\begin{pmatrix}\mathrm{A} & 0 \\ 0 & \mathrm{X}_2'\mathrm{P_W}\mathrm{M_X}\sigma^2\mathrm{M_X}\mathrm{P_W}\mathrm{X}_2\end{pmatrix}^{-1}(\mathrm{X}'\mathrm{P_W}\mathrm{X})\right] \tag{1.28}$$

$$\begin{aligned} test &= (0'\, u'\mathrm{M_X}\mathrm{P_W}\mathrm{X}_2)(\mathrm{X}'\mathrm{P_W}\mathrm{X})^{-1}\mathrm{E}\left[(\mathrm{X}'\mathrm{P_W}\mathrm{X})\begin{pmatrix}\mathrm{A}^{-1} & 0 \\ 0 & (\mathrm{X}_2'\mathrm{P_W}\mathrm{M_X}\sigma^2\mathrm{M_X}\mathrm{P_W}\mathrm{X}_2)^{-1}\end{pmatrix}(\mathrm{X}'\mathrm{P_W}\mathrm{X})\right] \\ &\quad (\mathrm{X}'\mathrm{P_W}\mathrm{X})^{-1}\begin{pmatrix}0 \\ \mathrm{X}_2'\mathrm{P_W}\mathrm{M_X}u\end{pmatrix} \\ &= (0'\, u'\mathrm{M_X}\mathrm{P_W}\mathrm{X}_2)\begin{pmatrix}\mathrm{A}^{-1} & 0 \\ 0 & (\mathrm{X}_2'\mathrm{P_W}\mathrm{M_X}\sigma^2\mathrm{M_X}\mathrm{P_W}\mathrm{X}_2)^{-1}\end{pmatrix}\begin{pmatrix}0 \\ \mathrm{X}_2'\mathrm{P_W}\mathrm{M_X}u\end{pmatrix} \\ &= u'\mathrm{M_X}\mathrm{P_W}\mathrm{X}_2(\mathrm{X}_2'\mathrm{P_W}\mathrm{M_X}\sigma^2\mathrm{M_X}\mathrm{P_W}\mathrm{X}_2)^{-1}\mathrm{X}_2'\mathrm{P_W}\mathrm{M_X}u \\ &= \frac{1}{\sigma^2}u'\mathrm{M_X}\mathrm{P_W}\mathrm{X}_2(\mathrm{X}_2'\mathrm{P_W}\mathrm{M_X}\mathrm{M_X}\mathrm{P_W}\mathrm{X}_2)^{-1}\mathrm{X}_2'\mathrm{P_W}\mathrm{M_X}u \end{aligned} \tag{1.29}$$

所以A是否为0并不重要

$$tr(\mathrm{M_X}\mathrm{P_W}\mathrm{X}_2(\mathrm{X}_2'\mathrm{P_W}\mathrm{M_X}\mathrm{M_X}\mathrm{P_W}\mathrm{X}_2)^{-1}\mathrm{X}_2'\mathrm{P_W}\mathrm{M_X}) = tr(\mathrm{X}_2'\mathrm{P_W}\mathrm{M_X}\mathrm{M_X}\mathrm{P_W}\mathrm{X}_2)^{-1}\mathrm{X}_2'\mathrm{P_W}\mathrm{M_X}\mathrm{M_X}\mathrm{P_W}\mathrm{X}_2) = k_2 \tag{1.30}$$

因此，Hausman Test

$$test \sim \chi^2(k_2) \tag{1.31}$$

第二章 广义矩估计

2.1 广义矩估计

Method of Moment MM/ Generalized Method of Moment GMM矩估计或者广义矩估计也是一种类似LS,2SLS,MLE的估计方法。其指导思想是先找到总体矩条件，再找样本矩条件(根据大数定理)，然后根据矩条件解方程计算未知参数。

$$\mathrm{E}(functionsofr.v.andpar) = 0 \tag{2.1}$$

就是poplution moment condition/restriciton(POPMC)

e.g. 总体矩条件$\mathrm{E}(\mathrm{X}-\mu) = 0$ 对应的样本矩条件sample moment condition(SMC)为$\frac{1}{n}\sum_{i=1} n(x_i-\mu) = 0$即$\bar{x}-\mu = 0 \Rightarrow \hat{\mu} = \bar{x}$

example 1

$$y = \mathrm{X}\beta + u \tag{2.2}$$

$$\mathrm{EX}'u = 0 \tag{2.3}$$

这个就是总体矩条件，(如果 $\mathrm{EX}'u \neq 0$可以用$\mathrm{EW}'u = 0$,见example 2)

$$\mathrm{EX}'u = 0 \tag{2.4}$$

$$\Rightarrow \mathrm{E}\sum_{i=1}^{n} x_i u_i = 0 \tag{2.5}$$

$$\Rightarrow \sum_{i=1}^{n} \mathrm{E}x_i u_i = 0 \tag{2.6}$$

$$\Rightarrow \cancel{n}\mathrm{E}x_i u_i = 0 \tag{2.7}$$

$$\Rightarrow \frac{1}{n}\sum_{i=1}^{n} = 0 \quad \mathrm{X}'u = 0 \tag{2.8}$$

即得到样本矩条件

$$X'(y-X\beta)=0 \tag{2.9}$$

$$X'y-X'X\beta=0 \tag{2.10}$$

$$\Rightarrow \hat{\beta}_{MM}=(X'X)^{-1}X'y=\hat{\beta}_{OLS} \tag{2.11}$$

example 2 $E(Xu)\neq 0$

$$y=X_1\beta+X_2\beta+u \tag{2.12}$$

$EX'u\neq 0$ but $EW'u=0$ 所以样本矩条件为

$$W'u=0 \tag{2.13}$$

$$W'(y-X\beta)=0 \tag{2.14}$$

$$\hat{\beta}_{MM}=(W'X)^{-1}W'y \tag{2.15}$$

W是$n\times l$的矩阵X是$n\times k$，所以有$l=k$

考虑$\hat{\beta}_{2SLS}$

$$\begin{aligned}\hat{\beta}_{2SLS}&=(X'P_WX)^{-1}X'P_Wy\\&=(X'W(W'W)^{-1}W'X)^{-1}XW(W'W)^{-1}W'y\end{aligned} \tag{2.16}$$

由于X′W,WW,W′X都是方阵，所以

$$(X'W(W'W)^{-1}W'X)^{-1}=(W'X)^{-1}(W'W)(X'W)^{-1} \tag{2.17}$$

$$\Rightarrow \hat{\beta}_{2SLS}=(W'X)^{-1}(W'W)(X'W)^{-1}XW'(W'W)^{-1}W'y \tag{2.18}$$

$$\Rightarrow \hat{\beta}_{2SLS}=(W'X)^{-1}W'y \tag{2.19}$$

$$X'u=0 \quad \hat{\beta}_{MM}=(X'X)^{-1}X'y=\hat{\beta}_{OLS}$$

$$W'u=0 \quad \hat{\beta}_{MM}=(W'X)^{-1}W'y=\hat{\beta}_{2SLS}$$

$$\max_{\theta} \mathrm{E} lnf(x,y|\theta) \Rightarrow \theta_0$$
$$\mathrm{E}(\frac{\partial lnf(x,y|\theta_0)}{\partial \theta}) = 0 \quad (popmc)$$
$$\frac{1}{n}\sum_{i=1}^{n}\frac{\partial lnf(x_i,y_i|\hat{\theta}_{\mathrm{MLE}})}{\partial \theta} = 0$$

GMM

$$g(z,\theta) = \begin{pmatrix} g_1(z,\theta) \\ g_2(z,\theta) \\ \vdots \\ g_{\mathrm{L}}(z,\theta) \end{pmatrix}_{\mathrm{L}\times 1} \tag{2.20}$$

$$\hat{g}_n(\theta) = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n} g_1(z_i,\theta) \\ \frac{1}{n}\sum_{i=1}^{n} g_2(z_i,\theta) \\ \vdots \\ \frac{1}{n}\sum_{i=1}^{n} g_{\mathrm{L}}(z_i,\theta) \end{pmatrix}_{\mathrm{L}\times 1} = \frac{1}{n}\sum_{i=1}^{n} g(z,\theta) \tag{2.21}$$

k: # of unknown parameters (未知参数的个数)

L: # of unknown independent restrictions (矩条件的个数)

$\mathrm{L} > k$ GMM

$\mathrm{L} = k$ MM

$\mathrm{X}'_{k\times u} u_{n\times 1} = 0_{k\times 1}$ $\beta_{k\times 1}$ $\mathrm{L} = k$ 用MM

如果$\mathrm{L} > k$ 不能找到满足所有矩条件都为0的参数，即$\hat{g}_n(\theta) \neq 0$

考虑GMM的目标函数

$$\mathrm{Q}_n^{\mathrm{W}}(\theta) = \underset{1\times \mathrm{L}}{\hat{g}_n(\theta)'} \underset{\mathrm{L}\times \mathrm{L}}{\mathrm{W}} \underset{\mathrm{L}\times 1}{\hat{g}_n(\theta)} \tag{2.22}$$

其中W是任意的正定矩阵，所以有

$$\hat{\theta}_{\mathrm{GMM}} = argmin\, \mathrm{Q}_n^{\mathrm{W}}(\theta) \tag{2.23}$$

$\mathrm{E}(r.v., para) = 0$ $\frac{1}{n}\sum() = \hat{g}_n(\theta)$

需要找到$\hat{\theta}_{\mathrm{GMM}}$的分布($\hat{\beta}_{\mathrm{OLS}} \sim \mathrm{N}(\beta, \sigma^2(\mathrm{X}'\mathrm{X})^{-1})$)

Theorem 2.1.1. ***Asymptotic normality of GMM Estimator***

Under appropriate conditions, we have

$$\sqrt{n}(\hat{\theta}_{\mathrm{GMM}}-\theta_0) \xrightarrow{d} \mathrm{N}(0,(\mathrm{G}'\mathrm{W}\mathrm{G})^{-1}\mathrm{G}'\mathrm{W}\Omega_0\mathrm{W}\mathrm{G}(\mathrm{G}'\mathrm{W}\mathrm{G})^{-1}) \tag{2.24}$$

where $\mathrm{G}(\theta)=\mathrm{E}(\nabla_\theta g(z,\theta)), \mathrm{G}=\mathrm{G}(\theta_0)_{\mathrm{L}\times k}, \Omega_0=\mathrm{E}[\underset{\mathrm{L}\times 1}{g(z,\theta_0)}\underset{1\times\mathrm{L}}{g(z,\theta_0)'}]$ *and,*

$$\nabla_\theta \underset{\mathrm{L}\times 1}{g(z,\theta)}=\frac{\partial g(z,\theta)}{\partial\theta'}=\begin{pmatrix}\frac{\partial g_1}{\partial\theta_1} & \cdots & \frac{\partial g_1}{\partial\theta_k}\\ \vdots & & \vdots\\ \frac{\partial g_\mathrm{L}}{\partial\theta_1} & \cdots & \frac{\partial g_\mathrm{L}}{\partial\theta_k}\end{pmatrix}_{\mathrm{L}\times k} \tag{2.25}$$

$$\nabla_{\theta'} g=\frac{\partial g'}{\partial\theta}=\left(\frac{\partial g}{\partial\theta}\right)'_{k\times\mathrm{L}} \tag{2.26}$$

考虑$\hat{\beta}_{\mathrm{OLS}}$的证明过程，利用$\hat{\beta}=argmin\,\mathrm{SSR}$先写出$\hat{\beta}$的表达式，同理利用$\hat{\theta}_{\mathrm{GMM}}=argmin\,\mathrm{Q}_n^{\mathrm{W}}(\theta)$

Proof

$$\mathrm{F.O.C}:\frac{\partial\mathrm{Q}_n^{\mathrm{W}}(\hat{\theta})}{\partial\theta}=\underset{k\times\mathrm{L}}{\frac{\partial\hat{g}'_{1\times\mathrm{L}}}{\partial\theta_{k\times 1}}}\mathrm{W}_{\mathrm{L}\times\mathrm{L}}\hat{g}_{\mathrm{L}\times 1}+\Bigg(\hat{g}'\mathrm{W}\underset{\substack{1\times\mathrm{L}\\ \mathrm{L}\times 1}}{\frac{\partial\hat{g}_{\mathrm{L}\times 1}}{\partial\theta'_{1\times k}}}\Bigg)' \tag{2.27}$$

$$=2\underset{k\times\mathrm{L}}{\nabla_{\theta'}\hat{g}_n(\hat{\theta})}\underset{\mathrm{L}\times\mathrm{L}}{\mathrm{W}}\underset{\mathrm{L}\times 1}{\hat{g}_n(\hat{\theta})}=0 \tag{2.28}$$

但是该一阶条件并不能显式求解$\hat{\theta}$,因此考虑Taylor Expansion

$$\hat{g}_n(\hat{\theta})=\hat{g}_n(\theta_0)+\nabla_\theta\hat{g}_n(\bar{\theta})(\hat{\theta}-\theta_0) \tag{2.29}$$

可以得到

$$\frac{\partial\hat{\mathrm{Q}}(\hat{\theta})}{\partial\theta}=2\nabla_{\theta'}\hat{g}_n\mathrm{W}[\hat{g}_n(\theta_0)+\nabla_\theta\hat{g}_n(\bar{\theta})(\hat{\theta}-\theta_0)]=0 \tag{2.30}$$

$$\sqrt{n}(\hat{\theta}-\theta_0)=-[\nabla_{\theta'}\hat{g}_n\mathrm{W}\nabla_\theta\hat{g}_n(\bar{\theta})]^{-1}\nabla_{\theta'}\hat{g}_n\mathrm{W}\sqrt{n}\nabla_\theta\hat{g}_n(\theta_0) \tag{2.31}$$

$$\nabla_{\theta'}\hat{g}_n(\hat{\theta})=\nabla_{\theta'}\frac{1}{n}\sum_{i=1}^n g(z_i,\theta)\to\mathrm{E}\nabla_{\theta'}g(z_i,\hat{\theta})=\mathrm{G}'(\hat{\theta})\to\mathrm{G}'(\theta_0) \tag{2.32}$$

$$\therefore\sqrt{n}(\hat{\theta}-\theta_0)=-[(\mathrm{G}'+o_p(1))\ldots]^{-1}\ldots \tag{2.33}$$

Review $\mathrm{O}(1), o(1), \mathrm{O}_p(1), o_p(1)$

$$\frac{1}{n} = o(1) \quad \mathrm{C} = \mathrm{O}(1)$$
$$a_n = \mathrm{O}(b_n) \quad \frac{a_n}{b_n} = \mathrm{O}(1)$$
$$a_n = o(b_n) \quad \frac{a_n}{b_n} = o(1)$$
$$\frac{1}{n^2} = o(\frac{1}{n}) \quad \frac{\frac{1}{n^2}}{\frac{1}{n}} = o(1)$$

$$\sqrt{n}(\hat{\theta} - \theta_0) = -[\mathrm{G}'\mathrm{W}\mathrm{G}]^{-1}\mathrm{G}'\mathrm{W}\sqrt{n}\hat{g}_n(\theta_0) \tag{2.34}$$

$$\sqrt{n}\hat{g}_n(\theta_0) = \sqrt{n}\frac{1}{n}\sum_{i=1}^{n} g(z_i, \theta_0) \sim \mathrm{N}(0, \mathrm{E}gg') = \mathrm{N}(0, \Omega_0) \tag{2.35}$$

$$\therefore \sqrt{n}(\hat{\theta} - \theta_0) = \mathrm{N}(0, (\mathrm{G}'\mathrm{W}\mathrm{G})^{-1}\mathrm{G}'\mathrm{W}\Omega_0\mathrm{W}\mathrm{G}(\mathrm{G}'\mathrm{W}\mathrm{G})^{-1}) \tag{2.36}$$

如果$\hat{\theta}_{\mathrm{GMM}} \nrightarrow \theta_0$ 不能使用Taylor Expansion,根据Uniform Weak Law of Large Number,if $\hat{\mathrm{Q}}_n(\theta) \to \mathrm{Q}_0(\theta)$,then $\hat{\theta}_{\mathrm{GMM}} = argmin\hat{\mathrm{Q}}_n(\theta) \to \theta_0 = argmin\mathrm{Q}_0(\theta)$

$$\hat{\mathrm{Q}}_n(\theta) = \hat{g}_n(\theta)'\mathrm{W}\hat{g}_n(\theta) = [\frac{1}{n}\sum_{i=1}^{n} g(z_i, \theta)]'\mathrm{W}[\frac{1}{n}\sum_{i=1}^{n} g(z_i, \theta)] \tag{2.37}$$

$$\mathrm{Q}_0(\theta) = \mathrm{E}g(z, \theta)'\mathrm{W}\mathrm{E}g(z, \theta) \tag{2.38}$$

令$\mathrm{W} = \Omega_0^{-1}$,有

$$(\mathrm{G}'\mathrm{W}\mathrm{G})^{-1}\mathrm{G}'\mathrm{W}\Omega_0\mathrm{W}\mathrm{G}(\mathrm{G}'\mathrm{W}\mathrm{G})^{-1} = (\mathrm{G}'\Omega_0^{-1}\mathrm{G})^{-1} \tag{2.39}$$

$\mathrm{W} = \Omega_0^{-1}$被称为 optimal weighting matrix,$(\mathrm{G}'\Omega_0^{-1}\mathrm{G})^{-1}$被称为optimal variance,$\hat{\theta}_{\mathrm{GMM}}$被称为 efficient GMM estimator。但由于$\mathrm{W} = \Omega_0^{-1}$是infeasible的,所以需要其他feasible的方法来进行估计。

1. two step feasible efficient GMM

step 1 estimate Ω_0 by $\hat{\Omega}$

$$\hat{\Omega} = \frac{1}{n}\sum_{i=1}^{n} g(z_i, \hat{\theta})g(z_i, \hat{\theta})' \tag{2.40}$$

$$\tilde{\theta} = \arg\min_{\theta \in \text{Ⓗ}}[\hat{g}_n'](\theta)\hat{g}_n(\theta)] = \arg\min_{\theta \in \mathrm{H}} \mathrm{Q}_n^{\mathrm{I}}(\theta) \tag{2.41}$$

因为无论取什么样的W，$\tilde{\theta}$都是一致的，所以取$\mathrm{W} = \mathrm{I}$

step 2 $\tilde{\theta}^*_{\text{GMM}} = \arg\min \hat{g}'_n(\theta)[\hat{\Omega}(\hat{\theta})]^{-1}\hat{g}_n(\theta)$

$$\mathrm{V}ar(\sqrt{n}(\tilde{\theta}^*_{\text{GMM}} - \theta_0)) = (\hat{\mathrm{G}}'(\tilde{\theta}^*_{\text{GMM}})\hat{\Omega}^{-1}(\tilde{\theta}^*_{\text{GMM}})\hat{\mathrm{G}}(\tilde{\theta}^*_{\text{GMM}}))^{-1} \tag{2.42}$$

2. continuous updating method

$$\hat{\theta} = \arg\min \mathrm{Q}_n(\theta) = \arg\min \hat{g}_n(\theta)'[\hat{\Omega}]^{-1}\hat{g}_n(\theta) \tag{2.43}$$

GMM的方差

1) $\mathrm{L} = k$ L: # moment conditions k: # of parameters

$$\text{F.O.C}: \frac{\partial\hat{\mathrm{Q}}(\theta)}{\partial\theta} = -2\nabla_{\theta'}\hat{g}_n(\hat{\theta})\mathrm{W}\hat{g}_{\mathrm{N}}(\theta) = 0 \tag{2.44}$$

在$\mathrm{L} = k$的情况下，$\nabla_{\theta'}\hat{g}_n(\hat{\theta})$和W都是满秩矩阵，所以$\hat{g}_n(\theta) = 0$和MM一致。

$$\begin{aligned}\mathrm{V}ar(\sqrt{n}\hat{\theta}_{\text{GMM}}) &= (\mathrm{G}'\mathrm{W}\mathrm{G})^{-1}\mathrm{G}'\mathrm{W}\Omega_0\mathrm{W}\mathrm{G}(\mathrm{G}'\mathrm{W}\mathrm{G})^{-1} \quad \text{笔记上记的是}\mathrm{V}ar(\hat{\theta}_{\text{GMM}}) \\ &= \mathrm{G}^{-1}\mathrm{W}^{-1}\mathrm{G}'^{-1}\mathrm{G}'\mathrm{W}\Omega_0\mathrm{W}\mathrm{G}\mathrm{G}^{-1}\mathrm{W}^{-1}\mathrm{G}'^{-1} \\ &= \mathrm{G}^{-1}\Omega_0\mathrm{G}'^{-1} = (\mathrm{G}'\Omega^{-1}\mathrm{G})^{-1}\end{aligned} \tag{2.45}$$

其中由于G是方阵，所以$(\mathrm{G}'\mathrm{W}\mathrm{G})^{-1}$可以直接展开。该过程说明当$\mathrm{L} = k$时，权重矩阵取什么无所谓。

e.g. OLS $\mathrm{L} = k \quad y = \mathrm{X}\beta + u$

使用GMM，样本矩条件为$\frac{1}{n}\mathrm{X}'u = 0$

$$\hat{g}_n(\beta) = \frac{1}{n}\mathrm{X}'(y - \mathrm{X}\beta) = 0 \tag{2.46}$$

$$\begin{aligned}\hat{\beta}_{\text{OLS}} = \hat{\beta}_{\text{GMM}} &= (\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}'y \\ &= \beta + (\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}'u\end{aligned} \tag{2.47}$$

根据之前的论述，under homo $\mathrm{V}ar(\hat{\beta}_{\text{OLS}}) = \sigma^2(\mathrm{X}'\mathrm{X})^{-1}$,因此验证GMM方法计算的方差

$$\begin{aligned}\mathrm{G} &= \mathrm{E}[\nabla_\beta g(x_i, \beta_0)] \\ &= \mathrm{E}[\nabla_\beta \frac{1}{n}\sum_{i=1}^{n} g(x_i, \beta_0)] \\ &= \mathrm{E}[\nabla_\beta \hat{g}_n(\beta_0)] \\ &= -\frac{1}{n}\mathrm{E}(\mathrm{X}'\mathrm{X})\end{aligned} \tag{2.48}$$

$$\begin{aligned}\Omega &= \mathrm{E}(g(x_i,\beta_0)g(x_i,\beta_0)') \\ &= \frac{1}{n}\mathrm{E}[\sum_{i=1}^{n} g(x_i,\beta_0)\sum_{j=1}^{n} g(x_j,\beta_0)'] \\ &(when\, i \neq j \mathrm{E}(gg') = 0, independent) \\ &= n\mathrm{E}[\frac{1}{n}\sum_{i=1}^{n} g(x_i,\beta_0)\frac{1}{n}\sum_{j=1}^{n} g(x_j,\beta_0)'] \\ &= n\mathrm{E}[\hat{g}_n(\beta_0)\hat{g}_n(\beta_0)'] \\ &= \frac{1}{n}\mathrm{E}[\mathrm{X}'uu'\mathrm{X}]\end{aligned} \tag{2.49}$$

$$\begin{aligned}\mathrm{V}ar(\sqrt{n}\hat{\beta}_{\mathrm{GMM}}) &= (\mathrm{G}'\Omega^{-1}\mathrm{G})^{-1} \\ &= [\mathrm{E}(\frac{\mathrm{X}'\mathrm{X}}{n})]^{-1}\mathrm{E}[\frac{\mathrm{X}'uu'\mathrm{X}}{n}][\mathrm{E}(\frac{\mathrm{X}'\mathrm{X}}{n})]^{-1} \\ &\sim (\frac{\mathrm{X}'\mathrm{X}}{n})^{-1}\frac{\mathrm{X}'\mathrm{E}(uu)'\mathrm{X}}{n}(\frac{\mathrm{X}'\mathrm{X}}{n})^{-1} \\ &= \sigma^2(\frac{\mathrm{X}'\mathrm{X}}{n})^{-1}\end{aligned} \tag{2.50}$$

e.g. IV Estimation $\mathrm{L} = k\,\mathrm{W} = (\mathrm{X}, \mathrm{X}_{\mathrm{IV}})$, weighting matrix is Z

$$\hat{\mathrm{Q}} = u'\mathrm{W}\mathrm{Z}^{-1}\mathrm{W}'u(\text{严格来说，这里好像漏了俩}\frac{1}{n}) \tag{2.51}$$

$$= (y-\mathrm{X}\beta)'\mathrm{W}\mathrm{Z}'\mathrm{W}'(y-\mathrm{X}\beta) \tag{2.52}$$

$$= \dots \tag{2.53}$$

$$\mathrm{F.O.C}: -2\mathrm{X}'\mathrm{W}\mathrm{Z}^{-1}\mathrm{W}'y + 2\mathrm{X}'\mathrm{W}\mathrm{Z}^{-1}\mathrm{W}'\mathrm{X}\hat{\beta} = 0 \tag{2.54}$$

$$\Rightarrow \hat{\beta} = (\mathrm{W}'\mathrm{X})^{-1}\mathrm{W}'y \quad (\mathrm{IV}estimation) \tag{2.55}$$

$$\begin{aligned}\mathrm{V}ar(\sqrt{n}\hat{\beta}_{\mathrm{GMM}}) &= (\mathrm{G}'\Omega^{-1}\mathrm{G}')^{-1} \\ &= ((-\frac{1}{n}\mathrm{E}\mathrm{W}'\mathrm{X})'(\frac{1}{n}\mathrm{E}\mathrm{W}'uu'\mathrm{W})^{-1}(-\frac{1}{n}\mathrm{E}\mathrm{W}'\mathrm{X}))^{-1} \\ &= \sigma^2(\frac{\mathrm{X}'\mathrm{P}_{\mathrm{W}}\mathrm{X}}{n})^{-1}\end{aligned} \tag{2.56}$$

$$\hat{\beta} = \beta + (\mathrm{W}'\mathrm{X})^{-1}\mathrm{W}'u \tag{2.57}$$

$$\sqrt{n}(\hat{\beta}_{2\mathrm{SLS}} - \beta) = (\frac{\mathrm{W}'\mathrm{X}}{n})^{-1}\frac{1}{\sqrt{n}}\mathrm{W}'u \tag{2.58}$$

$$(\frac{\mathrm{W}'\mathrm{X}}{n})^{-1} \to (\mathrm{E}\mathrm{W}_i\mathrm{X}_i')^{-1} \tag{2.59}$$

$$\frac{1}{\sqrt{n}}\mathrm{W}'u \to \mathrm{N}(0, \sigma^2\mathrm{E}\mathrm{W}_i\mathrm{W}_i') \tag{2.60}$$

$$\Rightarrow \sim \mathrm{N}(0, \sigma^2(\frac{\mathrm{X}'\mathrm{P}_\mathrm{W}\mathrm{X}}{n})^{-1}) \tag{2.61}$$

$\mathrm{L} > k$

$$\hat{\mathrm{Q}}_n = u'\mathrm{W}[\quad]\mathrm{W}'u \tag{2.62}$$

$$\begin{aligned}\Omega_0 &= \mathrm{V}ar(g(z_i, \theta_0)) \\ \mathrm{V}ar(\hat{g}_n(\theta)) &= \mathrm{V}ar(\frac{1}{n}\sum_{i=1}^{n} g(z_i, \theta)) \\ &= \frac{1}{n}\mathrm{V}ar(g(z_i, \theta))\end{aligned} \tag{2.63}$$

$$\begin{aligned}\hat{\mathrm{Q}}_n &= u'\mathrm{W}[\mathrm{V}ar(\mathrm{W}'u)]^{-1}\mathrm{W}u' \\ &= \frac{1}{\sigma^2}u'\mathrm{W}(\mathrm{W}'\mathrm{W})^{-1}\mathrm{W}'u \\ &= \frac{1}{\sigma^2}u'\mathrm{P}_\mathrm{W}u \\ &= \frac{1}{\sigma^2}(y - \mathrm{X}\beta)'\mathrm{P}_\mathrm{W}(y - \mathrm{X}\beta)\end{aligned} \tag{2.64}$$

$$\mathrm{F.O.C}: -2\mathrm{X}'\mathrm{P}_\mathrm{W}y + 2\mathrm{X}'\mathrm{P}_\mathrm{W}\mathrm{X}\hat{\beta} = 0 \tag{2.65}$$

$$\hat{\beta}_{\mathrm{GMM}} = (\mathrm{X}'\mathrm{P}_\mathrm{W}\mathrm{X})^{-1}\mathrm{X}'\mathrm{P}_\mathrm{W}y = \hat{\beta}_{2\mathrm{SLS}} \tag{2.66}$$

$$\mathrm{V}ar(\sqrt{n}\hat{\beta}_{\mathrm{GMM}}) = \sigma^2(\frac{\mathrm{X}'\mathrm{P}_\mathrm{W}\mathrm{X}}{n})^{-1} \tag{2.67}$$

2.2 基于广义矩估计的检验

3 Test parameter restrictions Wald LM LR(基于GMM的角度)

首先从根据一般教科书的惯例，从MLE的角度介绍三种检验的区别。

$$\mathrm{H}_0 : r(\theta_0) = 0 \text{非线性约束，线性约束为:} \mathrm{R}\beta = 0 \tag{2.68}$$

$$\mathrm{H}_1 : r(\theta_0) \neq 0 \tag{2.69}$$

其中$r : \mathscr{R}^k \to \mathscr{R}^q$
(TODO缺一张图，以后补)

Wald Test 先MLE最大化，得到参数的估计值，代入到$r(\theta)$,如果约束成立则统计量$\mathrm{W} = 0$(只需要求解无约束的最大化问题)

LM Test 求解有约束的最大化问题，得到$\hat{\theta}_{\mathrm{R}}$,比较有约束的Score和0的差异(无约束的Score为0,不用算,所以只需要求解有约束的最大化问题)。

LR Test 分别求解有约束和无约束的最大化问题，比较Likelihood的大小。

Wald Test

基本思路

$$\hat{\theta}_{\mathrm{GMM}} \sim \mathrm{N} \tag{2.70}$$

$$r(\hat{\theta}_{\mathrm{GMM}}) \sim \mathrm{N} \quad (\mathrm{D}elta\mathrm{M}ethod) \tag{2.71}$$

$$\Rightarrow \chi^2() \tag{2.72}$$

$$\sqrt{n}(\hat{\theta}_{unr} - \theta_0) \sim \mathrm{N}(0, \mathrm{V}_0) \tag{2.73}$$

$$\sqrt{n}(r(\hat{\theta}_{unr}) - r(\theta_0)) \sim \mathrm{N}(0, \mathrm{R}_0\mathrm{V}_0\mathrm{R}_0') \tag{2.74}$$

其中 V_0 是用GMM计算的方差, $\mathrm{R}_0 = \mathrm{R}(\theta_0)$ $\mathrm{R}(\theta) = \frac{\partial r(\theta)}{\partial \theta'}{}_{q\times k}$

$$\mathrm{W}ald = nr(\hat{\theta}_{unr})[\hat{\mathrm{R}}\hat{\mathrm{V}}\hat{\mathrm{R}}']^{-1} r(\hat{\theta}_{unr}) \sim \chi^2(q) \tag{2.75}$$

LM test

$$\frac{\partial \hat{Q}_n(\hat{\theta}_r)}{\partial \theta} = 2\nabla_{\theta'}\hat{g}_n(\theta)\hat{\Omega}^{-1}\hat{g}_n(\theta)\Big|_{\theta=\theta_0} \approx 0 \tag{2.76}$$

$$\mathrm{LM} = (\underset{1\times k}{\frac{\partial \hat{Q}}{\partial \theta}})'[\mathrm{V}ar(\underset{k\times k}{\frac{\partial \hat{Q}}{\partial \theta}})]^{-1}(\frac{\partial \hat{Q}}{\partial \theta}) \tag{2.77}$$

$$\frac{\partial \hat{Q}_n(\hat{\theta}_{res})}{\partial \theta} \to 2\mathrm{G}(\hat{\theta}_{res})'\hat{\Omega}^{-1}\hat{g}_n(\hat{\theta}_{res}) \tag{2.78}$$

$$\sqrt{n}\hat{g}_n(\hat{\theta}_{res}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} g(z_i,\hat{\theta}_{res}) \sim \mathrm{N}(0,\hat{\Omega}) \tag{2.79}$$

$$\begin{aligned} \mathrm{V}ar(\frac{\partial \hat{Q}_n}{\partial \theta}(\hat{\theta}_{res})) &= \frac{4}{n}\mathrm{G}'(\hat{\theta}_{res})\hat{\Omega}^{-1}\mathrm{V}ar(\sqrt{n}\hat{g}_n(\hat{\theta}_{res}))\hat{\Omega}^{-1}\mathrm{G}(\hat{\theta}_{res}) \\ &= \frac{4}{n}\underset{k\times \mathrm{L}}{\mathrm{G}'(\hat{\theta}_{res})} \quad \underset{\mathrm{L}\times \mathrm{L}}{\hat{\Omega}^{-1}} \quad \underset{\mathrm{L}\times k}{\mathrm{G}(\hat{\theta}_{res})} \end{aligned} \tag{2.80}$$

$$\mathrm{LM} = n[\underset{1\times k}{\hat{g}_n(\hat{\theta}_{res})'\hat{\Omega}^{-1}\mathrm{G}(\hat{\theta}_{res})}][\underset{k\times k}{\mathrm{G}'(\hat{\theta})_{res}\hat{\Omega}^{-1}\mathrm{G}(\hat{\theta}_{res})}]^{-1}[\underset{k\times 1}{\mathrm{G}'(\hat{\theta}_{res})\hat{\Omega}^{-1}\hat{g}_n(\hat{\theta}_{res})}] = \underset{1\times 1}{(\quad)} \sim \chi^2(q)(q<k) \tag{2.81}$$

LR

$$\hat{Q}_n(\theta) = \hat{g}_n(\theta)'\Omega^{-1}\hat{g}_n(\theta) \tag{2.82}$$

$$\Omega = \mathrm{E}gg' = \mathrm{V}ar(\sqrt{n}\hat{g}_n) = n\mathrm{V}ar(\hat{g}_n) \Rightarrow \Omega^{-1} = \frac{1}{n}\mathrm{V}ar^{-1} \to 0 \tag{2.83}$$

$$\therefore objective function is \quad n\hat{Q}(\theta) \tag{2.84}$$

无约束的GMM目标函数

$$n\hat{Q}(\theta) = (\sqrt{n}\hat{g}_n(\hat{\theta}))'[\mathrm{V}ar\sqrt{n}\hat{g}_n(\hat{\theta})]^{-1}(\sqrt{n}\hat{g}_n(\hat{\theta})) \sim \chi^2(l-k) \tag{2.85}$$

有约束的GMM目标函数

$$n\hat{Q}_n(\theta_{\mathrm{R}}) = (\sqrt{n}\hat{g}_n(\hat{\theta}_{\mathrm{R}}))'[\mathrm{V}ar\sqrt{n}\hat{g}_n(\hat{\theta}_{\mathrm{R}})]^{-1}(\sqrt{n}\hat{g}_n(\hat{\theta}_{\mathrm{R}})) \sim \chi^2(l-(k-q)) \tag{2.86}$$

$$\mathrm{LR} = n\hat{Q}_n(\hat{\theta}_{\mathrm{R}}) - n\hat{Q}_n(\hat{\theta}) \sim \chi^2(q) \tag{2.87}$$

Test of Moment Restrictions We begin by partitioning the moment restrictions into a set of k reliable moment conditions that identifies θ_0 $\mathrm{E}(g_l(z,\theta_0)) = 0$ for $l = 1,2,\ldots,k$ and a set of remaining questionable moment restrictions that comprise the H_0

$$
\begin{aligned}
\mathrm{H}_0 æ\mathrm{E}(g_l(z,\theta_0)) &= 0 \quad l = k+1, k+2, \ldots, \mathrm{L} \\
\mathrm{H}_1 : \mathrm{E}(g_l(z,\theta_0)) &\neq 0 \quad for\, some l = k+1, k+2, \ldots, \mathrm{L}
\end{aligned} \tag{2.88}
$$

Test of Moment Conditions 要求 $\mathrm{L} > k$

增广矩条件 augumented

$$
\begin{aligned}
g^a(z,\theta,\phi) = [&\underbrace{g_1(z,\theta), \ldots, g_l(z,\theta),}_{reliable \quad \mathrm{E}()=0} \\
&g_{k+1(z,\theta)} - \phi_1, \ldots, g_{\mathrm{L}}(z,\theta) - \phi_{\mathrm{L}-k}]
\end{aligned} \tag{2.89}
$$

将对矩条件的检验转换为对参数的检验

$$
\mathrm{H}_0 : \psi_j = 0, j = 1, 2, \ldots, \mathrm{L}-k \tag{2.90}
$$

$$
\begin{aligned}
\mathrm{LR} =& n[\hat{\mathrm{Q}}^a(\hat{\theta}_{res}, \overset{\phi_j=0}{0}) - \hat{\mathrm{Q}}^a(\hat{\theta}_{unr}, \hat{\phi}_{unr})] \\
&= n[\hat{\mathrm{Q}}_n(\hat{\theta}_{unr}) - 0] \\
&= n\hat{\mathrm{Q}}_n(\hat{\theta}) \sim \chi^2(l-k)
\end{aligned} \tag{2.91}
$$

2.3 M-Estimation

2.3.1 Estimation

An estimation of $\hat{\theta}$ is a M-estimator if there is an objective function $\hat{\mathrm{Q}}(w_i,\theta)$, where $w_i = y_i, x_i$ such that

$$
\hat{\theta} max/min \hat{\mathrm{Q}}(w_i,\theta)\, s.t. \theta \in Ⓗ \tag{2.92}
$$

1.Linear Regression

$$
y_i = x_i'\beta + u_i \tag{2.93}
$$

$$
min\,\mathrm{SSR} \tag{2.94}
$$

2.MLE

$$y_i \sim \mathrm{N}(x_i'\beta, \sigma^2(\theta, z_i)) \tag{2.95}$$

$$ln\mathrm{L}(\beta, \theta|x, y, z) = \sum_{i=1}^{n} lnf(y_i|x_i, z_i, \beta, \theta) \tag{2.96}$$

$$max \sum_{i=1}^{n} lnf(y_i|x_i, z_i, \beta, \theta) \tag{2.97}$$

3.nonlinear regression model

$$\begin{array}{cc} y_i = m(x_i, \theta) + u_i & y_i = m(x) + u_i \\ \text{参数模型}m()\text{已知} & \text{非参数模型}m()\text{未知} \end{array} \tag{2.98}$$

e.g.

$$m(\mathrm{X}, \theta) = exp(\mathrm{X}\theta) \tag{2.99}$$

$$m(\mathrm{X}, \theta) = \frac{exp(\mathrm{X}\theta)}{1 + exp(\mathrm{X}\theta)}\text{(logistic function)} \tag{2.100}$$

$$y = m(\mathrm{X}) + u \quad m(\mathrm{X}, \theta) = \mathrm{X}'\theta \tag{2.101}$$

$$\Leftrightarrow y_i = x_i'(\theta) + u_i \tag{2.102}$$

$$\mathrm{E}(u_i|x_i) = 0 \Leftrightarrow \mathrm{E}(y|\mathrm{X}) = x_i'\theta + \cancel{\mathrm{E}(u_i|\mathrm{X})} \tag{2.103}$$

NLS assumption 1:

For some $\theta_0 \in Ⓗ$, $\mathrm{E}(y|\mathrm{X}) = m(\mathrm{X}, \theta)$

$$\min_{\theta \in Ⓗ} \mathrm{E}(y - m(x, \theta))^2 \tag{2.104}$$

$$
\begin{aligned}
\mathrm{E}(y-m(x,\theta))^2 &= \mathrm{E}[y-\mathrm{E}(y|\mathrm{X})+\mathrm{E}(y|\mathrm{X})-m(\mathrm{X},\theta]^2 \\
&= \mathrm{E}[(y-m(\mathrm{X},\theta_0))+(m(\mathrm{X},\theta_0)-m(\mathrm{X},\theta))]^2 \\
&= \mathrm{E}[y-m(\mathrm{X},\theta_0)]^2+2\mathrm{E}[(y-m(\mathrm{X},\theta_0))(m(\mathrm{X},\theta_0)-m(\mathrm{X},\theta))]+\mathrm{E}[m(\mathrm{X},\theta_0)-m(\mathrm{X},\theta)]^2 \\
&= \mathrm{E}[y-m(\mathrm{X},\theta_0)]^2+2\mathrm{E}\left[\mathrm{E}[(y-m(\mathrm{X},\theta_0))(m(\mathrm{X},\theta_0)-m(\mathrm{X},\theta))|\mathrm{X}]\right]+\mathrm{E}[m(\mathrm{X},\theta_0)-m(\mathrm{X},\theta)]^2 \\
&= \mathrm{E}[y-m(\mathrm{X},\theta_0)]^2+2\mathrm{E}\left[\mathrm{E}[y-m(\mathrm{X},\theta_0)|\mathrm{X}](m(\mathrm{X},\theta_0)-m(\mathrm{X},\theta))\right]+\mathrm{E}[m(\mathrm{X},\theta_0)-m(\mathrm{X},\theta)]^2 \\
&= \mathrm{E}[y-m(\mathrm{X},\theta_0)]^2+2\mathrm{E}\left[(\underbrace{\mathrm{E}(y|\mathrm{X})}_{=m(\mathrm{X},\theta_0)}-m(\mathrm{X},\theta_0))(m(\mathrm{X},\theta_0)-m(\mathrm{X},\theta))\right]+\mathrm{E}[m(\mathrm{X},\theta_0)-m(\mathrm{X},\theta)]^2 \\
&= \mathrm{E}[y-m(\mathrm{X},\theta_0)]^2+\mathrm{E}[m(\mathrm{X},\theta_0)-m(\mathrm{X},\theta)]^2
\end{aligned}
\tag{2.105}
$$

1) if $\theta=\theta_0$ then $\theta_0=\arg\min_{\theta\in Ⓗ}\in \mathrm{E}(y-m(\mathrm{X},\theta_0))^2$

2) if $\theta\neq\theta_0$ then

$$\mathrm{E}[m(\mathrm{X},\theta_0)-m(\mathrm{X},\theta)]^2\geq 0 \tag{2.106}$$

$$\text{if } \mathrm{E}[m(\mathrm{X},\theta_0)-m(\mathrm{X},\theta)]^2=0 \tag{2.107}$$

$$\text{then } \mathrm{E}(y-m(\mathrm{X},\theta))^2=\mathrm{E}(y-m(\mathrm{X},\theta_0))^2 \tag{2.108}$$

$$\therefore \theta \text{ can't be uniquely identified}$$

$$\text{if } \mathrm{E}[m(\mathrm{X},\theta_0)-m(\mathrm{X},\theta)]^2>0 \tag{2.109}$$

$$\therefore \theta \text{ is uniquely identified}$$

NLS assumption 2:

$\mathrm{E}[m(\mathrm{X},\theta_0)-m(\mathrm{X},\theta)]^2>0$ for all $\theta\in Ⓗ\ \theta\neq\theta_0$

$$
\begin{aligned}
&\text{assume } m(\mathrm{X},\theta)=\mathrm{X}\theta \\
&\mathrm{E}(m(\mathrm{X},\theta_0)-m(\mathrm{X},\theta))^2 \\
=&\mathrm{E}[(\mathrm{X}\theta_0-\mathrm{X}\theta)'(\mathrm{X}\theta_0-\mathrm{X}\theta)] \\
=&\mathrm{E}[(\theta_0-\theta)'\mathrm{X}'\mathrm{X}(\theta_0-\theta)]>0
\end{aligned}
\tag{2.110}
$$

因此要求$\mathrm{X}'\mathrm{X}$positive definite,即X矩阵列满秩，$rank(\mathrm{E}(\mathrm{X}'\mathrm{X}))=k$(上课后来改成了$rank(\mathrm{E}(xx'))=k$,其中$x'$是X的行向量) NLS assumption 2 又叫 Identification Condition

一个不满足NLS 2的例子，假设真实模型为

$$m(x,\theta_0)=\theta_{10}+\theta_{20}x_2 \tag{2.111}$$

待估计的模型为

$$m(\mathrm{X},\theta)=\theta_1+\theta_2x_2+\theta_3x_3^{\theta_4} \tag{2.112}$$

$$\underset{\theta_1,\theta_2,\theta_3,\theta_4}{min}\ \mathrm{E}[y-m(\mathrm{X},\theta)]^2 \tag{2.113}$$

$$\theta_1=\theta_{10},\theta_2=\theta_{20},\theta_3=0,\theta_4=\text{any value} \tag{2.114}$$

所以NLS assumption 2 违背。

The General M-estimator can be expressed assume

$$\underset{\theta\in\Theta}{min}\ \mathrm{E}[q(w,\theta)]\ \text{e.g.}\ q(w,\theta)=[y-m(\mathrm{X},\theta)]^2 \tag{2.115}$$

The identification requires

$$\mathrm{E}[q(w,\theta_0)]<\mathrm{E}[q(w,\theta)]\forall\theta\in\Theta,\theta\neq\theta_0 \tag{2.116}$$

用算术平均值代替期望

$$\hat{\theta}=\underset{\theta\in\Theta}{min}\ \frac{1}{n}\sum_{i=1}^{n}q(w_i,\theta) \tag{2.117}$$

问题是在什么条件下，满足一致性条件$\hat{\theta}\xrightarrow{p}\theta_0$

类似于GMM中提到的的，如果目标函数是一致的，则他们的估计值也是一样的。

Theorem 2.3.1. ***Uniform Weak Law of Large Numbers*** *If*

1. *Data w_i is i.i.d*

2. *$\theta\in\Theta$, Θ is a compact set*

3. *for each w_i, $q(w)$ is continuous on Θ*

4. *$|q(w_i,\theta)|\leq b(w_i)\forall\theta\in\Theta\mathrm{E}(b(w_i))<\infty$*

Then

$$\frac{1}{n}\sum_{i=1}^{n}q(w_i,\theta)\xrightarrow{p}\mathrm{E}[q(w,\theta)] \tag{2.118}$$

Theorem 2.3.2. ***Consistency of M-estimator***

Under the assumption of Theorem 1 and assume identification assumption hold,then

$$\hat{\theta} \xrightarrow{p} \theta_0 \tag{2.119}$$

Proof of Theorem 2.2 see Newey and Mcfadden(1994).

If $\hat{\theta} \xrightarrow{p} \theta_0$ as $n \to \infty$, $\frac{1}{n}\sum_{i=1}^{n} r(w_i, \hat{\theta}) \xrightarrow{?} \mathrm{E}(r(w, \theta_0))$

Lemma 2.3.1. *Suppose that $\hat{\theta} \to \theta_0$ and assume any functions $r(w_i, \theta)$ satisfies the same assumption as in Theorem 2.2, then*

$$\frac{1}{n}\sum_{i=1}^{n} r(w_i, \hat{\theta}) \xrightarrow{p} \mathrm{E}(r(w, \theta_0)) \tag{2.120}$$

即只要$r(w, \theta)$连续,有界

然后要解决的问题是如何找到$\hat{\theta}$的分布，$\hat{\theta} \sim ?$

$$\min_{\theta \in \mathrm{H}} \frac{1}{n}\sum_{i=1}^{n} q(w_i, \theta) \tag{2.121}$$

$$\text{F.O.C: } \sum_{i=1}^{n} \frac{\partial q(w_i, \hat{\theta})}{\partial \theta} = \sum_{i=1}^{n} \frac{\partial q(w_i, \theta_0)}{\partial \theta} + \sum_{i=1}^{n} \frac{\partial^2 q(w_i, \bar{\theta})}{\partial \theta \partial \theta'}(\hat{\theta} - \theta_0) = 0 \tag{2.122}$$

$$\frac{1}{n}\sum_{i=1}^{n} \frac{\partial^2 q(w_i, \bar{\theta})}{\partial \theta \partial \theta'} \sqrt{n}(\hat{\theta} - \theta_0) = -\frac{1}{\sqrt{n}} \sum_{i=1}^{n}\sum_{i=1}^{n} \frac{\partial q(w_i, \theta_0)}{\partial \theta} \tag{2.123}$$

Let $\mathrm{H}_i = \mathrm{H}(w_i, \bar{\theta}) = \frac{\partial^2 q(w_i, \bar{\theta})}{\partial \theta \partial \theta'}$ be the Hessian matrix of the objective function,$\mathrm{S}(w_i, \theta_0) = \frac{\partial q(w_i, \theta_0)}{\partial \theta}$ be the Score of the objective function

$$\sqrt{n}(\hat{\theta} - \theta_0) = (\frac{1}{n}\sum_{i=1}^{n} \mathrm{H}_i)^{-1}(-\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \mathrm{S}(w_i, \theta_0)) \text{ as } n \to \infty \tag{2.124}$$

根据Lemma2.1有

$$\frac{1}{n}\sum_{i}^{n} \mathrm{H}_i = \frac{1}{n}\sum_{i=1}^{n} \mathrm{H}(w, \bar{\theta}) \xrightarrow{p} \mathrm{E}[\mathrm{H}(w, \theta_0)] \stackrel{def}{=} \mathrm{A}_0 \tag{2.125}$$

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} [-\mathrm{S}(w_i, \theta_0)] \sim \mathrm{N}(0, \underbrace{\mathrm{E}\mathrm{S}_i\mathrm{S}_i'}_{=\mathrm{B}_0}) \tag{2.126}$$

$$\sqrt{n}(\hat{\theta} - \theta_0) \sim \mathrm{N}(0, \mathrm{A}_0^{-1}\mathrm{B}_0\mathrm{A}_0^{-1}) \tag{2.127}$$

$$\sqrt{n}(\hat{\theta} - \theta_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} [-\mathrm{A}_0^{-1}\mathrm{S}_i(\theta_0)] + o_p(1) \tag{2.128}$$

$$\stackrel{def}{=} \frac{1}{\sqrt{n}}\sum_{i=1}^{n} e_i(\theta_0) + o_p(1) \tag{2.129}$$

which is the influnce function representation of $\hat{\theta}$, where $e(w_i, \theta_0)$ is called the influence function.

$N(0, A_0^{-1}B_0A_0^{-1})$ 要求

1) $E(S_i) = 0$

M-estimator 目标函数为Eq,F.O.C为$\frac{\partial Eq}{\partial \theta} = 0$，即$E\frac{\partial q}{\partial \theta} = 0$

2) A_0可逆

在最小化问题中，如果满足识别条件，则有Hessian矩阵正定，必然可逆

$$q(w, \theta_0) = \frac{1}{2}(y - m(x_i, \theta_0))^2 \tag{2.130}$$

$$S_i = -\frac{\partial m_i}{\partial \theta}(y_i - m_i) = -\nabla_{\theta_0} m_i'(y_i - m_i) \tag{2.131}$$

$$ES_i = E[E(S_i|X_i)], E(S_i|X_i) = 0 \tag{2.132}$$

example 1

$$m(X, \theta) = X\theta,\ A_0 = E(X'X) --- \textbf{full rank} \tag{2.133}$$

$$\tag{2.134}$$

example 2

$$m(X, \theta) = \theta_1 + \theta_2 X_2 + \theta_3 X_3^{\theta_4},\ \theta_3 = 0,\ \theta_4 \text{ can be any value} \tag{2.135}$$

$$H(w, \theta) = \nabla_\theta m(X, \theta)' \nabla_\theta m(X, \theta) - \nabla_\theta^2 m(X, \theta)(y - m(X, \theta)) \tag{2.136}$$

$$A_0 = E[H(w, \theta_0)] = E[\nabla_\theta m(X, \theta)' \nabla_\theta m(X, \theta)] \tag{2.137}$$

$$E\left[\begin{pmatrix} 1 \\ x_2 \\ x_3^{\theta_4} \\ \underbrace{\theta_3 x_3^{\theta_4} ln(x_3)}_{=0} \end{pmatrix} \begin{pmatrix} 1 & x_2 & x_3^{\theta_4} & \underbrace{\theta_3 x_3^{\theta_4} ln(x_3)}_{=0} \end{pmatrix}\right] = E\left[\begin{pmatrix} & & & 0 \\ & & & 0 \\ & & & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}\right] \tag{2.138}$$

行列式=0不可逆。

Two step M-estimation

$$\min \sum_{i=1}^{n} q(w_i, \theta, \hat{\gamma}) \tag{2.139}$$

Example: Weighted Nonlinear Least Square(WNLS)

$$y_i = m(x_i, \theta) + u_i, \mathrm{E}(u_i^2|x_i) = h(x_i, \gamma_0) \tag{2.140}$$

$$\frac{y_i}{\sqrt{h(x_i, \gamma_0)}} = \frac{m(x_i, \theta)}{\sqrt{h(x_i, \gamma_0)}} + \frac{u_i}{\sqrt{h(x_i, \gamma_0)}} \tag{2.141}$$

$$\min_{\theta \in Ⓗ} \frac{1}{2} \sum_{i=1}^{n} (y_i - m(x_i, \theta))^2 / h(x_i, \gamma_0) \tag{2.142}$$

$$\hat{u}_i{}^2 = h(x_i, \gamma) + error_i \overset{\text{M–E}stimation}{\Rightarrow} \hat{\gamma} \tag{2.143}$$

$$\min_{\theta \in Ⓗ} \frac{1}{2} \sum_{i=1}^{n} (y_i - m(x_i, \theta))^2 / h(x_i, \hat{\gamma}) \tag{2.144}$$

$$\tag{2.145}$$

WNLS Assumption 1

$$\mathrm{E}(u|\mathrm{X}) = 0 \tag{2.146}$$

$$\mathrm{E}\left(\frac{y_i}{\sqrt{h(x_i, \gamma_0)}}\right) = \mathrm{E}\left(\frac{m(x_i, \theta)}{\sqrt{h(x_i, \gamma_0)}}\right) + \mathrm{E}\left(\frac{u_i}{\sqrt{h(x_i, \gamma_0)}}\right) \tag{2.147}$$

which is the same as NLS Assumption 1.

WNLS Assumption 2

$$\mathrm{E}[(m(\mathrm{X}, \theta_0) - m(\mathrm{X}, \theta))^2 / h(\mathrm{X}, \gamma^*)] > 0, \textbf{for all } \theta \neq \theta_0,\ \theta \in Ⓗ \tag{2.148}$$

更一般的有

$$\mathrm{E}[q(\mathrm{W}, \theta_0, \gamma^*)] < \mathrm{E}[q(\mathrm{W}, \theta, \gamma^*)], \textbf{for all } \theta \neq \theta_0,\ \theta \in Ⓗ \tag{2.149}$$

在模型设定正确的情况下有，$\mathrm{V}ar(y|\mathrm{X}) = h(\mathrm{X}, \gamma_0), \hat{\gamma} \to \gamma_0$

在模型设定错误的情况下有，$\mathrm{V}ar(y|\mathrm{X}) = h(\mathrm{X}, \gamma_0), \hat{\gamma} \to \gamma_*$

$$\begin{cases} \frac{1}{n} \sum_{i=1}^{n} q(w_i, \theta, \gamma^*) \to \mathrm{E}q(w_i, \theta, \gamma^*) \quad \text{UWLLN} \\ \text{identification condition} \end{cases} \Rightarrow \text{consistency: } \hat{\theta} \xrightarrow{p} \theta_0 \tag{2.150}$$

Lemma 2.3.2. *Like Lemma 2.1, we have*

$$\frac{1}{n}\sum_{i=1}^{n} q(w_i,\hat{\theta},\hat{\gamma}) \to \mathrm{E}q(w_i,\theta,\gamma^*) \tag{2.151}$$

$$\sqrt{n}(\hat{\theta}-\theta_0) = \mathrm{A}_0^{-1}(-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathrm{S}_i(\theta_0,\hat{\gamma})) + o_p(1) \tag{2.152}$$

$$\text{as } n\to\infty, \hat{\gamma}\to\gamma^*, \alpha_0 \stackrel{def}{=} (\theta_0,\gamma^*)$$

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathrm{S}_i(\theta_0,\hat{\gamma}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathrm{S}_i(\theta_0,\gamma^*) + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial \mathrm{S}_i(\theta_0,\gamma^*)}{\partial\gamma^*}(\hat{\gamma}-\gamma^*) + o_p(1) \tag{2.153}$$

$$\mathrm{M}-estimation \Rightarrow \sqrt{n}(\hat{\gamma}-\gamma^*) \to \mathrm{N}(,) = \mathrm{O}_p(1) \tag{2.154}$$

$$\therefore \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial \mathrm{S}_i(\theta_0,\gamma^*)}{\partial\gamma^*}(\hat{\gamma}-\gamma^*) = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial \mathrm{S}_i(\theta_0,\gamma^*)}{\partial\gamma^*}\sqrt{n}(\hat{\gamma}-\gamma^*) \to \mathrm{E}\frac{\partial \mathrm{S}_i}{\partial r}\mathrm{O}_p(1) \tag{2.155}$$

$$\mathrm{E}[\nabla_\gamma \mathrm{S}(\mathrm{W},\theta_0,\gamma^*)] \stackrel{def}{=} \mathrm{F}_0 \tag{2.156}$$

if $\mathrm{F}_0 = 0$ 不影响结果 $\sqrt{n}(\hat{\theta}-\theta_0) \sim \mathrm{N}(0,\mathrm{A}_0^{-1}\mathrm{B}_0\mathrm{A}_0^{-1})$, $\mathrm{A}_0 = \mathrm{E}[\mathrm{H}(\mathrm{W},\theta_0,\gamma^*)]$, $\mathrm{B}_0 = \mathrm{E}[\mathrm{S}(\mathrm{W},\theta_0,\gamma^*)\mathrm{S}(\mathrm{W},\theta_0,\gamma^*)']$。

对于$y = m(\mathrm{X},\theta_0)+u, \mathrm{F}_0 = 0$

对于probit,tobit,$\mathrm{F}_0 \neq 0$ Influence function representation: $\sqrt{n}(\hat{\gamma}-\gamma^*) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} r_i(\gamma^*) + o_p(1)$

if $\mathrm{F}_0 \neq 0$

$$\sqrt{n}(\hat{\theta}-\theta_0) = -\mathrm{A}_0^{-1}[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathrm{S}_i(\theta_0,\gamma^*) + \mathrm{F}_0\sqrt{n}(\hat{\gamma}-\gamma^*)] + o_p(1) \tag{2.157}$$

$$= -\mathrm{A}_0^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[\mathrm{S}_i(\theta_0,\gamma^*) + \mathrm{F}_0 r(\gamma^*)] + o_p(1) \tag{2.158}$$

$\mathrm{E}[-g_i(\theta_0,\gamma^*) = 0], \mathrm{V}ar(-g_i(\theta_0,\gamma^*) = \mathrm{E}[g_i g_i'] \stackrel{def}{=} \mathrm{D}_0$

$$\sqrt{n}(\hat{\theta}-\theta_0) \sim \mathrm{N}(0,\mathrm{A}_0^{-1}\mathrm{D}_0\mathrm{A}_0^{-1}) \tag{2.159}$$

$$\mathrm{D}_0 = \mathrm{E}[\mathrm{S}_i(\theta_0,\gamma^*)\mathrm{S}_i(\theta_0,\gamma^*)'] + \mathrm{F}_0\mathrm{E}[r_i(\gamma*)r_i(\gamma^*)']\mathrm{F}_0' = \mathrm{B}_0 + \mathrm{P}ositive\mathrm{D}efinte > \mathrm{B}_0 \tag{2.160}$$

$$\mathrm{A}_0 = \mathrm{E}[\mathrm{H}(\mathrm{W},\theta_0)], \mathrm{B}_0 = \mathrm{E}[\mathrm{S}(\mathrm{W},\theta_0,\gamma^*)\mathrm{S}(\mathrm{W},\theta_0,\gamma^*)'] \tag{2.161}$$

H的形式可以知道，但W的总体分布未知，因此用样本算术平均代替

方法1

$$A = \frac{1}{n}\sum_{i=1}^{n} H(w_i, \hat{\theta}) \to A_0 \tag{2.162}$$

$$B = \frac{1}{n}\sum_{i=1}^{n} [S(w_i, \hat{\theta}) S(w_i, \hat{\theta})'] \to B_0 \tag{2.163}$$

但在实际情况中可能会出现二阶导计算非常麻烦的情况

方法2 在模型设定正确的情况下

$$A(X, \theta_0) = E[H(W, \theta_0)|X] \tag{2.164}$$

$$\text{For nonlinear model: } y = m(X, \theta_0) + u \tag{2.165}$$

$$H(W, \theta_0) = \underbrace{\nabla_\theta m(X, \theta_0)'}_{k\times 1} \underbrace{\nabla_\theta m(X, \theta_0)}_{1\times k} - \nabla_\theta^2 m(X, \theta_0) \underbrace{y - m(X, \theta_0)}_{assumption2 \Rightarrow E()=0} \tag{2.166}$$

$$A_0 = E[A(X, \theta_0)] \tag{2.167}$$

$$A(X, \theta_0) = \nabla_\theta m(X, \theta_0)' \nabla_\theta m(X, \theta_0) \tag{2.168}$$

$$A_0 = E[A(X, \theta_0)] = E[\nabla_\theta m(X, \theta_0)' \nabla_\theta m(X, \theta_0)] \tag{2.169}$$

$$\hat{A} = \frac{1}{n}\sum_{i=1}^{n} \underbrace{A(X_i, \hat{\theta})}_{\hat{A}_i} \to A_0 \tag{2.170}$$

$$\hat{B}_0 = \frac{1}{n}\sum_{i=1}^{n} S(w_i, \hat{\theta}) S(w_i, \hat{\theta})' \to B_0 \tag{2.171}$$

$$\sqrt{n}(\hat{\theta} - \theta_0) \to N(0, \hat{A}_0^{-1} \hat{B}_0 \hat{A}_0^{-1}) \tag{2.172}$$

$$\hat{Var}(\hat{\theta}) = \begin{cases} (\sum_{i=1}^{n} \hat{H}_i)^{-1} (\sum_{i=1}^{n} \hat{S}_i \hat{S}_i') (\hat{H}_i)^{-1} & \text{Fully Robust Estimator} \\ (\sum_{i=1}^{n} \hat{A}_i)^{-1} (\sum_{i=1}^{n} \hat{S}_i \hat{S}_i') (\hat{A}_i)^{-1} & \text{Semi Robust Estimator} \end{cases} \tag{2.173}$$

$$\hat{A}(X_i, \hat{\theta}) = \underbrace{\nabla_\theta \hat{m}_i'}_{k\times 1} \underbrace{\nabla_\theta \hat{m}_i}_{1\times k} \tag{2.174}$$

$$\hat{S}_i = -\nabla_\theta \hat{m}_i (y_i - \hat{m}_i) = -\nabla_\theta \hat{m}_i \hat{u}_i \tag{2.175}$$

$$\hat{Var}(\hat{\theta}) = (\sum_{i=1}^{n} \nabla_\theta \hat{m}_i' \hat{m}_i)^{-1} (\sum_{i=1}^{n} \hat{u}_i \nabla_\theta \hat{m}_i' \hat{m}_i) (\sum_{i=1}^{n} \nabla_\theta \hat{m}_i' \hat{m}_i)^{-1} \tag{2.176}$$

在STATA等统计软件中，robust 一般指semi rubust estimator, 即 Heteroskadasticity robust, 并不是模型正确设定与否的robust(fully robust).

NLS Assumption 3

$$\mathrm{V}ar(y|\mathrm{X}) = \mathrm{V}ar(u|\mathrm{X}) = \sigma_0^2 \tag{2.177}$$

$$\mathrm{B}_0 = \sigma_0^2 \mathrm{E}[\nabla_\theta m(\mathrm{X}, \theta_0)' \nabla_\theta m(\mathrm{X}, \theta_0)] = \sigma_0^2 \mathrm{A}_0,\ \mathrm{V}ar(\sqrt{n}\hat{\theta}) = \mathrm{A}_0^{-1}\mathrm{B}_0\mathrm{A}_0^{-1} = \sigma_0^2 \mathrm{A}_0^{-1} \tag{2.178}$$

$$\hat{\mathrm{V}ar}(\hat{\theta}) = \hat{\sigma}^2 (\sum_{i=1}^n \hat{\mathrm{H}}_i)^{-1} \text{ or } \hat{\sigma}^2 (\sum_{i=1}^n \hat{\mathrm{A}}_i)^{-1} \tag{2.179}$$

$$\hat{\sigma}^2 = \frac{1}{n-k} \sum_{i=1}^n \hat{u}_i^2 \tag{2.180}$$

Under NLS assumption 1-3 $\hat{\sigma}^2 (\sum_{i=1}^n \nabla_\theta m(x_i, \hat{\theta})' \nabla_\theta m(x_i, \hat{\theta}))^{-1}$

e.g. $y_i = exp(x_i \theta) + u_i$

$$\hat{\mathrm{V}ar}(\hat{\theta}) = \hat{\sigma}^2 (\sum_{i=1}^n exp(2x_i\hat{\theta}) x_i' x_i)^{-1} \tag{2.181}$$

Variance estimation for two step M-estimation

If $\mathrm{E}[\nabla_\gamma \mathrm{S}(\mathrm{W}, \theta_0, \gamma^*)] = 0$,$(\mathrm{F}_0 = 0)$

$$\hat{\mathrm{V}ar}(\hat{\theta}) = \begin{cases} (\sum_{i=1}^n \hat{\mathrm{H}}_i)^{-1} (\sum_{i=1}^n \hat{\mathrm{S}}_i \hat{\mathrm{S}}_i') (\sum_{i=1}^n \hat{\mathrm{H}}_i)^{-1} \\ (\sum_{i=1}^n \hat{\mathrm{A}}_i)^{-1} (\sum_{i=1}^n \hat{\mathrm{S}}_i \hat{\mathrm{S}}_i') (\sum_{i=1}^n \hat{\mathrm{A}}_i)^{-1} \end{cases} \tag{2.182}$$

$\hat{\mathrm{S}}_i, \hat{\mathrm{H}}_i, \hat{\mathrm{A}}_i$ depend on $\hat{\gamma}_i, \hat{\theta}_i$

Under NLS assumption 1-2

$$\mathrm{E}q(\mathrm{W}, \theta, \gamma^*) = \mathrm{E}\frac{1}{2}(y - m(\mathrm{X}, \theta))^2 / h(\mathrm{X}, \gamma^*) \tag{2.183}$$

$$\mathrm{S}(\mathrm{W}, \theta_0, \gamma^*) = \frac{\partial q(\mathrm{W}, \theta_0, \gamma^*)}{\partial \theta} = -\nabla_\theta m(\mathrm{X}, \theta_0)'(y - m(\mathrm{X}, \theta_0))/h(\mathrm{X}, \gamma^*) = -\nabla_\theta m' u / h \tag{2.184}$$

$$\mathrm{H}(\mathrm{W}, \theta_0, \gamma^*) = \frac{\partial^2 q}{\partial \theta \partial \theta'} = \nabla_\theta m' \nabla_\theta^2 m (y - m)/h \tag{2.185}$$

$$\mathrm{E}[\nabla_\gamma \mathrm{S}(\mathrm{W}, \theta_0, \gamma^*)] = 0 \tag{2.186}$$

$$\hat{\mathrm{V}ar}(\hat{\theta}) = (\sum_{i=1}^n \nabla_\theta m_i' \nabla_\theta m_i / h_i)^{-1} (\sum_{i=1}^n \nabla_\theta \hat{m}_i' \hat{u}_i^2 \nabla_\theta \hat{m}_i / \hat{h}_i)(\sum_{i=1}^n \nabla_\theta m_i' \nabla_\theta m_i / h_i)^{-1} \tag{2.187}$$

WNLS assumption 3 : $\mathrm{V}ar(y|\mathrm{X}) = \sigma_0^2 h(\mathrm{X}, \gamma_0)$

$$\mathrm{B}_0 = \sigma_0^2 \mathrm{E}[\nabla_\theta m' \nabla_\theta m/h] = \sigma_0^2 \mathrm{A}_0 \tag{2.188}$$

$$\mathrm{A}_0 = \mathrm{E}[\nabla_\theta m' \nabla_\theta m/h] \tag{2.189}$$

$$\hat{\mathrm{V}ar}(\hat{\theta}) = \hat{\sigma}^2 (\sum_{i=1}^{n} \nabla_\theta \hat{m}_i' \nabla_\theta \hat{m}_i / h_i)^{-1} \tag{2.190}$$

$$\hat{\sigma}^2 = \frac{1}{n-k} \sum_{i=1}^{n} (\frac{\hat{u}_i^2}{\sqrt{\hat{h}_i}})^2 \tag{2.191}$$

if $\mathrm{F}_0 \neq 0$, $\mathrm{E}[\nabla_\gamma \mathrm{S}(\mathrm{W}, \theta_0, \gamma^*)] \neq 0$

$$\hat{\mathrm{V}ar}(\hat{\theta}) = \begin{cases} (\sum_{i=1}^{n} \hat{\mathrm{H}}_i)^{-1} (\sum_{i=1}^{n} \hat{g}_i \hat{g}_i') (\sum_{i=1}^{n} \hat{\mathrm{H}}_i)^{-1} \\ (\sum_{i=1}^{n} \hat{\mathrm{A}}_i)^{-1} (\sum_{i=1}^{n} \hat{g}_i \hat{g}_i') (\sum_{i=1}^{n} \hat{\mathrm{A}}_i)^{-1} \end{cases} \tag{2.192}$$

$$\hat{g}_i = \hat{\mathrm{S}}_i + \hat{\mathrm{F}}_i + \hat{r}_i \tag{2.193}$$

$$\hat{\mathrm{F}}_i = \frac{1}{n} \sum_{i=1}^{n} \nabla_\gamma \mathrm{S}_i(\hat{\theta}, \hat{\gamma}) \tag{2.194}$$

2.3.2 Numerical Optimization

Newton-Raphson Method

$$\sum_{i=1}^{n} \mathrm{S}(\mathrm{W}_i, \hat{\theta}) = 0 \Rightarrow \hat{\theta} \tag{2.195}$$

$$\sum_{i=1}^{n} \mathrm{S}_i(\theta^{\{g+1\}}) = \sum_{i=1}^{n} \mathrm{S}_i(\theta^{\{g\}}) + [\sum_{i=1}^{n} \mathrm{H}_i(\theta^{\{g\}})](\theta^{\{g+1\}} - \theta^{\{g\}}) + r^{\{g\}} \tag{2.196}$$

$$\text{Let } \sum_{i=1}^{n} \mathrm{S}_i(\theta^{\{g+1\}}) = 0, r^{\{g\}} = 0 \tag{2.197}$$

$$\theta^{\{g+1\}} = \theta^{\{g\}} - [\sum_{i=1}^{n} \mathrm{H}_i(\theta^{\{g\}})]^{-1} [\sum_{i=1}^{n} \mathrm{S}_i(\theta^{\{g\}})] \quad \text{iterative method} \tag{2.198}$$

$\theta^{\{g+1\}}$ is very close to $\theta^{\{g\}}$

1. $|\theta_j^{\{g+1\}} - \theta_j^{\{g\}}|$, for $j = 1, 2, \ldots, k$ be smaller than some small constant.

2. largest percentage change is parameter values be smaller than some small constant.

3. quadratic form

$$[\sum_{i=1}^{n} \mathrm{S}_i(\theta^{\{g\}})]' [\sum_{i=1}^{n} \mathrm{H}_i(\theta^{\{g\}})] [\sum_{i=1}^{n} \mathrm{S}_i(\theta^{\{g\}})] \tag{2.199}$$

drawbacks:1. H_i second derivative 2. H_i is not PD

Berndt Hall Hall and Hausman Method

$$\theta^{\{g+1\}} = \theta^{\{g\}} - r[\sum_{i=1}^{n} S_i(\theta^{\{g\}})S_i(\theta^{\{g\}})']^{-1}[\sum_{i=1}^{n} S_i(\theta^{\{g\}})] \tag{2.200}$$

where r is step size

NL Model $B_0 = \sigma_0^2 A_0$ GIME(Generalized Information Matrix Equality)

$$\sum_{i=1}^{n} S_i(\theta^{\{g\}})S_i(\theta^{\{g\}})' = \sigma_0^2 \sum_{i=1}^{n} H_i(\theta^{\{g\}}) \tag{2.201}$$

stopping rule

$$[\sum_{i=1}^{n} S_i(\theta^{\{g\}})]'[\sum_{i=1}^{n} S_i(\theta^{\{g\}})S_i(\theta^{\{g\}})']^{-1}[\sum_{i=1}^{n} S_i(\theta^{\{g\}})] \sim \chi^2(k) \tag{2.202}$$

其中 k 是 S_i的维度，由CLT $\frac{1}{\sqrt{n}}\sum_{i=1}^{n} S_i(\theta^{\{g\}}) \sim N()$, 可以通过查表对$H_0 : \sum_{i=1}^{n} S_i(\theta^{\{g\}}) = 0$进行检验。

或者 reg 1 on $S_i(\theta^{\{g\}})'$ R^2 is uncentered R^2, 则该检验与$nR^2 = \frac{SSE}{SST} * n$一致。

Gauss-Newton Method

$$\theta^{\{g+1\}} = \theta^{\{g\}} - r[\sum_{i=1} nA_i(\theta^{\{g\}})]^{-1}[\sum_{i=1}^{n} S_i(\theta^{\{g\}})] \tag{2.203}$$

$$A(X_i, \theta_0) = E[H(W_i, \theta_0)|X_i] \tag{2.204}$$

e.g. $y = m(X_i, \theta_0) + u_i$

$$\theta^{\{g+1\}} = \theta^{\{g\}} - r[\sum_{i=1}^{n} \nabla_\theta m(X_i, \theta^{\{g\}})' \sum_{i=1}^{n} \nabla_\theta m(X_i, \theta^{\{g\}})]^{-1}[\sum_{i=1}^{n} \nabla_\theta m(X_i, \theta^{\{g\}})' u_i^{\{g\}}] \tag{2.205}$$

类似之前的reg 1 on $S_i(\theta^{\{g\}})'$, reg $u_i^{\{g\}}$ on $\nabla_\theta m(X_i, \theta^{\{g\}})$, $nR^2 \sim \chi^2(k)$

考虑$y = m(\mathrm{X}_i, \theta_0) + u_i$，Taylor展开有

$$m(\mathrm{X}, \theta^{\{2\}}) \approx m(\mathrm{X}, \theta^{\{1\}}) + \nabla_\theta m(\mathrm{X}, \theta^{\{1\}})(\theta^{\{2\}} - \theta^{\{1\}}) \tag{2.206}$$

$$y - m(\mathrm{X}, \theta^{\{1\}}) \approx \nabla_\theta m(\mathrm{X}, \theta^{\{1\}})(\theta^{\{2\}} - \theta^{\{1\}}) + y - m(\mathrm{X}, \theta^{\{2\}}) \quad \text{regression} \tag{2.207}$$

$$\mathrm{H}_0 : b = \theta^{\{2\}} - \theta^{\{1\}} = 0 \tag{2.208}$$

$$\text{if} b \neq 0, \theta^{\{2\}} = b + \theta^{\{1\}} \tag{2.209}$$

$$\vdots$$

$$\text{until } \theta^{\{i+1\}} - \theta^{\{i\}} = 0 \tag{2.210}$$

e.g. $y_i = \beta_1 x_{1i} + \beta_2 x_{2i}^{\beta_3} + u_i$

$$\nabla_\beta m(x_i, \beta) = (x_1, x_2^{\beta_3}, \beta_2 x_2^{\beta_2}\, ln x_2) \tag{2.211}$$

$$\text{initial value:}(\beta_1, \beta_2, \beta_3) = (1, 1, 1) \tag{2.212}$$

reg $y - x_1 - x_2$ on x_1, x_2, $x_2\, ln x_2$

$$b = \begin{pmatrix} \beta_1^{\{2\}} - \beta_1^{\{1\}} \\ \beta_2^{\{2\}} - \beta_2^{\{1\}} \\ \beta_3^{\{2\}} - \beta_3^{\{1\}} \end{pmatrix} \stackrel{?}{=} 0 \tag{2.213}$$

如何检验$\mathrm{H}_0 : \beta_3 = 1$

1. 用M-estimation 估计 β_3

2. 用t, F, Wald, LM, LR等方法检验。

M-estimation

$$\hat{\beta} \sim \mathrm{N}(\beta_0, \sigma_0^2 \mathrm{A}_0^{-1}/n) \tag{2.214}$$

$$\mathrm{V}\hat{ar}(\hat{\beta}) = \hat{\sigma}^2 (\sum_{i=1}^n \nabla_\beta m(x_i, \hat{\beta})' \nabla_\beta m(x_i, \hat{\beta}))^{-1} \tag{2.215}$$

$$\hat{\sigma}^2 = \frac{1}{n-k} \sum_{i=1}^n \hat{u}_i^2 \tag{2.216}$$

$$t = \frac{\hat{\beta}_{m-estimation} - \beta_0}{\mathrm{SE}(\hat{\beta})} \text{ e.g. } \frac{\hat{\beta}_3 - 1}{\mathrm{SE}(\hat{\beta}_3)} \tag{2.217}$$

另一种检验的方法

$\mathrm{H}_0 : \beta_3 = 1$ impose this restriction reg $y = \tilde{\beta}_1 x_1 + \tilde{\beta}_2 x_2 + u$

$$\tilde{\beta} = (\tilde{\beta}_1, \tilde{\beta}_2, 1) \tag{2.218}$$

$$\tilde{u} = y - \tilde{\beta}_1 x_1 = \tilde{\beta}_2 x_2 \tag{2.219}$$

$$\nabla_\beta m(x,\beta) = (x_1, x_2, \tilde{\beta}_2 x_2 \, ln x_2) \tag{2.220}$$

$$\tag{2.221}$$

reg $\tilde{u}$ on $\nabla_\beta m(x,\beta) \Leftrightarrow$ reg $\tilde{u} = \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 \tilde{\beta}_2 x_2 \, ln x_2$

$\alpha_3 = 0 \overset{?}{\Leftrightarrow} \mathrm{H}_0 : \beta_3 = 1$

将$\tilde{\beta}$作为某一次迭代过程的得到的值，则该回归方程的系数表示两次迭代的差，所以$\alpha_3 = 0$表示$\beta_3^{\{i+1\}} = \beta_3^{\{i\}}$

2.4 Quantile Regression

Let y_i denote a random draw from a population

$0 < \tau < 1$ $q(\tau)$ is a τ-th quantitle, if $\mathrm{P}(y_i \le q(\tau)) = \tau$ $\mathrm{P}(y_i \ge q(\tau)) = 1 - \tau$

$\mathrm{Q}uantile_\tau(y_i)$: τ-th quantile of y_i

$\mathrm{Q}uantile_\tau(y_i|x_i) = \beta_0(\tau) + x_i\beta_1$

考虑最小二乘法 $min \sum_{i=1}^n (y_i - q)^2$

F.O.C $\sum_{i=1}^n 2(y_i - q)(-1) = 0 \Rightarrow, q = \frac{1}{n}\sum_{i=1}^n y_i \to \mathrm{E}y_i, q_{x_0} = \mathrm{E}(y|x = x_0) = \beta x_0$

考虑Least Absolute Deviation(LAD) $\min_q \sum_{i=1}^n |y_i - q|$

$$|y_i - q| = 1(y_i \ge q)(y_i - q) + 1(y_i < q)(q - y_i) \tag{2.222}$$

$$\begin{aligned} \text{F.O.C} \quad & \sum_{i=1}^n (-1)\, 1(y_i \ge q) + 1(y_i < q) \\ &= \sum_{i=1}^n (-1)[1 - 1(y_i < q)] + \sum_{i=1}^n 1(y_i < q) \\ &= -n + 2\sum_{i=1}^n 1(y_i < q) = 0 \end{aligned} \tag{2.223}$$

$$\Rightarrow \frac{1}{n}\sum_{i=1}^n 1(y_i < q) = 2 \tag{2.224}$$

$$\mathrm{E}\, 1(y_i < q) = \frac{1}{2} \tag{2.225}$$

$$\mathrm{P}(y_i < q) = \frac{1}{2}, q = \mathrm{M}edian(y_i) \tag{2.226}$$

对于其他quantile,给绝对值>0和<0分配不同得权重

$$\mathrm{L}(e) = \mathrm{E}e^2 \quad \mathrm{E}(y|x) \tag{2.227}$$

$$\mathrm{L}(e) = \mathrm{E}|e| \quad \mathrm{M}ed(y|x) \tag{2.228}$$

$$\mathrm{L}(e) = \begin{cases} \mathrm{E}(1-\tau)|e| & e < 0 \\ \mathrm{E}\tau|e| & e \geq 0 \end{cases} \quad \mathrm{Q}uant_\tau(y|x) \min \mathrm{L}|e| = \min \mathrm{E}\{[\tau\, 1(y_i - q \geq 0) + (1-\tau)\, 1(y_i - q < 0)]|y_i - q|\} \tag{2.229}$$

$$\mathrm{C}_\tau(e) = [\tau\, 1(e) \geq 0) + (1-\tau)\, 1(e < 0)]|e| \tag{2.230}$$

$$\min_{\alpha,\beta} \sum_{i=1}^{n} \mathrm{C}_\tau(y_i - \alpha - x_i\beta) \tag{2.231}$$

$$\text{in M-estimation } \hat{\theta}(\tau) = (\hat{\alpha}(\tau), \hat{\beta}(\tau)') \to \theta_0(\tau) = (\alpha_0(\tau), \beta_0(\tau)') \tag{2.232}$$

题外话：关于LAD

OLS is sensitive to changes in data points **no-robust**(Mean受到outliers的影响)

LAD is insensitive to changes in data points **robust** (Median不受outliers的影响)

$y_i = \alpha_0 + x_i\beta_i + u_i$

assumption ①

$\mathrm{D}(y_i|x_i) = \alpha_0 + x_i\beta_i + \mathrm{D}(u|x_i)$ if $\mathrm{D}(u_i|x_i)$ is symmetric about zero

$\mathrm{E}(u_i|x_i) = \mathrm{M}ed(u_i|x_i) = 0$

$\mathrm{E}(y_i|x_i) = \alpha_0 + x_i\beta_0 = \mathrm{M}ed(y_i|x_i)$

assumption ②

$\mathrm{D}(u_i|x_i) = \mathrm{D}(u_i)$ and $\mathrm{E}u_i = 0$

$\mathrm{E}(y_i|x_i) = \alpha_0 + x_i\beta_0 + \cancel{\mathrm{E}(u_i|x_i)}$

$\mathrm{M}ed(y_i|x_i) = \alpha_0 + x_i\beta_0 + \mathrm{M}ed(u_i|\cancel{x_i})$ let $\mathrm{M}ed(u_i) = \eta_0$ $\mathrm{M}ed(y_i|x_i) = (\alpha_0 + \eta_0) + x_i\beta_0$

LAD M-estimation 并没有假设分布,通常情况下assumption①和②不被满足，但有一种不满足假设，但可以用LAD代替OLS的方法。

考虑收入y，通常是一个右偏的分布，因此take $\ln y$,$\mathrm{E}[\ln y_i|x_i] = \alpha_0 + x_i\beta_0 + \mathrm{E}[u_i|x_i]$

under assumption ①

$$\mathrm{E}[\ln y_i|x_i] = \alpha_0 + x_i\beta_0 \tag{2.233}$$

$$e^{\ln y_i} = e^(\alpha_0 + x_i\beta_0 + u_i) = y_i \tag{2.234}$$

$$\mathrm{E}[y_i|x_i] = e^(\alpha_0 + x_i\beta_0)\mathrm{E}(e^{u_i}|x_i) \tag{2.235}$$

$\mathrm{E}(e^{u_i}|x_i)$起算起来很麻烦

LAD得到

$$\mathrm{M}ed(\ln y_i|x_i) = \alpha_0 + x_i\beta_0 \tag{2.236}$$

LAD无法规避计算均值$\mathrm{E}(e^{u_i}|x_i)$的问题

under assumption ②

$$\mathrm{E}[\ln y_i|x_i] = \alpha_0 + x_i\beta_0 + \cancel{\mathrm{E}[u_i|x_i]} \tag{2.237}$$

$$\mathrm{E}(y_i|x_i) = e^{\alpha_0+x_i\beta_0}\mathrm{E}(e^{u_i}|\cancel{x_i}) \tag{2.238}$$

不需要计算$\mathrm{E}(e^{u_i}|x_i)$，只需要计算$\mathrm{E}(e^{u_i})(\mathrm{E}(e^{\hat{u}_i}))$

$\mathrm{M}ed(\ln y_i|x_i) = \alpha_0 + x_i\beta_0 + \mathrm{M}ed(u_i|x_i) = (\alpha_0 + \eta_0) + x_i\beta_0)$

$\alpha_0 + \eta_0$是一起估计出来的。

$$u_i = \mathrm{M}ed(u_i) + \tilde{u}_i = \eta_0 + \tilde{u}_i \tag{2.239}$$

$$u_i - \mathrm{M}ed(u_i) = \tilde{u}_i \quad \mathrm{M}ed(\tilde{u}_i) = 0 \tag{2.240}$$

$$e^{u_i} = e^{\eta_0}e^{\tilde{u}_i} \tag{2.241}$$

where $\tilde{u}_i$ is the error term in $\ln y_i = \alpha_0 + \eta_0 + x_i\beta_0 + \tilde{u}_i$ by using LAD

$$\mathrm{E}e^{u_i} = e^{\eta_0}\mathrm{E}e^{\tilde{u}_i} \tag{2.242}$$

$$e^{\alpha_0+x_i\beta_0}\mathrm{E}(e^{u_i}) = e^{\alpha_0+x_i\beta_0}e^{\eta_0}\mathrm{E}e^{\tilde{u}_i} \tag{2.243}$$

虽然无法完全规避计算均值的影响，但$\mathrm{E}(y_i|x_i) = e^{\alpha_0+x_i\beta_0}e^{\eta_0}\mathrm{E}e^{\tilde{u}_i}$可以用LAD计算得到。

上述讨论表明试图take log 并完全用LAD的方法，规避E(.)是做不到的

OLS的优点是可以使用Law of Iterated Expectation $\mathrm{E}(x_iy_i) = \mathrm{E}[x_i\mathrm{E}(y_i|x_i)]$ 但是 $\mathrm{M}ed(x_iy_i) = \mathrm{M}ed[x_i\mathrm{M}ed(y_i|x_i)]$,其次M$ed$不能进行线性计算。

考虑$y_i = a_i + x_ib_i$ a_i, b_i are random and independent of x_i

$$\mathrm{E}(y_i|x_i) = \mathrm{E}(a_i|\cancel{x_i}) + x_i\mathrm{E}(b_i|\cancel{x_i}) \tag{2.244}$$

$$= \alpha_0 + x_i\beta_0 \quad \text{OLS average partial effect} \tag{2.245}$$

$\mathrm{M}ed(y_i|x_i)$ =$\mathrm{M}ed(a_i|x_i)$+$x_i\mathrm{M}ed(b_i|x_i)$

=$\mathrm{M}ed(a_i)$ +x_i $\mathrm{M}ed(b_i)$

y_i	a_i	x_i	b_i
3.1	2.1	1	1
4	2	1	2
2.1	0	1	1.1

$\mathrm{L}eftside = 3.1 \neq \mathrm{R}ightside = 4$

题外话结束

$$y_i = x_i\theta_0 + u_i, \quad \mathrm{Q}uant_\tau(u_i|x_i) = 0 \tag{2.246}$$

$$q(w_i,\theta) = \tau\,1(y_i - x_i\theta \geq 0)(y - x_i\theta) - (1-\tau)\,1(y_i - x_i\theta < 0)(y_i - x_i\theta) \tag{2.247}$$

之前求解的过程中，在尖点的导数是错误的但是$0 = y_i - x_i\theta_0 = u_i$, $\mathrm{P}(u_i = 0) = 0$在错误点求导的概率是0，因此$\hat{\theta} \to \theta_0$ as $n \to \infty$

$$\mathrm{S}_i(\theta) = -x_i'\{\tau\,1(y_i - x_i\theta \geq 0) - (1-\tau)\,1(y_i - x_i\theta < 0)\} \tag{2.248}$$

$$\mathrm{H}(x_i,\theta) = \frac{\partial \mathrm{S}_i}{\partial \theta'} = 0 \tag{2.249}$$

所以有A_i不可逆，在之前的求解过程中包括$\frac{\partial \mathrm{E}q}{\partial \theta} = \mathrm{E}\frac{\partial q}{\partial \theta}$当q连续时，E,$\partial$可交换，但在这里有S不连续，所以不能交换。

重新考虑计算$\mathrm{E}[\mathrm{S}_i(\theta)]$,利用$\mathrm{E}(\mathrm{E}(\,|x))$,首先计算$\mathrm{E}[\mathrm{S}_i(\theta)|x_i]$

$$\begin{aligned}
\mathrm{E}[\mathrm{S}_i(\theta)|x_i] &= -x_i'\{\tau\mathrm{P}(y_i - x_i\theta \geq 0|x_i) - (1-\tau)\mathrm{P}(y_i - x_i\theta < 0|x_i)\} \\
&= -x_i'\{\tau\mathrm{P}(u_i \geq x_i(\theta-\theta_0)|x_i) - (1-\tau)\mathrm{P}(u_i < x_i(\theta-\theta_0)|x_i)\} \\
&= -x_i'\{\tau[1 - \mathrm{F}_u(x_i(\theta-\theta_0)|x_i)] - (1-\tau)\mathrm{F}_u(x_i(\theta-\theta_0)|x_i)]\} \\
&= -x_i'\{\tau - \mathrm{F}_u(x_i(\theta-\theta_0)|x_i)\}
\end{aligned}$$

因为F_u连续，所以$\mathrm{E}\frac{\partial \mathrm{E}(|x)}{\partial \theta'} = \frac{\partial \mathrm{E}[\mathrm{E}(|x)]}{\partial \theta'} = \frac{\partial \mathrm{E}()}{\partial \theta'}$

$$\frac{\partial \mathrm{E}[\mathrm{S}_i(\theta|x_i)]}{\partial \theta'} = f_u(x_i(\theta-\theta_0)|x_i)x_i'x_i \tag{2.250}$$

$$\mathrm{A}_0 = \mathrm{A}(\theta_0) = \mathrm{E}[f_u(0|x_i)x_i'x_i] \tag{2.251}$$

$$\sqrt{n}(\hat{\theta}-\theta_0) = \mathrm{A}_0^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathrm{S}_i(\theta_0)+o_p(1) \sim \mathrm{N}(0, \mathrm{A}_0^{-1}\mathrm{B}_0\mathrm{A}_0^{-1}) \tag{2.252}$$

对于$\mathrm{E}[\mathrm{S}_i(\theta_0)|x_i] = -x_i[\tau-\mathrm{F}_u(0|x_i)]$,因为$\mathrm{F}_u = \mathrm{P}(u_i < x_i(\theta-\theta_0)|x_i), \mathrm{F}_u(0) = \mathrm{P}(y_i - x_i\theta_0 < 0|x_i)$，$x_i\theta_0$ is the τ percentile of y, 所以$\mathrm{P}(y_i < x_i\theta_0|x_i) = \tau$,即满足$\mathrm{E}[\mathrm{S}_i(\theta_0)] = \mathrm{E}[\mathrm{E}[\mathrm{S}_i(\theta_0)|x_i]] = \mathrm{E}[0] = 0$

$$\begin{aligned}
\mathrm{B}_0 &= \mathrm{E}[\mathrm{S}_i(\theta_0)\mathrm{S}_i(\theta_0)'] \\
&= \mathrm{E}[x_i'x'(\tau^2 1(y \geq x_i\theta_0)+(1-\tau)^2 1(y_i < x_i\theta_0)] \\
&= \mathrm{E}[x_i'x_i(\tau^2(1-\tau)+(1-\tau)^2\tau)] \\
&= \tau(1-\tau)\mathrm{E}[x_i'x_i]
\end{aligned} \tag{2.253}$$

$$\hat{\mathrm{B}}_0 = \tau(1-\tau)\frac{1}{n}\sum_{i=1}^{n}x_i'x' \tag{2.254}$$

f_u不好求，所以根据导数的定义

$$\begin{aligned}
f_u(0|x_i) &\approx [\mathrm{F}_u(h|x_i)-\mathrm{F}_u(-h|x_i)]/2h \\
&= \mathrm{P}(-h \leq u_i \leq h|x)_i)/2h \\
&= \mathrm{P}(|u_i| \leq h|x_i)/2h \\
&= \mathrm{E}[1(|u_i| \leq h)|x_i]/2h
\end{aligned}$$

$$\begin{aligned}
\mathrm{A}_0 &= \mathrm{E}[f_u(0|x_i)x_i'x_i] \\
&= \mathrm{E}\{\mathrm{E}[1(|u_i| \leq h)|x_i]x_i'x_i\}/2h \\
&= \frac{1}{2h}\mathrm{E}\{1(|u_i| \leq h)x_i'x_i\} \\
\hat{\mathrm{A}}_0 &= \frac{1}{2nh}\sum_{i=1}^{n}1(|\hat{u}_i| < h)x_i'x_i
\end{aligned}$$

h is called bandwith or smoothing parameter. 利用非参估计，不假设pdf = f(x)。

如果有$f_u(0|x_i)=f_i(0)$即u和x_i独立,有

$$\hat{Var}(\sqrt{n}(\hat{\theta}))=\frac{\tau(1-\tau)}{[\hat{f}_u(0)]^2}(\frac{1}{n}\sum_{i=1}^{n}x_i'x')^{-1} \tag{2.255}$$

$$\hat{f}_u(0)=\frac{1}{2nh}\sum_{i=1}^{n}1[|u_i|\leq h] \tag{2.256}$$

2.5 Time Series

目标：

1. 假定一个probability model来表示时间序列数据
2. 估计模型的参数
3. 时间序列关注模型的R^2(截面数据通常不需要做预测，所以只关注变量之间的因果关系，不要求高R^2)
4. 用模型解释数据，帮助我们加深对数据的理解
5. 预测

Definition 2.5.1. *Strictly Stationary*

$(y_{t1},\ldots,y_{tk})$ $(y_{t1+h},\ldots,y_{tk+h})$ 分布相同 $\forall(t1,\ldots,tk)$ *and* $k,h=1,2,3,\ldots$

Definition 2.5.2. *Weekly Stationary (Covariance Stationary)*

对于时间序列$\{x_t\}$，满足

$$\begin{cases}\mathrm{E}(x_t)=\mu\\ \mathrm{V}ar(x_t)=\gamma(0)<\infty\\ \mathrm{C}ov(x_{t+h},x_t)=\gamma(h)<\infty,\forall h=\pm1\pm2,\ldots\end{cases}$$

w.n. (White Noise)

x_t is white noise If i)E$x_t=0$ ii) E$x_t^2=\sigma^2$ iii)E$x_tx_s=0 \quad \forall s\neq t$

Tread Stationary

$y_t=\alpha+\beta t+z_t$ $z_t\sim w.n.(0,\sigma^2)$ E$y_t=\alpha+\beta t$

Random Walk

$y_t=y_{t-1}+z_t$ $z_t\sim w..n.(0,\sigma^2)$ then $y_t=y_{t-2}+z_{t-1}+z_t=z_1+z_2+\cdots+z_t$(assume $y_0=0$)

E$y_t=0$ V$ar(y_t)=t\sigma^2$

Random Walk with drift

$y_t = \mu + y_{t-1} + z_t = z_1 + z_2 + \cdots + z_t + t\mu$

$\mathrm{E}y_t = t\mu$ $\mathrm{V}ar(y_t) = t\sigma^2$

定义ACVF(auto covariance function)$\gamma(h) = cov(y_{t+h}, y_t)$ $\forall h = 0, \pm1, \pm2, \ldots$

$\gamma(0) = \mathrm{V}ar(y_t) = \mathrm{V}ar(y_{t+h})$

ACF(auto correlation function) $\rho(h) = \frac{\gamma(h)}{\gamma(0)} = corr(y_{t+h}, y_t)$ $\rho(0) = 1$

PACF(partial auto correlation function)ρ^*, y_{t+h}与y_t的直接关系(ACF包含了两个变量的直接关系和间接关系) $\rho^*(h) = corr[y_t - \mathrm{E}(y_t|y_{t-1}, \ldots, y_{t-h+1}), y_{t-h} - \mathrm{E}(y_{t-h}|y_{t-1}, \ldots, y_{t-h+1})]$

$$y_t = \beta_1 y_{t-1} + \beta_2 y_{t-2} + \cdots + \beta_{h-1} y_{t-h+1} + \rho^*(h) y_{t-h} + error \tag{2.257}$$

$$\text{in OLS} : y = x_1\beta_1 + x_2\beta_2 + u \tag{2.258}$$

$$\mathrm{M}_{x_1} y = \mathrm{M}_{x_1} x_1 \beta_1 + \mathrm{M}_{x_1} x_2 \beta_2 + \mathrm{M}_{x_1} u \tag{2.259}$$

$$\mathrm{M}_{x_1} y = \mathrm{M}_{x_1} x_2 \beta_2 + \mathrm{M}_{x_1} u \tag{2.260}$$

$$y_t - \mathrm{E}(y|y_{t-1}, \ldots, y_{t-h+1}) = (y_{t-h} - \mathrm{E}(y_{t-h}|y_{t-1}, \ldots, y_{t-h+1})\beta \tag{2.261}$$

$$\beta = [(y_{t-h} - \mathrm{E}(y_{t-h}|y_{t-1}, \ldots, y_{t-h+1})'(y_{t-h} - \mathrm{E}(y_{t-h}|.)]^{-1}(y_{t-h} - \mathrm{E}(y_{t-h}|.)'(y_t - \mathrm{E}(y|.)) \tag{2.262}$$

$$= \mathrm{V}ar^{-1}\mathrm{C}ov = \rho^*(h) \tag{2.263}$$

因此求$\rho^*(h)$,只需要做OLS

$$(\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_{h-1}, \rho^*(h)) = \Gamma_h^{-1}\gamma_h \tag{2.264}$$

$$\Gamma_h = (\gamma(i-j))_{i,j=1}^h, \gamma_h = (\gamma(1), \gamma(2), \ldots, \gamma(h))' \tag{2.265}$$

$$= \begin{pmatrix} \gamma(0) & \gamma(1) & \ldots & \gamma(h-1) \\ \gamma(1) & \gamma(0) & & \vdots \\ \vdots & & \ddots & \\ \gamma(h-1) & \gamma(h-2) & \ldots & \gamma(0) \end{pmatrix} \tag{2.266}$$

类似OLS $\hat{\beta} = (\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}'y$, $\mathrm{X}'\mathrm{X}$是var-cov matirx,$\mathrm{X}'y$是Cov

$|\rho(h)| \to 0$ as $h \to \infty$,一般情况下，如果$\not\to 0$则不平稳。

根据趋向于0速度的差异可以分为

1. short term memory time series data
2. medium term memory time series data
3. long term memory time series data (不讲)

short term memory if $\rho(h) \neq 0$ until h>q, where q is a finite integer medium term memory if $|\rho(h)| = O(|\xi|^h)$, where $|\xi| < 1$

ARMA(p,q) Model

$$x_t = \overbrace{\phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + z_t}^{\mathrm{AR}(p)} \underbrace{+ \theta_1 z_{t-1} + \cdots + \theta_q z_{t-q}}_{\mathrm{MA}(q)} \tag{2.267}$$

AR(p) auto regressive

MA(q) moving average

ARMA平稳性要求AR部分特征方程的特征根>1,落在单位圆外,MA的特征根不影响ARMA的平稳性，其特征根>1表示MA可逆，转化为AR。

AR(p)

$$0 = \phi(z) = 1 - \phi_1(z) - \cdots - \phi_p z^p \tag{2.268}$$

is the characteristic form of the AR part of ARMA model

MA(q)

$$0 = \theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q \tag{2.269}$$

is the characteristic form of the MA part of ARMA model

ARMA(0) $\Leftrightarrow$ AR(p)

$$x_t = \psi_1 x_{t-1} + \cdots + \psi_p x_{t-p} + z_t\, z_t \sim \mathrm{WN}(0, \sigma^2) \tag{2.270}$$

ARMA(0,q) $\Leftrightarrow$ MA(q)

$$x_t = z_t + \theta_1 z_{t-1} + \cdots + \theta_q z_{t-q}\, z_t \sim \mathrm{WN}(0, \sigma^2) \tag{2.271}$$

An ARMA process is stationary if $\phi(z) = 0$ only when $|z| > 1$

An ARMA process is invertible if $\theta(z) =$ only when $|z| > 1$

下面讨论ARMA过程是short term还是medium term stationary

(1) AR(1):$x_t = \phi x_{t-1} + z_t$ $|\psi| < 1$ $z_t \sim \mathrm{WN}(0, \sigma^2)$

$$\mathrm{E}z_t x_t = \phi \mathrm{E}x_{t-1}z_t + \mathrm{E}z_t^2 = \sigma^2 \tag{2.272}$$

$$\mathrm{E}x_t^2 = \mathrm{E}\phi x_t x_{t-1} + \mathrm{E}x_t z_t \tag{2.273}$$

$$\gamma(0) = \phi\gamma(1) + \sigma^2 \tag{2.274}$$

$$\mathrm{E}x_{t-1}x_t = \phi \mathrm{E}x_{t-1}^2 + \mathrm{E}x_{t-1}z_t \tag{2.275}$$

$$\gamma(1) = \phi\gamma(0) + 0 \tag{2.276}$$

$$\vdots \tag{2.277}$$

$$\gamma(h) = \phi\gamma(h-1) = \phi^h\gamma(0) \tag{2.278}$$

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)} = \phi^h \tag{2.279}$$

AR is medium-term memory

(2) MA(1) $x_t = z_t + \theta z_{t-1}$ $z_t \sim \mathrm{WN}(0, \sigma^2)$

$$\rho(h) = \begin{cases} \dfrac{\theta}{1+\theta^2} & \text{if } h = 1 \\ 0 & \text{if } h > 1 \end{cases} \tag{2.280}$$

(3) MRMA(1,1) $x_t = \phi x_{t-1} + z_t + \theta z_{t-1}$ $z_t \sim \mathrm{WN}(0, \sigma^2)$

$$\rho(h) = \begin{cases} \dfrac{\theta+\phi+\theta^2\phi+\theta\phi^2}{1+\theta^2+2\theta\phi} & \text{if } h = 1 \\ \phi^{h-1}\dfrac{\theta+\phi+\theta^2\phi+\theta\phi^2}{1+\theta^2+2\theta\phi} & \text{if } h > 1 \end{cases} \tag{2.281}$$

AR(p),MA(q),ARMA(p,q)

AR(p) ACF:$\rho(h) \to 0$ as $h \to \infty$, PACF:$\rho^*(h) = 0$ if $h > p$

MA(q) ACF:$\rho(h) = 0$ if $h > p$, PACF:$\rho^*(h) \to 0$ as $h \to \infty$

关于MA的PACF$x_t = \theta_1 x_{t-1} + \theta_2 x_{t-2} + \dots$，考虑MA(1)

$$x_t = z_t - \theta z_{t-1} \tag{2.282}$$

$$x_{t-1} = z_{t-1} - \theta z_{t-2} \tag{2.283}$$

$$\vdots$$

$$\mathrm{MA}(1): x_t = -\theta x_{t-1} - \theta^2 x_{t-2} - \cdots - \theta^n x_{t-n} \; |\theta| < 1 \tag{2.284}$$

AR(3)不能很好的拟合模型，但是AR(50)可以，即PACF$\to 0$，可以考虑用类似ARMA(2,2)来拟合，

同理MA(50)，ACF$\rightarrow 0$,也可以考虑用ARMA(p,q)拟合。

if we find a AR(p) model fit data well when p is very large, we can add MA part to fit the data.

Estimation of ARMA(p,q)

(1) OLS: AR(p) 可以 ARMA(1,1) 不行 $y_t = \phi y_{t-1} + z_t + \theta z_{t-1}$ y_t y_{t-1}相关

(2) Method of Moments:

$$\begin{cases} \gamma(0) = \dots \\ \gamma(1) = \dots \end{cases} \tag{2.285}$$

(3) MLE: ARMA(p,q) $x_t = \phi_1 x_{t-1} + \cdots + \phi_q x_{t-q} + z_t + \theta_1 z_{t-1} + \cdots + \theta_q z_{t-q}$ $z_t \sim \mathrm{N}(0, \sigma^2) \Rightarrow x_t \sim \mathrm{N}$

(4) NLS: example: MA(1)

$$x_t = \mu + z_t - \alpha_1 z_{t-1} \tag{2.286}$$

$$x_1 = \mu - \alpha_1 z_0 + z_1 \tag{2.287}$$

$$x_2 = \mu - \alpha_1 z_1 + z_2 \tag{2.288}$$

$$= (\mu + \alpha_1 \mu) - \alpha_1 x_1 - \alpha_1^2 z_0 + z_2 \tag{2.289}$$

$$x_t = (\sum_{s=0}^{t-1} \alpha_1^s)\mu - \sum_{s=1}^{t-1} \alpha_1^s x_{t-s} - \alpha^t z_0 + z_t \tag{2.290}$$

Distribution

if $\{y_t\}_{t=1}^n$ is an AR(p) with $z_t \sim iid(0, \sigma^2)$ $\hat{\Phi}_p$ is the estimation of Φ_p, then

$$\sqrt{n}(\hat{\Phi}_p - \Phi_p) \sim \mathrm{N}(0, \sigma^2 \Gamma_p^{-1}) \tag{2.291}$$

where Γ_p is the covariance matrix, $[\gamma(i-j)]_{i,j=1}^p$

If $\{y_t\}_{t=1}^n$ is an AR(p) with $z_t \sim iid(0, \sigma^2)$ and If $\hat{\Phi}_h = (\hat{\phi}_1, \hat{\phi}_2, \dots, \hat{\phi}_h)' = \hat{\Gamma}_h^{-1} \gamma_h, h > p$

$$\sqrt{n}(\hat{\Phi}_h - \Phi_h) \sim \mathrm{N}(0, \sigma^2 \Gamma_h^{-1}) \tag{2.292}$$

for $h > p$ $\sqrt{n}\hat{\phi}_h \sim \mathrm{N}(0, 1)$

e.g. $y_t = \rho y_{t-1} + u_t \quad u_t \sim \mathrm{WN}(0, \sigma^2)$ true model is $\rho = 0$

$$\hat{\rho} = \frac{\sum_t y_{t-1} y_t}{\sum_t y_{t-1}^2} = \rho + \frac{\sum_t y_{t-1} u_t}{\sum_t y_{t-1}^2} \tag{2.293}$$

$$\sqrt{n}(\hat{\rho} - \rho) = \frac{\frac{1}{\sqrt{n}} \sum_t y_{t-1} u_t}{\frac{1}{n} \sum_t y_{t-1}^2} \tag{2.294}$$

一般情况下对分子使用CLT,对分母使用LLN,可以得到分布，但在现在的假设条件下，分子和分母不是iid的,不能依概率收敛到总体分布,因此考虑依MSE收敛。

$$\mathrm{E}y_{t-1}^2 = \sigma_y^2 \tag{2.295}$$

$$\mathrm{E}\frac{1}{n}\sum_t y_t^2 = \sigma_y^2 = \mathrm{E}y_{t-1}^2 \tag{2.296}$$

$\mathrm{B}ias = 0, \mathrm{B}ias^2 = 0$

$|y|^4 = \mathrm{C} < \infty$

$$\mathrm{V}ar(\frac{1}{n}\sum_t y_{t-1}^2) = \frac{1}{n^2}[\sum_t \mathrm{V}ar(y_{t-1}^2) + 2\sum_t\sum_{s>t}\mathrm{C}ov(y_{t-1}^2, y_{s-1}^2)] \tag{2.297}$$

$$= \mathrm{O}(\frac{1}{n}) + (?) \tag{2.298}$$

$$y_s^2 = \rho^2 y_{s-1}^2 + u_s^2 + 2\rho y_{s-1}u_s \tag{2.299}$$

$$\begin{aligned}\mathrm{C}ov(y_s^2, y_{s-1}^2) &= \mathrm{C}ov(\rho^2 y_{s-1}^2, y_{s-1}^2) + 0 \\ &= \rho^2 \mathrm{V}ar(y_{s-1}^2)\end{aligned} \tag{2.300}$$

$$\mathrm{C}ov(y_s^2, y_{s-t}^2) = \rho^{2t}\mathrm{C} \tag{2.301}$$

$$\begin{aligned}\mathrm{V}ar(\frac{1}{n}\sum_t y_{t-1}^2) &= \frac{1}{n^2}[n\mathrm{C} + 2\sum_t\sum_{s>t}\mathrm{C}\rho^{2(s-t)}] \\ &= \mathrm{O}(\frac{1}{n}) + \mathrm{O}(\frac{1}{n}) = \mathrm{O}(\frac{1}{n}) \to 0\end{aligned} \tag{2.302}$$

$\mathrm{B}ias^2 \to 0, \mathrm{V}ar \to 0$ 所以分母收敛到$\mathrm{E}y_{t-1}^2 = \sigma_y^2$

考虑分子，假设分子是Martigale Difference Process,即有$\mathrm{E}(y_{t-1}u_t|\mathrm{I}_{t-1}) = 0\ (y_{t-1})\mathrm{E}[u_t|\mathrm{I}_{t-1}] = 0, \mathrm{I}_{t-1}$是t-1期及之前的信息集

Theorem 2.5.1. *Martingale Difference CLT*

分子 $\frac{1}{\sqrt{n}}\sum y_{t-1}u_t \sim \mathrm{N}(0, (\mathrm{E}y_{t-1})^2(\mathrm{E}u_t)^2) = \mathrm{N}(0, \sigma_y^2\sigma_u^2)(y_{t-1}$和$u_t$独立)

所以 $\sqrt{n}(\hat{\rho} - \rho) \to \mathrm{N}(0, \frac{\sigma_u^2}{\sigma_y^2})$

$$y_t = \rho y_{t-1} + u_t$$

$$\sigma_y^2 = \rho^2\sigma_y^2 + \sigma_u^2$$

$$\frac{\sigma_u^2}{\sigma_y^2} = 1 - \rho^2 \sqrt{n}(\hat{\rho} - \rho) \to \mathrm{N}(0, 1-\rho)$$

under true model $\rho = 0\ \sqrt{n}\hat{\rho} \sim \mathrm{N}(0,1)$

上述讨论为了区分AR(p)和MA(q)

Theorem 2.5.2. *If the true data follows the AR(p) process, then we can distinguish it from a MA process by testing whether* $\rho^*(h) = 0$ *for h>p*

$$\sqrt{n}(\hat{\rho}^*_h - \rho^*_h) \sim \mathrm{N}(0,1) \tag{2.303}$$

Theorem 2.5.3. *If the true data follows the MA(q) process, then we can distinguish it from a AR process by testing whether* $\rho(h) = 0$ *for h>q*

$$\sqrt{n}(\hat{\rho}_h - \rho_h) \sim \mathrm{N}(0,\mathrm{W}) \tag{2.304}$$

$$\hat{\rho}_h = (\hat{\rho}(1),\dots,\hat{\rho}(h))'$$

$$\rho_h = (\rho(1),\dots,\rho(h))'$$

$$\mathrm{W} = \sum_{k=-\infty}^{\infty}\{\rho(k+i)\rho(k+j)+\rho(k-j)\rho(k+j)+2\rho(i)\rho(j)\rho(k)^2-2\rho(i)\rho(k)\rho(k+j)+2\rho(j)\rho(k)\rho(k+i)\}$$

$$\sqrt{n}\hat{\rho}(h) \sim \mathrm{N}(0,\mathrm{V}), \quad \mathrm{V} = 1+2\sum_{s=1}^{q}\rho^2(s)$$

对于假设 $\mathrm{H}_0 : \rho(h) = 0 \mathrm{H}_1 : \rho(h) \neq 0$ for each h>0 $y_t = \underset{=0?}{\rho}\, y_{t-1} + u_t$

(1) $\sqrt{n}\hat{\rho}(h) \sim \mathrm{N}(0,1)$

(2) $\sqrt{n}\hat{\rho}(h) \sim \mathrm{N}(0,\mathrm{V})$, $\mathrm{V} = 1+2\sum_{s=1}^{q}\rho^2(s) = 1+2\rho^2(1) = 1$

(3) (1)(2)通过回归方程构造ρ的表达式，(3)根据ρ的定义$\hat{\rho}(1) = \gamma(1) = \frac{\frac{1}{\sqrt{n}}\sum y_t y_{t-1}}{\frac{1}{n}\sum y_t^2}$

$\mathrm{E}[y_t y_{t-1}|\mathrm{I}_{t-1}] = y_{t-1}\mathrm{E}[y_t|\mathrm{I}_{t-1}] = 0$, 根据MDCLT

$$\frac{1}{\sqrt{n}}\sum y_t y_{t-1} \sim \mathrm{N}(0,\sigma_y^4) \tag{2.305}$$

$$\sqrt{n}\hat{\rho}(1) \sim \mathrm{N}(0,1) \tag{2.306}$$

上述讨论检验的是1个h,问题是如何检验多个h

$\mathrm{H}_0 : \rho(h) = 0, h = 1,2,\dots,p$

$\mathrm{H}_1 : \rho(h) \neq 0$, for some h

$$\mathrm{Q} = n\sum_{h=1}^{o}\hat{\rho}(h)^2 \sim \chi^2(p) \tag{2.307}$$

上述内容讨论的是data的序列相关性，下面讨论residual/error的序列相关性，Test error serial correlation。

问,如果error没有序列相关性，说明模型完备，是否可以用刚才的方法检验error的序列相关性。

核心在于error的分布

可用的情况：

$$y_t = x_t'\beta + u_t \quad y_t = x_t'\hat{\beta} + \hat{u}_t \quad \mathrm{E}[u_t|x_1, x_2, \ldots, x_t] = 0 \tag{2.308}$$

$$\gamma(1) = \mathrm{E}[u_{t-1}u_t] \overset{?}{=} 0 \tag{2.309}$$

$$\hat{u}_t = y_t - x_t'\beta = u_t - x_t'(\hat{\beta} - \beta) \tag{2.310}$$

$$\hat{u}_{t-1} = y_{t-1} - x_{t-1}'\beta = u_{t-1} - x_{t-1}'(\hat{\beta} - \beta) \tag{2.311}$$

$$\hat{\gamma}(1) = \frac{1}{n}\sum_t \hat{u}_{t-1}\hat{u}_t = \frac{1}{n}\sum[u_t u_{t-1} - u_{t-1}x_t'(\hat{\beta}-\beta) - u_t x_{t-1}'(\hat{\beta}-\beta) + (\hat{\beta}-\beta)x_{t-1}'x_t(\hat{\beta}-\beta)] \tag{2.312}$$

$$= \mathrm{A}_1 - \mathrm{A}_2 - \mathrm{A}_3 + \mathrm{A}_4 \tag{2.313}$$

if $\mathrm{A} = \mathrm{O}_p(\frac{1}{n}), \mathrm{B} = \mathrm{O}_p(\frac{1}{n^2})$，then A is leading term, B is s.o.(smaller order) term

$$\mathrm{A}_1 = \frac{1}{n}\sum_t u_t u_{t-1} = \mathrm{O}_p(?) \tag{2.314}$$

对于$a_n = \mathrm{O}_p(b_n)$, 如果有$\mathrm{E}|a_n| = \mathrm{O}(\frac{1}{n})$ 则$a_n = \mathrm{O}_p(\frac{1}{n})$,如果有$\mathrm{E}(a_n)^2 = \mathrm{O}(\frac{1}{n^2})$ 则 $a_n = \mathrm{O}_p(\frac{1}{n})$

原假设下u_t, u_{t-1}独立,$\mathrm{E}u_t = 0$,有$\mathrm{EA}_1 = 0$

所以 $\mathrm{EA}_1^2 = \mathrm{V}ar(\mathrm{A}_1)$ 根据MDCLT

$$\sqrt{n}\mathrm{A}_1 = \frac{1}{\sqrt{n}}\sum_t u_t u_{t-1} \sim \mathrm{N}(0, \sigma_u^4) \tag{2.315}$$

$$\mathrm{V}ar(\sqrt{n}\mathrm{A}_1) = \sigma^4 \Rightarrow \mathrm{V}ar(\mathrm{A}_1) = \frac{\sigma_u^4}{n} \tag{2.316}$$

$$\mathrm{EA}_1^2 = \mathrm{O}(\frac{1}{n}) \Rightarrow = \mathrm{O}_p(\frac{1}{\sqrt{n}}) \tag{2.317}$$

$$\mathrm{A}_2 = [\frac{1}{n}\sum_t u_{t-1}x_t'](\hat{\beta}-\beta) = \underset{1\times k}{\mathrm{B}_2}\mathrm{C}_2 \tag{2.318}$$

$$[\sqrt{n}(\hat{\beta}-\beta) \sim \mathrm{N} = \mathrm{O}_p(\frac{1}{\sqrt{n}})] \tag{2.319}$$

$$\mathrm{V}ar(\mathrm{B}_2') = \frac{1}{n^2}[n\mathrm{V}ar(x_t u_{t-1}) + 2\cancel{\sum_t\sum_{s>t}\mathrm{C}ov(x_t u_{t-1}, x_s u_{s-1})}] \tag{2.320}$$

$$= \frac{1}{n}\mathrm{E}(x_t u_{t-1}^2 x_t') = \frac{1}{n}\underset{\sigma_u^2}{\mathrm{E}(u_{t-1}^2)}\underset{limit}{\mathrm{E}(x_t x_{t-1})} \tag{2.321}$$

$$\mathrm{B}_2 = \mathrm{O}_p(\frac{1}{\sqrt{n}}) \tag{2.322}$$

$$\mathrm{A}_2 = \mathrm{B}_2\mathrm{C}_2 = \mathrm{O}_p(\frac{1}{\sqrt{n}})\mathrm{O}_p(\frac{1}{\sqrt{n}}) = \mathrm{O}_p(\frac{1}{n}) \tag{2.323}$$

$$\mathrm{A}_3 = \mathrm{O}_p(\frac{1}{n}) \tag{2.324}$$

$$A_4 = \frac{1}{n}\sum(\hat{\beta}-\beta)x_{t-1}x_t'(\hat{\beta}-\beta) \tag{2.325}$$

$$= (\hat{\beta}-\beta)\frac{1}{n}\sum x_{t-1}x_t'(\hat{\beta}-\beta) \tag{2.326}$$

$$= O_p(\frac{1}{\sqrt{n}})(?)O_p(\frac{1}{\sqrt{n}}) \tag{2.327}$$

$$Var(\frac{1}{n}) \to 0 \tag{2.328}$$

$$Var(\frac{1}{n}\sum x_{t-1}x_t') = \underbrace{E(\frac{1}{n}\sum x_{t-1}x_t')^2}_{\xrightarrow{p}(Ex_{t-1}x_t')^2} - \underbrace{(Ex_{t-1}x_t')^2}_{O_p(1)} = O_p(1) \tag{2.329}$$

$$\Rightarrow E(\frac{1}{n}\sum x_{t-1}x_t')^2 = O_p(1) \tag{2.330}$$

$$\Rightarrow \frac{1}{n}\sum x_{t-1}x_t' = O_p(\sqrt{1}) = O_p(1) \tag{2.331}$$

$$A_4 = O_p(\frac{1}{\sqrt{n}})O_p(1)O_p(\frac{1}{\sqrt{n}}) = O_p(\frac{1}{n}) \tag{2.332}$$

so A_1 is leading term,$-A_2 - A_3 + A_4$ is s.o. term

$$\hat{\gamma}(1) = O_p(\frac{1}{\sqrt{n}}) + O_p(\frac{1}{n}) \tag{2.333}$$

$$\sqrt{n}\hat{\gamma}(1) = \sqrt{n}A_1 + s.o. \sim N(0,\sigma_u^4) \tag{2.334}$$

$$\hat{u}_t = u_t^2 - 2u_t x_t'(\hat{\beta}-\beta) + x_t'(\hat{\beta}-\beta)(\hat{\beta}-\beta)'x_t \tag{2.335}$$

$$\hat{\gamma}(0) = \frac{1}{n}\sum\hat{u}_t = \underbrace{\frac{1}{n}\sum_t u_t^2}_{\to Eu_t^2} - \underbrace{\frac{2}{n}\sum_t u_t x_t'}_{O_p(1)}(\hat{\beta}-\beta) + (\hat{\beta}-\beta)\underbrace{\frac{1}{n}\sum_t x_t'x_t}_{O_p(1)}(\hat{\beta}-\beta) \tag{2.336}$$

$$\hat{\gamma}(0) \to Eu_t^2 = \sigma_u^2 \tag{2.337}$$

$$\sqrt{n}\hat{\rho}(1) = \sqrt{n}\frac{\hat{\gamma}(1)}{\hat{\gamma}(0)} \sim N(0,1) \tag{2.338}$$

不能使用的情况

$y_t = \phi y_{t-1} + u_t, u_t \sim WN(0,\sigma^2)$

Test $\gamma(1) = E(u_{t-1}u_t) = 0$ or not

$$\begin{aligned}
\hat{\gamma}(1) &= \frac{1}{n}\sum_t \hat{u}_t\hat{u}_{t-1} \\
&= \dots \\
&= \frac{1}{n}\sum_t[u_t u_{t-1} - u_{t-1}y_{t-1}(\hat{\phi}-\phi) - u_t u_{t-2}(\hat{\phi}-\phi) + y_{t-1}y_{t-2}(\hat{\phi}-\phi)^2] \\
&= A_1 - A_2 - A_3 + A_4
\end{aligned}$$

类似的，有$A_1 = O_p(\frac{1}{\sqrt{n}}, A_3 = O_p(\frac{1}{n}), A_4 = O_p(\frac{1}{n}))$, 唯一不一样的是$A_2$

$$A_2 = [\frac{1}{n}\sum_t y_{t-1}u_{t-1}](\hat{\phi}-\phi) \tag{2.339}$$

$$Var([\frac{1}{n}\sum_t y_{t-1}u_{t-1}]) \to 0 \tag{2.340}$$

$$E(y_{t-1}u_{t-1})^2 \to O_p(1) \tag{2.341}$$

$$E(\frac{1}{n}\sum_t y_{t-1}u_{t-1})^2 \to E(y_{t-1}u_{t-1})^2 = O_p(1) \tag{2.342}$$

$$\Rightarrow A_2 = O_p(\frac{1}{\sqrt{1}})O_p(\frac{1}{\sqrt{n}}) \tag{2.343}$$

$$\begin{aligned}
A_2 &= [\frac{1}{n}\sum_t y_{t-1}u_{t-1}](\hat{\phi}-\phi)\\
&= [E(y_{t-1}u_{t-1}) + o_p(1)][\frac{1}{n}\sum y_{t-1}^2]^{-1}[\frac{1}{n}\sum y_{t-1}u_t]\\
&= E(y_{t-1}u_{t-1})(Ey_{t-1}^2)^{-1}[\frac{1}{n}\sum y_{t-1}u_t] + o_p(1)\\
&= \sigma_u^2(\frac{1-\phi^2}{\sigma_u^2})[\frac{1}{n}\sum y_{t-1}u_t] + s.o.\\
&= (1-\phi^2)[\frac{1}{n}\sum y_{t-1}u_t] + s.o.
\end{aligned}$$

$$\begin{aligned}
\sqrt{n}\hat{\gamma}(1) &= \sqrt{n}(A_1 - A_2)\\
&= \sqrt{n}(\frac{1}{n}\sum u_t u_{t-1} - (1-\phi^2)\frac{1}{n}\sum y_{t-1}u_t) + s.o.\\
&= \frac{1}{\sqrt{n}}\sum_t u_t[u_{t-1} - (1-\phi^2)y_{t-1}] + s.o.
\end{aligned}$$

根据MDCLT, $Eu_t[u_{t-1} - (1-\phi^2)y_{t-1}] = 0$

$$Var(u_t[u_{t-1} - (1-\phi^2)y_{t-1}]) = E(u_t^2[u_{t-1} - (1-\phi^2)y_{t-1}]^2) = \sigma_u^2\phi^2 \tag{2.344}$$

$$\Rightarrow \sqrt{n}\hat{\gamma}(1) \sim N(0, \sigma_u^4\phi^2) \tag{2.345}$$

$$\hat{\gamma}(0) \to Eu_t^2 = \sigma_u^2 \tag{2.346}$$

$$\sqrt{n}\hat{\rho} = \sqrt{n}\frac{\hat{\gamma}(1)}{\hat{\gamma}(0)} \overset{H_0}{\sim} N(0, \phi^2) \tag{2.347}$$

Weakly Exogenous $E(u_t|x_t) = 0$

Predetermined $E(u_t|x_{t-1}, x_{t-2}, \ldots, x_1)$

Strongly Exogenous $E(u_t|x_1, \ldots, x_n) = 0$

x	Time	y
	t^*-1	0
$x\uparrow 1$	t^*	β_0
	t^*+1	$\beta_0+\beta_1+\gamma_1\beta_0$
	t^*+2	$\beta_0+\beta_1+\beta_2+\gamma_2(\beta_0+\beta_1+\gamma_1\beta_0)$

ARDL Auto Regressive Distributional Lag Model

$$\underbrace{y_t=\mu+\gamma_1 y_{t-1}+\cdots+\gamma_p y_{t-p}}_{autoregressive}+\underbrace{\beta_0 x_t+\beta_1 x_{t-1}+\cdots+\beta_r x_{t-r}+\varepsilon_t}_{distribution\,lag} \tag{2.348}$$

$$\Delta \text{ in } x\begin{cases} \text{shock}\begin{cases}\text{SR} & \text{YES}\\ \text{LR} & \text{NO}\end{cases}\\ \text{permenant}\begin{cases}\text{SR} & \text{YES}\\ \text{LR} & \text{YES}\end{cases}\end{cases} \tag{2.349}$$

$$y^*=\mu+\gamma_1 y^*+\cdots+\gamma_p y^*+\beta_0 x^*+\beta_x^*+\cdots+\beta_r x^*+\varepsilon \tag{2.350}$$

$$y^*=\frac{\hat{\mu}^2}{1-\sum_{i=1}^{p}\hat{\gamma}_i}+\frac{\sum_{i=0}^{r}\hat{\beta}_j}{1-\sum_{i=1}^{p}\hat{\gamma}_i}x^* \tag{2.351}$$

L:Lag Operator

$$\mathrm{L}y_t=y_{t-1} \tag{2.352}$$

$$\mathrm{L}^2 y_t=\mathrm{L}(\mathrm{L}y_t)=y_{t-2} \tag{2.353}$$

$$(1-\mathrm{L})y_t=y_t-y_{t-1}=\Delta y_t \tag{2.354}$$

Define

$$\mathrm{B(L)}=\beta_0+\beta_1\mathrm{L}+\beta_2\mathrm{L}^2+\cdots+\beta_r\mathrm{L}^r \tag{2.355}$$

$$\mathrm{C(L)}=1-\gamma_1\mathrm{L}-\cdots-\gamma_p\mathrm{L}^p \tag{2.356}$$

$$\mathrm{ARDL}:\mathrm{C(L)}y_t=\mu+\mathrm{B(L)}x_t+\varepsilon_t \tag{2.357}$$

Partial Adjustment Model

$$y_t^* = \alpha + \beta x_t + \delta w_t + \varepsilon_t \tag{2.358}$$

$$y_t - y_{t-1} = (1-\lambda)(y_t^* - y_{t-1}) \tag{2.359}$$

$$\Rightarrow y_t = \alpha(1-\lambda) + \beta(1-\lambda)x_t + \delta(1-\lambda)w_t + \lambda y_{t-1} + (1-\lambda)\varepsilon_t \tag{2.360}$$

$$y_t = \alpha' + \beta' x_t + \delta' w_t + \lambda y_{t-1} + \varepsilon_t' \tag{2.361}$$

$$\mathrm{C(L)}y_t = \alpha' + \beta' x_t + \delta' w_t + \varepsilon_t' \quad \mathrm{C(L)} = 1 - \lambda \mathrm{L} \tag{2.362}$$

$$\frac{1}{\mathrm{C(L)}} = \frac{1}{1-\lambda \mathrm{L}} = 1 + \lambda \mathrm{L} + (\lambda \mathrm{L})^2 + (\lambda \mathrm{L})^3 + \dots \quad |\lambda| < 1 \tag{2.363}$$

$$y_t = [\alpha' + \lambda\alpha' + \lambda^2\alpha' + \dots] + \beta'[x_t + \lambda x_{t-1} + \lambda^2 x_{t-2} + \dots] + \delta'[w_t + \lambda w_{t-1} + \dots] + [\varepsilon_t' + \lambda\varepsilon_{t-1}' + \dots] \tag{2.364}$$

Common Factor Restriction

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t \tag{2.365}$$

$$\varepsilon_t = \rho\varepsilon_{t-1} + u_t \tag{2.366}$$

$$(1-\rho \mathrm{L})\varepsilon_t = u_t \quad \varepsilon_t = \frac{u_t}{1-\rho \mathrm{L}} \tag{2.367}$$

$$y_t = \beta_0 + \beta_1 x_t + \frac{u_t}{1-\rho \mathrm{L}} \tag{2.368}$$

$$1-\rho \mathrm{L} y_t = 1-\rho \mathrm{L}\beta_0 + 1-\rho \mathrm{L}\beta_1 x_t + u_t \tag{2.369}$$

$$y_t - \rho y_{t-1} = (\beta_0 - \rho\beta_0) + \beta_1 x_t - \beta_1\rho x_{t-1} + u_t \tag{2.370}$$

$$y_t = \beta_0' + \beta_1 x_t - \beta_1\rho x_{t-1} + \rho y_{t-1} + u_t \tag{2.371}$$

$$y_t = \gamma_0 + \gamma_1 x_t + \gamma_2 x_{t-1} + \gamma_3 y_{t-1} + u_t \tag{2.372}$$

common factor restriction if $\gamma_2 = -\gamma_1\gamma_3$ then $y_t = \gamma_0 + \gamma_1 x_t + \gamma_2 x_{t-1} + \gamma_3 y_{t-1} + u_t$ can be converted to $y_t = \beta_0 + \beta_1 x_t + \varepsilon_t$ with AR(1) error.

β_1 is long run effect γ_2 is short run effect

反过来，这个方法也可以处理AR(1) error

由于模型中存在y_{t-1},所以不能使用DW对残差($\hat{u}_t$)进行检验,因此可以使用Durbins h test和 BP test处理存在y_{t-1}的情况。

问题: if y_{t-1} on RHS，除了Durbins h 和 BP test还有没有其他方法。（好像忘了回答了）

in ARDL

$$y_t = lagy_t's + \beta x_t + lagx_t's + \underset{serialcorr}{error} \tag{2.373}$$

如果只控制了2期，但发现存在序列相关，可以增价滞后项。

e.g. $\varepsilon_t = \rho_1\varepsilon_{t-1} + \rho_2\varepsilon_{t-2} + u_t, \varepsilon_t = \frac{u_t}{1-\rho_1 \mathrm{L}-\rho_2 \mathrm{L}^2}$, 将$1-\rho_1\mathrm{L}-\rho_2\mathrm{L}^2$乘到等号两边即可

VAR

$$y_t = \mu + \Gamma_1 y_{t-1} + \Gamma_2 y_{t-2} + \cdots + \Gamma_p y_{t-p} + \varepsilon_t, \quad \varepsilon_t \text{ w.n.} \tag{2.374}$$

e.g. m=4 $y_t = \begin{pmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \\ y_{4t} \end{pmatrix}$

(1) $\mathrm{E}\varepsilon_t = 0$, (2) $\mathrm{E}\varepsilon_t\varepsilon_s' = 0$(error term 无序列相关), (3) $\mathrm{E}\varepsilon_t\varepsilon_t' = \Omega = \begin{pmatrix} 4 & 2 & -1 & \\ 2 & 3 & 0 & \vdots \\ -1 & 0 & 1 & \\ & \dots & & 6 \end{pmatrix}$ 非对角线元素可以不为0

i. GLS 在只需要考虑Ω参数，跨期为0的情况下，GLS可行

ii.分开执行OLS separate OLS $\Rightarrow$ GLS 证明:GLS和$\otimes$ 见课本

irf, MA representation

$$y_t = \hat{\mu} + \hat{\Gamma}_1 y_{t-1} + \hat{\Gamma}_2 y_{t-2} + e_t \tag{2.375}$$

$$y_{t-1} = \hat{\mu} + \hat{\Gamma}_1 y_{t-2} + \hat{\Gamma}_2 y_{t-3} + e_t \tag{2.376}$$

$$y_t = (\hat{\mu} + \hat{\Gamma}_1\hat{\mu}) + (e_t + \hat{\Gamma}_1 e_{t-1}) + (\hat{\Gamma}_1^{\,2} + \hat{\Gamma}_2) y_{t-2} + \hat{\Gamma}_1\hat{\Gamma}_2 y_{t-3} \tag{2.377}$$

$$\vdots \tag{2.378}$$

Non Stationary Process

(1) $y_t = \alpha + \beta t + u_t$ trend stationary

(2) $y_t = y_{t-1} + u_t$ random walk

(1) **trend**

$$y_t=\alpha+\beta t+u_t\text{OLS}:\quad \mathrm{X}=\begin{pmatrix}1&1\\1&2\\\vdots&\vdots\\1&\mathrm{T}\end{pmatrix} \tag{2.379}$$

$$\begin{pmatrix}\sqrt{n}(\hat{\alpha}-\alpha)\\n^{\frac{3}{2}}(\hat{\beta}-\beta)\end{pmatrix}=\begin{pmatrix}\sqrt{n}&0\\0&n^{\frac{3}{2}}\end{pmatrix}(\mathrm{X}'\mathrm{X})^{-1}\begin{pmatrix}\sqrt{n}&0\\0&n^{\frac{3}{2}}\end{pmatrix}\begin{pmatrix}\frac{\sum u_t}{\sqrt{n}}\\\frac{\sum tu_t}{n^{\frac{3}{2}}}\end{pmatrix} \tag{2.380}$$

$$\tag{2.381}$$

$$\mathrm{A}=\begin{pmatrix}\frac{\sum u_t}{\sqrt{n}}\\\frac{\sum tu_t}{n^{\frac{3}{2}}}\end{pmatrix},\quad \mathrm{EA}=0 \tag{2.382}$$

$$\mathrm{EAA}'=\mathrm{E}\begin{pmatrix}\frac{\sum u_t}{\sqrt{n}}\frac{\sum u_t}{\sqrt{n}}&\frac{\sum u_t}{\sqrt{n}}\frac{\sum tu_t}{n^{\frac{3}{2}}}\\\frac{\sum u_t}{\sqrt{n}}\frac{\sum tu_t}{n^{\frac{3}{2}}}&\frac{\sum u_t}{\sqrt{n}}\frac{\sum tu_t}{n^{\frac{3}{2}}}\frac{\sum u_t}{\sqrt{n}}\frac{\sum tu_t}{n^{\frac{3}{2}}}\end{pmatrix}=\begin{pmatrix}\sigma^2&\frac{1}{n^2}\sigma^2\frac{n(n+1)}{2}\\\frac{1}{n^2}\sigma^2\frac{n(n+1)}{2}&\frac{\sigma^2 n(n+1)(2n+1)}{6n^3}\end{pmatrix} \tag{2.383}$$

$$\mathrm{A}\sim\mathrm{N}(\begin{pmatrix}0\\0\end{pmatrix},\sigma^2\begin{pmatrix}1&\frac{1}{2}\\\frac{1}{2}&\frac{1}{3}\end{pmatrix}) \tag{2.384}$$

$$\text{Lindberg Feller CLT:}\quad\begin{pmatrix}\sqrt{n}(\hat{\alpha}-\alpha)\\\underbrace{n^{\frac{3}{2}}}_{\text{super consistent}}(\hat{\beta}-\beta)\end{pmatrix}=\mathrm{N}(\begin{pmatrix}0\\0\end{pmatrix},\sigma^2\begin{pmatrix}4&-6\\-6&12\end{pmatrix}) \tag{2.385}$$

(2) **Unit Root Process**

$$y_t=\rho y_{t-1}+u_t,\quad \rho=1 \tag{2.386}$$

$$\sqrt{n}(\hat{\rho}-\rho)\sim\mathrm{N}(0,1-\rho^2)=\mathrm{N}(0,1-1)=0 \tag{2.387}$$

$$\sqrt{n}\text{速度不够} \tag{2.388}$$

LLN,CLT ?$(\hat{\rho}-\rho)$

Browian Motion: $\mathrm{W}(r)$ for $r\in[0,1]$

$$\mathrm{W}(0)=0,r>s,\mathrm{W}(r)-\mathrm{W}(s)\sim\mathrm{N}(0,r-s) \tag{2.389}$$

$$r_2>r_1\quad[\mathrm{W}(r_2)-\mathrm{W}(r_1)]\perp\mathrm{W}(r_1)\quad\text{增量独立} \tag{2.390}$$

$$y_1, y_2, \ldots, y_n, [nr]\text{取整} \tag{2.391}$$

$$r \in [0,1], [nr] \in 0 \text{ to n} \tag{2.392}$$

$$y_n(r) = \begin{cases} 0, & 0 \le r < \frac{1}{n} \\ u_1, & \frac{1}{n} \le r < \frac{2}{n} \\ u_1 + u_2, & \frac{2}{n} \le r < \frac{3}{n} \\ \vdots & \vdots \\ u_1 + u_2 + \cdots + u_n, & r = 1 \end{cases} \tag{2.393}$$

for any fixed $r \in [0,1]$, we can show $\frac{1}{\sqrt{n}} y_n(r) \sim \mathrm{N}(0, r\sigma^2)$

$$\frac{1}{\sqrt{n}} y_n(r) = \frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} u_t = \frac{\sqrt{[nr]}}{\sqrt{n}} \left(\frac{1}{\sqrt{[nr]}} \sum_{t=1}^{[nr]} u_t\right) \to \sqrt{r}\mathrm{N}(0, \sigma^2) = \sigma \mathrm{N}(0, r) = \sigma \mathrm{W}(r) \tag{2.394}$$

$$\mathrm{C}_n(r + \Delta r) - \mathrm{C}_n(r) \perp \mathrm{C}_n(r)$$

$$\sum_{t=1}^{[n(r+\Delta r)]} - \sum_{t=1}^{[n(r)]} \tag{2.395}$$

$$\sum_{t=[nr]+1}^{[n(r+\Delta r)]} u_t \perp \sum_{t=1}^{[n(r)]} u_t$$

$$\text{if } r = 1, \quad \frac{1}{\sqrt{n}} y_n(1)/\sigma \to \mathrm{W}(1) \tag{2.396}$$

Review 积分

$$\int_0^1 g(x)dx = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1} \underbrace{g(x_{i-1})}_{g(x_i)\checkmark} \tag{2.397}$$

随机变量积分

$$\int_0^1 g(w(s))dw(s) = \lim_{n\to\infty} \sum_{i=1}^{n} \underbrace{g(w(s_i))}_{g(w(s_{i+1}))\times} \underbrace{(w(s_{i+1}) - w(s_i))}_{\text{对应}\frac{1}{n}} \tag{2.398}$$

$w(s_i)$不能取$w(s_{i+1})$

$$\frac{1}{n} \sim dr \tag{2.399}$$

$$\sum_i^n \sim \int_0^1 \tag{2.400}$$

$$\frac{1}{\sqrt{n}} \mathrm{Y}_{t-1} = \frac{1}{\sqrt{n}} \sum_{s=1}^{[nr]} u_s \sim \sigma_u w(r), \quad t-1 \leq nr \leq t \tag{2.401}$$

$$u_t = \mathrm{Y}_t - \mathrm{Y}_{t-1} \tag{2.402}$$

$$(1)\frac{u_t}{\sqrt{n}} = \frac{\mathrm{Y}_t - \mathrm{Y}_{t-1}}{\sqrt{n}} \sim d\sigma_u w_u(r) \tag{2.403}$$

$$(2)\frac{1}{\sqrt{n}} \mathrm{Y}_{t-1} \sim \sigma_u w_u(r) \tag{2.404}$$

$$(3)\frac{1}{n} \sim dr \tag{2.405}$$

$$(4)\sum_i^n \sim \int_0^1 \tag{2.406}$$

$$n(\hat{\rho} - 1) = \frac{n \sum \mathrm{Y}_{t-1} u_t}{\sum \mathrm{Y}_{t-1}^2} = \frac{\sum \frac{\mathrm{Y}_{t-1}}{\sqrt{n}} \frac{u_t}{\sqrt{n}}}{\frac{1}{n} \sum (\frac{\mathrm{Y}_{t-1}}{\sqrt{n}})^2} \tag{2.407}$$

$$\sim \frac{\int_0^1 \sigma_u^2 w_u(r) dw_u(r)}{\int_0^1 \sigma_u^2 w_u^2(r) dr} \quad \text{有限的} \not\to \infty \not\to 0 \tag{2.408}$$

$$y_t = x_t \beta + u \tag{2.409}$$

$$\left.\begin{aligned} x_t &= x_{t-1} + v_t \\ y_t &= y_{t-1} + \varepsilon_t \end{aligned}\right\} \text{random walk} \tag{2.410}$$

$$\hat{\beta} - \beta = (\mathrm{X}'\mathrm{X})^{-1} \mathrm{X}' u \tag{2.411}$$

$$n(\hat{\beta} - \beta) = (\frac{1}{n^2} \sum x_t^2)^{-1} \frac{1}{n} \sum x_t u_t \tag{2.412}$$

$$= (\frac{1}{n^2} \sum x_t^2)^{-1} (\frac{1}{n} \sum x_{t-1} u_t + \frac{1}{n} \sum v_t u_t) \tag{2.413}$$

$$\frac{1}{n} \sum (\frac{x_t}{\sqrt{n}})^2 \to \int_0^1 \sigma_v^2 w_r^2(r) dr \tag{2.414}$$

$$\sum \frac{x_{t-1}}{\sqrt{n}} \frac{u_t}{\sqrt{n}} \to \int_0^1 \sigma_v w_v(r) d\sigma_u w_u(r) \tag{2.415}$$

$$\frac{1}{n} \sum v_t u_t \to \mathrm{E}[v_t u_t] \overset{assume}{=} 0 \tag{2.416}$$

$$n(\hat{\beta} - \beta) \sim [\int_0^1 \sigma_v^2 w_r^2(r) dr]^{-1} [\int_0^1 \sigma_v w_v(r) \sigma_u dw_u(r)] \tag{2.417}$$

如果不用左端点计算,结果会有问题，考虑

$$y_t u_t = y_{t-1} u_t + u_t^2 \tag{2.418}$$

$$\frac{1}{n}\sum y_t u_t = \frac{1}{n}\sum y_{t-1} u_t + \frac{1}{n}\sum u_t^2 \tag{2.419}$$

$$\frac{1}{n}\sum y_t u_t \to \int_0^1 \sigma_u^2 w_u(r) dw_u(r) \tag{2.420}$$

$$\frac{1}{n}\sum y_{t-1} u_t \to \int_0^1 \sigma_u^2 w_u(r) dw_u(r) \tag{2.421}$$

$$\frac{1}{n}\sum u_t^2 \to \sigma^2 \tag{2.422}$$

$$\frac{1}{n}\sum y_t u_t < \frac{1}{n}\sum y_{t-1} u_t + \frac{1}{n}\sum u_t^2 \text{fi} \quad \text{矛盾} \tag{2.423}$$

2.6 Panel Data

$$y_{it} = x_{it}'\beta + c_i + u_{it} \tag{2.424}$$

$$\tag{2.425}$$

1) Pooled OLS 不存在c_i

2) Fixed Effect regressor $x_{it}'\beta + c_i$，c_i存在且和x_i有关

3) Random Effect error term $v_{it} = c_i + u_{it}$，c_i存在但是和x_i无关

个体存在观测(比如性别)和不可观测(比如 personal taste)的个体信息

2.6.1 Fixed Effect

$$\underset{T\times 1}{y_1} = \begin{pmatrix} y_{11} \\ \vdots \\ y_{1T} \end{pmatrix}, \underset{T\times 1}{i} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \tag{2.426}$$

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \beta + \underset{d_1 \dots d_n}{\begin{pmatrix} i & & 0 \\ & \ddots & \\ 0 & & i \end{pmatrix}} \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} + \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix} \tag{2.427}$$

$$y_i = x_i\beta + i\alpha_i + u_i \tag{2.428}$$

$$\mathrm{FE}: y = \mathrm{X}\beta + \mathrm{D}\alpha + u \tag{2.429}$$

LSDV(Least Square Dummy Variable)

$$\mathrm{M_D}y = \mathrm{M_D X}\beta + \underset{=0}{\mathrm{M_D D}}\alpha + u \tag{2.430}$$

$$\mathrm{M_D} = \underset{n\mathrm{T}\times n\mathrm{T}}{\mathrm{I}} - \mathrm{D(D'D)^{-1}D'} = \mathrm{I} - \begin{pmatrix} \frac{1}{\mathrm{T}}ii' & & 0 \\ & \vdots & \\ 0 & & \frac{1}{\mathrm{T}}ii' \end{pmatrix} = \begin{pmatrix} \mathrm{M}^0 & & 0 \\ & \ddots & \\ 0 & & \mathrm{M}^0 \end{pmatrix} \tag{2.431}$$

$$\mathrm{M}^0 = \mathrm{I} - \frac{1}{\mathrm{T}}ii' \tag{2.432}$$

$$\mathrm{M_D}y = \begin{pmatrix} \mathrm{M}^0 y_1 & & \\ & \ddots & \\ & & \mathrm{M}^0 y_n \end{pmatrix} \tag{2.433}$$

$$x_{it} - \bar{x}_{i\cdot} = \ddot{x}_{it} \tag{2.434}$$

存在参考组，drop d_1,则α表示其他组和参考组的差别 $\mathrm{H}_0: \alpha_2 = \cdots = \alpha_n = 0$

上面的方法叫做one-way fixed effect,类似的也可以控制t

$$y_{it} = x_{it}'\beta + c_i + u_{it} \tag{2.435}$$

$$y_{*it} = y_{it} - \bar{y}_{i\cdot} - (\bar{y}_{\cdot t} - \bar{\bar{y}}) \tag{2.436}$$

$$x_{*it} = x_{it} - \bar{y}_{i\cdot} - (\bar{x}_{\cdot t} - \bar{\bar{x}}) \tag{2.437}$$

这个方法称为two-way fixed effect, c_i, g_t可能会导致多重共线性，造成估计不精确

2.6.2 Random Effect

$$y_{it} = x_{it}'\beta + c_i + u_{it} \tag{2.438}$$

$$= x_{it}'\beta + v_{it} \tag{2.439}$$

$$\tag{2.440}$$

误差项的协方差矩阵不是对角阵，所以要用(F)GLS估计。

$$\mathrm{E}(c_i|\mathrm{X}) = 0 \tag{2.441}$$

$$\mathrm{E}(u_{it}|\mathrm{X}) = 0 \tag{2.442}$$

$$\mathrm{E}(u_{it}^2|\mathrm{X}) = \sigma_u^2 \tag{2.443}$$

$$\mathrm{E}(c_i^2|\mathrm{X}) = \sigma_c \tag{2.444}$$

$$\mathrm{E}(u_{it}c_j|\mathrm{X}) = 0 \tag{2.445}$$

$$\mathrm{E}(u_{it}u_{js}|\mathrm{X}) = 0, i \neq j \text{ or } s \neq t \tag{2.446}$$

$$\mathrm{E}(c_i c_j|\mathrm{X}) = 0 \text{if } i \neq j \tag{2.447}$$

$$v_i = \begin{pmatrix} v_{i1} \\ \vdots \\ v_{i\mathrm{T}} \end{pmatrix}, \Sigma = \mathrm{E}v_i v_i' = \begin{pmatrix} \sigma_c^2 + \sigma_u^2 & \sigma_c^2 & \dots & \sigma_c^2 \\ \sigma_c^2 & \ddots & & \\ & & \ddots & \\ \sigma_c^2 & \ddots & \ddots & \sigma_c^2 + \sigma_u^2 \end{pmatrix} \tag{2.448}$$

$$\Omega = \mathrm{EVV}' = \mathrm{I}_n \otimes \Sigma = \begin{pmatrix} \Sigma & & 0 \\ & \ddots & \\ 0 & & \Sigma \end{pmatrix} \tag{2.449}$$

$$\hat{\beta}_{\mathrm{GLS}} = (\mathrm{X}'\Omega^{-1}\mathrm{X})^{-1}\mathrm{X}'\Omega^{-1}y \tag{2.450}$$

$$= (\sum_{i=1}^{n} x_i'\Sigma^{-1}x_i)^{-1}(\sum_{i=1}^{n} x_i\Sigma^{-1}y_i) \tag{2.451}$$

即对原数据做$\Omega^{-\frac{1}{2}}$Transformation

$$\Sigma^{-\frac{1}{2}} = \frac{1}{\sigma_u}[\mathrm{I} - \frac{\theta}{\mathrm{T}} ii'] \tag{2.452}$$

$$\theta = 1 - \frac{\sigma_u}{\sqrt{\sigma_u^2 + \mathrm{T}\sigma_c^2}} \tag{2.453}$$

$$\Sigma^{-\frac{1}{2}} y_i = \frac{1}{\sigma_u} \begin{pmatrix} y_{i1} - \theta \bar{y}_{i\cdot} \\ \vdots \\ y_{i\mathrm{T}} - \theta \bar{y}_{i\cdot} \end{pmatrix} \tag{2.454}$$

有点类似FE，$\theta = 1$ FE,$\sigma_c = 0$ Pooled

$\sigma_u^2 + \sigma_c^2$的估计

$$y_{it} = x_{it}'\beta + c_i + u_{it} \quad \text{OLS} \tag{2.455}$$

$$= x_{it}'\hat{\beta} + \hat{v}_{it} \tag{2.456}$$

$$\frac{1}{n}\sum \hat{v}_{it}^2 = \sigma_u^2 + \sigma_c^2 \tag{2.457}$$

$\mathrm{E}\ddot{u}_{it}^2$ demean之后c_i消掉了，即

$$\bar{y}_{i\cdot} = \bar{x}_{i\cdot}' + c_i + \bar{u}_{i\cdot} \tag{2.458}$$

$$y_{it} - \bar{y}_{i\cdot} = (x_{it}' - \bar{x}_{i\cdot}')\hat{\beta} + u_{it} - \bar{u}_{i\cdot} \tag{2.459}$$

$$\ddot{y}_{it} = \ddot{x}_{it}\hat{\beta} + \hat{\ddot{u}}_{it} \tag{2.460}$$

$$\mathrm{E}\ddot{u}_{it}^2 = \mathrm{E}(u_{it}^2 - 2u_{it}\bar{u}_{i\cdot} + \bar{u}_{i\cdot}^2) = \frac{\mathrm{T}-1}{\mathrm{T}}\sigma_u^2 \tag{2.461}$$

$$\sum_{i=1}^{n}\sum_{t=1}^{\mathrm{T}} \mathrm{E}\ddot{u}_{it}^2 = n(\mathrm{T}-1)\sigma_u^2 \tag{2.462}$$

$$\sigma_u^2 = \mathrm{E}[\sum_{i=1}^{n}\sum_{t=1}^{\mathrm{T}} \frac{\ddot{u}_{it}^2}{n(\mathrm{T}-1)}] \tag{2.463}$$

$\sum_{i=1}^{n}\sum_{t=1}^{\mathrm{T}} \frac{\ddot{u}_{it}^2}{n(\mathrm{T}-1)}$是$\sigma_u^2$的无偏估计，修正后

$$\frac{1}{n(\mathrm{T}-1)-k}\sum_{i=1}^{n}\sum_{t=1}^{\mathrm{T}} \ddot{u}_{it}^2 = \hat{\sigma}_u^2 \tag{2.464}$$

$$\hat{\beta}_{\mathrm{RE/FGLS}} = (\mathrm{X}'\hat{\Omega}^{-1}\mathrm{X})^{-1}\mathrm{X}'\hat{\Omega}^{-1}y \tag{2.465}$$

$$= (\sum_{i=1}^{n} x_i'\hat{\Sigma}^{-1}x_i)^{-1}(\sum_{i=1}^{n} x_i\hat{\Sigma}^{-1}y_i) \tag{2.466}$$

有关系用FE,无关RE和FE都可以，但是RE更有效

对于Nonlinear Panel c_i不容易被消除

总结

RE1:$\mathrm{E}(u_{it}|x_i,c_i)=0,\mathrm{E}(c_i|x_i)=\mathrm{E}(c_i)=0$

RE2:$rank\mathrm{E}(x_i'\Omega^{-1}x_i)=k$

RE3:$\mathrm{E}(u_iu_i'|x_i,c_i)=\sigma_u^2\mathrm{I_T},\mathrm{E}(c_i^2|x_i)=\sigma_c^2$

FE1:$\mathrm{E}(u_{it}|x_i,c_i)=0$

FE2:$rank\mathrm{E}(\ddot{x}_i'\ddot{x}_i)=k$

FE3:$\mathrm{E}(u_iu_i'|x_i,c_i)=\sigma_u^2\mathrm{I_T}$

$\mathrm{E}(u_{it}|x_i)$不是x_{it}即严格外生假设，所有时期都无关，因为估计的时候用了demean处理,所以要求所有数据都无关。

if RE3 is violated

$$\sqrt{n}(\hat{\beta}_{\mathrm{RE}}-\beta)=(\frac{1}{n}\sum_{i=1}^{n}x_i'\Sigma^{-1}x_i)^{-1}(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}x_i'\Sigma^{-1}v_i) \tag{2.467}$$

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}x_i'\Sigma^{-1}v_i\sim\mathrm{N}(0,\mathrm{E}[x_i'\Sigma^{-1}\mathrm{E}(v_iv_i'|x_i)\Sigma^{-1}x_i]) \tag{2.468}$$

$$\hat{Var}(\sqrt{n}\hat{\beta}_{\mathrm{RE}})=(\frac{1}{n}\sum_{i=1}^{n}x_i'\Sigma^{-1}x_i)^{-1}(\frac{1}{n}\sum_{i=1}^{n}x_i'\Sigma^{-1}\hat{v}_i\hat{v}_i'\Sigma^{-1}x_i)(\frac{1}{n}\sum_{i=1}^{n}x_i'\Sigma^{-1}x_i)^{-1} \tag{2.469}$$

which is Robust Variance Matrix Estimator

$\hat{\Sigma}=\frac{1}{n}\sum_{i=1}^{n}\hat{v}_i\hat{v}_i'$RE OLS,都是一致的

General Feasible Generalized LS

$$\hat{\beta}_{\mathrm{GFGLS}}=(\sum_{i=1}^{n}x_i'\hat{\Sigma}^{-1}x_i)^{-1}(\sum_{i=1}^{n})^{-1}(\sum_{i=1}^{n}x_i'\hat{\Sigma}y_i) \tag{2.470}$$

$$\hat{Var}(\sqrt{n}\hat{\beta}_{\mathrm{GFGLS}})=(\sum_{i=1}^{n}x_i'\hat{\Sigma}^{-1}x_i)^{-1} \tag{2.471}$$

if FE3 is violated

$$\hat{Var}(\sqrt{n}\hat{\mathrm{FE}})=(\frac{1}{n}\sum_{i=1}^{n}\ddot{x}_i'\ddot{x}_i)^{-1}(\frac{1}{n}\sum_{i=1}^{n}\ddot{x}_i'\hat{\ddot{u}}_i\hat{\ddot{u}}_i'\ddot{x}_i)(\frac{1}{n}\sum_{i=1}^{n}\ddot{x}_i'\ddot{x}_i)^{-1} \tag{2.472}$$

which is Robust Var Matrix Estimator

FE3 要求对角阵

FEGLS3 $\mathrm{E}(u_iu_i'|x_i,\mathrm{C}_i)=\Lambda$ 任意形式

$\ddot{u}_i = \mathrm{Q_T} u_i$, $\mathrm{Q_T}$ is demean Transformation

$$\mathrm{E}(\ddot{u}_i\ddot{u}_i'|\ddot{x}_i) = \mathrm{E}(\ddot{u}_i\ddot{u}_i') \tag{2.473}$$
$$= \mathrm{Q_T}\mathrm{E}(u_iu_i')\mathrm{Q_T} \tag{2.474}$$
$$= \mathrm{Q_T}\Lambda\mathrm{Q_T} \overset{def}{=} \Sigma \tag{2.475}$$
$$\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n}\hat{\ddot{u}}_i\hat{\ddot{u}}_i' \tag{2.476}$$
$$\hat{\Sigma}^{-\frac{1}{2}}\ddot{y}_i = \hat{\Sigma}^{-\frac{1}{2}}\sum_{i=1}^{n}\ddot{x}_i\beta + \hat{\Sigma}^{-\frac{1}{2}}\ddot{u}_i \tag{2.477}$$
$$\mathrm{V}ar(\sqrt{n}\hat{\beta}_{\mathrm{FEGLS}}) = (\frac{1}{n}\sum_{i=1}^{n}\ddot{x}_i'\hat{\Sigma}\ddot{x}_i)^{-1} \tag{2.478}$$

由于对$\ddot{u}_i$做了demean处理，$\hat{\Sigma}$是$\mathrm{T}-1\times\mathrm{T}-1$矩阵，去掉了某一个时间的值。

内生性 z_i是IV

REIV

$$\mathrm{E}(u_{it}|x_i) \neq 0 \tag{2.479}$$
$$\mathrm{E}(u_{it}|z_i) = 0 \tag{2.480}$$
$$\mathrm{E}(c_i|z_i) = 0 \tag{2.481}$$
$$\Sigma^{-\frac{1}{2}}y_i = \Sigma^{-\frac{1}{2}}x_i\beta + \Sigma^{-\frac{1}{2}}v_i \tag{2.482}$$
$$(\mathrm{I}_n\otimes\Sigma^{-\frac{1}{2}}) \quad \Omega^{-\frac{1}{2}}y = \Omega^{-\frac{1}{2}}\mathrm{X}\beta + \Omega^{-\frac{1}{2}}v \tag{2.483}$$
$$\Omega^{-\frac{1}{2}}y = \Omega^{-\frac{1}{2}}\mathrm{Z}\beta + \Omega^{-\frac{1}{2}}v \quad \text{(2SLS)} \tag{2.484}$$

First stage

$$\Omega^{-\frac{1}{2}}\mathrm{X} = \Omega^{-\frac{1}{2}}\mathrm{Z}\delta + error \tag{2.485}$$

Second stage

$$\Omega^{-\frac{1}{2}}y = \Omega^{-\frac{1}{2}}\mathrm{Z}(\mathrm{Z}'\Omega^{-1}\mathrm{Z})^{-1}\mathrm{Z}'\Omega^{-1}\mathrm{X}\beta + error \tag{2.486}$$
$$\hat{\beta}_{\mathrm{REIV}} = (\mathrm{X}\Omega^{-1}\mathrm{Z}(\mathrm{Z}'\Omega^{-1}\mathrm{Z})^{-1}\mathrm{Z}'\Omega^{-1}\mathrm{X})^{-1}\mathrm{X}'\Omega^{-1}\mathrm{Z}(\mathrm{Z}'\Omega^{-1}\mathrm{Z})^{-1}\mathrm{Z}'\Omega^{-1}y \tag{2.487}$$
$$\sqrt{n}(\hat{\beta}_{\mathrm{REIV}}-\beta) \sim \mathrm{N}(0,[\frac{\mathrm{X}'\Omega^{-1}\mathrm{Z}}{n}(\frac{\mathrm{Z}'\Omega^{-1}\mathrm{Z}}{n})^{-1}\frac{\mathrm{Z}'\Omega\mathrm{X}}{n}]^{-1}) \tag{2.488}$$

Test of endogeneity

$$y_{it1} = z_{it1}\delta + y_{it2}\alpha_1 + y_{it3}\gamma_1 + \underbrace{c_{i1} + u_{it1}}_{v_{it1}} \tag{2.489}$$

$$\mathrm{H}_0: \quad \mathrm{E}(y_{it3}|v_{it1}) = 0 \quad \forall s = 1, \ldots, \mathrm{T} \tag{2.490}$$

$$y_{it3} = z_{it}\Pi_3 + v_{it3} \Rightarrow \hat{v}_{it3} \quad z_{it} = (\overset{z_{it1}}{exo}, \overset{z_{it2}}{iv}) \tag{2.491}$$

$$y_{it1} = z_{it1}\delta + y_{it2}\alpha_1 + y_{it3}\gamma_1 + v_{it3}\rho_1 + error \tag{2.492}$$

$$\mathrm{H}_0: \rho_1 = 0 \tag{2.493}$$

$\rho_1 = 0$,y_{it3}, y_{it1}无关

FEIV

$$\mathrm{E}(u_{it}|x_i) = 0 \tag{2.494}$$

$$\mathrm{E}(u_{it}|z_i) = 0 \tag{2.495}$$

$$\tag{2.496}$$

First stage

$$\ddot{x}_i = \ddot{z}_i\delta + error \tag{2.497}$$

Second stage

$$\ddot{y}_i = \hat{\ddot{x}}_i\beta + error \tag{2.498}$$

$$\hat{\beta}_{\mathrm{FEIV}} = \tag{2.499}$$

$$\sqrt{n}(\hat{\beta}_{\mathrm{FEIV}} - \beta) \sim \mathrm{N}(0, \sigma^2[(\mathrm{E}\ddot{x}_i'\ddot{z}_i)(\mathrm{E}\ddot{z}_i'\ddot{z}_i)^{-1}(\mathrm{E}\ddot{z}_i'x_i)]) \tag{2.500}$$

Test of endogeneity

$$y_{it3} = z_{it}\Pi_3 + v_{it3} \Rightarrow \hat{v}_{it3} \quad z_{it} \tag{2.501}$$

$$\mathrm{FEIV}(z_{it}, y_{it3}, \hat{v}_{it3}) \tag{2.502}$$

First Difference　FD

$$y_{it} = x'_{it}\beta + c_i + u_{it} \tag{2.503}$$

$$\Delta y_{it} = \Delta x'_{it}\beta + \Delta u_{it} \tag{2.504}$$

$$\Delta x_{it} \quad \Delta u_{it} \quad \text{无关} \tag{2.505}$$

$$x_{it} - x_{it-1} \quad u_{it} - u_{it-1} \quad \text{无关}$$

$$\text{和同期未来一期滞后一期都无关}$$

$\mathrm{E}(u_{it}|z_i) = 0$, use w_{it} as IV $\mathrm{E}(w'_{it}\Delta u_{it}) = 0$, w_{it}和t相关

$$\mathrm{E}(\begin{pmatrix} w'_{i2} & & \\ & \ddots & \\ & & w'_{i\mathrm{T}} \end{pmatrix} \begin{pmatrix} \Delta u_{i2} \\ \vdots \\ \Delta u_{i\mathrm{T}} \end{pmatrix}) = 0 \tag{2.506}$$

即对每个时间t找一个工具变量，不需要严格外生假定

$$\left.\begin{aligned} \Delta x'_{it} \xleftarrow{\text{IV}} w_{it}, \quad i = 1, \ldots, n \\ \Delta x'_{it-1} \xleftarrow{\text{IV}} w_{it-1}, \quad i = 1, \ldots, n \\ \vdots \qquad\qquad \end{aligned}\right\} \text{T-1 seperate OLS} \tag{2.507}$$

先做T-1个First stage，再做Second stage，叫做System 2SLS

问题：如何找到w_{it}

sequential exogeneity

$$y_{it} = x'_{it}\beta + c_i + u_{it} \tag{2.508}$$

$$\mathrm{E}(u_{it}|x'_{it}, x'_{it-1}, \ldots, x'_{i1}) = 0 \tag{2.509}$$

$$\tag{2.510}$$

e.g. $y_{it} = z_{it}\gamma + \delta h_{it-1} + c_i + u_{it}$,y is percentage of flights cancelled, z is strictly exos, h is profit

$x_{it} = (z_{it}, h_{it-1})$

动态面板, $y_{it} = \rho y_{it-1} + c_i + u_i t$满足sequential exogeneous

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

$$\left.\begin{array}{l} t = 2y_{i0} \\ t = 3y_{i0}, y_{i1} \\ t = \mathrm{T}y_{i0}, \ldots, y_{i\mathrm{T}-2} \end{array}\right\} \text{IVs,System 2SLS} \Delta y_{it} = \Delta x_{it}'\beta + \Delta u_{it} \tag{2.511}$$

$$\left.\begin{array}{l} t = 2x_{i1} \\ t = 3x_{i1}, x_{i2} \\ t = \mathrm{T}x_{i1}, \ldots, y_{i\mathrm{T}-1} \end{array}\right\} \text{IVs,System 2SLS} \tag{2.512}$$

2.7 Nonlinear Model

2.7.1 Binary Choice

Linear Probability Model

$$y_i = x_i'\beta + \varepsilon_i \tag{2.513}$$

$$\mathrm{E}(\varepsilon_i|x_i) = 0 \tag{2.514}$$

$$y = 1 or 0 \tag{2.515}$$

$$\mathrm{E}(y_i|x_i) = x_i'\beta = 1 \times \mathrm{P}rob(y_i = 1|x_i) + \cancel{0 \times \mathrm{P}rob(y_i = 0|x_i)} \tag{2.516}$$

$$\mathrm{V}ar(\varepsilon_i|x_i) = \mathrm{E}(\varepsilon_i^2|x_i) \tag{2.517}$$

$$= (1 - x_i'\beta)^2 x_i\beta + (-x_i\beta)^2(1 - x_i\beta) \tag{2.518}$$

$$= (1 - x_i'\beta)x_i'\beta \tag{2.519}$$

弊端是可能为负

$$\mathrm{U}_i^{rent} = x_i'\beta_r + \varepsilon_{ir} \tag{2.520}$$

$$\mathrm{U}_i^{buy} = x_i'\beta_b + \varepsilon_{ib} \tag{2.521}$$

$$y_i^* = \mathrm{U}_i^r - \mathrm{U}_i^b > 0 \quad y_i = 1 \tag{2.522}$$

$$y_i^* = \mathrm{U}_i^r - \mathrm{U}_i^b \leq 0 \quad y_i = 0 \tag{2.523}$$

$$\tag{2.524}$$

y_i^* is latent Variable

$$P(y_i = 1|x_i) = P(y_i^* > 0|x_i) \tag{2.525}$$
$$= P(x_i'\beta + \varepsilon_i > 0|x_i) \tag{2.526}$$
$$= P(\varepsilon_i > -x_i'\beta|x_i) \quad \text{如果}\varepsilon_i\text{是对称分布} \tag{2.527}$$
$$= \underset{\text{CDF}}{F}(x_i'\beta) \tag{2.528}$$

if $\varepsilon_i \sim N(0,1)$ standard normal $F = \Phi$, probit model,if ε_i is logistic, $F = \Lambda(x_i'\beta)$ logit model

use MLE to estimate parameters

semi-parametrix 半参数single index，让数据自己产生分布，只假设β, 对ε非参假设

2.7.2 Probit model with endog var

two stage

$$y_1^* = z_1\delta_1 + \alpha_1 y_2 + u_1 \tag{2.529}$$
$$y_1 = 1(y_1^* > 0) \quad Probit \quad u_i \sim N(0,1) \tag{2.530}$$
$$y_2\text{连续内生} \tag{2.531}$$
$$y_2 = z_1\delta_{21} + z_2\delta_{22} + v_2 \tag{2.532}$$
$$\text{first stage: } = z\delta_2 + v_2 \tag{2.533}$$
$$y_2|Z \sim N(z\delta_2, \tau_2^2) \quad \tau_2^2 = Var(v_2) \tag{2.534}$$
$$u_1 = \theta_1 v_2 + e_1 \tag{2.535}$$
$$\theta_1 = \frac{Cov(u_1, v_2)}{Var(v_2)} \overset{def}{=} \frac{\eta_1}{\tau_2^2} \tag{2.536}$$
$$\text{second stage: } y_1^* = z_1\delta_1 + \alpha_1 y_2 + \theta_1 v_2 + e_1 \tag{2.537}$$

e_1不是标准正态

$$\mathrm{E}(e_1) = \mathrm{E}(u_1 - \theta_1 v_2) = 0 \tag{2.538}$$

$$\mathrm{V}ar(u_1) = \theta_1^2 \mathrm{V}ar(v_2) + \mathrm{V}ar(e_1) \tag{2.539}$$

$$\mathrm{V}ar(e_1) = 1 - \frac{\eta_1^2}{\tau_2^2} \overset{def}{=} 1 - \rho_1^2 \tag{2.540}$$

$$\rho_1 = \frac{cov(u_1, v_2)}{\sqrt{var(u_1)var(v_2)}} = \frac{\eta_1}{\tau_2} \tag{2.541}$$

$$e_1 \sim \mathrm{N}(0, 1 - \rho_1^2) \tag{2.542}$$

$$\mathrm{P}(y_1 = 1 | z, y_2, \hat{v}_2) = \Phi[(z_1 \delta_1 + \alpha_1 y_2 + \theta_1 \hat{y}_2)/(1 - \hat{\rho}_1^2)^{\frac{1}{2}}] \tag{2.543}$$

$$\rho_1 = \theta_1 \tau_2 \tag{2.544}$$

$$\Rightarrow \hat{\delta}_1, \hat{\alpha}_1, \ldots \tag{2.545}$$

MLE

$$\text{MLE:}\ f(y_1, y_2 | z) = f(y_1 | y_2, z) f(y_2 | z) \quad \text{连续用pdf f()表示} \tag{2.546}$$

$$= \mathrm{P}(y_1 | y_2, z) f(y_2 | z) \quad \text{离散用分布函数 P()表示} \tag{2.547}$$

$$\mathrm{P}(y_1 = 1 | y_2, z) = \Phi[(z_1 \delta_1 + \alpha_1 y_2 + \frac{\rho_1}{\tau_2}(y_2 - z\delta_2))/(1 - \rho_1^2)^{\frac{1}{2}}] \tag{2.548}$$

$$f(y_2 | z) = \frac{1}{\sqrt{2\pi}\tau} e^{\frac{(y_2 - z\delta_2)^2}{2\tau_2^2}} \tag{2.549}$$

$$\Rightarrow \hat{\delta}_1, \hat{\alpha}_1 \tag{2.550}$$

内生变量是离散内生的情况

$$y_1 = 1[z_1 \delta_1 + \alpha_1 y_2 + u_1 > 0] \tag{2.551}$$

$$y_2 = 1[z\delta_2 + v_2 > 0] \quad \text{离散内生} \tag{2.552}$$

$$u_1, v_2 \sim \mathrm{N}(0, 1) \tag{2.553}$$

two stage 不能使用 $\hat{v}_2$算不出来，属于一个区间

MLE:$\mathrm{P}(y_1 = i, y_2 = j) = \mathrm{P}(y_1 = i|y_2 = j, z)\mathrm{P}(y_2 = j|z), \quad i, j = 0, 1$,e.g.

$$\mathrm{P}(y_1 = 1, y_2 = 1) = \mathrm{P}(y_1 = 1|y_2 = 1, z)\mathrm{P}(y_2 = 1|z) \tag{2.554}$$

$$\mathrm{P}(y_1 = 1|v_2, z) = \Phi[(z_1\delta_1 + \alpha_1 y_2 + \rho_1 v_2)/(1-\rho_1^2)^{\frac{1}{2}}] \tag{2.555}$$

$$\mathrm{E}[\mathrm{P}(y_1 = 1|v_2, z)|y_2 = 1, z] \tag{2.556}$$

$$\text{对}v_2\text{积分} = \int_{-z\delta_2}^{\infty} \mathrm{P}(y_1 = 1|v_2, z) f(v_2|y_2 = 1, z) dv_2 \tag{2.557}$$

$$= \int_{-z\delta_2}^{\infty} f(y_1 = 1, v_2|y_2 = 1, z) dv_2 \tag{2.558}$$

$$= \mathrm{P}(y_1 = 1|y_2 = 1, z) \tag{2.559}$$

$$(2.557) = \int_{-z\delta_2}^{\infty} \Phi[(z_1\delta_1 + \alpha_1 y_2 + \rho_1 v_2)/(1-\rho_1^2)^{\frac{1}{2}}] \frac{\phi(v_2)}{\mathrm{P}(v_2 > -z\delta_2)} dv_2 \tag{2.560}$$

$$= \frac{1}{\Phi(z\delta_2)} \int_{-z\delta_2}^{\infty} \Phi[(z_1\delta_1 + \alpha_1 y_2 + \rho_1 v_2)/(1-\rho_1^2)^{\frac{1}{2}}]\phi(v_2) dv_2 \tag{2.561}$$

$$\mathrm{P}(y_2 = 1|z) = \Phi(z\delta_2) \tag{2.562}$$

$$\mathrm{P}(y_1 = 1, y_2 = 1|z) = \mathrm{P}(y_1 = 1|y_2 = 1, z)\mathrm{P}(y_2 = 1|z) = \int_{-z\delta_2}^{\infty} \int_{-z_1\delta_1 - \alpha_1 y_2}^{\infty} \frac{1}{2\pi\sqrt{1-\rho_1^2}} e^{-\frac{1}{2(1-\rho_1^2)}[u_1^2 - 2\rho_1 u_1 u_2 + v_2^2]} \tag{2.563}$$

是二元正态分布分布函数 bivariate normal density.

in STATA, IV probit适用于连续变量，离散内生变量应该用biprobit解决内生性问题

bivariate probit model

$$y_1 = 1[x_1\beta_1 + e_1 > 0] \tag{2.564}$$

$$y_2 = 1[x_2\beta_2 + e_2 > 0] \tag{2.565}$$

$$\begin{pmatrix} e_1 \\ e_2 \end{pmatrix} \sim \mathrm{N}(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}) \tag{2.566}$$

if $\rho = 0$,separate probit, joint MLE都一致有效

if $\rho \neq 0$ 参考式(2.563)

其他情况 1 ordered probit 多个选择，latent variable $y*$ 分成多个区间。

无序的情况见课本

2.7.3 Truncated Data 受限数据

没有影响的时候，x受到外生因素的缺少受限

如果y受限 e.g. truncation model,truncated normal distribution

$$\mathrm{E}(y_i|x_i, y \leq a) \neq x_i'\beta \tag{2.567}$$

$$f(x|x > a) = \frac{f(x)}{\mathrm{P}(x > a)} \tag{2.568}$$

试图用大于a的数据，获得全样本的结果(因此前提是假设了数据的分布)

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{1}{2}(\frac{x-\mu}{\sigma})^2} \quad x \sim \mathrm{N}(\mu, \sigma^2) \tag{2.569}$$

$$\mathrm{P(X} > a) = 1 - \Phi(\frac{a-\mu}{\sigma}) \tag{2.570}$$

$$f(x|x > a) = \frac{\frac{1}{\sigma}\phi(\frac{x-\mu}{\sigma})}{1-\Phi(\frac{a-\mu}{\sigma})} \tag{2.571}$$

$$\mathrm{E}[x|x > a] = \int_a^\infty x f(x|x > a) dx \tag{2.572}$$

$$= \int_a^\infty x \frac{\frac{1}{\sigma}\phi(\frac{x-\mu}{\sigma})}{1-\Phi(\frac{a-\mu}{\sigma})} dx \tag{2.573}$$

$$\frac{x-\mu}{\sigma} \overset{def}{==} v \quad \frac{a-\mu}{\sigma} \overset{def}{=} \alpha \tag{2.574}$$

$$= \int_\alpha^\infty (\sigma v + \mu) \frac{\cancel{\frac{1}{\sigma}}\phi(v)}{1-\Phi(\alpha)} \cancel{\sigma} dv \tag{2.575}$$

$$\phi'(x) = -x\phi(x) \tag{2.576}$$

$$= \frac{\sigma}{1-\Phi(\alpha)} \int_\alpha^\infty (-\phi'(x)) dv + \frac{\mu}{1-\Phi(\alpha)} \int_\alpha^\infty \phi(v) dv \tag{2.577}$$

$$= \mu + \sigma \frac{\phi(\frac{a-\mu}{\sigma})}{1-\Phi(\alpha)} \tag{2.578}$$

$$y_i = x_i'\beta + \varepsilon_i, \quad \varepsilon_i \sim \mathrm{N}(0, \sigma^2) \tag{2.579}$$

$$y_i|x_i \sim \mathrm{N}(x_i'\beta, \sigma^2) \tag{2.580}$$

$$\mathrm{E}[y_i|y_i > a] = x_i'\beta + \sigma \frac{\phi(\frac{a-x_i'\beta}{\sigma})}{1-\Phi(\frac{a-x_i'\beta}{\sigma})} \tag{2.581}$$

$$\text{MLE: } \mathrm{L}(\beta, \sigma; x, y) = \prod_{i=1}^n \frac{\frac{1}{\sigma}\phi(\frac{y_i - x_i'\beta}{\sigma})}{1-\Phi(\frac{a-x_i'\beta}{\sigma})} \tag{2.582}$$

2.8 Non Paramtric Model

参数模型$y = m(x) + u \quad \mathrm{E}(y|x)$, $\hat{\theta} \to \theta$, $\sqrt{n}$ consistent if model is true,$\hat{m} \to m$, $\sqrt{n}$,但在模型设定不正确时不收敛。

非参数模型收敛的速度慢，但能正确收敛

$$f(x) = \frac{d\mathrm{F}(x)}{dx} \tag{2.583}$$

$$= \lim_{h\to 0} \frac{\mathrm{F}(x+h) - \mathrm{F}(x-h)}{2h} \tag{2.584}$$

$$= \lim_{h\to 0} \frac{x-h \le rv \le x+h}{2h} \tag{2.585}$$

$$= \lim_{h\to 0} \frac{\{\text{\# of } x_i'\text{s falling in the interval [x-h,x+h]}\}}{2h\dot{n}} \tag{2.586}$$

$$\text{Def } k(\frac{x_i - x}{h}) = k(z) = \{ \begin{matrix} \frac{1}{2} & if|z| < 1 \\ 0 & \text{otherwise} \end{matrix} \tag{2.587}$$

$$= \lim_{h\to 0} \frac{1}{nh} \sum_{i=1}^{n} k(\frac{x_i - x}{h}) \tag{2.588}$$

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{h} k(\frac{x_i - x}{h}) \tag{2.589}$$

k Kernel function 距离中心点越远越小

2nd order kernel function

$$\begin{cases} \int k(v)dv = 1 \\ \int vk(v)dv = 0 \quad \text{奇函数} \\ \int v^2 k(v)dv = k_2 > 0 \end{cases} \tag{2.590}$$

4th order kernel function

$$\begin{cases} \int k(v)dv = 1 \\ \int vk(v)dv = 0 \\ \int v^2 k(v)dv = 0 \\ \int v^3 k(v)dv = 0 \\ \int v^4 k(v)dv = k_4 \neq 0 \end{cases} \tag{2.591}$$

h is bandwith $n, h, \mathrm{X}, k(\dot{})$已知

$\xrightarrow{\text{MSE}}$

$$\mathrm{E}\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n}\mathrm{E}k(\frac{x_i - x}{h}) \quad \text{k is 2nd kernel} \tag{2.592}$$
$$= \frac{1}{h}\mathrm{E}k(\frac{x_i - x}{h}) \tag{2.593}$$
$$\mathrm{E}k(\frac{x_i - x}{h}) = \int f(x_i)k(\frac{x_i - x}{h})dx_i \quad \frac{x_i - x}{h} = v, x_i = x + hv \tag{2.594}$$
$$= \int f(x + hv)k(v)hdv \tag{2.595}$$
$$\text{Taylor} = \int [f(x) + f'(x)hv + \frac{1}{2}f''(x)h^2x^2 + \dots]k(v)hdv \tag{2.596}$$
$$= hf(x) + \frac{h^3}{2}f''(x)\int v^2k(v)dv + o(h^3) \tag{2.597}$$
$$\mathrm{E}\hat{f}(x) = f(x) + \frac{h^2}{2}f''(x)\mathrm{K}_2 + o(h^2) \tag{2.598}$$
$$\text{bias: } \mathrm{E}\hat{f} - f = \frac{h^2}{2}f''(x)\mathrm{K}_2 + o(h^2) \tag{2.599}$$

$$var(\hat{f}) = \frac{1}{nh^2}var(k(\frac{x_i - x}{h})) \tag{2.600}$$
$$var(k(\frac{x_i - x}{h})) = \mathrm{E}k^2(\frac{x_i - x}{h}) - [\mathrm{E}k(\frac{x_i - x}{h})]^2 \tag{2.601}$$
$$\mathrm{E}k^2(\frac{x_i - x}{h}) = \int f(x_i)k^2(\frac{x_i - x}{h})dx \tag{2.602}$$
$$= \int f(x + hv)k^2(v)hdv \tag{2.603}$$
$$= \int [f(x) + f'(x)hv + \frac{f''(x)}{2}h^2v^2 + o(h^2)]k^2(v)hdv \tag{2.604}$$
$$= hf(x)\int k^2(v)dv + o(h) \tag{2.605}$$
$$var(\hat{f}(x)) = \frac{1}{nh}f(x)\int k^2(v)dv + o(\frac{1}{nh}) = \frac{1}{nh}f\mathrm{K} + o(\frac{1}{nh}) \tag{2.606}$$
$$bias^2 + var = \xrightarrow{\text{MSE}} 0 \Rightarrow \xrightarrow{p} 0 \tag{2.607}$$
$$bias^2 + var \xrightarrow{?} 0 \tag{2.608}$$
$$\mathrm{MSE}\hat{f} = \frac{h^4}{4}(\mathrm{K}_2f'')^2 + \frac{fk}{nh} + o(h^4 + \frac{1}{nh}) \tag{2.609}$$
$$\frac{\partial \mathrm{MSE}}{\partial h}\hat{f} = 0 \Rightarrow h_{opt} = [\frac{kf}{(\mathrm{K}_2f'')^2}]^{\frac{1}{5}}n^{-\frac{1}{5}} \stackrel{def}{=} \mathrm{C}(x)n^{-\frac{1}{5}} \tag{2.610}$$
$$\mathrm{C}(x) \quad \text{finite} \tag{2.611}$$

assume $x \sim \mathrm{N}(\mu, \sigma^2)$, $\quad v \sim \mathrm{N}(0,1)$, $\quad h \propto n^{-\frac{1}{5}}$

$$\frac{fk}{nh} = \mathrm{O}(n^{-\frac{4}{5}}), \frac{h^4}{4}(k_2f'')^2 = \mathrm{O}(n^{-\frac{4}{5}}) \tag{2.612}$$

$$g(x) = \mathrm{E}(y|x) \tag{2.613}$$
$$= \int yf(y|x)dy \tag{2.614}$$
$$= \int y\frac{f(y,x)}{f(x)}dy \tag{2.615}$$
$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n}k(\frac{x_i-x}{h}) \tag{2.616}$$
$$\hat{f}(y,x) = \frac{1}{nh_xh_y}\sum_{i=1}^{n}k(\frac{x_i-x}{h_x})k(\frac{y_i-y}{h_y}) \tag{2.617}$$

$$\hat{g}(x) = \frac{\int y\hat{f}(y,x)dy}{\hat{f}(x)} \overset{def}{=} \frac{\hat{m}(x)}{\hat{f}(x)} \tag{2.618}$$
$$\hat{m}(x) = \int y\frac{1}{nh_xh_y}\sum_{i=1}^{n}k(\frac{x_i-x}{h_x})k(\frac{y_i-y}{h_y})dy \tag{2.619}$$
$$\int yk(\underbrace{\frac{y_i-y}{h_y}}_{def\, v})dy, \quad y_i = y+h_yv \tag{2.620}$$
$$= \int_{+\infty}^{-\infty}(y_i-h_yv)k(v)(-h_y)dv \tag{2.621}$$
$$= \int_{-\infty}^{+\infty}(y_i-h_yv)k(v)(h_y)dv \tag{2.622}$$
$$= h_y\int_{-\infty}^{+\infty}(y_i-h_yv)k(v)dv \tag{2.623}$$
$$= h_yy_i \tag{2.624}$$

$$\hat{m}(x) = \frac{1}{nh_x}\sum_{i=1}^{n}y_ik(\frac{x_i-x}{h_x}) \tag{2.625}$$
$$\hat{g}(x) = \frac{\cancel{\frac{1}{nh}}\sum_{i=1}^{n}y_ik(\frac{x_i-x}{h_x})}{\cancel{\frac{1}{nh}}\sum_{i=1}^{n}k(\frac{x_i-x}{h})} = \frac{\hat{m}(x)}{\hat{f}(x)} \tag{2.626}$$
$$= \frac{\mathrm{E}\hat{m}+\hat{m}-\mathrm{E}\hat{m}}{\mathrm{E}\hat{f}+\hat{f}-\mathrm{E}\hat{f}} = \frac{\mathrm{E}\hat{m}+\hat{m}-\mathrm{E}\hat{m}}{\mathrm{E}\hat{f}[1+\frac{\hat{f}-\mathrm{E}\hat{f}}{\mathrm{E}\hat{f}}]} \tag{2.627}$$
$$= \frac{\mathrm{E}\hat{m}+\hat{m}-\mathrm{E}\hat{m}}{\mathrm{E}\hat{f}}[1-\frac{\hat{f}-\mathrm{E}\hat{f}}{\mathrm{E}\hat{f}}+()^2-()^3+\ldots] \tag{2.628}$$
$$= \frac{\mathrm{E}\hat{m}}{\mathrm{E}\hat{f}}+\frac{\hat{m}-\mathrm{E}\hat{m}}{\mathrm{E}\hat{f}}-\frac{\mathrm{E}\hat{m}(\hat{f}-\mathrm{E}\hat{f})}{(\mathrm{E}\hat{f})^2}-\frac{(\hat{m}-\mathrm{E}\hat{m})(\hat{f}-\mathrm{E}\hat{f})}{(\mathrm{E}\hat{f})^2}+\frac{\mathrm{E}\hat{m}(\hat{f}-\mathrm{E}\hat{f})^2}{\mathrm{E}(\hat{f})^3}+\ldots \tag{2.629}$$

$$\mathrm{E}\hat{m} = \mathrm{E}\{\mathrm{E}(\hat{m}(x)|x_i)\} \tag{2.630}$$
$$= \mathrm{E}\{\mathrm{E}(\frac{1}{nh}\sum y_i k(\frac{x_i - x}{h})|x_i)\} \tag{2.631}$$
$$= \mathrm{E}\{\frac{1}{nh}\sum \mathrm{E}(y_i|x_i)k(\frac{x_i - x}{h})\} \tag{2.632}$$
$$= \mathrm{E}\{\frac{1}{nh}\sum g(x_i)k(\frac{x_i - x}{h})\} \tag{2.633}$$
$$= \frac{1}{h}\mathrm{E}g(x_i)k(\frac{x_i - x}{h}) \tag{2.634}$$
$$= \frac{1}{h}\int g(x_i)k(\frac{x_i - x}{h})f(x_i)dx_i \tag{2.635}$$
$$= \frac{1}{\cancel{h}}\int g(x+hv)k(v)f(x+hv)\cancel{h}dv \tag{2.636}$$
$$= \int [g + g'hv + \frac{g''}{2}h^2v^2 + o(h^2)]k(v)[f + f'hv + \frac{f''}{2}h^2v^2 + o(h^2)]dv \tag{2.637}$$
$$= gf + \frac{h^2}{2}[2g'f' + gf'' + g''f]\mathrm{K}_2 + o(h^2) \tag{2.638}$$

$$\mathrm{V}ar\hat{m} = \mathrm{E}\{\underbrace{\mathrm{V}ar(\hat{m}|x_i)}_{\mathrm{A}}\} + \underbrace{\mathrm{V}ar\{\mathrm{E}(\hat{m}|x_i)\}}_{\mathrm{B}} \tag{2.639}$$

law of total variance/variance decomposion/conditional variance /law of iterated variance, 也可以用回归来理解

$$\mathrm{A} = \mathrm{E}\{\mathrm{V}ar[\frac{1}{nh}\sum y_i k(\frac{x_i - x}{h})|x_i]\} \tag{2.640}$$
$$= \mathrm{E}\{\mathrm{V}ar[\frac{1}{nh}\sum y_i k(\frac{x_i - x}{h})|x_i]\} \tag{2.641}$$
$$= \mathrm{E}\{\frac{1}{n^2h^2}n\underbrace{\mathrm{V}ar(y_i|x_i)}_{=\sigma^2}k^2(\frac{x_i - x}{h}) + \underbrace{\mathrm{C}ov}_{=0}\} \tag{2.642}$$
$$= \frac{\sigma^2}{nh^2}\mathrm{E}k^2(\frac{x_i - x}{h}) = \frac{\sigma^2}{nh}f\int k^2(v)dv \tag{2.643}$$
$$= \frac{\sigma^2}{nh}f\mathrm{K} + o(\frac{1}{nh}) \tag{2.644}$$

$$\mathrm{B} = \mathrm{V}ar\{\mathrm{E}[\frac{1}{nh}\sum y_i k(\frac{x_i - x}{h})|x_i]\} \tag{2.645}$$

$$= \mathrm{V}ar\{\frac{1}{nh}\sum g(x_i)k(\frac{x_i - x}{h})\} \tag{2.646}$$

$$= \frac{1}{n^{\cancel{2}}h^2}\cancel{n}\mathrm{V}ar(g(x_i)k(\frac{x_i - x}{h})) \tag{2.647}$$

$$= \frac{1}{nh^2}\{\mathrm{E}g^2k^2(\frac{x_i - x}{h}) - \underbrace{[\mathrm{E}gk(\frac{x_i - x}{h})]^2}_{o_p(1)}\} \tag{2.648}$$

$$= \frac{1}{nh^2}g^2(f + o(\frac{1}{nh}))\int k^2(v)dv \tag{2.649}$$

$$= \frac{1}{nh}g^2 f\mathrm{K} + o(\frac{1}{nh}) \tag{2.650}$$

$$\mathrm{V}ar(\hat{m}(x)) = \frac{1}{nh}[\sigma^2 f + g^2 f]\mathrm{K} + o(\frac{1}{nh}) \tag{2.651}$$

$$\mathrm{C}ov(\hat{m},\hat{f}) = \mathrm{E}\{(\hat{m} - \mathrm{E}\hat{m})(\hat{f} - \mathrm{E}\hat{f})\} \tag{2.652}$$

$$= \mathrm{E}\{\mathrm{E}(\hat{m}|x_i)(\hat{f} - \mathrm{E}\hat{f})\} \tag{2.653}$$

$$= \mathrm{E}\{\frac{1}{nh}\sum g(x_i)k(\frac{x_i - x}{h})\} \tag{2.654}$$

$$= \mathrm{E}\{\frac{1}{nh}\sum g(x_i)k(\frac{x_i - x}{h})[\frac{1}{nh}]\} \tag{2.655}$$

$$= \frac{1}{nh^2}\{\mathrm{E}g(x_i)k^2(\frac{x_i - x}{h}) - \mathrm{E}g(x_i)k(\frac{x_i - x}{h})\mathrm{E}k(\frac{x_i - x}{h})\} \tag{2.656}$$

$$= \frac{1}{nh}gf\mathrm{K} + o(\frac{1}{nh}) \tag{2.657}$$

$$\mathrm{E}\hat{g}(x) = \frac{\mathrm{E}\hat{m}}{\mathrm{E}\hat{f}} - \underbrace{\frac{\mathrm{C}ov(\hat{m},\hat{f})}{(\mathrm{E}f)^2}}_{\mathrm{O}(\frac{1}{nh})} + \underbrace{\frac{\mathrm{E}\hat{m}\mathrm{V}ar\hat{f}}{(\mathrm{E}\hat{f})^3}}_{\mathrm{O}(\frac{1}{nh})} \tag{2.658}$$

$$\approx \frac{\{gh + \frac{h^2}{2}[g'' + gf'' + 2g'f']\int v^2k(v)dv\}}{f + \frac{h^2f''}{2}\int v^2k(v)dv} \tag{2.659}$$

$$= \frac{\{\}}{f[1 + \frac{h^2f''}{2f}\mathrm{K}_2]} \tag{2.660}$$

$$= \frac{\{\}}{f}[1 - (\frac{h^2f''}{2f}\mathrm{K}_2) + ()^2 - ()^3 + \dots] \tag{2.661}$$

$$= g + \frac{h^2}{2f}[g''f + gf'' + 2g'f']\mathrm{K}^2 - \frac{h^2gf''}{2f}\mathrm{K}_2 + o(h^2) \tag{2.662}$$

$$= g + \frac{h^2}{2f}[g''f + 2g'f']\mathrm{K}_2 + o(h^2) \tag{2.663}$$

$$\mathrm{V}ar\hat{g} = \mathrm{E}[\hat{g} - \mathrm{E}\hat{g}]^2 \tag{2.664}$$

$$\hat{g} - \mathrm{E}\hat{g} \approx \frac{\hat{m} - \mathrm{E}\hat{m}}{\mathrm{E}\hat{f}} - \frac{\mathrm{E}\hat{m}(\hat{f} - \mathrm{E}\hat{f})}{(\mathrm{E}\hat{f})^2} \tag{2.665}$$

$$\mathrm{V}ar\hat{g} = \frac{\mathrm{V}ar(\hat{m})}{(\mathrm{E}\hat{f})^2} + \frac{(\mathrm{E}\hat{m})^2\mathrm{V}ar\hat{f}}{(\mathrm{E}\hat{f})^4} - \frac{2\mathrm{E}\hat{m}\mathrm{C}ov(\hat{m},\hat{f})}{(\mathrm{E}\hat{f})^3} \stackrel{def}{=} \mathrm{A} + \mathrm{B} - \mathrm{C} \tag{2.666}$$

$$\mathrm{V}ar\hat{g}(x) = \mathrm{A} + \mathrm{B} - \mathrm{C} \tag{2.667}$$

$$= \frac{1}{nhf}[\sigma^2 + g^2 + g^2 - 2g^2] \int k^2(v)dv + o(\frac{1}{nh}) \tag{2.668}$$

$$= \frac{\sigma^2}{nhf} \int k^2(v)dv + o(\frac{1}{nh}) \tag{2.669}$$

$$\mathrm{MSE}\hat{g} = \frac{h^4}{4f^2(x)}[g''f + 2g'f']^2\mathrm{K}_2^2 + \frac{\sigma^2\mathrm{K}}{nhf} + o(h^4 + \frac{1}{nh}) \tag{2.670}$$

$$h^4 \propto \frac{1}{nh} \Rightarrow h \propto n^{-\frac{1}{5}} \tag{2.671}$$

$$\mathrm{MSE}\hat{g} \propto n^{-\frac{4}{5}} \tag{2.672}$$

$$\hat{f} = f + o_p(n^{-\frac{2}{5}}) \tag{2.673}$$

$$\hat{g} = g + o_p(n^{-\frac{2}{5}}) \tag{2.674}$$

多维Kernel

$$\hat{f}(x_1, x_2, \ldots, x_q) = \frac{1}{nh_1, h_2, \ldots, h_q} \sum_{i=1}^{n} k(\frac{x_{i1} - x_1}{h_1} \frac{x_{i2} - x_2}{h_2} \ldots \frac{x_{iq} - x_q}{h_q}) \tag{2.675}$$

2nd,4th,6th,...kernel 有区别，假设2nd order kernel

$$\mathrm{E}\hat{f} = \frac{1}{h_1 \ldots h_q} \mathrm{E}\frac{x_{i1} - x_1}{h_1} \ldots \frac{x_{iq} - x_q}{h_q} \tag{2.676}$$

$$\int \frac{x_{i1} - x_1}{h_1} \ldots \frac{x_{iq} - x_q}{h_q} f(x_{i1}, \ldots, x_{iq}) dx_{i1}, \ldots, dx_{iq} \tag{2.677}$$

$$= \int k(v_1) \ldots k(v_q) f(x_1 + h_1v_1, \ldots, x_q + h_qv_q) h_1 \ldots h_q dv_1 \ldots dv_q \tag{2.678}$$

$$= \int k(v_1) \ldots k(v_q) \{f(x_1, \ldots, x_q) + \sum_{s=1}^{q} h_sv_sf_s + \frac{1}{2}\sum_s\sum_t h_sh_tv_sv_tf_{st} + (s.o.)\} h_1 \ldots h_q dv_1 \ldots dv_q \tag{2.679}$$

$$= h_1 \ldots h_q f(x_1, \ldots, x_q) + h_1 \ldots h_q \frac{\sum_s h_s^2}{2} f_{ss} \int v_s^2 k(vs) ds + (s.o.) \tag{2.680}$$

$$\mathrm{E}\hat{f} = f + \frac{\sum_s h_s^2 f_{ss}}{2}\mathrm{K}_2 + (s.o.) \tag{2.681}$$

$$\mathrm{V}ar\hat{f} = \frac{1}{nh_1^2 \ldots h_q^2}\{\mathrm{E}k^2(\frac{x_{i1}-x_1}{h_1})\ldots k^2(\frac{x_{iq}-x_q}{h_q})[\mathrm{E}k(\frac{x_{i1}-x_1}{h_1})\ldots k(\frac{x_{iq}-x_q}{h_q})]\} \tag{2.682}$$

$$= \frac{1}{nh_1 \ldots h_q} f(x_1,\ldots,x_q)(\int k^2(v_i)dv)^q + (s.o.) \tag{2.683}$$

$$\mathrm{E}\hat{f}, \mathrm{V}ar\hat{f} \tag{2.684}$$

$$\sqrt{nh}(\hat{f} - f - \frac{h^2}{2}f''\mathrm{K}_2) \sim \mathrm{N}(0, f\mathrm{K}) \tag{2.685}$$