高计笔记

赵凡青

2020 年 9 月 6 日

摘要

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。

1 Review

$$y_i = \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + u_i \quad (i = 1, 2, \ldots, n) \tag{1}$$

$$y = X\beta + u \tag{2}$$

Under ideal conditions, we have BLUE(Best Linear Unbiased Estimator).

$$\min_{\beta} (y - \underset{1*n}{X\beta})'(y - \underset{n*1}{X'\beta}) = \min_{\beta} S(\beta) \tag{3}$$

$$\begin{aligned} F.O.C: \frac{\partial S(\beta)}{\partial \beta} &= \frac{\partial}{\partial \beta}(y'y - \beta' X' y - y' X\beta + \beta' X' X\beta) \\ &= -X'y - X'y + (X'X + X'X)\beta \qquad \Rightarrow \beta = (X'X)^{-1}X'y \\ &= -2X'y + 2X'X\beta \\ &= 0 \end{aligned} \tag{4}$$

unbias: $E(\hat{\beta}) = \beta$ best: $Var(\hat{\beta}) = \sigma^2 (x'x)^{-1}$ asymptotic variance: $\hat{\sigma}^2 = s^2 = \frac{\hat{u}'\hat{u}}{n-k}$
notation:

$$\begin{aligned} y &= X\beta + u \\ &= X\hat{\beta} + \hat{u} \\ &= X\beta + \varepsilon \\ &= X\hat{\beta} + e \end{aligned} \tag{5}$$

$$X\hat{\beta} = \hat{y} \quad y = \hat{y} + \hat{u} = \hat{y} + e \tag{6}$$

$$\begin{aligned} \hat{u} = y - X\beta &= y - X(X'X)^{-1}X'y \\ &= \underbrace{(I_n - x(x'x)^{-1}x')y}_{=M} \\ \hat{u} &= My \\ 0 &= MX \end{aligned} \tag{7}$$

$$P = I_n - M = X(X'X)^{-1}X'X\hat{\beta} = \hat{y} = Py \tag{8}$$

$$\begin{aligned} y - \hat{y} &= \hat{u} \\ y - Py &= My \\ \hat{u} = My &= M(X\beta + u) = Mu \end{aligned} \tag{9}$$

矩阵 M 是将向量变换到与 x 张成的平面垂直的平面上所以 $MX=0$，矩阵 P 是将向量变换到与 x 张成的平面平行的向量 (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)

$$E(\hat{\sigma}^2)=E(\frac{\hat{u}'\hat{u}}{n-k}) \tag{10}$$

$$\hat{u}'\hat{u}=u'M'Mu=u'Mu \quad (symetricidempotent) \tag{11}$$

$$\begin{aligned}
\underset{1*1}{E(u'Mu)} &= tr(E(u'Mu)) \\
&= E(tr(u'Mu)) \\
&= E(tr(Muu')) \Leftarrow tr(AB)=tr(BA) \\
&= tr(E(uMu')) \\
&= tr[E(ME(uu'|x))] \\
underhomo &= tr[E(M\sigma^2)] \\
&= \sigma^2 E[tr(M)] \\
&= \sigma^2(n-k)
\end{aligned} \tag{12}$$

$$\begin{aligned}
tr(M) &= tr(I_n - X(X'X)^{-1}X') \\
&= tr(I_n) - tr((X'X)^{-1}X'X) \Leftarrow tr(AB)=tr(BA), tr(A+B)=tr(A)+tr(B) \\
&= n - tr(I_k) = n-k
\end{aligned} \tag{13}$$

$$\therefore E(\hat{u}'\hat{u})=\sigma^2(n-k) \quad E(\frac{\hat{u}'\hat{u}}{n-k})=\sigma^2 \tag{14}$$

$$\begin{aligned}
&\underset{n*1}{u} \sim N(0,\sigma^2 I_n) \\
&\hat{\beta} \sim N \Leftarrow \hat{\beta}=(X'X)^{-1}X'y=\beta+(X'X)^{-1}X'u \\
&\hat{\beta} \sim N(\beta, \sigma^2(X'X)^{-1})
\end{aligned} \tag{15}$$

$$z_j=\frac{\hat{\beta}_j-\beta_j}{\sqrt{\sigma^2(X'X_{jj}^{-1})}} \sim N(0,1) \tag{16}$$

but σ^2 is unknown, so we need to use $\hat{\sigma}^2$ instead.

$$\begin{aligned}
\frac{\hat{\beta}_j-\beta_j}{\sqrt{s^2(X'X)_{jj}^{-1}}} &= \frac{(\hat{\beta}_j-\beta_j)/\sqrt{\sigma^2(X'X)_{jj}^{-1}}}{\sqrt{[(n-k)s^2/\sigma^2]/(n-k)}} \\
&= \frac{N(0,1)}{\sqrt{\chi^2_{(n-k)}/(n-k)}} \sim t-distribution
\end{aligned} \tag{17}$$

$$(n-k)s^2/\sigma^2=\frac{\hat{u}'\hat{u}}{\sigma^2}=\frac{u'Mu}{\sigma^2}=(\frac{u}{\sigma})'M(\frac{u}{\sigma}) \tag{18}$$

where $u \sim N(0,\sigma^2), so\, \frac{u}{\sigma} \sim N(0,1)$

Theorem 1.1 *if* $\underset{m*1}{Z} \sim N(0,I_m)$*,* A *is a* $m*m$ *symetric idempotent matrix,then*

$$Z'AZ \sim \chi^2(tr(A)) \tag{19}$$

then $(\frac{u}{\sigma})'M(\frac{u}{\sigma}) \sim \chi^2(\underset{n-k}{tr(M)})$ we need to prove the independence of numerator and denominator of equ.(17)

F test $H_0: \underset{J*k}{R} \cdot \underset{k*1}{\beta} = \underset{J*1}{q} \Rightarrow d = R\hat{\beta} - q \sim N$
$E(d) = RE(\hat{\beta}) - q \overset{H_0}{=} 0$(假设检验是在 H_0 成立的情况下进行检验)
$Var(d) = RVar(\hat{\beta})R' = \sigma^2 R(X'X)^{-1}R' \Rightarrow d = R\hat{\beta} - q \sim N(0, \sigma^2 R(X'X)^{-1}R')$
$d'd = \sum_{i=1}^{J} d_i^2$
卡方分布：χ^2 是标准正态的平方和，自由度表示有几个独立的标准正态。
Wald test:$W = d'[Var(d)]^{-1}d = (R\hat{\beta} - q)'[\sigma^2 R(X'X)^{-1}R']^{-1}(R\hat{\beta} - q) \sim \chi^2(J)$

Theorem 1.2 χ^2
If $Z_{m\times 1} \sim N(0,1)$ *then*

$$(Z-\mu)'\Sigma^{-1}(Z-\mu) \sim \chi^2(m) \tag{20}$$

类似对 β 的检验 (β, σ^2) unknown,Σ 中 σ^2 unknown

$$F = \frac{(R\hat{\beta} - q)'[\sigma^2 R(X'X)^{-1}R']^{-1}(R\hat{\beta} - q)/J}{\frac{s^2(n-k)}{\sigma^2}/(n-k)} \sim F(J, n-k) \tag{21}$$

$F = \frac{\chi_1^2/n_1}{\chi_2^2/n_2}$ 要求分子分母独立 (与 t 检验一样)。通常独立 $\Rightarrow$ 不相关，不相关 $\nRightarrow$ 独立, 但是正态分布下，独立 $\Leftrightarrow$ 不相关。为了证明不相关，考虑分子

$$[R\hat{\beta} - q]/\sigma = R(\hat{\beta} - \beta)/\sigma = R(X'X)^{-1}X'u/\sigma = Au/\sigma$$

分母

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

因此只需要考虑 Au/σ 和 Mu/σ 的相关性

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

所以相关性为 0，分子分母独立，服从 $F(J, n-k)$ F 检验的另一种形式表示为，在 $H_0: R\beta = q$ 下

$$F = \frac{(e^{*\prime}e^* - e'e)/J}{e'e/(n-k)} = \frac{(SSR^* - SSR)/J}{SSR/(n-k)} \tag{22}$$

$$\begin{aligned} SSR^*: \min \quad & (y - X\beta)'(y - X\beta) \\ s.t. \quad & R\beta - q = 0 \end{aligned}$$

$$\mathcal{L}(\beta, \lambda) = (y - X\beta)'(y - X\beta) + 2\lambda(R\beta - q) \tag{23}$$

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

保证 $\frac{\partial \mathcal{L}}{\partial \beta}$ 的维度和微分的维度一致 β 是 $k*1$ 维 β' 是 $1*k$ 维

$$\begin{cases} \frac{\partial \mathcal{L}}{\partial \beta}|_{\beta=\hat{\beta}^*} = -2X'(y - X\hat{\beta}^*) + 2R'\hat{\lambda}^* = 0 & (24.1) \\ \frac{\partial \mathcal{L}}{\partial \lambda}|_{\lambda=\hat{\lambda}^*} = 2(R\hat{\beta} - q) = 0 & (24.2) \end{cases} \tag{24}$$

$$\begin{aligned} (24.1) \Rightarrow R'\hat{\lambda}^* &= X'y - X'X\hat{\beta}^* \\ &\Rightarrow R(X'X)^{-1}R'\hat{\lambda}^* = R(X'X)^{-1}X'y - R(X'X)^{-1}X'X\hat{\beta}^* \\ (y = X\hat{\beta} + e) &= R\hat{\beta} + R(X'X)^{-1}X'e - R\hat{\beta}^* \\ &= R(\hat{\beta} - \hat{\beta}^*) = R\hat{\beta} - q \\ &(F.O.C: X'e = 0 \quad or \quad X'e = XMu = 0) \end{aligned} \tag{25}$$

$$
\begin{aligned}
\hat{\lambda}^* &= [R(X'X)^{-1}R']^{-1}(R\hat{\beta}-q) \\
R'\hat{\lambda}^* &= R'[R(X'X)^{-1}R']^{-1}(R\hat{\beta}-q) \\
(\ 24.1) &= X'y - X'X\hat{\beta}^* \\
&= X'(X\hat{\beta}+e) - X'X\hat{\beta}^* \\
&= X'X(\hat{\beta}-\hat{\beta}^*)
\end{aligned} \tag{26}
$$

$$
\Rightarrow \hat{\beta}-\hat{\beta}^* = (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}(R\hat{\beta}-q) \tag{27}
$$

需要计算 $SSR^* = e^{*\prime}e^*$ 和 $SSR = e'e$, 现在已经计算出了 $\hat{\beta}-\hat{\beta}^*$

$$
\begin{aligned}
e^* &= y - X\hat{\beta}^* = e + X(\hat{\beta}-\hat{\beta}^*) \\
e^{*\prime}e^* &= e'e + (\hat{\beta}-\hat{\beta}^*)'X'X(\hat{\beta}-\hat{\beta}^*) \\
&= e'e + (R\hat{\beta}-q)'\cancel{[R(X'X)^{-1}R']^{-1}}R(X'X)^{-1}\cancel{X'X(X'X)^{-1}}R'[R(X'X)^{-1}R']^{-1}(R\hat{\beta}-q) \\
&= e'e + (R\hat{\beta}-q)'[R(X'X)^{-1}R']^{-1}(R\hat{\beta}-q)
\end{aligned} \tag{28}
$$

所以

$$
\begin{aligned}
&\frac{(SSR^*-SSR)/J}{SSR/(n-k)} \\
=&\frac{(R\hat{\beta}-q)'[R(X'X)^{-1}R']^{-1}(R\hat{\beta}-q)/J}{\hat{e}'\hat{e}/(n-K)} \\
=&\frac{(R\hat{\beta}-q)'[R(X'X)^{-1}R']^{-1}(R\hat{\beta}-q)/J}{s^2}
\end{aligned} \tag{29}
$$

化简后式 (21) 和式 (29) 一致。在之前的所有讨论中，包括了 $u \sim N(0,\sigma^2)$,X is non stochastic 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)，所以这些假设过于严格。当数据是随机的时候，需要用到以下工具。
if X is stochastic

Theorem 1.3 *Law of Large Number, LLN*
Let z_i be i.i.d $M \times p$ matrix of observations, $E(z_i)=\mu$, assume $E|z_i|^2$ is finite, then

$$
\frac{1}{n}\sum_{i=1}^{n} z_i \xrightarrow{p} E(z_i) = \mu \tag{30}
$$

LLN 说明样本均值依概率收敛到总体均值

Theorem 1.4 *Central Limit Theory, CLT*
Let z_i be i.i.d $M \times p$ vector of obs, $E(z_i)=\mu$ and $Var(z_i)=\Omega$, then

$$
Z_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(z_i-\mu) \xrightarrow{d} N(0,\Omega) \tag{31}
$$

CLT 说明样本均值的抽样分布依分布收敛到正态分布

Definition 1.1 *Converge in Mean Square Error,$\xrightarrow{MSE}$(也有叫 Converge in Mean Square,$\xrightarrow{L^2}$)*
We Say a sequence of r.v. x_n converges to a constant θ in mean square error(MSE) if $\lim_{n\to\infty} E(x_n-\theta)^2 = 0$,then $x_n \xrightarrow{MSE} \theta$

$$
\begin{aligned}
&E(x_n-\theta)^2 \\
=&E(x_n - E(x_n) + E(x_n) - \theta)^2 \\
=&E[(x_n - Ex_n)^2 + (Ex_n-\theta)^2 + 2(x_n - Ex_n)(Ex_n - \theta)] \\
=&E(x_n - Ex_n)^2 + (Ex_n-\theta)^2 \\
=&Var(x_n) + Bias(x_n)^2
\end{aligned} \tag{32}
$$

$\xrightarrow{MSE} \Leftrightarrow Var(x_n)=0$ and $Bias(x_n)=0$, 所以 $\xrightarrow{MSE} \Rightarrow \xrightarrow{p}$, 一般用 $\xrightarrow{p}$ 表示 consistent，所以 $\xrightarrow{MSE} \Rightarrow$consistent。

$$\hat{\beta}_{OLS}=(X'X)^{-1}X'y=\beta+(X'X)^{-1}X'u \tag{33}$$

$$\hat{\beta}_{OLS}-\beta=(X'X)^{-1}X'u \tag{34}$$

$$\sqrt{n}(\hat{\beta}_{OLS}-\beta)=(\frac{1}{n}X'X)^{-1}\frac{1}{\sqrt{n}}X'u \tag{35}$$

$$(\frac{1}{n}X'X)^{-1}\frac{1}{\sqrt{n}}X'u=(\frac{1}{n}\sum_i^n x_ix_i')^{-1}\frac{1}{\sqrt{n}}\sum_i^n x_iu_i \tag{36}$$

$$LLN \Rightarrow \frac{1}{n}\sum_i^n x_ix_i' \xrightarrow{p} Ex_ix_i' \tag{37}$$

$$CLT \Rightarrow \frac{1}{\sqrt{n}}\sum_i^n x_iu_i \xrightarrow{d} N(0,\sigma^2E(x_ix_i')) \tag{38}$$

$$\sqrt{n}(\hat{\beta}_{OLS}-\beta) \xrightarrow{d} N(0,\sigma^2[E(x_ix_i')]^{-1}) \tag{39}$$

$$\Rightarrow E(\hat{\beta}_{OLS})=0, \quad Var(\sqrt{n}\hat{\beta}_{OLS})=\sigma^2(\frac{1}{n}X'X)^{-1} \tag{40}$$

$$\frac{\hat{\beta}_j-\beta_j}{\sqrt{\sigma^2(X'X)^{-1}_{jj}}} \sim N(0,1) \tag{41}$$

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

在经典假设下

$$Var(u)=Eu'u=\begin{pmatrix}\sigma^2 & \dots & 0\\ \vdots & \ddots & \vdots\\ 0 & \dots & \sigma^2\end{pmatrix}_{n\times n} \tag{42}$$

放宽假设, 如果 $\sigma_1^2\neq\sigma_2^2\neq\ldots\neq\sigma_n^2$ 且对角线以外的元素不为零, 会导致 $Var(\hat{\beta})=\hat{\sigma}^2(X'X)^{-1}$ 对方差的估计错误，但一般认为这个估计不会产生其他错误，即仍旧满足一致性 $\hat{\beta}_{OLS}\xrightarrow{p}\beta$

在不满足经典假设的情况下 $\hat{\beta}_{OLS}$ 不是最优的估计量 (not best), 所以可以使用 GLS。使用 GLS 的前提是估计 $Var(u)$ 的结构。但是 $Var(u)$ 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。

(1) **Heteroskadasticity** 需要估计 n 个参数

$$\begin{pmatrix}\sigma_1^2 & \dots & 0\\ \vdots & \sigma_i^2 & \vdots\\ 0 & \dots & \sigma_n^2\end{pmatrix}_{n\times n} \tag{43}$$

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

$$\begin{pmatrix}\sigma_1^2 & \sigma_{12}^2 & \dots & \sigma_{1n}^2\\ \sigma_{21}^2 & \sigma_{22}^2 & \dots & \sigma_{2n}^2\\ \vdots & \sigma_{ij}^2 & \ddots & \vdots\\ \sigma_{n1}^2 & \dots & \dots & \sigma_{nn}^2\end{pmatrix}_{n\times n} \tag{44}$$

(3) 其他参数估计方法 e.g. $\sigma_i^2=\sigma^2\exp(z_i'\alpha)$,$\alpha$ 是 $k\times 1$ 向量 $k<<n$, 或者记为 $\sigma_i^2=\sigma_i^2(\theta),\theta(\sigma^2,\alpha)$

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

$$\ln L(\beta,\theta|x,y,z)=-\frac{n}{2}\ln(2\pi)-\frac{1}{2}\sum_{i=1}n\ln\sigma_i^2(\theta)-\frac{1}{2}\sum_{i=1}^n\frac{(y_i-x_i'\beta)^2}{\sigma_i^2(\theta)} \tag{45}$$

If homos $\sigma_i^2(\theta)=\sigma^2$ MLE $\Leftrightarrow$ OLS now we have $\frac{1}{\sigma_i^2(\theta)}$ MLE $\Leftrightarrow$ GLS. Under hetero GLS:$Var(\hat{\beta}_{GLS})\downarrow$

Group *i.i.d* data 顺序可变

$$\begin{pmatrix} \sigma_B^2 & & & \\ & \sigma_S^2 & & \\ & & \sigma_B^2 & \\ & & & \ddots \end{pmatrix} \Rightarrow \begin{pmatrix} \sigma_B^2 & & & \\ & \sigma_B^2 & & \\ & & \sigma_S^2 & \\ & & & \ddots \end{pmatrix} \tag{46}$$

Groupwise HET

$$\hat{\beta}_{GLS} = [\sum_{g=1}^{G} \frac{1}{\hat{\sigma}_g^2} X_g' X_g]^{-1} [\sum_{g=1}^{G} (\frac{1}{\hat{\sigma}_g^2} X_g' y_g)] \tag{47}$$

$$X = \begin{pmatrix} X_1 \\ \vdots \\ X_g \end{pmatrix}, \hat{\sigma}_g^2 = \frac{e_g' e_g}{n_g} \tag{48}$$

在 $\sigma_i^2 = \sigma^2 \exp(z_i'\alpha)$ 中,Z 是分组 dummy Matrix

$$Z = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ & \vdots & & \end{pmatrix} \tag{49}$$

假设 R 是 Transformation Matrix

$$y = X\beta + u \quad Var(u) \tag{50}$$

$$Ry = RX\beta + Ru \to homo \tag{51}$$

$$y^* = X^*\beta + u^* \tag{52}$$

$$\Rightarrow \hat{\beta}_{OLS} = (X'R'RX)^{-1}X'R'Ry, \quad R'R = \frac{1}{\hat{\sigma_g^2}} \tag{53}$$

异方差稳健标准误 Heteroskadasticity Robust Standard Error

$$y = X\beta + u \tag{54}$$

$$\hat{\beta}_{OLS} = \beta + (X'X)^{-1}X'u \tag{55}$$

$$Var(\hat{\beta}_{OLS}) = (X'X)^{-1}X'Euu'X(X'X)^{-1} \tag{56}$$

$$= (X'X)^{-1}X'\Sigma'X(X'X)^{-1} \tag{57}$$

$\hat{\Sigma}$ 仍旧有 n 个待估参数，所以直接估计 $X'\Sigma X_{k\times k}$ 一般 $k << n$

$$X'\Sigma X = \Sigma_i \Sigma_j \sigma_{ij} x_j x_j' \tag{58}$$

σ_{ij} 是 Σ 的第 i,j 个元素，在对角线以外元素都为 0 的情况下

$$X'\Sigma X = \Sigma_i \sigma_i^2 x_i x_i' \tag{59}$$

$$\begin{aligned} E(u_i^2 x_i x_i') &= E[E(u_i^2|x_i)x_i x_i'] \\ &= E[\sigma_i^2 x_i x_i'] \\ \frac{1}{n}\sum_{i=1}^{n} \hat{u}_i{}^2 x_i x_i' &\to \frac{1}{n}\sum_{i=1}^{n} \sigma_i^2 x_i x_i' \end{aligned} \tag{60}$$

$$\hat{Var}(\hat{\beta}_{OLS}) = (X'X)^{-1}X' \begin{pmatrix} \hat{u_1}^2 & & \\ & \ddots & \\ & & \hat{u_n}^2 \end{pmatrix} X(X'X)^{-1} \tag{61}$$

在 STATA 等软件中会用到 (虽然看上去只是把 Σ 矩阵替换了, 但由于 $\hat{u}^2 \neq \hat{\sigma}^2$, 所以本质上是替换了 $X'\Sigma X$ 矩阵)

关于自相关，考虑

$$u_t = \rho u_{t-1} + \varepsilon_t, \quad |\rho| < 1 \tag{62}$$

$$\Rightarrow \sigma_u^2 = \rho\sigma_u^2 + \sigma_\varepsilon^2 \tag{63}$$

OLS 估计 ρ $\varepsilon\, is\, w.n.$

$$\sigma_u^2 = \frac{\sigma_\varepsilon^2}{1-\rho} \tag{64}$$

$$E(u_t u_{t-1}) = \rho\sigma_u^2 \tag{65}$$

$$E(u_t u_{t-2}) = \rho^2\sigma_u^2 \tag{66}$$

$$E(uu') = \frac{\sigma_\varepsilon^2}{1-\rho} \begin{pmatrix} 1 & \rho & \rho^2 & \dots \\ \rho & 1 & \rho & \vdots \\ \rho^2 & \rho & \ddots & \vdots \\ \vdots & \dots & \dots & 1 \end{pmatrix} \tag{67}$$

Transformation Matrix

$$R = \begin{pmatrix} \sqrt{1-\rho^2} & 0 & \dots & 0 \\ -\rho & 1 & \ddots & \vdots \\ 0 & -\rho & 1 & \vdots \\ \ddots & \dots & -\rho & 1 \end{pmatrix} \tag{68}$$

$$Ru = \begin{pmatrix} \sqrt{1-\rho^2}u_1 \\ u_2 - \rho u_1 \\ u_3 - \rho u_2 \\ \vdots \\ u_n - \rho u_{n-1} \end{pmatrix} \tag{69}$$

因此方程 $y = X\beta + u$ 转化为 $Ry = Rx + \underset{n\times n}{R}\underset{n\times 1}{u}$, $Var(u)$ 是一个对角线以外元素不为 0 的方阵，而 $Var(Ru) = RVar(u)R'$ 是一个 $\begin{pmatrix}\ddots\end{pmatrix}$ 对角阵，且对角线元素相等

Newey-West 对于自相关，考虑 $(X'X)^{-1}X'\Sigma'X(X'X)^{-1}$, 尝试用

$$\Sigma = \hat{u}\hat{u}' = \begin{pmatrix} \hat{u_1}^2 & \hat{u}_1\hat{u}_2 & \dots & \hat{u}_1\hat{u}_n \\ \hat{u}_2\hat{u}_1 & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \dots & \dots & \dots & \hat{u_n}^2 \end{pmatrix} \tag{70}$$

但是 $(X'X)^{-1}X'\hat{u}\hat{u}'X(X'X)^{-1} = 0$(因为 $X'\hat{u} = 0$), 所以有问题。

$$\begin{pmatrix} \hat{u_1}^2 & \frac{3}{4}\hat{u}_1\hat{u}_2 & \frac{1}{2}\hat{u}_1\hat{u}_3 & \frac{1}{4}\hat{u}_1\hat{u}_4 & 0 & 0 & \dots \\ & & & \vdots & & & \end{pmatrix} \tag{71}$$

$$W = \begin{cases} 1 - \dfrac{l}{L} & l \le L \\ 0 & l > L \end{cases} \tag{72}$$

其中 $L = floor(n^{\frac{1}{4}}) \quad e.g. n = 100 \quad L = 3 \quad l = |i-j|$

在之前的假设中, 大样本放松了 $u \sim N$ 的假设，异方差放松了 $E(uu') = \sigma^2 I$ 的假设。还有一条假设为 u 和 x 不相关

$$y = X\beta + u \tag{73}$$

$$E(u|X) = 0 \tag{74}$$

因此放松该假设到允许 X, u 相关，此时 $\hat{\beta} = \beta + (X'X)^{-1}X'u$ 是有偏的。所以接下来的问题是如果 $E(u_i|x_i) = 0$, $E(u_i|x_{i+1}) \neq 0$ 即 $E(u|X) \neq 0$ 怎么办。

$$\begin{aligned} \hat{\beta} - \beta &= (\frac{X'X}{n})^{-1}(\frac{X'u}{n}) \\ &= (\frac{1}{n}\sum x_i x_i')^{-1}\frac{1}{n}x_i u_i \\ &= (Ex_i x_i')^{-1}(Ex_i u_i) \quad E(x_i u_i) = E(x_i E(\underbrace{u_i|x_i}_{=0})) = 0 \end{aligned} \tag{75}$$

因此 $\hat{\beta}$ 虽然不是无偏的但是再大样本情况下是一致的 $\hat{\beta} - \beta \neq 0$, 但是有 $\hat{\beta} - \beta \xrightarrow{p} 0$

unbiased vs consistent

1. 考虑 $E(y_t) = \mu$,$\hat{\mu}_1 = \frac{1}{n+1}\sum_{t=1}^n y_t$,$E\hat{\mu}_1 = \frac{n}{n+1}\mu \neq 0$, 但是 $p\lim_{n\to\infty}\frac{n}{n+1}\frac{1}{n}\sum_{t=1}^n y_t = \mu$。所以是有偏但是一致。

2. 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 (y_1 仍旧是随机变量不等于 μ)。

$$\hat{\beta} = \beta + (Ex_i x_i')^{-1}(Ex_i u_i) \tag{76}$$

consistency 要求 $(Ex_i u_i) \to 0$, 但是当存在内生变量时，$Ex_i u_i$ 不是零向量

1. Measurement Error

$$y_t^0 = \beta_1 + \beta_2 x_t^0 + u_t^0 \tag{77}$$

$$x_t = x_t^0 + v_{1t} \qquad yt = y_t^0 + v_{2t} \tag{78}$$

$$y_t = \beta_1 + \beta_2(x_t - v_{1t}) + u_t^0 + v_{2t} \tag{79}$$

$$= \beta_1 + \beta_2 x_t + [\underbrace{u_t^0 + v_{2t} - \beta_2 v_{1t}}_{u_t}] \tag{80}$$

$\sigma^2(X'X)^{-1}$ 变大，导致 $V\hat{a}r(\hat{\beta}_O LS)$ 变大，是的估计不精确。(估计不有效 not efficient 表示存在方差更小的估计方法)

$$Cov(x_t, u_t) = E(x_t u_t) = E[(x_t^0 + v_{1t})(u_t^0 + v_{2t} - \beta_2 v_{1t})] = -\beta Var(v_{1t}) \tag{81}$$

考虑收入和消费的关系

$$\Uparrow\downarrow y_t = \beta_1 + \beta_2 x_t \Uparrow + u_t \downarrow \tag{82}$$

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)。

2. Simultaneous Equation

$$\begin{cases} q_t = \gamma_d p_t + x_t^d \beta_d + u_t^d \\ q_t = \gamma_s p_t + x_t^s \beta_s + u_t^s \end{cases} \tag{83}$$

p_t q_t 与 u_t^d, u_t^s

$$\begin{pmatrix} q_t \\ p_t \end{pmatrix} = \begin{pmatrix} 1 & -\gamma_d \\ 1 & -\gamma_s \end{pmatrix}^{-1} + \left[\begin{pmatrix} x_t^d & \beta_d \\ x_t^s & \beta_s \end{pmatrix} \begin{pmatrix} u_t^d \\ u_t^s \end{pmatrix} \right] \tag{84}$$

IV

$$X = (X_1, \underset{n\times k_2}{X_2}) \underset{n\times l}{W} = (\underset{n\times k_1}{X_1}, \underset{n\times (l-k_1)}{X_{IV}}) \tag{85}$$

$$y = X_1\gamma_1 + X_2\gamma_2 + u \qquad \rho(X_{IV}, u) = 0 \quad l - k_1 \geq k_2 \tag{86}$$

其中 X_1 是外生变量，X_2 是内生变量

2SLS

fisrt stage

$$\hat{X}_2 = X_1\hat{\Pi}_1 + X_{IV}\hat{\Pi}_2 \tag{87}$$

其中要求 $\hat{\Pi}_2$ 显著 (sigificant)，且 $F > 10$, 如果 $F \leq 10$，则是 weak IV。

second stage

$$y = X_1\gamma_1 + \hat{X}_2\gamma_2 + u \tag{88}$$

$$\hat{X}_2 = (X_1, X_{IV}) \begin{pmatrix} \hat{\Pi}_1 \\ \hat{\Pi}_2 \end{pmatrix} = W \begin{pmatrix} \hat{\Pi}_1 \\ \hat{\Pi}_2 \end{pmatrix} \tag{89}$$

$$\hat{X}_2 = P_W X_2 \tag{90}$$

$$X_1 = \hat{X}_1 = P_w X_1 \tag{91}$$

$$(X_1 = X_1\delta_1 + X_2\delta_2)$$

所以

$$\hat{X} = (\hat{X}_1, \hat{X}_2) = P_w X \tag{92}$$

$$\tag{93}$$

$$\begin{aligned} y &= X_1\gamma_1 + X_2\gamma_2 + u \\ &= (\hat{X}_1, \hat{X}_2) \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} + u \\ &= \hat{X}\delta + u \end{aligned} \tag{94}$$

$$\begin{aligned} \hat{\delta_{2SLS}} &= (\hat{X}'\hat{X})^{-1}\hat{X}'y \\ &= (X'P_W'P_W X)^{-1}X'P_W'y \\ &= (X'P_W X)^{-1}X'P_W y \end{aligned} \tag{95}$$

see «Econometric Theory and Methods» for more information.

An Example

$$CigarettePricePaid = \beta_1 BC + \beta_2 X + u_t \tag{96}$$

距离边境的距离是一个和 Border Crossing 相关但是和 price paid 无关的工具变量。

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

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 X_1, X_2 里，所以若有关，则必定与 u 相关)

Test of Overidentifying Restrictions (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。

$$\begin{aligned} &H_0 : EW'u = 0 \quad W = (X_1, X_{IV}) \\ &H_1 : EW'u \neq 0 \end{aligned} \tag{97}$$

考虑回归方程

$$y = X_1\beta_1 + X_2\beta_2 + u \tag{98}$$

$$W = (X_1, X_{IV}) \tag{99}$$

利用 2SLS

$$y = X_1\hat{\beta}_{2SLS,1} + X_2\hat{\beta}_{2SLS,2} + \hat{u} \tag{100}$$

regres $\hat{u}$ on W $\hat{u} = W\hat{b} + error$ 可以计算出 R^2,R^2 越高，说明相关性越强。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 nR^2

$$test = nR^2 = n\frac{SSE}{SST} = \frac{\hat{b}'W'W\hat{b}}{\frac{1}{n}\hat{u}'\hat{u}} \tag{101}$$

根据之前的笔记 $\hat{\sigma}^2 = \frac{\hat{u}'\hat{u}}{n-k} \quad y = X\beta + u \quad My = Mu = \hat{u} \quad Py = \hat{y} \quad y = \hat{y} + \hat{u}$ 所以 $W\hat{b} = P_W\hat{u}$, 式 (101) 可以写作

$$= \frac{\hat{u}'P_W'P_W\hat{u}}{\frac{1}{n}\hat{u}'\hat{u}} \sim \chi^2(q) \tag{102}$$

$$\begin{aligned} \underset{n\times l}{W} &= (X_1, X_{IV}) \\ \underset{n\times k}{X} &= (X_1, X_2) \\ \hat{\beta}_{2SLS} &= (X'P_WX)^{-1}X'P_Wy \end{aligned}$$

$$\begin{aligned} \hat{u} &= y - X\hat{\beta}_{2SLS} \\ &= [I - X(X'P_WX)^{-1}X'P_W](X\beta + u) \\ &= [I - X(X'P_WX)^{-1}X'P_W]u \end{aligned} \tag{103}$$

$$\begin{aligned} \hat{u}'P_W\hat{u} &= u'[I - P_WX(X'P_WX)^{-1}X']P_W[I - X(X'P_WX)^{-1}X'P_W]u \\ &= u'[P_W - P_WX(X'P_WX)^{-1}P_W]u \sim \chi^2(?) \end{aligned} \tag{104}$$

$$tr(P_W - P_WX(X'P_WX)^{-1}P_W) = tr(P_W) - tr(P_WX(X'P_WX)^{-1}X'P_W) \tag{105}$$

$$= tr(W(W'W)^{-1}W') - tr(P_WX(X'P_WX)^{-1}X'P_W) \tag{106}$$

$$= tr(W'W)^{-1}W'W) - tr((X'P_WX)^{-1}X'P_WP_WX) \tag{107}$$

$$= tr(I_l) - tr(I_k) = l - k \tag{108}$$

$$\hat{u}'P_W\hat{u} \sim \chi^2(l - k) \tag{109}$$

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。(如果恰好识别则不能检验。)

Hausman Test Hausman Test 是找到两个不同的估计量，在一个估计量在 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

H_0 : OLS is consistent, 2SLS is consistent ,but not efficient

H_1 : OLS is not consistent,2SLS is consistent 因此想法将 $\hat{\beta}_{OLS}-\hat{\beta}_{2SLS}$ 变成一个标准的分布，记 $\hat{\beta}_{IV}=\hat{\beta}_{2SLS}$ 已知

$$\hat{\beta}_{OLS}=(X'X)^{-1}X'y$$
$$\hat{\beta}_{IV}=(X'P_WX)^{-1}X'P_Wy$$

$$\begin{aligned}\hat{\beta}_{IV}-\hat{\beta}_{OLS}&=(X'P_WX)^{-1}X'P_Wy-(X'X)^{-1}X'y\\&=(X'P_WX)^{-1}X'P_W(X\hat{\beta}_{OLS}+\hat{u}_{OLS})-\hat{\beta}_{OLS}\\&=(X'P_WX)^{-1}X'P_WM_Xy\\&=(X'P_WX)^{-1}X'P_WM_Xu\end{aligned}\tag{110}$$

转换成含 u 的表达式后，思路是找 χ^2

$$test=(\hat{\beta}_{IV}-\hat{\beta}_{OLS})'[Var(\hat{\beta}_{IV}-\hat{\beta}_{OLS})]^{-1}(\hat{\beta}_{IV}-\hat{\beta}_{OLS})\tag{111}$$
$$=\chi^2(?)\tag{112}$$

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

$$\begin{aligned}\hat{\beta}_{IV}-\hat{\beta}_{OLS}&=(X'P_WX)^{-1}\begin{pmatrix}X_1'\\X_2'\end{pmatrix}P_WM_Xu\\&=(X'P_WX)^{-1}\begin{pmatrix}X_1'P_WM_Xu\\X_2'P_WM_Xu\end{pmatrix}\qquad(X_1'P_W=X_1,X_1M_X=0)\\&=(X'P_WX)^{-1}\begin{pmatrix}0\\X_2'P_WM_Xu\end{pmatrix}\end{aligned}\tag{113}$$

$$\begin{aligned}Var(\hat{\beta}_{IV}-\hat{\beta}_{OLS})&=E\left[(X'P_WX)^{-1}\begin{pmatrix}0\\X_2'P_WM_Xu\end{pmatrix}(0'\;u'M_XP_WX_2)(X'P_WX)\right]\\&=E\left[(X'P_WX)^{-1}\begin{pmatrix}0&0\\0&X_2'P_WM_Xuu'M_XP_WX_2\end{pmatrix}(X'P_WX)^{-1}\right]\end{aligned}\tag{114}$$

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

$$Var(\hat{\beta}_{IV}-\hat{\beta}_{OLS})=E\left[(X'P_WX)^{-1}\begin{pmatrix}A&0\\0&X_2'P_WM_X\sigma^2M_XP_WX_2\end{pmatrix}(X'P_WX)^{-1}\right]\tag{115}$$

$$Var(\hat{\beta}_{IV}-\hat{\beta}_{OLS})^{-1}=E\left[(X'P_WX)\begin{pmatrix}A&0\\0&X_2'P_WM_X\sigma^2M_XP_WX_2\end{pmatrix}^{-1}(X'P_WX)\right]\tag{116}$$

$$
\begin{aligned}
test &= (0'\, u'M_XP_WX_2)(X'P_WX)^{-1}E\left[(X'P_WX)\begin{pmatrix}A^{-1} & 0\\ 0 & (X_2'P_WM_X\sigma^2M_XP_WX_2)^{-1}\end{pmatrix}(X'P_WX)\right] \\
&(X'P_WX)^{-1}\begin{pmatrix}0\\ X_2'P_WM_Xu\end{pmatrix} \\
&= (0'\, u'M_XP_WX_2)\begin{pmatrix}A^{-1} & 0\\ 0 & (X_2'P_WM_X\sigma^2M_XP_WX_2)^{-1}\end{pmatrix}\begin{pmatrix}0\\ X_2'P_WM_Xu\end{pmatrix} \\
&= u'M_XP_WX_2(X_2'P_WM_X\sigma^2M_XP_WX_2)^{-1}X_2'P_WM_Xu \\
&= \frac{1}{\sigma^2}u'M_XP_WX_2(X_2'P_WM_XM_XP_WX_2)^{-1}X_2'P_WM_Xu
\end{aligned} \tag{117}
$$

所以 A 是否为 0 并不重要

$$
tr(M_XP_WX_2(X_2'P_WM_XM_XP_WX_2)^{-1}X_2'P_WM_X) = tr(X_2'P_WM_XM_XP_WX_2)^{-1}X_2'P_WM_XM_XP_WX_2) = k_2 \tag{118}
$$

因此，Hausman Test

$$
test \sim \chi^2(k_2) \tag{119}
$$

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。

$$
E(functions of r.v. and par) = 0 \tag{120}
$$

就是 poplution moment condition/restriciton(POPMC)

e.g. 总体矩条件 $E(X-\mu)=0$ 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

$$
y = X\beta + u \tag{121}
$$

$$
EX'u = 0 \tag{122}
$$

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

$$
EX'u = 0 \tag{123}
$$

$$
\Rightarrow E\sum_{i=1}^{n} x_iu_i = 0 \tag{124}
$$

$$
\Rightarrow \sum_{i=1}^{n} Ex_iu_i = 0 \tag{125}
$$

$$
\Rightarrow \cancel{n}Ex_iu_i = 0 \tag{126}
$$

$$
\Rightarrow \frac{1}{n}\sum_{i=1}^{n} = 0 \quad X'u = 0 \tag{127}
$$

即得到样本矩条件

$$
X'(y-X\beta) = 0 \tag{128}
$$

$$
X'y - X'X\beta = 0 \tag{129}
$$

$$
\Rightarrow \hat{\beta}_{MM} = (X'X)^{-1}X'y = \hat{\beta}_{OLS} \tag{130}
$$

example 2 $E(Xu) \neq 0$

$$y = X_1\beta + X_2\beta + u \tag{131}$$

$EX'u \neq 0$ but $EW'u = 0$ 所以样本矩条件为

$$W'u = 0 \tag{132}$$

$$W'(y - X\beta) = 0 \tag{133}$$

$$\hat{\beta}_{MM} = (W'X)^{-1}W'y \tag{134}$$

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

考虑 $\hat{\beta}_{2SLS}$

$$\begin{aligned}\hat{\beta}_{2SLS} &= (X'P_WX)^{-1}X'P_Wy \\ &= (X'W(W'W)^{-1}W'X)^{-1}XW(W'W)^{-1}W'y\end{aligned} \tag{135}$$

由于 $X'W$,WW,$W'X$ 都是方阵，所以

$$(X'W(W'W)^{-1}W'X)^{-1} = (W'X)^{-1}(W'W)(X'W)^{-1} \tag{136}$$

$$\Rightarrow \hat{\beta}_{2SLS} = (W'X)^{-1}(W'W)(X'W)^{-1}XW'(W'W)^{-1}W'y \tag{137}$$

$$\Rightarrow \hat{\beta}_{2SLS} = (W'X)^{-1}W'y \tag{138}$$

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

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

GMM

$$g(z, \theta) = \begin{pmatrix} g_1(z, \theta) \\ g_2(z, \theta) \\ \vdots \\ g_L(z, \theta) \end{pmatrix}_{L\times 1} \tag{139}$$

$$\hat{g}_n(\theta) = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n} g_1(z_i, \theta) \\ \frac{1}{n}\sum_{i=1}^{n} g_2(z_i, \theta) \\ \vdots \\ \frac{1}{n}\sum_{i=1}^{n} g_L(z_i, \theta) \end{pmatrix}_{L\times 1} = \frac{1}{n}\sum_{i=1}^{n} g(z, \theta) \tag{140}$$

k: # of unknown parameters (未知参数的个数)

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

$L > k$ GMM

$L = k$ MM

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

如果 $L > k$ 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

$$Q_n^W(\theta) = \underset{1\times L}{\hat{g}_n(\theta)'} \quad \underset{L\times L}{W} \quad \underset{L\times 1}{\hat{g}_n(\theta)} \tag{141}$$

其中 W 是任意的正定矩阵，所以有

$$\hat{\theta}_{GMM} = argmin\, Q_n^W(\theta) \tag{142}$$

$E(r.v., para) = 0$ $\frac{1}{n}\sum() = \hat{g}_n(\theta)$
需要找到 $\hat{\theta}_{GMM}$ 的分布 $(\hat{\beta}_{OLS} \sim N(\beta, \sigma^2(X'X)^{-1}))$

Theorem 1.5 ***Asymptotic normality of GMM Estimator***

Under appropriate conditions, we have

$$\sqrt{n}(\hat{\theta}_{GMM} - \theta_0) \overset{d}{\to} N(0, (G'WG)^{-1}G'W\Omega_0 WG(G'WG)^{-1}) \tag{143}$$

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

$$\underset{L\times 1}{\nabla_\theta g(z,\theta)} = \frac{\partial g(z,\theta)}{\partial \theta'} = \begin{pmatrix} \frac{\partial g_1}{\partial \theta_1} & \cdots & \frac{\partial g_1}{\partial \theta_k} \\ \vdots & & \vdots \\ \frac{\partial g_L}{\partial \theta_1} & \cdots & \frac{\partial g_L}{\partial \theta_k} \end{pmatrix}_{L\times k} \tag{144}$$

$$\nabla_{\theta'} g = \frac{\partial g'}{\partial \theta} = \left(\frac{\partial g}{\partial \theta}\right)'_{k\times L} \tag{145}$$

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

Proof

$$F.O.C: \frac{\partial Q_n^W(\hat{\theta})}{\partial \theta} = \underset{k\times L}{\frac{\partial \hat{g}'_{1\times L}}{\partial \theta_{k\times 1}}} W_{L\times L}\hat{g}_{L\times 1} + \underset{L\times 1}{\underset{1\times L}{\left(\hat{g}'W\frac{\partial \hat{g}_{L\times 1}}{\partial \theta'_{1\times k}}\right)'}} \tag{146}$$

$$= 2\underset{k\times L}{\nabla_{\theta'}\hat{g}_n(\hat{\theta})}\,\underset{L\times L}{W}\,\underset{L\times 1}{\hat{g}_n(\hat{\theta})} = 0 \tag{147}$$

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

$$\hat{g}_n(\hat{\theta}) = \hat{g}_n(\theta_0) + \nabla_\theta \hat{g}_n(\bar{\theta})(\hat{\theta} - \theta_0) \tag{148}$$

可以得到

$$\frac{\partial \hat{Q}(\hat{\theta})}{\partial \theta} = 2\nabla_{\theta'}\hat{g}_n W[\hat{g}_n(\theta_0) + \nabla_\theta \hat{g}_n(\bar{\theta})(\hat{\theta} - \theta_0)] = 0 \tag{149}$$

$$\sqrt{n}(\hat{\theta} - \theta_0) = -[\nabla_{\theta'}\hat{g}_n W \nabla_\theta \hat{g}_n(\bar{\theta})]^{-1}\nabla_{\theta'}\hat{g}_n W\sqrt{n}\nabla_\theta \hat{g}_n(\theta_0) \tag{150}$$

$$\nabla_{\theta'}\hat{g_n}(\hat{\theta}) = \nabla_{\theta'}\frac{1}{n}\sum_{i=1}^{n} g(z_i,\theta) \to E\nabla_{\theta'} g(z_i,\hat{\theta}) = G'(\hat{\theta}) \to G'(\theta_0) \tag{151}$$

$$\therefore \sqrt{n}(\hat{\theta} - \theta_0) = -[(G' + o_p(1))\ldots]^{-1}\ldots \tag{152}$$

Review $O(1), o(1), O_p(1), o_p(1)$

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

$$\sqrt{n}(\hat{\theta}-\theta_0) = -[G'WG]^{-1}G'W\sqrt{n}\hat{g}_n(\theta_0) \tag{153}$$

$$\sqrt{n}\hat{g}_n(\theta_0) = \sqrt{n}\frac{1}{n}\sum_{i=1}^{n} g(z_i,\theta_0) \sim N(0, Egg') = N(0,\Omega_0) \tag{154}$$

$$\therefore \sqrt{n}(\hat{\theta}-\theta_0) = N(0,(G'WG)^{-1}G'W\Omega_0 WG(G'WG)^{-1}) \tag{155}$$

如果 $\hat{\theta}_{GMM} \nrightarrow \theta_0$ 不能使用 Taylor Expansion, 根据 Uniform Weak Law of Large Number,if $\hat{Q}_n(\theta) \to Q_0(\theta)$,then $\hat{\theta}_{GMM} = argmin\hat{Q}_n(\theta) \to \theta_0 = argminQ_0(\theta)$

$$\hat{Q}_n(\theta) = \hat{g}_n(\theta)'W\hat{g}_n(\theta) = [\frac{1}{n}\sum_{i=1}^{n} g(z_i,\theta)]'W[\frac{1}{n}\sum_{i=1}^{n} g(z_i,\theta)] \tag{156}$$

$$Q_0(\theta) = Eg(z,\theta)'WEg(z,\theta) \tag{157}$$

令 $W=\Omega_0^{-1}$, 有

$$(G'WG)^{-1}G'W\Omega_0WG(G'WG)^{-1} = (G'\Omega_0^{-1}G)^{-1} \tag{158}$$

$W=\Omega_0^{-1}$ 被称为 optimal weighting matrix,$(G'\Omega_0^{-1}G)^{-1}$ 被称为 optimal variance,$\hat{\theta}_{GMM}$ 被称为 efficient GMM estimator。但由于 $W=\Omega_0^{-1}$ 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 $\hat{\Omega}$

$$\hat{\Omega} = \frac{1}{n}\sum_{i=1}^{n} g(z_i,\hat{\theta})g(z_i,\hat{\theta})' \tag{159}$$

$$\tilde{\theta} = \arg\min_{\theta\in \textcircled{H}}[\hat{g}_n'](\theta)\hat{g}_n(\theta)] = \arg\min_{\theta\in H} Q_n^I(\theta) \tag{160}$$

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

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

$$Var(\sqrt{n}(\tilde{\theta}^*_{GMM}-\theta_0)) = (\hat{G}'(\tilde{\theta}^*_{GMM})\hat{\Omega}^{-1}(\tilde{\theta}^*_{GMM})\hat{G}(\tilde{\theta}^*_{GMM}))^{-1} \tag{161}$$

2. continuous updating method

$$\hat{\theta} = \arg\min Q_n(\theta) = \arg\min \hat{g}_n(\theta)'[\hat{\Omega}]^{-1}\hat{g}_n(\theta) \tag{162}$$

GMM 的方差

1) $L=k$ L: # moment conditions k: # of parameters

$$F.O.C: \frac{\partial\hat{Q}(\theta)}{\partial\theta} = -2\nabla_{\theta'}\hat{g}_n(\hat{\theta})W\hat{g}_N(\theta) = 0 \tag{163}$$

在 $L=k$ 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。

$$\begin{aligned} Var(\sqrt{n}\hat{\theta}_{GMM}) &= (G'WG)^{-1}G'W\Omega_0WG(G'WG)^{-1} \quad \text{笔记上记的是} Var(\hat{\theta}_{GMM}) \\ &= G^{-1}W^{-1}G'^{-1}G'W\Omega_0WGG^{-1}W^{-1}G'^{-1} \\ &= G^{-1}\Omega_0G'^{-1} = (G'\Omega^{-1}G)^{-1} \end{aligned} \tag{164}$$

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

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

使用 GMM，样本矩条件为 $\frac{1}{n}X'u=0$

$$\hat{g}_n(\beta) = \frac{1}{n}X'(y-X\beta) = 0 \tag{165}$$

$$\begin{aligned}\hat{\beta}_{OLS} = \hat{\beta}_{GMM} &= (X'X)^{-1}X'y \\ &= \beta + (X'X)^{-1}X'u\end{aligned} \tag{166}$$

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

$$\begin{aligned}G &= E[\nabla_\beta g(x_i, \beta_0)] \\ &= E[\nabla_\beta \frac{1}{n}\sum_{i=1}^{n} g(x_i, \beta_0)] \\ &= E[\nabla_\beta \hat{g}_n(\beta_0)] \\ &= -\frac{1}{n}E(X'X)\end{aligned} \tag{167}$$

$$\begin{aligned}\Omega &= E(g(x_i, \beta_0)g(x_i, \beta_0)') \\ &= \frac{1}{n}E[\sum_{i=1}^{n} g(x_i, \beta_0)\sum_{j=1}^{n} g(x_j, \beta_0)'] \\ &(when\, i \neq j E(gg') = 0, independent) \\ &= nE[\frac{1}{n}\sum_{i=1}^{n} g(x_i, \beta_0)\frac{1}{n}\sum_{j=1}^{n} g(x_j, \beta_0)'] \\ &= nE[\hat{g}_n(\beta_0)\hat{g}_n(\beta_0)'] \\ &= \frac{1}{n}E[X'uu'X]\end{aligned} \tag{168}$$

$$\begin{aligned}Var(\sqrt{n}\hat{\beta}_{GMM}) &= (G'\Omega^{-1}G)^{-1} \\ &= [E(\frac{X'X}{n})]^{-1}E[\frac{X'uu'X}{n}][E(\frac{X'X}{n})]^{-1} \\ &\sim (\frac{X'X}{n})^{-1}\frac{X'E(uu)'X}{n}(\frac{X'X}{n})^{-1} \\ &= \sigma^2(\frac{X'X}{n})^{-1}\end{aligned} \tag{169}$$

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

$$\hat{Q} = u'WZ^{-1}W'u(\text{严格来说，这里好像漏了俩 } \frac{1}{n}) \tag{170}$$

$$= (y - X\beta)'WZ'W'(y - X\beta) \tag{171}$$

$$= \dots \tag{172}$$

$$F.O.C : -2X'WZ^{-1}W'y + 2X'WZ^{-1}W'X\hat{\beta} = 0 \tag{173}$$

$$\Rightarrow \hat{\beta} = (W'X)^{-1}W'y \quad (IV estimation) \tag{174}$$

$$\begin{aligned}Var(\sqrt{n}\hat{\beta}_{GMM}) &= (G'\Omega^{-1}G')^{-1} \\ &= ((-\frac{1}{n}EW'X)'(\frac{1}{n}EW'uu'W)^{-1}(-\frac{1}{n}EW'X))^{-1} \\ &= \sigma^2(\frac{X'P_W X}{n})^{-1}\end{aligned} \tag{175}$$

$$\hat{\beta} = \beta + (W'X)^{-1}W'u \tag{176}$$

$$\sqrt{n}(\hat{\beta}_{2SLS} - \beta) = (\frac{W'X}{n})^{-1}\frac{1}{\sqrt{n}}W'u \tag{177}$$

$$(\frac{W'X}{n})^{-1} \to (EW_iX_i')^{-1} \tag{178}$$

$$\frac{1}{\sqrt{n}}W'u \to N(0, \sigma^2 EW_iW_i') \tag{179}$$

$$\Rightarrow \sim N(0, \sigma^2(\frac{X'P_WX}{n})^{-1}) \tag{180}$$

$L > k$

$$\hat{Q}_n = u'W[\quad]W'u \tag{181}$$

$$\begin{aligned} \Omega_0 &= Var(g(z_i, \theta_0)) \\ Var(\hat{g}_n(\theta)) &= Var(\frac{1}{n}\sum_{i=1}^{n} g(z_i, \theta)) \\ &= \frac{1}{n}Var(g(z_i, \theta)) \end{aligned} \tag{182}$$

$$\begin{aligned} \hat{Q}_n &= u'W[Var(W'u)]^{-1}Wu' \\ &= \frac{1}{\sigma^2}u'W(W'W)^{-1}W'u \\ &= \frac{1}{\sigma^2}u'P_Wu \\ &= \frac{1}{\sigma^2}(y - X\beta)'P_W(y - X\beta) \end{aligned} \tag{183}$$

$$F.O.C: -2X'P_Wy + 2X'P_WX\hat{\beta} = 0 \tag{184}$$

$$\hat{\beta}_{GMM} = (X'P_WX)^{-1}X'P_Wy = \hat{\beta}_{2SLS} \tag{185}$$

$$Var(\sqrt{n}\hat{\beta}_{GMM}) = \sigma^2(\frac{X'P_WX}{n})^{-1} \tag{186}$$

三个检验 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。

$$H_0 : r(\theta_0) = 0\text{非线性约束，线性约束为:}R\beta = 0 \tag{187}$$

$$H_1 : r(\theta_0) \neq 0 \tag{188}$$

其中 $r : \mathcal{R}^k \to \mathcal{R}^q$

(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 $W = 0$(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 $\hat{\theta}_R$, 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

$$\hat{\theta}_{GMM} \sim N \tag{189}$$

$$r(\hat{\theta}_{GMM}) \sim N \quad (DeltaMethod) \tag{190}$$

$$\Rightarrow \chi^2() \tag{191}$$

$$\sqrt{n}(\hat{\theta}_{unr}-\theta_0)\sim N(0,V_0) \tag{192}$$
$$\sqrt{n}(r(\hat{\theta}_{unr})-r(\theta_0))\sim N(0,R_0V_0R_0') \tag{193}$$

其中 V_0 是用 GMM 计算的方差, $R_0=R(\theta_0)$ $R(\theta)=\frac{\partial r(\theta)}{\partial\theta'}_{q\times k}$

$$Wald=nr(\hat{\theta}_{unr})[\hat{R}\hat{V}\hat{R}']^{-1}r(\hat{\theta}_{unr})\sim\chi^2(q) \tag{194}$$

LM test

$$\frac{\partial\hat{Q}_n(\hat{\theta}_r)}{\partial\theta}=2\nabla_{\theta'}\hat{g}_n(\theta)\hat{\Omega}^{-1}\hat{g}_n(\theta)\Big|_{\theta=\theta_0}\approx 0 \tag{195}$$
$$LM=(\underset{1\times k}{\frac{\partial\hat{Q}}{\partial\theta}})'[Var(\underset{k\times k}{\frac{\partial\hat{Q}}{\partial\theta}})]^{-1}(\frac{\partial\hat{Q}}{\partial\theta}) \tag{196}$$
$$\frac{\partial\hat{Q}_n(\hat{\theta}_{res})}{\partial\theta}\to 2G(\hat{\theta}_{res})'\hat{\Omega}^{-1}\hat{g}_n(\hat{\theta}_{res}) \tag{197}$$
$$\sqrt{n}\hat{g}_n(\hat{\theta}_{res})=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}g(z_i,\hat{\theta}_{res})\sim N(0,\hat{\Omega}) \tag{198}$$

$$\begin{aligned}Var(\frac{\partial\hat{Q}_n}{\partial\theta}(\hat{\theta}_{res}))&=\frac{4}{n}G'(\hat{\theta}_{res})\hat{\Omega}^{-1}Var(\sqrt{n}\hat{g}_n(\hat{\theta}_{res}))\hat{\Omega}^{-1}G(\hat{\theta}_{res})\\&=\frac{4}{n}\underset{k\times L}{G'(\hat{\theta}_{res})}\quad\underset{L\times L}{\hat{\Omega}^{-1}}\quad\underset{L\times k}{G(\hat{\theta}_{res})}\end{aligned} \tag{199}$$

$$LM=n[\hat{g}_n(\hat{\theta}_{res})'\underset{1\times k}{\hat{\Omega}^{-1}}G(\hat{\theta}_{res})][G'(\hat{\theta})_{res}\underset{k\times k}{\hat{\Omega}^{-1}}G(\hat{\theta}_{res})]^{-1}[G'(\hat{\theta}_{res})\underset{k\times 1}{\hat{\Omega}^{-1}}\hat{g}_n(\hat{\theta}_{res})]=\underset{1\times 1}{(\quad)}\sim\chi^2(q)(q<k) \tag{200}$$

LR

$$\hat{Q}_n(\theta)=\hat{g}_n(\theta)'\Omega^{-1}\hat{g}_n(\theta) \tag{201}$$
$$\Omega=Egg'=Var(\sqrt{n}\hat{g}_n)=nVar(\hat{g}_n)\Rightarrow\Omega^{-1}=\frac{1}{n}Var^{-1}\to 0 \tag{202}$$
$$\therefore objective\,function\,is\quad n\hat{Q}(\theta) \tag{203}$$

无约束的 GMM 目标函数

$$n\hat{Q}(\theta)=(\sqrt{n}\hat{g}_n(\hat{\theta}))'[Var\sqrt{n}\hat{g}_n(\hat{\theta})]^{-1}(\sqrt{n}\hat{g}_n(\hat{\theta}))\sim\chi^2(l-k) \tag{204}$$

有约束的 GMM 目标函数

$$n\hat{Q}_n(\theta_R)=(\sqrt{n}\hat{g}_n(\hat{\theta}_R))'[Var\sqrt{n}\hat{g}_n(\hat{\theta}_R)]^{-1}(\sqrt{n}\hat{g}_n(\hat{\theta}_R))\sim\chi^2(l-(k-q)) \tag{205}$$

$$LR=n\hat{Q}_n(\hat{\theta}_R)-n\hat{Q}_n(\hat{\theta})\sim\chi^2(q) \tag{206}$$

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

$$\begin{aligned}&H_0\ E(g_l(z,\theta_0))=0\quad l=k+1,k+2,\ldots,L\\&H_1:E(g_l(z,\theta_0))\neq 0\quad for\,some l=k+1,k+2,\ldots,L\end{aligned} \tag{207}$$

Test of Moment Conditions 要求 $L>k$

增广矩条件 augumented

$$\begin{aligned} g^a(z,\theta,\phi) = [&\underbrace{g_1(z,\theta),\ldots,g_l(z,\theta)}_{reliable \quad E()=0}, \\ &g_{k+1(z,\theta)}-\phi_1,\ldots,g_L(z,\theta)-\phi_{L-k}] \end{aligned} \tag{208}$$

将对矩条件的检验转换为对参数的检验

$$H_0 : \psi_j = 0,\, j = 1, 2, \ldots, L-k \tag{209}$$

$$\begin{aligned} LR =& n[\hat{Q}^a(\hat{\theta}_{res}, \overset{\phi_j=0}{0}) - \hat{Q}^a(\hat{\theta}_{unr}, \hat{\phi}_{unr})] \\ &= n[\hat{Q}_n(\hat{\theta}_{unr}) - 0] \\ &= n\hat{Q}_n(\hat{\theta}) \sim \chi^2(l-k) \end{aligned} \tag{210}$$

2 M-Estimation

2.1 Estimation

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

$$\hat{\theta} max/min \hat{Q}(w_i,\theta)\, s.t. \theta \in Ⓗ \tag{211}$$

1.Linear Regression

$$y_i = x_i'\beta + u_i \tag{212}$$

$$min\, SSR \tag{213}$$

2.MLE

$$y_i \sim N(x_i'\beta, \sigma^2(\theta, z_i)) \tag{214}$$

$$lnL(\beta,\theta|x,y,z) = \sum_{i=1}^{n} lnf(y_i|x_i,z_i,\beta,\theta) \tag{215}$$

$$max \sum_{i=1}^{n} lnf(y_i|x_i,z_i,\beta,\theta) \tag{216}$$

3.nonlinear regression model

$$\begin{matrix} y_i = m(x_i,\theta) + u_i & y_i = m(x) + u_i \\ \text{参数模型}m()\text{已知} & \text{非参数模型}m()\text{未知} \end{matrix} \tag{217}$$

e.g.

$$m(X,\theta) = exp(X\theta) \tag{218}$$

$$m(X,\theta) = \frac{exp(X\theta)}{1+exp(X\theta)} (\text{logistic function}) \tag{219}$$

$$y = m(X) + u \quad m(X,\theta) = X'\theta \tag{220}$$

$$\Leftrightarrow y_i = x_i'(\theta) + u_i \tag{221}$$

$$E(u_i|x_i) = 0 \Leftrightarrow E(y|X) = x_i'\theta + \cancel{E(u_i|X)} \tag{222}$$

NLS assumption 1:

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

$$\min_{\theta \in Ⓗ} E(y - m(x,\theta))^2 \tag{223}$$

$$\begin{aligned}
E(y - m(x,\theta))^2 &= E[y - E(y|X) + E(y|X) - m(X,\theta]^2 \\
&= E[(y - m(X,\theta_0)) + (m(X,\theta_0) - m(X,\theta))]^2 \\
&= E[y - m(X,\theta_0)]^2 + 2E[(y - m(X,\theta_0))(m(X,\theta_0) - m(X,\theta))] + E[m(X,\theta_0) - m(X,\theta)]^2 \\
&= E[y - m(X,\theta_0)]^2 + 2E\Big[E[(y - m(X,\theta_0))(m(X,\theta_0) - m(X,\theta))|X]\Big] + E[m(X,\theta_0) - m(X,\theta)]^2 \\
&= E[y - m(X,\theta_0)]^2 + 2E\Big[E[y - m(X,\theta_0)|X](m(X,\theta_0) - m(X,\theta))\Big] + E[m(X,\theta_0) - m(X,\theta)]^2 \\
&= E[y - m(X,\theta_0)]^2 + 2E\Big[(\underbrace{E(y|X)}_{=m(X,\theta_0)} - m(X,\theta_0))(m(X,\theta_0) - m(X,\theta))\Big] + E[m(X,\theta_0) - m(X,\theta)]^2 \\
&= E[y - m(X,\theta_0)]^2 + E[m(X,\theta_0) - m(X,\theta)]^2
\end{aligned} \tag{224}$$

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

$$E[m(X,\theta_0) - m(X,\theta)]^2 \geq 0 \tag{225}$$

$$\text{if } E[m(X,\theta_0) - m(X,\theta)]^2 = 0 \tag{226}$$

$$\text{then } E(y - m(X,\theta))^2 = E(y - m(X,\theta_0))^2 \tag{227}$$

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

$$\text{if } E[m(X,\theta_0) - m(X,\theta)]^2 > 0 \tag{228}$$

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

NLS assumption 2:

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

$$\begin{aligned}
&\text{assume } m(X,\theta) = X\theta \\
&E(m(X,\theta_0) - m(X,\theta))^2 \\
=&E[(X\theta_0 - X\theta)'(X\theta_0 - X\theta)] \\
=&E[(\theta_0 - \theta)'X'X(\theta_0 - \theta)] > 0
\end{aligned} \tag{229}$$

因此要求 $X'X$positive definite, 即 X 矩阵列满秩，$rank(E(X'X)) = k$(上课后来改成了 $rank(E(xx')) = k$, 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

$$m(x,\theta_0) = \theta_{10} + \theta_{20}x_2 \tag{230}$$

待估计的模型为

$$m(X,\theta) = \theta_1 + \theta_2 x_2 + \theta_3 x_3^{\theta_4} \tag{231}$$

$$\min_{\theta_1,\theta_2,\theta_3,\theta_4} E[y - m(X,\theta)]^2 \tag{232}$$

$$\theta_1 = \theta_{10}, \theta_2 = \theta_{20}, \theta_3 = 0, \theta_4 = \text{any value} \tag{233}$$

所以 NLS assumption 2 违背。

The General M-estimator can be expressed assume

$$\underset{\theta\in\textcircled{H}}{min}\, E[q(w,\theta)] \text{ e.g. } q(w,\theta) = [y - m(X,\theta)]^2 \tag{234}$$

The identification requires

$$E[q(w,\theta_0)] < E[q(w,\theta)]\, \forall\theta \in \textcircled{H}, \theta \neq \theta_0 \tag{235}$$

用算术平均值代替期望

$$\hat{\theta} = \underset{\theta\in\textcircled{H}}{min} \frac{1}{n}\sum_{i=1}^{n} q(w_i,\theta) \tag{236}$$

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 2.1 ***Uniform Weak Law of Large Numbers*** *If*

1. *Data w_i is i.i.d*
2. *$\theta \in \textcircled{H}$, $\textcircled{H}$ is a compact set*
3. *for each w_i, $q(w)$ is continuous on $\textcircled{H}$*
4. *$|q(w_i,\theta)| \leq b(w_i)\, \forall\theta \in \textcircled{H}\, E(b(w_i)) < \infty$*

Then

$$\frac{1}{n}\sum_{i=1}^{n} q(w_i,\theta) \xrightarrow{p} E[q(w,\theta)] \tag{237}$$

Theorem 2.2 ***Consistency of M-estimator***

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

$$\hat{\theta} \xrightarrow{p} \theta_0 \tag{238}$$

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

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

Lemma 2.1 *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*

$$\frac{1}{n}\sum_{i=1}^{n} r(w_i,\hat{\theta}) \xrightarrow{p} E(r(w,\theta_0)) \tag{239}$$

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，$\hat{\theta} \sim ?$

$$\min_{\theta\in\textcircled{H}} \frac{1}{n}\sum_{i=1}^{n} q(w_i,\theta) \tag{240}$$

$$\text{F.O.C:} \sum_{i=1}^{n}\frac{\partial q(w_i,\hat{\theta})}{\partial\theta} = \sum_{i=1}^{n}\frac{\partial q(w_i,\theta_0)}{\partial\theta} + \sum_{i=1}^{n}\frac{\partial^2 q(w_i,\bar{\theta})}{\partial\theta\partial\theta'}(\hat{\theta}-\theta_0) = 0 \tag{241}$$

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2 q(w_i,\bar{\theta})}{\partial\theta\partial\theta'}\sqrt{n}(\hat{\theta}-\theta_0) = -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{i=1}^{n}\frac{\partial q(w_i,\theta_0)}{\partial\theta} \tag{242}$$

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

$$\sqrt{n}(\hat{\theta}-\theta_0) = (\frac{1}{n}\sum_{i=1}^{n} H_i)^{-1}(-\frac{1}{\sqrt{n}}\sum_{i=1}^{n} S(w_i,\theta_0)) \text{ as } n \to \infty \tag{243}$$

根据 Lemma2.1 有

$$\frac{1}{n}\sum_{i}^{n} H_i = \frac{1}{n}\sum_{i=1}^{n} H(w,\bar{\theta}) \xrightarrow{p} E[H(w,\theta_0)] \overset{def}{=} A_0 \tag{244}$$

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[-S(w_i,\theta_0)] \sim N(0,\underbrace{ES_iS_i'}_{=B_0}) \tag{245}$$

$$\sqrt{n}(\hat{\theta}-\theta_0) \sim N(0,A_0^{-1}B_0A_0^{-1}) \tag{246}$$

$$\sqrt{n}(\hat{\theta}-\theta_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}[-A_0^{-1}S_i(\theta_0)] + o_p(1) \tag{247}$$

$$\overset{def}{=} \frac{1}{\sqrt{n}}\sum_{i=1}^{n} e_i(\theta_0) + o_p(1) \tag{248}$$

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

$$q(w,\theta_0)=\frac{1}{2}(y-m(x_i,\theta_0))^2 \tag{249}$$

$$S_i=-\frac{\partial m_i}{\partial\theta}(y_i-m_i)=-\nabla_{\theta_0}m_i'(y_i-m_i) \tag{250}$$

$$ES_i=E[E(S_i|X_i)]\,,E(S_i|X_i)=0 \tag{251}$$

example 1

$$m(X,\theta)=X\theta,\ A_0=E(X'X)---\textbf{full rank} \tag{252}$$

$$\tag{253}$$

example 2

$$m(X,\theta)=\theta_1+\theta_2X_2+\theta_3X_3^{\theta_4},\ \theta_3=0,\ \theta_4 \text{ can be any value} \tag{254}$$

$$H(w,\theta)=\nabla_\theta m(X,\theta)'\nabla_\theta m(X,\theta)-\nabla_\theta^2 m(X,\theta)(y-m(X,\theta)) \tag{255}$$

$$A_0=E[H(w,\theta_0)]=E[\nabla_\theta m(X,\theta)'\nabla_\theta m(X,\theta)] \tag{256}$$

$$E\left[\begin{pmatrix}1\\ x_2\\ x_3^{\theta_4}\\ \underbrace{\theta_3x_3^{\theta_4}ln(x_3)}_{=0}\end{pmatrix}\begin{pmatrix}1\ x_2\ x_3^{\theta_4}\ \underbrace{\theta_3x_3^{\theta_4}ln(x_3)}_{=0}\end{pmatrix}\right]=E\left[\begin{pmatrix} & & & 0\\ & & & 0\\ & & & 0\\ 0 & 0 & 0 & 0\end{pmatrix}\right] \tag{257}$$

行列式 =0 不可逆。

Two step M-estimation

$$\min\sum_{i=1}^{n} q(w_i,\theta,\hat{\gamma}) \tag{258}$$

Example: Weighted Nonlinear Least Square(WNLS)

$$y_i = m(x_i,\theta) + u_i,\ E(u_i^2|x_i) = h(x_i,\gamma_0) \tag{259}$$

$$\frac{y_i}{\sqrt{h(x_i,\gamma_0)}} = \frac{m(x_i,\theta)}{\sqrt{h(x_i,\gamma_0)}} + \frac{u_i}{\sqrt{h(x_i,\gamma_0)}} \tag{260}$$

$$\min_{\theta\in Ⓗ} \frac{1}{2}\sum_{i=1}^{n}(y_i - m(x_i,\theta))^2/h(x_i,\gamma_0) \tag{261}$$

$$\hat{u_i}^2 = h(x_i,\gamma) + error_i \overset{M-Estimation}{\Rightarrow} \hat{\gamma} \tag{262}$$

$$\min_{\theta\in Ⓗ} \frac{1}{2}\sum_{i=1}^{n}(y_i - m(x_i,\theta))^2/h(x_i,\hat{\gamma}) \tag{263}$$

$$\tag{264}$$

WNLS Assumption 1

$$E(u|X) = 0 \tag{265}$$

$$E(\frac{y_i}{\sqrt{h(x_i,\gamma_0)}}) = E(\frac{m(x_i,\theta)}{\sqrt{h(x_i,\gamma_0)}}) + E(\frac{u_i}{\sqrt{h(x_i,\gamma_0)}}) \tag{266}$$

which is the same as NLS Assumption 1.

WNLS Assumption 2

$$E[(m(X,\theta_0) - m(X,\theta))^2/h(X,\gamma^*)] > 0, \textbf{for all } \theta \neq \theta_0,\ \theta\in Ⓗ \tag{267}$$

更一般的有

$$E[q(W,\theta_0,\gamma^*)] < E[q(W,\theta,\gamma^*)], \textbf{for all } \theta \neq \theta_0,\ \theta\in Ⓗ \tag{268}$$

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

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

$$\begin{cases} \frac{1}{n}\sum_{i=1}^{n} q(w_i,\theta,\gamma^*) \to Eq(w_i,\theta,\gamma^*) \quad \text{UWLLN} \\ \text{identification condition} \end{cases} \Rightarrow \text{consistency: } \hat{\theta} \xrightarrow{p} \theta_0 \tag{269}$$

Lemma 2.2 *Like Lemma 2.1, we have*

$$\frac{1}{n}\sum_{i=1}^{n} q(w_i,\hat{\theta},\hat{\gamma}) \to Eq(w_i,\theta,\gamma^*) \tag{270}$$

$$\sqrt{n}(\hat{\theta}-\theta_0) = A_0^{-1}(-\frac{1}{\sqrt{n}}\sum_{i=1}^{n} S_i(\theta_0,\hat{\gamma})) + o_p(1) \tag{271}$$

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

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} S_i(\theta_0,\hat{\gamma}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} S_i(\theta_0,\gamma^*) + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial S_i(\theta_0,\gamma^*)}{\partial\gamma^*}(\hat{\gamma}-\gamma^*) + o_p(1) \tag{272}$$

$$M - estimation \Rightarrow \sqrt{n}(\hat{\gamma}-\gamma^*) \to N(,) = O_p(1) \tag{273}$$

$$\therefore \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial S_i(\theta_0,\gamma^*)}{\partial\gamma^*}(\hat{\gamma}-\gamma^*) = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial S_i(\theta_0,\gamma^*)}{\partial\gamma^*}\sqrt{n}(\hat{\gamma}-\gamma^*) \to E\frac{\partial S_i}{\partial r}O_p(1) \tag{274}$$

$$E[\nabla_\gamma S(W,\theta_0,\gamma^*)] \overset{def}{=} F_0 \tag{275}$$

if $F_0=0$ 不影响结果 $\sqrt{n}(\hat{\theta}-\theta_0)\sim N(0,A_0^{-1}B_0A_0^{-1})$, $A_0=E[H(W,\theta_0,\gamma^*)]$, $B_0=E[S(W,\theta_0,\gamma^*)S(W,\theta_0,\gamma^*)']$。

对于 $y=m(X,\theta_0)+u, F_0=0$

对于 probit,tobit,$F_0\neq 0$ Influence function representation: $\sqrt{n}(\hat{\gamma}-\gamma^*)=\frac{1}{\sqrt{n}}\sum_{i=1}^n r_i(\gamma^*)+o_p(1)$

if $F_0\neq 0$

$$\sqrt{n}(\hat{\theta}-\theta_0)=-A_0^{-1}[\frac{1}{\sqrt{n}}\sum_{i=1}^n S_i(\theta_0,\gamma^*)+F_0\sqrt{n}(\hat{\gamma}-\gamma^*)]+o_p(1) \tag{276}$$

$$=-A_0^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n[S_i(\theta_0,\gamma^*)+F_0r(\gamma^*)]+o_p(1) \tag{277}$$

$E[-g_i(\theta_0,\gamma^*)=0], Var(-g_i(\theta_0,\gamma^*)=E[g_ig_i']\overset{def}{=}D_0$

$$\sqrt{n}(\hat{\theta}-\theta_0)\sim N(0,A_0^{-1}D_0A_0^{-1}) \tag{278}$$

$$D_0=E[S_i(\theta_0,\gamma^*)S_i(\theta_0,\gamma^*)']+F_0E[r_i(\gamma*)r_i(\gamma^*)']F_0'=B_0+PositiveDefinte>B_0 \tag{279}$$

$$A_0=E[H(W,\theta_0)],\ B_0=E[S(W,\theta_0,\gamma^*)S(W,\theta_0,\gamma^*)'] \tag{280}$$

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

$$A=\frac{1}{n}\sum_{i=1}^n H(w_i,\hat{\theta})\to A_0 \tag{281}$$

$$B=\frac{1}{n}\sum_{i=1}^n[S(w_i,\hat{\theta})S(w_i,\hat{\theta})']\to B_0 \tag{282}$$

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

$$A(X,\theta_0)=E[H(W,\theta_0)|X] \tag{283}$$

$$\text{For nonlinear model: } y=m(X,\theta_0)+u \tag{284}$$

$$H(W,\theta_0)=\underbrace{\nabla_\theta m(X,\theta_0)'}_{k\times 1}\underbrace{\nabla_\theta m(X,\theta_0)}_{1\times k}-\nabla_\theta^2 m(X,\theta_0)\underbrace{y-m(X,\theta_0)}_{assumption2\Rightarrow E()=0} \tag{285}$$

$$A_0=E[A(X,\theta_0)] \tag{286}$$

$$A(X,\theta_0)=\nabla_\theta m(X,\theta_0)'\nabla_\theta m(X,\theta_0) \tag{287}$$

$$A_0=E[A(X,\theta_0)]=E[\nabla_\theta m(X,\theta_0)'\nabla_\theta m(X,\theta_0)] \tag{288}$$

$$\hat{A}=\frac{1}{n}\sum_{i=1}^n \underbrace{A(X_i,\hat{\theta})}_{\hat{A}_i}\to A_0 \tag{289}$$

$$\hat{B}_0=\frac{1}{n}\sum_{i=1}^n S(w_i,\hat{\theta})S(w_i,\hat{\theta})'\to B_0 \tag{290}$$

$$\sqrt{n}(\hat{\theta}-\theta_0)\to N(0,\hat{A}_0^{-1}\hat{B}_0\hat{A}_0^{-1}) \tag{291}$$

$$V\hat{a}r(\hat{\theta})=\begin{cases}(\sum_{i=1}^n\hat{H}_i)^{-1}(\sum_{i=1}^n\hat{S}_i\hat{S}_i')(\hat{H}_i)^{-1} & \text{Fully Robust Estimator}\\ (\sum_{i=1}^n\hat{A}_i)^{-1}(\sum_{i=1}^n\hat{S}_i\hat{S}_i')(\hat{A}_i)^{-1} & \text{Semi Robust Estimator}\end{cases} \tag{292}$$

$$\hat{A}(X_i,\hat{\theta})=\underbrace{\nabla_\theta\hat{m}_i'}_{k\times 1}\underbrace{\nabla_\theta\hat{m}_i}_{1\times k} \tag{293}$$

$$\hat{S}_i=-\nabla_\theta\hat{m}_i(y_i-\hat{m}_i)=-\nabla_\theta\hat{m}_i\hat{u}_i \tag{294}$$

$$V\hat{a}r(\hat{\theta})=(\sum_{i=1}^n\nabla_\theta\hat{m}_i'\hat{m}_i)^{-1}(\sum_{i=1}^n\hat{u}_i\nabla_\theta\hat{m}_i'\hat{m}_i)(\sum_{i=1}^n\nabla_\theta\hat{m}_i'\hat{m}_i)^{-1} \tag{295}$$

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

$$Var(y|X) = Var(u|X) = \sigma_0^2 \tag{296}$$

$$B_0 = \sigma_0^2 E[\nabla_\theta m(X,\theta_0)'\nabla_\theta m(X,\theta_0)] = \sigma_0^2 A_0,\ Var(\sqrt{n}\hat{\theta}) = A_0^{-1}B_0A_0^{-1} = \sigma_0^2 A_0^{-1} \tag{297}$$

$$V\hat{ar}(\hat{\theta}) = \hat{\sigma}^2(\sum_{i=1}^n \hat{H}_i)^{-1} \text{ or } \hat{\sigma}^2(\sum_{i=1}^n \hat{A}_i)^{-1} \tag{298}$$

$$\hat{\sigma}^2 = \frac{1}{n-k}\sum_{i=1}^n \hat{u}_i^2 \tag{299}$$

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

$$V\hat{ar}(\hat{\theta}) = \hat{\sigma}^2(\sum_{i=1}^n exp(2x_i\hat{\theta})x_i'x_i)^{-1} \tag{300}$$

Variance estimation for two step M-estimation

If $E[\nabla_\gamma S(W,\theta_0,\gamma^*)] = 0$,($F_0 = 0$)

$$V\hat{ar}(\hat{\theta}) = \begin{cases} (\sum_{i=1}^n \hat{H}_i)^{-1}(\sum_{i=1}^n \hat{S}_i\hat{S}_i')(\sum_{i=1}^n \hat{H}_i)^{-1} \\ (\sum_{i=1}^n \hat{A}_i)^{-1}(\sum_{i=1}^n \hat{S}_i\hat{S}_i')(\sum_{i=1}^n \hat{A}_i)^{-1} \end{cases} \tag{301}$$

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

Under NLS assumption 1-2

$$Eq(W,\theta,\gamma^*) = E\frac{1}{2}(y - m(X,\theta))^2/h(X,\gamma^*) \tag{302}$$

$$S(W,\theta_0,\gamma^*) = \frac{\partial q(W,\theta_0,\gamma^*)}{\partial\theta} = -\nabla_\theta m(X,\theta_0)'(y - m(X,\theta_0))/h(X,\gamma^*) = -\nabla_\theta m'u/h \tag{303}$$

$$H(W,\theta_0,\gamma^*) = \frac{\partial^2 q}{\partial\theta\partial\theta'} = \nabla_\theta m'\nabla_\theta^2 m(y-m)/h \tag{304}$$

$$E[\nabla_\gamma S(W,\theta_0,\gamma^*)] = 0 \tag{305}$$

$$V\hat{ar}(\hat{\theta}) = (\sum_{i=1}^n \nabla_\theta m_i'\nabla_\theta m_i/h_i)^{-1}(\sum_{i=1}^n \nabla_\theta \hat{m}_i'\hat{u}_i^2\nabla_\theta\hat{m}_i/\hat{h}_i)(\sum_{i=1}^n \nabla_\theta m_i'\nabla_\theta m_i/h_i)^{-1} \tag{306}$$

WNLS assumption 3 : $Var(y|X) = \sigma_0^2 h(X,\gamma_0)$

$$B_0 = \sigma_0^2 E[\nabla_\theta m'\nabla_\theta m/h] = \sigma_0^2 A_0 \tag{307}$$

$$A_0 = E[\nabla_\theta m'\nabla_\theta m/h] \tag{308}$$

$$V\hat{ar}(\hat{\theta}) = \hat{\sigma}^2(\sum_{i=1}^n \nabla_\theta\hat{m}_i'\nabla_\theta\hat{m}_i/h_i)^{-1} \tag{309}$$

$$\hat{\sigma}^2 = \frac{1}{n-k}\sum_{i=1}^n (\frac{\hat{u}_i^2}{\sqrt{\hat{h}_i}})^2 \tag{310}$$

if $F_0 \neq 0$, $E[\nabla_\gamma S(W, \theta_0, \gamma^*)] \neq 0$

$$\hat{Var}(\hat{\theta}) = \begin{cases} (\sum_{i=1}^{n} \hat{H}_i)^{-1}(\sum_{i=1}^{n} \hat{g}_i\hat{g}_i')(\sum_{i=1}^{n} \hat{H}_i)^{-1} \\ (\sum_{i=1}^{n} \hat{A}_i)^{-1}(\sum_{i=1}^{n} \hat{g}_i\hat{g}_i')(\sum_{i=1}^{n} \hat{A}_i)^{-1} \end{cases} \tag{311}$$

$$\hat{g}_i = \hat{S}_i + \hat{F}_i + \hat{r}_i \tag{312}$$

$$\hat{F}_i = \frac{1}{n}\sum_{i=1}^{n} \nabla_\gamma S_i(\hat{\theta}, \hat{\gamma}) \tag{313}$$

2.2 Numerical Optimization

2.2.1 Newton-Raphson Method

$$\sum_{i=1}^{n} S(W_i, \hat{\theta}) = 0 \Rightarrow \hat{\theta} \tag{314}$$

$$\sum_{i=1}^{n} S_i(\theta^{\{g+1\}}) = \sum_{i=1}^{n} S_i(\theta^{\{g\}}) + [\sum_{i=1}^{n} H_i(\theta^{\{g\}})](\theta^{\{g+1\}} - \theta^{\{g\}}) + r^{\{g\}} \tag{315}$$

$$\text{Let } \sum_{i=1}^{n} S_i(\theta^{\{g+1\}}) = 0, r^{\{g\}} = 0 \tag{316}$$

$$\theta^{\{g+1\}} = \theta^{\{g\}} - [\sum_{i=1}^{n} H_i(\theta^{\{g\}})]^{-1}[\sum_{i=1}^{n} S_i(\theta^{\{g\}})] \quad \text{iterative method} \tag{317}$$

$\theta^{\{g+1\}}$ 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

$$[\sum_{i=1}^{n} S_i(\theta^{\{g\}})]'[\sum_{i=1}^{n} H_i(\theta^{\{g\}})][\sum_{i=1}^{n} S_i(\theta^{\{g\}})] \tag{318}$$

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

2.2.2 Berndt Hall Hall and Hausman Method

$$\theta^{\{g+1\}} = \theta^{\{g\}} - r[\sum_{i=1}^{n} S_i(\theta^{\{g\}})S_i(\theta^{\{g\}})']^{-1}[\sum_{i=1}^{n} S_i(\theta^{\{g\}})] \tag{319}$$

where r is step size

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

$$\sum_{i=1}^{n} S_i(\theta^{\{g\}})S_i(\theta^{\{g\}})' = \sigma_0^2 \sum_{i=1}^{n} H_i(\theta^{\{g\}}) \tag{320}$$

stopping rule

$$[\sum_{i=1}^{n} S_i(\theta^{\{g\}})]'[\sum_{i=1}^{n} S_i(\theta^{\{g\}})S_i(\theta^{\{g\}})']^{-1}[\sum_{i=1}^{n} S_i(\theta^{\{g\}})] \sim \chi^2(k) \tag{321}$$

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。

2.2.3 Gauss-Newton Method

$$\theta^{\{g+1\}} = \theta^{\{g\}} - r[\sum_{i=1} nA_i(\theta^{\{g\}})]^{-1}[\sum_{i=1}^{n} S_i(\theta^{\{g\}})] \tag{322}$$

$$A(X_i, \theta_0) = E[H(W_i, \theta_0)|X_i] \tag{323}$$

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

$$\theta^{\{g+1\}} = \theta^{\{g\}} - r[\sum_{i=1}^{n} \nabla_\theta m(X_i, \theta^{\{g\}})' \sum_{i=1}^{n} \nabla_\theta m(X_i, \theta^{\{g\}})]^{-1}[\sum_{i=1}^{n} \nabla_\theta m(X_i, \theta^{\{g\}})' u_i^{\{g\}}] \tag{324}$$

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 $y = m(X_i, \theta_0) + u_i$，Taylor 展开有

$$m(X, \theta^{\{2\}}) \approx m(X, \theta^{\{1\}}) + \nabla_\theta m(X, \theta^{\{1\}})(\theta^{\{2\}} - \theta^{\{1\}}) \tag{325}$$

$$y - m(X, \theta^{\{1\}}) \approx \nabla_\theta m(X, \theta^{\{1\}})(\theta^{\{2\}} - \theta^{\{1\}}) + y - m(X, \theta^{\{2\}}) \quad \text{regression} \tag{326}$$

$$H_0 : b = \theta^{\{2\}} - \theta^{\{1\}} = 0 \tag{327}$$

$$\text{if} b \neq 0, \theta^{\{2\}} = b + \theta^{\{1\}} \tag{328}$$

$$\vdots$$

$$\text{until } \theta^{\{i+1\}} - \theta^{\{i\}} = 0 \tag{329}$$

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

$$\nabla_\beta m(x_i, \beta) = (x_1, x_2^{\beta_3}, \beta_2 x_2^{\beta_2} \, ln \, x_2) \tag{330}$$

$$\text{initial value:}(\beta_1, \beta_2, \beta_3) = (1, 1, 1) \tag{331}$$

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

$$b = \begin{pmatrix} \beta_1^{\{2\}} - \beta_1^{\{1\}} \\ \beta_2^{\{2\}} - \beta_2^{\{1\}} \\ \beta_3^{\{2\}} - \beta_3^{\{1\}} \end{pmatrix} \stackrel{?}{=} 0 \tag{332}$$

如何检验 $H_0 : \beta_3 = 1$

1. 用 M-estimation 估计 β_3

2. 用 t, F, Wald, LM, LR 等方法检验。

M-estimation

$$\hat{\beta} \sim N(\beta_0, \sigma_0^2 A_0^{-1}/n) \tag{333}$$

$$V\hat{a}r(\hat{\beta}) = \hat{\sigma}^2(\sum_{i=1}^{n} \nabla_\beta m(x_i, \hat{\beta})' \nabla_\beta m(x_i, \hat{\beta}))^{-1} \tag{334}$$

$$\hat{\sigma}^2 = \frac{1}{n-k} \sum_{i=1}^{n} \hat{u}_i^2 \tag{335}$$

$$t = \frac{\hat{\beta}_{m-estimation} - \beta_0}{SE(\hat{\beta})} \text{ e.g. } \frac{\hat{\beta}_3 - 1}{SE(\hat{\beta}_3)} \tag{336}$$

另一种检验的方法

$H_0 : \beta_3 = 1$ impose this restriction reg $y = \tilde{\beta}_1 x_1 + \tilde{\beta}_2 x_2 + u$

$$\tilde{\beta} = (\tilde{\beta}_1, \tilde{\beta}_2, 1) \tag{337}$$

$$\tilde{u} = y - \tilde{\beta}_1 x_1 = \tilde{\beta}_2 x_2 \tag{338}$$

$$\nabla_\beta m(x,\beta) = (x_1, x_2, \tilde{\beta}_2 x_2 \, ln\, x_2) \tag{339}$$

$$ \tag{340}$$

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

$\alpha_3 = 0 \overset{?}{\Leftrightarrow} H_0 : \beta_3 = 1$

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 $\beta_3^{\{i+1\}} = \beta_3^{\{i\}}$

3 Quantile Regression

Let y_i denote a random draw from a population

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

$Quantile_\tau(y_i)$: τ-th quantile of y_i

$Quantile_\tau(y_i|x_i) = \beta_0(\tau) + x_i\beta_1$

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

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

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

$$|y_i - q| = 1(y_i \ge q)(y_i - q) + 1(y_i < q)(q - y_i) \tag{341}$$

$$\begin{aligned} F.O.C \quad & \sum_{i=1}^n (-1)\, 1(y_i \ge q) + 1(y_i < q) \\ &= \sum_{i=1}^n (-1)[1 - 1(y_i < q)] + \sum_{i=1}^n 1(y_i < q) \end{aligned} \tag{342}$$

$$= -n + 2\sum_{i=1}^n 1(y_i < q) = 0$$

$$\Rightarrow \frac{1}{n}\sum_{i=1}^n 1(y_i < q) = 2 \tag{343}$$

$$E\, 1(y_i < q) = \frac{1}{2} \tag{344}$$

$$P(y_i < q) = \frac{1}{2}, q = Median(y_i) \tag{345}$$

对于其他 quantile, 给绝对值 >0 和 <0 分配不同得权重

$$L(e) = Ee^2 \quad E(y|x) \tag{346}$$

$$L(e) = E|e| \quad Med(y|x) \tag{347}$$

$$L(e) = \begin{cases} E(1-\tau)|e| & e < 0 \\ E\tau|e| & e \ge 0 \end{cases} \quad Quant_\tau(y|x) \min L|e| = \min E\{[\tau\, 1(y_i - q \ge 0) + (1-\tau)\, 1(y_i - q < 0)]|y_i - q|\} \tag{348}$$

$$C_\tau(e) = [\tau\, 1(e) \ge 0) + (1-\tau)\, 1(e < 0)]|e| \tag{349}$$

$$\min_{\alpha,\beta} \sum_{i=1}^n C_\tau(y_i - \alpha - x_i\beta) \tag{350}$$

$$\text{in M-estimation } \hat{\theta}(\tau) = (\hat{\alpha}(\tau), \hat{\beta}(\tau)') \to \theta_0(\tau) = (\alpha_0(\tau), \beta_0(\tau)') \tag{351}$$

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

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

$E(u_i|x_i) = Med(u_i|x_i) = 0$

$E(y_i|x_i) = \alpha_0 + x_i\beta_0 = Med(y_i|x_i)$

assumption ②

$D(u_i|x_i) = D(u_i)$ and $Eu_i = 0$

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

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

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 $\ln y$,$E[\ln y_i|x_i] = \alpha_0 + x_i\beta_0 + E[u_i|x_i]$

under assumption ①

$$E[\ln y_i|x_i] = \alpha_0 + x_i\beta_0 \tag{352}$$

$$e^{\ln y_i} = e^(\alpha_0 + x_i\beta_0 + u_i) = y_i \tag{353}$$

$$E[y_i|x_i] = e^(\alpha_0 + x_i\beta_0)E(e^{u_i}|x_i) \tag{354}$$

$E(e^{u_i}|x_i)$ 起算起来很麻烦

LAD 得到

$$Med(\ln y_i|x_i) = \alpha_0 + x_i\beta_0 \tag{355}$$

LAD 无法规避计算均值 $E(e^{u_i}|x_i)$ 的问题

under assumption ②

$$E[\ln y_i|x_i] = \alpha_0 + x_i\beta_0 + \cancel{E[u_i|x_i]} \tag{356}$$

$$E(y_i|x_i) = e^{\alpha_0 + x_i\beta_0}E(e^{u_i}|\cancel{x_i}) \tag{357}$$

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

$Med(\ln y_i|x_i) = \alpha_0 + x_i\beta_0 + Med(u_i|x_i) = (\alpha_0 + \eta_0) + x_i\beta_0)$

$\alpha_0 + \eta_0$ 是一起估计出来的。

$$u_i = Med(u_i) + \tilde{u}_i = \eta_0 + \tilde{u}_i \tag{358}$$

$$u_i - Med(u_i) = \tilde{u_i} \quad Med(\tilde{u}_i) = 0 \tag{359}$$

$$e^{u_i} = e^{\eta_0}e^{\tilde{u}_i} \tag{360}$$

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

$$Ee^{u_i} = e^{\eta_0}Ee^{\tilde{u}_i} \tag{361}$$

$$e^{\alpha_0 + x_i\beta_0}E(e^{u_i}) = e^{\alpha_0 + x_i\beta_0}e^{\eta_0}Ee^{\tilde{u}_i} \tag{362}$$

虽然无法完全规避计算均值的影响，但 $E(y_i|x_i) = e^{\alpha_0 + x_i\beta_0}e^{\eta_0}Ee^{\tilde{u}_i}$ 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 $E(x_iy_i) = E[x_iE(y_i|x_i)]$ 但是 $Med(x_iy_i) = Med[x_iMed(y_i|x_i)]$, 其次 Med 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

$$E(y_i|x_i) = E(a_i|\cancel{x_i}) + x_iE(b_i|\cancel{x_i}) \tag{363}$$

$$= \alpha_0 + x_i\beta_0 \quad \text{OLS average partial effect} \tag{364}$$

$$\begin{aligned} Med(y_i|x_i) =& Med(a_i|x_i) + x_i Med(b_i|x_i) \\ =& Med(a_i) \quad + x_i \quad Med(b_i) \end{aligned}$$

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

$Leftside = 3.1 \neq Rightside = 4$

题外话结束

$$y_i = x_i\theta_0 + u_i, \quad Quant_\tau(u_i|x_i) = 0 \tag{365}$$

$$q(w_i,\theta) = \tau\, 1(y_i - x_i\theta \geq 0)(y - x_i\theta) - (1-\tau)\, 1(y_i - x_i\theta < 0)(y_i - x_i\theta) \tag{366}$$

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

$$S_i(\theta) = -x_i'\{\tau\, 1(y_i - x_i\theta \geq 0) - (1-\tau)\, 1(y_i - x_i\theta < 0)\} \tag{367}$$

$$H(x_i,\theta) = \frac{\partial S_i}{\partial \theta'} = 0 \tag{368}$$

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

重新考虑计算 $E[S_i(\theta)]$, 利用 $E(E(\,|x))$, 首先计算 $E[S_i(\theta)|x_i]$

$$\begin{aligned} E[S_i(\theta)|x_i] &= -x_i'\{\tau\, P(y_i - x_i\theta \geq 0|x_i) - (1-\tau)\, P(y_i - x_i\theta < 0|x_i)\} \\ &= -x_i'\{\tau\, P(u_i \geq x_i(\theta-\theta_0)|x_i) - (1-\tau)\, P(u_i < x_i(\theta-\theta_0)|x_i)\} \\ &= -x_i'\{\tau[1 - F_u(x_i(\theta-\theta_0)|x_i)] - (1-\tau)F_u(x_i(\theta-\theta_0)|x_i)]\} \\ &= -x_i'\{\tau - F_u(x_i(\theta-\theta_0)|x_i)\} \end{aligned}$$

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

$$\frac{\partial E[S_i(\theta|x_i)]}{\partial \theta'} = f_u(x_i(\theta-\theta_0)|x_i)x_i'x_i \tag{369}$$

$$A_0 = A(\theta_0) = E[f_u(0|x_i)x_i'x_i] \tag{370}$$

$$\sqrt{n}(\hat{\theta} - \theta_0) = A_0^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} S_i(\theta_0) + o_p(1) \sim N(0, A_0^{-1}B_0A_0^{-1}) \tag{371}$$

对于 $E[S_i(\theta_0)|x_i] = -x_i[\tau - F_u(0|x_i)]$, 因为 $F_u = P(u_i < x_i(\theta-\theta_0)|x_i), F_u(0) = P(y_i - x_i\theta_0 < 0|x_i)$，$x_i\theta_0$ is the τ percentile of y, 所以 $P(y_i < x_i\theta_0|x_i) = \tau$, 即满足 $E[S_i(\theta_0)] = E[E[S_i(\theta_0)|x_i]] = E[0] = 0$

$$\begin{aligned} B_0 &= E[S_i(\theta_0)S_i(\theta_0)'] \\ &= E[x_i'x'(\tau^2\, 1(y \geq x_i\theta_0) + (1-\tau)^2\, 1(y_i < x_i\theta_0)] \\ &= E[x_i'x_i(\tau^2(1-\tau) + (1-\tau)^2\tau)] \\ &= \tau(1-\tau)E[x_i'x_i] \end{aligned} \tag{372}$$

$$\hat{B}_0 = \tau(1-\tau)\frac{1}{n}\sum_{i=1}^{n} x_i'x' \tag{373}$$

f_u 不好求，所以根据导数的定义

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

$$\begin{aligned} A_0 &= E[f_u(0|x_i)x_i'x_i] \\ &= E\{E[1(|u_i| \le h)|x_i]x_i'x_i\}/2h \\ &= \frac{1}{2h}E\{1(|u_i| \le h)x_i'x_i\} \\ \hat{A}_0 &= \frac{1}{2nh}\sum_{i=1}^{n} 1(|\hat{u}_i| < h)x_i'x_i \end{aligned}$$

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

$$V\hat{a}r(\sqrt{n}(\hat{\theta})) = \frac{\tau(1-\tau)}{[\hat{f}_u(0)]^2}(\frac{1}{n}\sum_{i=1}^{n} x_i'x')^{-1} \tag{374}$$

$$\hat{f}_u(0) = \frac{1}{2nh}\sum_{i=1}^{n} 1[|u_i| \le h] \tag{375}$$

4 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 4.1 *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 4.2 *Weekly Stationary (Covariance Stationary)*
对于时间序列 $\{x_t\}$，满足

$$\begin{cases} E(x_t) = \mu \\ Var(x_t) = \gamma(0) < \infty \\ Cov(x_{t+h}, x_t) = \gamma(h) < \infty, \forall h = \pm 1 \pm 2, \ldots \end{cases}$$

w.n. (White Noise)
x_t is white noise If i)$Ex_t = 0$ ii) $Ex_t^2 = \sigma^2$ iii)$Ex_tx_s = 0 \quad \forall s \neq t$
Tread Stationary
$y_t = \alpha + \beta t + z_t$ $z_t \sim w.n.(0,\sigma^2)$ $Ey_t = \alpha + \beta t$
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 $y_0 = 0$)
$Ey_t = 0$ $Var(y_t) = t\sigma^2$

Random Walk with drift

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

$Ey_t = t\mu \; Var(y_t) = t\sigma^2$

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

$\gamma(0) = Var(y_t) = Var(y_{t+h})$

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) $\rho^*(h) = corr[y_t - E(y_t|y_{t-1}, \ldots, y_{t-h+1}), y_{t-h} - E(y_{t-h}|y_{t-1}, \ldots, y_{t-h+1})]$

$$y_t = \beta_1 y_{t-1} + \beta_2 y_{t-2} + \cdots + \beta_{h-1} y_{t-h+1} + \rho^*(h) y_{t-h} + error \tag{376}$$

$$\text{in OLS} : y = x_1\beta_1 + x_2\beta_2 + u \tag{377}$$

$$M_{x_1} y = M_{x_1} x_1 \beta_1 + M_{x_1} x_2 \beta_2 + M_{x_1} u \tag{378}$$

$$M_{x_1} y = M_{x_1} x_2 \beta_2 + M_{x_1} u \tag{379}$$

$$y_t - E(y|y_{t-1}, \ldots, y_{t-h+1}) = (y_{t-h} - E(y_{t-h}|y_{t-1}, \ldots, y_{t-h+1})\beta \tag{380}$$

$$\beta = [(y_{t-h} - E(y_{t-h}|y_{t-1}, \ldots, y_{t-h+1})'(y_{t-h} - E(y_{t-h}|.)]^{-1}(y_{t-h} - E(y_{t-h}|.)'(y_t - E(y|.)) \tag{381}$$

$$= Var^{-1}Cov = \rho^*(h) \tag{382}$$

因此求 $\rho^*(h)$, 只需要做 OLS

$$(\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_{h-1}, \rho^*(h)) = \Gamma_h^{-1}\gamma_h \tag{383}$$

$$\Gamma_h = (\gamma(i-j))_{i,j=1}^{h}, \gamma_h = (\gamma(1), \gamma(2), \ldots, \gamma(h))' \tag{384}$$

$$= \begin{pmatrix} \gamma(0) & \gamma(1) & \ldots & \gamma(h-1) \\ \gamma(1) & \gamma(0) & & \vdots \\ \vdots & & \ddots & \\ \gamma(h-1) & \gamma(h-2) & \ldots & \gamma(0) \end{pmatrix} \tag{385}$$

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

$$x_t = \overbrace{\phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + z_t}^{AR(p)} \!\!\!\!\underbrace{ + \theta_1 z_{t-1} + \cdots + \theta_q z_{t-q}}_{MA(q)} \tag{386}$$

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)

$$0 = \phi(z) = 1 - \phi_1(z) - \cdots - \phi_p z^p \tag{387}$$

is the characteristic form of the AR part of ARMA model

MA(q)

$$0 = \theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q \tag{388}$$

is the characteristic form of the MA part of ARMA model

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

$$x_t = \psi_1 x_{t-1} + \cdots + \psi_p x_{t-p} + z_t \, z_t \sim WN(0, \sigma^2) \tag{389}$$

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

$$x_t = z_t + \theta_1 z_{t-1} + \cdots + \theta_q z_{t-q} \, z_t \sim WN(0, \sigma^2) \tag{390}$$

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

$$Ez_t x_t = \phi E x_{t-1} z_t + E z_t^2 = \sigma^2 \tag{391}$$
$$E x_t^2 = E \phi x_t x_{t-1} + E x_t z_t \tag{392}$$
$$\gamma(0) = \phi\gamma(1) + \sigma^2 \tag{393}$$
$$E x_{t-1} x_t = \phi E x_{t-1}^2 + E x_{t-1} z_t \tag{394}$$
$$\gamma(1) = \phi\gamma(0) + 0 \tag{395}$$
$$\vdots \tag{396}$$
$$\gamma(h) = \phi\gamma(h-1) = \phi^h \gamma(0) \tag{397}$$
$$\rho(h) = \frac{\gamma(h)}{\gamma(0)} = \phi^h \tag{398}$$

AR is medium-term memory

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

$$\rho(h) = \begin{cases} \dfrac{\theta}{1+\theta^2} & \text{if } h = 1 \\ 0 & \text{if } h > 1 \end{cases} \tag{399}$$

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

$$\rho(h) = \begin{cases} \dfrac{\theta + \phi + \theta^2\phi + \theta\phi^2}{1 + \theta^2 + 2\theta\phi} & \text{if } h = 1 \\ \phi^{h-1} \dfrac{\theta + \phi + \theta^2\phi + \theta\phi^2}{1 + \theta^2 + 2\theta\phi} & \text{if } h > 1 \end{cases} \tag{400}$$

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$x_t = \theta_1 x_{t-1} + \theta_2 x_{t-2} + \ldots$，考虑 MA(1)

$$x_t = z_t - \theta z_{t-1} \tag{401}$$
$$x_{t-1} = z_{t-1} - \theta z_{t-2} \tag{402}$$
$$\vdots$$
$$MA(1) : x_t = -\theta x_{t-1} - \theta^2 x_{t-2} - \cdots - \theta^n x_{t-n} \, |\theta| < 1 \tag{403}$$

AR(3) 不能很好的拟合模型，但是 AR(50) 可以，即 PACF→ 0，可以考虑用类似 ARMA(2,2) 来拟合，同理 MA(50)，ACF→ 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:

$$\begin{cases} \gamma(0) = \dots \\ \gamma(1) = \dots \end{cases} \tag{404}$$

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

(4) NLS: example: MA(1)

$$x_t = \mu + z_t - \alpha_1 z_{t-1} \tag{405}$$

$$x_1 = \mu - \alpha_1 z_0 + z_1 \tag{406}$$

$$x_2 = \mu - \alpha_1 z_1 + z_2 \tag{407}$$

$$= (\mu + \alpha_1 \mu) - \alpha_1 x_1 - \alpha_1^2 z_0 + z_2 \tag{408}$$

$$x_t = (\sum_{s=0}^{t-1} \alpha_1^s)\mu - \sum_{s=1}^{t-1} \alpha_1^s x_{t-s} - \alpha^t z_0 + z_t \tag{409}$$

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

$$\sqrt{n}(\hat{\Phi}_p - \Phi_p) \sim N(0, \sigma^2 \Gamma_p^{-1}) \tag{410}$$

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 $\hat{\Phi}_h = (\hat{\phi}_1, \hat{\phi}_2, \dots, \hat{\phi}_h)' = \hat{\Gamma}_h^{-1} \gamma_h, h > p$

$$\sqrt{n}(\hat{\Phi}_h - \Phi_h) \sim N(0, \sigma^2 \Gamma_h^{-1}) \tag{411}$$

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

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

$$\hat{\rho} = \frac{\sum_t y_{t-1} y_t}{\sum_t y_{t-1}^2} = \rho + \frac{\sum_t y_{t-1} u_t}{\sum_t y_{t-1}^2} \tag{412}$$

$$\sqrt{n}(\hat{\rho} - \rho) = \frac{\frac{1}{\sqrt{n}} \sum_t y_{t-1} u_t}{\frac{1}{n} \sum_t y_{t-1}^2} \tag{413}$$

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。

$$E y_{t-1}^2 = \sigma_y^2 \tag{414}$$

$$E \frac{1}{n} \sum_t y_t^2 = \sigma_y^2 = E y_{t-1}^2 \tag{415}$$

$Bias = 0, Bias^2 = 0$

$|y|^4 = C < \infty$

$$Var(\frac{1}{n}\sum_t y_{t-1}^2) = \frac{1}{n^2}[\sum_t Var(y_{t-1}^2) + 2\sum_t \sum_{s>t} Cov(y_{t-1}^2, y_{s-1}^2)] \tag{416}$$

$$= O(\frac{1}{n}) + (?) \tag{417}$$

$$y_s^2 = \rho^2 y_{s-1}^2 + u_s^2 + 2\rho y_{s-1}u_s \tag{418}$$

$$\begin{aligned} Cov(y_s^2, y_{s-1}^2) &= Cov(\rho^2 y_{s-1}^2, y_{s-1}^2) + 0 \\ &= \rho^2 Var(y_{s-1}^2) \end{aligned} \tag{419}$$

$$Cov(y_s^2, y_{s-t}^2) = \rho^{2t}C \tag{420}$$

$$\begin{aligned} Var(\frac{1}{n}\sum_t y_{t-1}^2) &= \frac{1}{n^2}[nC + 2\sum_t \sum_{s>t} C\rho^{2(s-t)}] \\ &= O(\frac{1}{n}) + O(\frac{1}{n}) = O(\frac{1}{n}) \to 0 \end{aligned} \tag{421}$$

$Bias^2 \to 0, Var \to 0$ 所以分母收敛到 $Ey_{t-1}^2 = \sigma_y^2$

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

Theorem 4.1 *Martingale Difference CLT*

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

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

$$\begin{aligned} y_t &= \rho y_{t-1} + u_t \\ \sigma_y^2 &= \rho^2\sigma_y^2 + \sigma_u^2 \\ \frac{\sigma_u^2}{\sigma_y^2} &= 1 - \rho^2 \sqrt{n}(\hat{\rho} - \rho) \to N(0, 1-\rho) \end{aligned}$$

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

上述讨论为了区分 AR(p) 和 MA(q)

Theorem 4.2 *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*

$$\sqrt{n}(\hat{\rho}_h^* - \rho_h^*) \sim N(0, \mathbf{1}) \tag{422}$$

Theorem 4.3 *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*

$$\sqrt{n}(\hat{\rho}_h - \rho_h) \sim N(0, W) \tag{423}$$

$$\begin{aligned} \hat{\rho}_h &= (\hat{\rho}(1), \ldots, \hat{\rho}(h))' \\ \rho_h &= (\rho(1), \ldots, \rho(h))' \end{aligned}$$

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

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

对于假设 $H_0: \rho(h) = 0 H_1: \rho(h) \neq 0$ for each h>0 $y_t = \underset{=0?}{\rho}\, y_{t-1} + u_t$

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

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

(3) (1)(2) 通过回归方程构造 ρ 的表达式，(3) 根据 ρ 的定义 $\hat{\rho}(1) = \gamma(1) = \frac{\frac{1}{\sqrt{n}}\sum y_t y_{t-1}}{\frac{1}{n}\sum y_t^2}$
$E[y_t y_{t-1}|I_{t-1}] = y_{t-1}E[y_t|I_{t-1}] = 0$, 根据 MDCLT

$$\frac{1}{\sqrt{n}}\sum y_t y_{t-1} \sim N(0, \sigma_y^4) \tag{424}$$

$$\sqrt{n}\hat{\rho}(1) \sim N(0,1) \tag{425}$$

上述讨论检验的是 1 个 h, 问题是如何检验多个 h
$H_0 : \rho(h) = 0, h = 1, 2, \ldots, p$
$H_1 : \rho(h) \neq 0,$ for some h

$$Q = n\sum_{h=1}^{o} \hat{\rho}(h)^2 \sim \chi^2(p) \tag{426}$$

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：

$$y_t = x_t'\beta + u_t \quad y_t = x_t'\hat{\beta} + \hat{u}_t \quad E[u_t|x_1, x_2, \ldots, x_t] = 0 \tag{427}$$

$$\gamma(1) = E[u_{t-1}u_t] \stackrel{?}{=} 0 \tag{428}$$

$$\hat{u}_t = y_t - x_t'\beta = u_t - x_t'(\hat{\beta} - \beta) \tag{429}$$

$$\hat{u}_{t-1} = y_{t-1} - x_{t-1}'\beta = u_{t-1} - x_{t-1}'(\hat{\beta} - \beta) \tag{430}$$

$$\hat{\gamma}(1) = \frac{1}{n}\sum_t \hat{u}_{t-1}\hat{u}_t = \frac{1}{n}\sum [u_t u_{t-1} - u_{t-1}x_t'(\hat{\beta} - \beta) - u_t x_{t-1}'(\hat{\beta} - \beta) + (\hat{\beta} - \beta)x_{t-1}'x_t(\hat{\beta} - \beta)] \tag{431}$$

$$= A_1 - A_2 - A_3 + A_4 \tag{432}$$

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

$$A_1 = \frac{1}{n}\sum_t u_t u_{t-1} = O_p(?) \tag{433}$$

对于 $a_n = O_p(b_n)$, 如果有 $E|a_n| = O(\frac{1}{n})$ 则 $a_n = O_p(\frac{1}{n})$, 如果有 $E(a_n)^2 = O(\frac{1}{n^2})$ 则 $a_n = O_p(\frac{1}{n})$
原假设下 u_t, u_{t-1} 独立,$Eu_t = 0$, 有 $EA_1 = 0$
所以 $EA_1^2 = Var(A_1)$ 根据 MDCLT

$$\sqrt{n}A_1 = \frac{1}{\sqrt{n}}\sum_t u_t u_{t-1} \sim N(0, \sigma_u^4) \tag{434}$$

$$Var(\sqrt{n}A_1) = \sigma^4 \Rightarrow Var(A_1) = \frac{\sigma_u^4}{n} \tag{435}$$

$$EA_1^2 = O(\frac{1}{n}) \Rightarrow = O_p(\frac{1}{\sqrt{n}}) \tag{436}$$

$$A_2 = [\frac{1}{n}\sum_t u_{t-1}x_t'](\hat{\beta}-\beta) = \underset{1\times k}{B_2}C_2 \tag{437}$$

$$[\sqrt{n}(\hat{\beta}-\beta) \sim N = O_p(\frac{1}{\sqrt{n}})] \tag{438}$$

$$Var(B_2') = \frac{1}{n^2}[nVar(x_t u_{t-1}) + 2\cancel{\sum_t\sum_{s>t} Cov(x_t u_{t-1}, x_s u_{s-1})}] \tag{439}$$

$$= \frac{1}{n}E(x_t u_{t-1}^2 x_t') = \frac{1}{n}E(\underset{\sigma_u^2}{u_{t-1}^2})E(\underset{limit}{x_t x_{t-1}}) \tag{440}$$

$$B_2 = O_p(\frac{1}{\sqrt{n}}) \tag{441}$$

$$A_2 = B_2C_2 = O_p(\frac{1}{\sqrt{n}})O_p(\frac{1}{\sqrt{n}}) = O_p(\frac{1}{n}) \tag{442}$$

$$A_3 = O_p(\frac{1}{n}) \tag{443}$$

$$A_4 = \frac{1}{n}\sum(\hat{\beta}-\beta)x_{t-1}x_t'(\hat{\beta}-\beta) \tag{444}$$

$$= (\hat{\beta}-\beta)\frac{1}{n}\sum x_{t-1}x_t'(\hat{\beta}-\beta) \tag{445}$$

$$= O_p(\frac{1}{\sqrt{n}})(?)O_p(\frac{1}{\sqrt{n}}) \tag{446}$$

$$Var(\frac{1}{n}) \to 0 \tag{447}$$

$$Var(\frac{1}{n}\sum x_{t-1}x_t') = \underset{\xrightarrow{p}(Ex_{t-1}x_t')^2}{E(\frac{1}{n}\sum x_{t-1}x_t')^2} - (\underset{O_p(1)}{Ex_{t-1}x_t'})^2 = O_p(1) \tag{448}$$

$$\Rightarrow E(\frac{1}{n}\sum x_{t-1}x_t')^2 = O_p(1) \tag{449}$$

$$\Rightarrow \frac{1}{n}\sum x_{t-1}x_t' = O_p(\sqrt{1}) = O_p(1) \tag{450}$$

$$A_4 = O_p(\frac{1}{\sqrt{n}})O_p(1)O_p(\frac{1}{\sqrt{n}}) = O_p(\frac{1}{n}) \tag{451}$$

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

$$\hat{\gamma}(1) = O_p(\frac{1}{\sqrt{n}}) + O_p(\frac{1}{n}) \tag{452}$$

$$\sqrt{n}\hat{\gamma}(1) = \sqrt{n}A_1 + s.o. \sim N(0, \sigma_u^4) \tag{453}$$

$$\hat{u}_t = u_t^2 - 2u_t x_t'(\hat{\beta}-\beta) + x_t'(\hat{\beta}-\beta)(\hat{\beta}-\beta)'x_t \tag{454}$$

$$\hat{\gamma}(0) = \frac{1}{n}\sum \hat{u}_t = \underset{\to Eu_t^2}{\frac{1}{n}\sum_t u_t^2} - \underset{O_p(1)}{\frac{2}{n}\sum_t u_t x_t'}(\hat{\beta}-\beta) + (\hat{\beta}-\beta)\underset{O_p(1)}{\frac{1}{n}\sum_t x_t' x_t}(\hat{\beta}-\beta) \tag{455}$$

$$\hat{\gamma}(0) \to Eu_t^2 = \sigma_u^2 \tag{456}$$

$$\sqrt{n}\hat{\rho}(1) = \sqrt{n}\frac{\hat{\gamma}(1)}{\hat{\gamma}(0)} \sim N(0,1) \tag{457}$$

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

$$A_2 = [\frac{1}{n}\sum_t y_{t-1}u_{t-1}](\hat{\phi} - \phi) \tag{458}$$

$$Var([\frac{1}{n}\sum_t y_{t-1}u_{t-1}]) \to 0 \tag{459}$$

$$E(y_{t-1}u_{t-1})^2 \to O_p(1) \tag{460}$$

$$E(\frac{1}{n}\sum_t y_{t-1}u_{t-1})^2 \to E(y_{t-1}u_{t-1})^2 = O_p(1) \tag{461}$$

$$\Rightarrow A_2 = O_p(\frac{1}{\sqrt{1}})O_p(\frac{1}{\sqrt{n}}) \tag{462}$$

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

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

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

$$Var(u_t[u_{t-1} - (1-\phi^2)y_{t-1}]) = E(u_t^2[u_{t-1} - (1-\phi^2)y_{t-1}]^2) = \sigma_u^2\phi^2 \tag{463}$$

$$\Rightarrow \sqrt{n}\hat{\gamma}(1) \sim N(0, \sigma_u^4\phi^2) \tag{464}$$

$$\hat{\gamma}(0) \to Eu_t^2 = \sigma_u^2 \tag{465}$$

$$\sqrt{n}\hat{\rho} = \sqrt{n}\frac{\hat{\gamma}(1)}{\hat{\gamma}(0)} \overset{H_0}{\sim} N(0, \phi^2) \tag{466}$$

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

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

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

ARDL Auto Regressive Distributional Lag Model

$$\underbrace{y_t = \mu + \gamma_1 y_{t-1} + \cdots + \gamma_p y_{t-p}}_{autoregressive} + \underbrace{\beta_0 x_t + \beta_1 x_{t-1} + \cdots + \beta_r x_{t-r} + \varepsilon_t}_{distribution\,lag} \tag{467}$$

$$\Delta \text{ in } x \begin{cases} \text{shock} \begin{cases} SR & YES \\ LR & NO \end{cases} \\ \text{permenant} \begin{cases} SR & YES \\ LR & YES \end{cases} \end{cases} \tag{468}$$

$$y^* = \mu + \gamma_1 y^* + \cdots + \gamma_p y^* + \beta_0 x^* + \beta_x^* + \cdots + \beta_r x^* + \varepsilon \tag{469}$$

$$y^* = \frac{\hat{\mu}^2}{1 - \sum_{i=1}^p \hat{\gamma}_i} + \frac{\sum_{i=0}^r \hat{\beta}_j}{1 - \sum_{i=1}^p \hat{\gamma}_i}x^* \tag{470}$$

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

L:Lag Operator

$$Ly_t = y_{t-1} \tag{471}$$
$$L^2y_t = L(Ly_t) = y_{t-2} \tag{472}$$
$$(1-L)y_t = y_t - y_{t-1} = \Delta y_t \tag{473}$$

Define

$$B(L) = \beta_0 + \beta_1 L + \beta_2 L^2 + \cdots + \beta_r L^r \tag{474}$$
$$C(L) = 1 - \gamma_1 L - \cdots - \gamma_p L^p \tag{475}$$
$$ARDL : C(L)y_t = \mu + B(L)x_t + \varepsilon_t \tag{476}$$

Partial Adjustment Model

$$y_t^* = \alpha + \beta x_t + \delta w_t + \varepsilon_t \tag{477}$$
$$y_t - y_{t-1} = (1-\lambda)(y_t^* - y_{t-1}) \tag{478}$$
$$\Rightarrow y_t = \alpha(1-\lambda) + \beta(1-\lambda)x_t + \delta(1-\lambda)w_t + \lambda y_{t-1} + (1-\lambda)\varepsilon_t \tag{479}$$
$$y_t = \alpha' + \beta' x_t + \delta' w_t + \lambda y_{t-1} + \varepsilon_t' \tag{480}$$
$$C(L)y_t = \alpha' + \beta' x_t + \delta' w_t + \varepsilon_t' \quad C(L) = 1 - \lambda L \tag{481}$$
$$\frac{1}{C(L)} = \frac{1}{1-\lambda L} = 1 + \lambda L + (\lambda L)^2 + (\lambda L)^3 + \ldots \quad |\lambda| < 1 \tag{482}$$
$$y_t = [\alpha' + \lambda\alpha' + \lambda^2\alpha' + \ldots] + \beta'[x_t + \lambda x_{t-1} + \lambda^2 x_{t-2} + \ldots] + \delta'[w_t + \lambda w_{t-1} + \ldots] + [\varepsilon_t' + \lambda\varepsilon_{t-1}' + \ldots] \tag{483}$$

Common Factor Restriction

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t \tag{484}$$
$$\varepsilon_t = \rho\varepsilon_{t-1} + u_t \tag{485}$$
$$(1-\rho L)\varepsilon_t = u_t \quad \varepsilon_t = \frac{u_t}{1-\rho L} \tag{486}$$
$$y_t = \beta_0 + \beta_1 x_t + \frac{u_t}{1-\rho L} \tag{487}$$
$$1 - \rho L y_t = 1 - \rho L\beta_0 + 1 - \rho L\beta_1 x_t + u_t \tag{488}$$
$$y_t - \rho y_{t-1} = (\beta_0 - \rho\beta_0) + \beta_1 x_t - \beta_1\rho x_{t-1} + u_t \tag{489}$$
$$y_t = \beta_0' + \beta_1 x_t - \beta_1\rho x_{t-1} + \rho y_{t-1} + u_t \tag{490}$$
$$y_t = \gamma_0 + \gamma_1 x_t + \gamma_2 x_{t-1} + \gamma_3 y_{t-1} + u_t \tag{491}$$

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

$$y_t = lagy_t's + \beta x_t + lagx_t's + \underset{serialcorr}{error} \tag{492}$$

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

VAR

$$y_t = \mu + \Gamma_1 y_{t-1} + \Gamma_2 y_{t-2} + \cdots + \Gamma_p y_{t-p} + \varepsilon_t, \quad \varepsilon_t \text{ w.n.} \tag{493}$$

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

(1) $E\varepsilon_t = 0$, (2) $E\varepsilon_t\varepsilon_s' = 0$(error term 无序列相关), (3) $E\varepsilon_t\varepsilon_t' = \Omega = \begin{pmatrix} 4 & 2 & -1 & \\ 2 & 3 & 0 & \vdots \\ -1 & 0 & 1 & \\ & \cdots & & 6 \end{pmatrix}$ 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

$$y_t = \hat{\mu} + \hat{\Gamma}_1 y_{t-1} + \hat{\Gamma}_2 y_{t-2} + e_t \tag{494}$$

$$y_{t-1} = \hat{\mu} + \hat{\Gamma}_1 y_{t-2} + \hat{\Gamma}_2 y_{t-3} + e_t \tag{495}$$

$$y_t = (\hat{\mu} + \hat{\Gamma}_1\hat{\mu}) + (e_t + \hat{\Gamma}_1 e_{t-1}) + (\hat{\Gamma}_1^{\;2} + \hat{\Gamma}_2) y_{t-2} + \hat{\Gamma}_1\hat{\Gamma}_2 y_{t-3} \tag{496}$$

$$\vdots \tag{497}$$

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**

$$y_t = \alpha + \beta t + u_t OLS: \quad X = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ \vdots & \vdots \\ 1 & T \end{pmatrix} \tag{498}$$

$$\begin{pmatrix} \sqrt{n}(\hat{\alpha}-\alpha) \\ n^{\frac{3}{2}}(\hat{\beta}-\beta) \end{pmatrix} = \begin{pmatrix} \sqrt{n} & 0 \\ 0 & n^{\frac{3}{2}} \end{pmatrix} (X'X)^{-1} \begin{pmatrix} \sqrt{n} & 0 \\ 0 & n^{\frac{3}{2}} \end{pmatrix} \begin{pmatrix} \frac{\sum u_t}{\sqrt{n}} \\ \frac{\sum t u_t}{n^{\frac{3}{2}}} \end{pmatrix} \tag{499}$$

$$\tag{500}$$

$$A = \begin{pmatrix} \frac{\sum u_t}{\sqrt{n}} \\ \frac{\sum t u_t}{n^{\frac{3}{2}}} \end{pmatrix}, \quad EA = 0 \tag{501}$$

$$EAA' = E\begin{pmatrix} \frac{\sum u_t}{\sqrt{n}} \frac{\sum u_t}{\sqrt{n}} & \frac{\sum u_t}{\sqrt{n}} \frac{\sum t u_t}{n^{\frac{3}{2}}} \\ \frac{\sum u_t}{\sqrt{n}} \frac{\sum t u_t}{n^{\frac{3}{2}}} & \frac{\sum u_t}{\sqrt{n}} \frac{\sum t u_t}{n^{\frac{3}{2}}} \frac{\sum u_t}{\sqrt{n}} \frac{\sum t u_t}{n^{\frac{3}{2}}} \end{pmatrix} = \begin{pmatrix} \sigma^2 & \frac{1}{n^2}\sigma^2 \frac{n(n+1)}{2} \\ \frac{1}{n^2}\sigma^2 \frac{n(n+1)}{2} & \frac{\sigma^2 n(n+1)(2n+1)}{6n^3} \end{pmatrix} \tag{502}$$

$$A \sim N(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \sigma^2 \begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{3} \end{pmatrix}) \tag{503}$$

$$\text{Lindberg Feller CLT: } \begin{pmatrix} \sqrt{n}(\hat{\alpha} - \alpha) \\ \underbrace{n^{\frac{3}{2}}}_{\text{super consistent}}(\hat{\beta} - \beta) \end{pmatrix} = N(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \sigma^2 \begin{pmatrix} 4 & -6 \\ -6 & 12 \end{pmatrix}) \tag{504}$$

(2) **Unit Root Process**

$$y_t = \rho y_{t-1} + u_t, \quad \rho = 1 \tag{505}$$

$$\sqrt{n}(\hat{\rho} - \rho) \sim N(0, 1 - \rho^2) = N(0, 1 - 1) = 0 \tag{506}$$

$$\sqrt{n}\text{速度不够} \tag{507}$$

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

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

$$W(0) = 0, r > s, W(r) - W(s) \sim N(0, r - s) \tag{508}$$

$$r_2 > r_1 \quad [W(r_2) - W(r_1)] \perp W(r_1) \quad \text{增量独立} \tag{509}$$

$$y_1, y_2, \ldots, y_n, [nr]\text{取整} \tag{510}$$

$$r \in [0, 1], [nr] \in 0 \text{ to n} \tag{511}$$

$$y_n(r) = \begin{cases} 0, & 0 \le r < \frac{1}{n} \\ u_1, & \frac{1}{n} \le r < \frac{2}{n} \\ u_1 + u_2, & \frac{2}{n} \le r < \frac{3}{n} \\ \vdots & \vdots \\ u_1 + u_2 + \cdots + u_n, & r = 1 \end{cases} \tag{512}$$

for any fixed $r \in [0, 1]$, we can show $\frac{1}{\sqrt{n}} y_n(r) \sim N(0, r\sigma^2)$

$$\frac{1}{\sqrt{n}} y_n(r) = \frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} u_t = \frac{\sqrt{[nr]}}{\sqrt{n}} (\frac{1}{\sqrt{[nr]}} \sum_{t=1}^{[nr]} u_t) \to \sqrt{r} N(0, \sigma^2) = \sigma N(0, r) = \sigma W(r) \tag{513}$$

$$\begin{gathered} C_n(r + \Delta r) - C_n(r) \perp C_n(r) \\ \sum_{t=1}^{[n(r+\Delta r)]} - \sum_{t=1}^{[n(r)]} \\ \sum_{t=[nr]+1}^{[n(r+\Delta r)]} u_t \perp \sum_{t=1}^{[n(r)]} u_t \end{gathered} \tag{514}$$

$$\text{if } r = 1, \quad \frac{1}{\sqrt{n}} y_n(1)/\sigma \to W(1) \tag{515}$$

Review 积分

$$\int_0^1 g(x)dx = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1} \underbrace{g(x_{i-1})}_{g(x_i)\checkmark} \tag{516}$$

随机变量积分

$$\int_0^1 g(w(s))dw(s) = \lim_{n\to\infty}\sum_{i=1}^{n} \underbrace{g(w(s_i))}_{g(w(s_{i+1}))\times} \underbrace{(w(s_{i+1}) - w(s_i))}_{\text{对应}\frac{1}{n}} \tag{517}$$

$w(s_i)$ 不能取 $w(s_{i+1})$

$$\frac{1}{n} \sim dr \tag{518}$$

$$\sum_i^n \sim \int_0^1 \tag{519}$$

$$\frac{1}{\sqrt{n}}Y_{t-1} = \frac{1}{\sqrt{n}}\sum_{s=1}^{[nr]} u_s \sim \sigma_u w(r), \quad t-1 \le nr \le t \tag{520}$$

$$u_t = Y_t - Y_{t-1} \tag{521}$$

$$(1)\frac{u_t}{\sqrt{n}} = \frac{Y_t - Y_{t-1}}{\sqrt{n}} \sim d\sigma_u w_u(r) \tag{522}$$

$$(2)\frac{1}{\sqrt{n}}Y_{t-1} \sim \sigma_u w_u(r) \tag{523}$$

$$(3)\frac{1}{n} \sim dr \tag{524}$$

$$(4)\sum_i^n \sim \int_0^1 \tag{525}$$

$$n(\hat{\rho}-1) = \frac{n\sum Y_{t-1}u_t}{\sum Y_{t-1}^2} = \frac{\sum \frac{Y_{t-1}}{\sqrt{n}}\frac{u_t}{\sqrt{n}}}{\frac{1}{n}\sum(\frac{Y_{t-1}}{\sqrt{n}})^2} \tag{526}$$

$$\sim \frac{\int_0^1 \sigma_u^2 w_u(r)dw_u(r)}{\int_0^1 \sigma_u^2 w_u^2(r)dr} \quad \text{有限的} \not\to \infty \not\to 0 \tag{527}$$

$$y_t = x_t\beta + u \tag{528}$$

$$\left.\begin{aligned} x_t &= x_{t-1} + v_t \\ y_t &= y_{t-1} + \varepsilon_t \end{aligned}\right\} \text{random walk} \tag{529}$$

$$\hat{\beta} - \beta = (X'X)^{-1}X'u \tag{530}$$

$$n(\hat{\beta}-\beta) = (\frac{1}{n^2}\sum x_t^2)^{-1}\frac{1}{n}\sum x_t u_t \tag{531}$$

$$= (\frac{1}{n^2}\sum x_t^2)^{-1}(\frac{1}{n}\sum x_{t-1}u_t + \frac{1}{n}\sum v_t u_t) \tag{532}$$

$$\frac{1}{n}\sum(\frac{x_t}{\sqrt{n}})^2 \to \int_0^1 \sigma_v^2 w_r^2(r)dr \tag{533}$$

$$\sum \frac{x_{t-1}}{\sqrt{n}}\frac{u_t}{\sqrt{n}} \to \int_0^1 \sigma_v w_v(r) d\sigma_u w_u(r) \tag{534}$$

$$\frac{1}{n}\sum v_t u_t \to E[v_t u_t] \overset{assume}{=} 0 \tag{535}$$

$$n(\hat{\beta}-\beta) \sim [\int_0^1 \sigma_v^2 w_r^2(r)dr]^{-1}[\int_0^1 \sigma_v w_v(r)\sigma_u dw_u(r)] \tag{536}$$

如果不用左端点计算, 结果会有问题，考虑

$$y_t u_t = y_{t-1}u_t + u_t^2 \tag{537}$$
$$\frac{1}{n}\sum y_t u_t = \frac{1}{n}\sum y_{t-1}u_t + \frac{1}{n}\sum u_t^2 \tag{538}$$
$$\frac{1}{n}\sum y_t u_t \to \int_0^1 \sigma_u^2 w_u(r)dw_u(r) \tag{539}$$
$$\frac{1}{n}\sum y_{t-1} u_t \to \int_0^1 \sigma_u^2 w_u(r)dw_u(r) \tag{540}$$
$$\frac{1}{n}\sum u_t^2 \to \sigma^2 \tag{541}$$
$$\frac{1}{n}\sum y_t u_t < \frac{1}{n}\sum y_{t-1}u_t + \frac{1}{n}\sum u_t^2 \quad \text{矛盾} \tag{542}$$

5 Panel Data

$$y_{it} = x_{it}'\beta + c_i + u_{it} \tag{543}$$
$$\tag{544}$$

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

5.1 Fixed Effect

$$\underset{T\times 1}{y_1} = \begin{pmatrix} y_{11} \\ \vdots \\ y_{1T} \end{pmatrix}, \underset{T\times 1}{i} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \tag{545}$$
$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}\beta + \underset{d_1 \dots d_n}{\begin{pmatrix} i & & 0 \\ & \ddots & \\ 0 & & i \end{pmatrix}} \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} + \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix} \tag{546}$$
$$y_i = x_i\beta + i\alpha_i + u_i \tag{547}$$
$$FE: y = X\beta + D\boldsymbol{\alpha} + u \tag{548}$$

LSDV(Least Square Dummy Variable)

$$M_D y = M_D X\beta + \underset{=0}{M_D D\boldsymbol{\alpha}} + u \tag{549}$$
$$M_D = \underset{nT\times nT}{I} - D(D'D)^{-1}D' = I - \begin{pmatrix} \frac{1}{T}ii' & & 0 \\ & \vdots & \\ 0 & & \frac{1}{T}ii' \end{pmatrix} = \begin{pmatrix} M^0 & & 0 \\ & \ddots & \\ 0 & & M^0 \end{pmatrix} \tag{550}$$
$$M^0 = I - \frac{1}{T}ii' \tag{551}$$

$$M_D y = \begin{pmatrix} M^0 y_1 & & \\ & \ddots & \\ & & M^0 y_n \end{pmatrix} \tag{552}$$

$$x_{it} - \bar{x}_{i\cdot} = \ddot{x}_{it} \tag{553}$$

存在参考组，drop d_1, 则 α 表示其他组和参考组的差别 $H_0 : \alpha_2 = \cdots = \alpha_n = 0$
上面的方法叫做 one-way fixed effect, 类似的也可以控制 t

$$y_{it} = x'_{it}\beta + c_i + u_{it} \tag{554}$$

$$y_{*it} = y_{it} - \bar{y}_{i\cdot} - (\bar{y}_{\cdot t} - \bar{\bar{y}}) \tag{555}$$

$$x_{*it} = x_{it} - \bar{y}_{i\cdot} - (\bar{x}_{\cdot t} - \bar{\bar{x}}) \tag{556}$$

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

5.2 Random Effect

$$y_{it} = x'_{it}\beta + c_i + u_{it} \tag{557}$$

$$= x'_{it}\beta + v_{it} \tag{558}$$

$$\tag{559}$$

误差项的协方差矩阵不是对角阵，所以要用 (F)GLS 估计。

$$E(c_i|X) = 0 \tag{560}$$

$$E(u_{it}|X) = 0 \tag{561}$$

$$E(u_{it}^2|X) = \sigma_u^2 \tag{562}$$

$$E(c_i^2|X) = \sigma_c \tag{563}$$

$$E(u_{it}c_j|X) = 0 \tag{564}$$

$$E(u_{it}u_{js}|X) = 0, i \neq j \text{ or } s \neq t \tag{565}$$

$$E(c_i c_j|X) = 0 \text{if } i \neq j \tag{566}$$

$$v_i = \begin{pmatrix} v_{i1} \\ \vdots \\ v_{iT} \end{pmatrix}, \Sigma = E v_i v_i' = \begin{pmatrix} \sigma_c^2 + \sigma_u^2 & \sigma_c^2 & \dots & \sigma_c^2 \\ \sigma_c^2 & \ddots & & \\ & & \ddots & \\ \sigma_c^2 & \ddots & \ddots & \sigma_c^2 + \sigma_u^2 \end{pmatrix} \tag{567}$$

$$\Omega = EVV' = I_n \otimes \Sigma = \begin{pmatrix} \Sigma & & 0 \\ & \ddots & \\ 0 & & \Sigma \end{pmatrix} \tag{568}$$

$$\hat{\beta}_{GLS} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y \tag{569}$$

$$= (\sum_{i=1}^{n} x_i'\Sigma^{-1}x_i)^{-1}(\sum_{i=1}^{n} x_i\Sigma^{-1}y_i) \tag{570}$$

即对原数据做 $\Omega^{-\frac{1}{2}}$Transformation

$$\Sigma^{-\frac{1}{2}} = \frac{1}{\sigma_u}[I - \frac{\theta}{T}ii'] \tag{571}$$

$$\theta = 1 - \frac{\sigma_u}{\sqrt{\sigma_u^2 + T\sigma_c^2}} \tag{572}$$

$$\Sigma^{-\frac{1}{2}}y_i = \frac{1}{\sigma_u}\begin{pmatrix} y_{i1} - \theta\bar{y}_{i\cdot} \\ \vdots \\ y_{iT} - \theta\bar{y}_{i\cdot} \end{pmatrix} \tag{573}$$

有点类似 FE，$\theta = 1$ FE,$\sigma_c = 0$ Pooled

$\sigma_u^2 + \sigma_c^2$ 的估计

$$y_{it} = x_{it}'\beta + c_i + u_{it} \quad \text{OLS} \tag{574}$$

$$= x_{it}'\hat{\beta} + \hat{v}_{it} \tag{575}$$

$$\frac{1}{n}\sum \hat{v}_{it}^2 = \sigma_u^2 + \sigma_c^2 \tag{576}$$

$E\ddot{u}_{it}^2$ demean 之后 c_i 消掉了，即

$$\bar{y}_{i\cdot} = \bar{x}_{i\cdot}' + c_i + \bar{u}_{i\cdot} \tag{577}$$

$$y_{it} - \bar{y}_{i\cdot} = (x_{it}' - \bar{x}_{i\cdot}')\hat{\beta} + u_{it} - \bar{u}_{i\cdot} \tag{578}$$

$$\ddot{y}_{it} = \ddot{x}_{it}\hat{\beta} + \hat{\ddot{u}}_{it} \tag{579}$$

$$E\ddot{u}_{it}^2 = E(u_{it}^2 - 2u_{it}\bar{u}_{i\cdot} + \bar{u}_{i\cdot}^2) = \frac{T-1}{T}\sigma_u^2 \tag{580}$$

$$\sum_{i=1}^{n}\sum_{t=1}^{T} E\ddot{u}_{it}^2 = n(T-1)\sigma_u^2 \tag{581}$$

$$\sigma_u^2 = E[\sum_{i=1}^{n}\sum_{t=1}^{T}\frac{\ddot{u}_{it}^2}{n(T-1)}] \tag{582}$$

$\sum_{i=1}^{n}\sum_{t=1}^{T}\frac{\ddot{u}_{it}^2}{n(T-1)}$ 是 σ_u^2 的无偏估计，修正后

$$\frac{1}{n(T-1)-k}\sum_{i=1}^{n}\sum_{t=1}^{T}\ddot{u}_{it}^2 = \hat{\sigma}_u^2 \tag{583}$$

$$\hat{\beta}_{RE/FGLS} = (X'\hat{\Omega}^{-1}X)^{-1}X'\hat{\Omega}^{-1}y \tag{584}$$

$$= (\sum_{i=1}^{n} x_i'\hat{\Sigma}^{-1}x_i)^{-1}(\sum_{i=1}^{n} x_i\hat{\Sigma}^{-1}y_i) \tag{585}$$

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:$E(u_{it}|x_i, c_i) = 0, E(c_i|x_i) = E(c_i) = 0$

RE2:$rankE(x_i'\Omega^{-1}x_i) = k$

RE3:$E(u_iu_i'|x_i, c_i) = \sigma_u^2 I_T, E(c_i^2|x_i) = \sigma_c^2$

FE1:$E(u_{it}|x_i, c_i) = 0$

FE2:$rankE(\ddot{x}_i'\ddot{x}_i) = k$

FE3:$E(u_iu_i'|x_i, c_i) = \sigma_u^2 I_T$

$E(u_{it}|x_i)$ 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

$$\sqrt{n}(\hat{\beta}_{RE}-\beta)=(\frac{1}{n}\sum_{i=1}^{n}x_i'\Sigma^{-1}x_i)^{-1}(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}x_i'\Sigma^{-1}v_i) \tag{586}$$

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}x_i'\Sigma^{-1}v_i\sim N(0,E[x_i'\Sigma^{-1}E(v_iv_i'|x_i)\Sigma^{-1}x_i]) \tag{587}$$

$$V\hat{a}r(\sqrt{n}\hat{\beta}_{RE})=(\frac{1}{n}\sum_{i=1}^{n}x_i'\Sigma^{-1}x_i)^{-1}(\frac{1}{n}\sum_{i=1}^{n}x_i'\Sigma^{-1}\hat{v}_i\hat{v}_i'\Sigma^{-1}x_i)(\frac{1}{n}\sum_{i=1}^{n}x_i'\Sigma^{-1}x_i)^{-1} \tag{588}$$

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

$$\hat{\beta}_{GFGLS}=(\sum_{i=1}^{n}x_i'\hat{\Sigma}^{-1}x_i)^{-1}(\sum_{i=1}^{n})^{-1}(\sum_{i=1}^{n}x_i'\hat{\Sigma}y_i) \tag{589}$$

$$V\hat{a}r(\sqrt{n}\hat{\beta}_{GFGLS})=(\sum_{i=1}^{n}x_i'\hat{\Sigma}^{-1}x_i)^{-1} \tag{590}$$

if FE3 is violated

$$V\hat{a}r(\sqrt{n}\hat{FE})=(\frac{1}{n}\sum_{i=1}^{n}\ddot{x}_i'\ddot{x}_i)^{-1}(\frac{1}{n}\sum_{i=1}^{n}\ddot{x}_i'\hat{\ddot{u}}_i\hat{\ddot{u}}_i'\ddot{x}_i)(\frac{1}{n}\sum_{i=1}^{n}\ddot{x}_i'\ddot{x}_i)^{-1} \tag{591}$$

which is Robust Var Matrix Estimator

FE3 要求对角阵

FEGLS3 $E(u_iu_i'|x_i,C_i)=\Lambda$ 任意形式

$\ddot{u}_i=Q_Tu_i$, Q_T is demean Transformation

$$E(\ddot{u}_i\ddot{u}_i'|\ddot{x}_i)=E(\ddot{u}_i\ddot{u}_i') \tag{592}$$

$$=Q_TE(u_iu_i')Q_T \tag{593}$$

$$=Q_T\Lambda Q_T\overset{def}{=}\Sigma \tag{594}$$

$$\hat{\Sigma}=\frac{1}{n}\sum_{i=1}^{n}\hat{\ddot{u}}_i\hat{\ddot{u}}_i' \tag{595}$$

$$\hat{\Sigma}^{-\frac{1}{2}}\ddot{y}_i=\hat{\Sigma}^{-\frac{1}{2}}\sum_{i=1}^{n}\ddot{x}_i\beta+\hat{\Sigma}^{-\frac{1}{2}}\ddot{u}_i \tag{596}$$

$$Var(\sqrt{n}\hat{\beta}_{FEGLS})=(\frac{1}{n}\sum_{i=1}^{n}\ddot{x}_i'\hat{\Sigma}\ddot{x}_i)^{-1} \tag{597}$$

由于对 $\ddot{u}_i$ 做了 demean 处理，$\hat{\Sigma}$ 是 $T-1\times T-1$ 矩阵，去掉了某一个时间的值。

内生性 z_i 是 IV

REIV

$$E(u_{it}|x_i)\neq 0 \tag{598}$$

$$E(u_{it}|z_i)=0 \tag{599}$$

$$E(c_i|z_i)=0 \tag{600}$$

$$\Sigma^{-\frac{1}{2}}y_i=\Sigma^{-\frac{1}{2}}x_i\beta+\Sigma^{-\frac{1}{2}}v_i \tag{601}$$

$$(I_n\otimes\Sigma^{-\frac{1}{2}})\quad\Omega^{-\frac{1}{2}}y=\Omega^{-\frac{1}{2}}X\beta+\Omega^{-\frac{1}{2}}v \tag{602}$$

$$\Omega^{-\frac{1}{2}}y=\Omega^{-\frac{1}{2}}Z\beta+\Omega^{-\frac{1}{2}}v\quad(2SLS) \tag{603}$$

First stage

$$\Omega^{-\frac{1}{2}}X = \Omega^{-\frac{1}{2}}Z\delta + error \tag{604}$$

Second stage

$$\Omega^{-\frac{1}{2}}y = \Omega^{-\frac{1}{2}}Z(Z'\Omega^{-1}Z)^{-1}Z'\Omega^{-1}X\beta + error \tag{605}$$

$$\hat{\beta}_{REIV} = (X\Omega^{-1}Z(Z'\Omega^{-1}Z)^{-1}Z'\Omega^{-1}X)^{-1}X'\Omega^{-1}Z(Z'\Omega^{-1}Z)^{-1}Z'\Omega^{-1}y \tag{606}$$

$$\sqrt{n}(\hat{\beta}_{REIV} - \beta) \sim N(0, [\frac{X'\Omega^{-1}Z}{n}(\frac{Z'\Omega^{-1}Z}{n})^{-1}\frac{Z'\Omega X}{n}]^{-1}) \tag{607}$$

Test of endogeneity

$$y_{it1} = z_{it1}\delta + y_{it2}\alpha_1 + y_{it3}\gamma_1 + \underbrace{c_{i1} + u_{it1}}_{v_{it1}} \tag{608}$$

$$H_0: \quad E(y_{it3}|v_{it1}) = 0 \quad \forall s = 1, \ldots, T \tag{609}$$

$$y_{it3} = z_{it}\Pi_3 + v_{it3} \Rightarrow \hat{v}_{it3} \quad z_{it} = (\overset{z_{it1}}{exo}, \overset{z_{it2}}{iv}) \tag{610}$$

$$y_{it1} = z_{it1}\delta + y_{it2}\alpha_1 + y_{it3}\gamma_1 + v_{it3}\rho_1 + error \tag{611}$$

$$H_0: \rho_1 = 0 \tag{612}$$

$\rho_1 = 0$,y_{it3}, y_{it1} 无关
FEIV

$$E(u_{it}|x_i) = 0 \tag{613}$$

$$E(u_{it}|z_i) = 0 \tag{614}$$

$$\tag{615}$$

First stage

$$\ddot{x}_i = \ddot{z}_i\delta + error \tag{616}$$

Second stage

$$\ddot{y}_i = \hat{\ddot{x}}_i\beta + error \tag{617}$$

$$\hat{\beta}_{FEIV} = \tag{618}$$

$$\sqrt{n}(\hat{\beta}_{FEIV} - \beta) \sim N(0, \sigma^2[(E\ddot{x}_i'\ddot{z}_i)(E\ddot{z}_i'\ddot{z}_i)^{-1}(E\ddot{z}_i'x_i)]) \tag{619}$$

Test of endogeneity

$$y_{it3} = z_{it}\Pi_3 + v_{it3} \Rightarrow \hat{v}_{it3} \quad z_{it} \tag{620}$$

$$FEIV(z_{it}, y_{it3}, \hat{v}_{it3}) \tag{621}$$

First Difference FD

$$y_{it} = x_{it}'\beta + c_i + u_{it} \tag{622}$$

$$\Delta y_{it} = \Delta x_{it}'\beta + \Delta u_{it} \tag{623}$$

$$\Delta x_{it} \quad \Delta u_{it} \quad \text{无关} \tag{624}$$

$$x_{it} - x_{it-1} \quad u_{it} - u_{it-1} \quad \text{无关}$$

和同期未来一期滞后一期都无关

$E(u_{it}|z_i) = 0$, use w_{it} as IV $E(w'_{it}\Delta u_{it}) = 0$, w_{it} 和 t 相关

$$E(\begin{pmatrix} w'_{i2} & & \\ & \ddots & \\ & & w'_{iT} \end{pmatrix} \begin{pmatrix} \Delta u_{i2} \\ \vdots \\ \Delta u_{iT} \end{pmatrix}) = 0 \tag{625}$$

即对每个时间 t 找一个工具变量，不需要严格外生假定

$$\left.\begin{aligned} \Delta x'_{it} \overset{IV}{\leftarrow} w_{it}, \quad i = 1, \ldots, n \\ \Delta x'_{it-1} \overset{IV}{\leftarrow} w_{it-1}, \quad i = 1, \ldots, n \\ \vdots \end{aligned}\right\} \text{T-1 seperate OLS} \tag{626}$$

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

$$y_{it} = x'_{it}\beta + c_i + u_{it} \tag{627}$$
$$E(u_{it}|x'_{it}, x'_{it-1}, \ldots, x'_{i1}) = 0 \tag{628}$$
$$\tag{629}$$

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

$$\left.\begin{aligned} t = 2 y_{i0} \\ t = 3 y_{i0}, y_{i1} \\ t = T y_{i0}, \ldots, y_{iT-2} \end{aligned}\right\} \text{IVs,System 2SLS} \Delta y_{it} = \Delta x'_{it}\beta + \Delta u_{it} \tag{630}$$

$$\left.\begin{aligned} t = 2 x_{i1} \\ t = 3 x_{i1}, x_{i2} \\ t = T x_{i1}, \ldots, y_{iT-1} \end{aligned}\right\} \text{IVs,System 2SLS} \tag{631}$$

6 Nonlinear Model

6.1 Binary Choice

Linear Probability Model

$$y_i = x'_i\beta + \varepsilon_i \tag{632}$$
$$E(\varepsilon_i|x_i) = 0 \tag{633}$$
$$y = 1 or 0 \tag{634}$$
$$E(y_i|x_i) = x'_i\beta = 1 \times Prob(y_i = 1|x_i) + \cancel{0 \times Prob(y_i = 0|x_i)} \tag{635}$$
$$Var(\varepsilon_i|x_i) = E(\varepsilon_i^2|x_i) \tag{636}$$
$$= (1 - x'_i\beta)^2 x_i\beta + (-x_i\beta)^2(1 - x_i\beta) \tag{637}$$
$$= (1 - x'_i\beta)x'_i\beta \tag{638}$$

弊端是可能为负

$$U_i^{rent} = x_i'\beta_r + \varepsilon_{ir} \tag{639}$$
$$U_i^{buy} = x_i'\beta_b + \varepsilon_{ib} \tag{640}$$
$$y_i^* = U_i^r - U_i^b > 0 \quad y_i = 1 \tag{641}$$
$$y_i^* = U_i^r - U_i^b \leq 0 \quad y_i = 0 \tag{642}$$
$$\tag{643}$$

y_i^* is latent Variable

$$P(y_i = 1|x_i) = P(y_i^* > 0|x_i) \tag{644}$$
$$= P(x_i'\beta + \varepsilon_i > 0|x_i) \tag{645}$$
$$= P(\varepsilon_i > -x_i'\beta|x_i) \quad \text{如果 } \varepsilon_i \text{ 是对称分布} \tag{646}$$
$$= \underset{CDF}{F}(x_i'\beta) \tag{647}$$

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

6.2 Probit model with endog var

two stage

$$y_1^* = z_1\delta_1 + \alpha_1 y_2 + u_1 \tag{648}$$
$$y_1 = 1(y_1^* > 0) \quad Probit \quad u_i \sim N(0,1) \tag{649}$$
$$y_2\text{连续内生} \tag{650}$$
$$y_2 = z_1\delta_{21} + z_2\delta_{22} + v_2 \tag{651}$$
$$\text{first stage: } = z\delta_2 + v_2 \tag{652}$$
$$y_2|Z \sim N(z\delta_2, \tau_2^2) \quad \tau_2^2 = Var(v_2) \tag{653}$$
$$u_1 = \theta_1 v_2 + e_1 \tag{654}$$
$$\theta_1 = \frac{Cov(u_1, v_2)}{Var(v_2)} \overset{def}{=} \frac{\eta_1}{\tau_2^2} \tag{655}$$
$$\text{second stage: } y_1^* = z_1\delta_1 + \alpha_1 y_2 + \theta_1 v_2 + e_1 \tag{656}$$

e_1 不是标准正态

$$E(e_1) = E(u_1 - \theta_1 v_2) = 0 \tag{657}$$
$$Var(u_1) = \theta_1^2 Var(v_2) + Var(e_1) \tag{658}$$
$$Var(e_1) = 1 - \frac{\eta_1^2}{\tau_2^2} \overset{def}{=} 1 - \rho_1^2 \tag{659}$$
$$\rho_1 = \frac{cov(u_1, v_2)}{\sqrt{var(u_1)var(v_2)}} = \frac{\eta_1}{\tau_2} \tag{660}$$
$$e_1 \sim N(0, 1 - \rho_1^2) \tag{661}$$
$$P(y_1 = 1|z, y_2, \hat{v}_2) = \Phi[(z_1\delta_1 + \alpha_1 y_2 + \theta_1 \hat{y}_2)/(1 - \hat{\rho}_1^2)^{\frac{1}{2}}] \tag{662}$$
$$\rho_1 = \theta_1 \tau_2 \tag{663}$$
$$\Rightarrow \hat{\delta}_1, \hat{\alpha_1}, \ldots \tag{664}$$

MLE

$$\text{MLE: } f(y_1, y_2|z) = f(y_1|y_2, z)f(y_2|z) \quad \text{连续用 pdf f() 表示} \tag{665}$$

$$= P(y_1|y_2, z)f(y_2|z) \quad \text{离散用分布函数 P() 表示} \tag{666}$$

$$P(y_1 = 1|y_2, z) = \Phi[(z_1\delta_1 + \alpha_1 y_2 + \frac{\rho_1}{\tau_2}(y_2 - z\delta_2))/(1-\rho_1^2)^{\frac{1}{2}}] \tag{667}$$

$$f(y_2|z) = \frac{1}{\sqrt{2\pi}\tau}e^{\frac{(y_2-z\delta_2)^2}{2\tau_2^2}} \tag{668}$$

$$\Rightarrow \hat{\delta}_1, \hat{\alpha}_1 \tag{669}$$

内生变量是离散内生的情况

$$y_1 = 1[z_1\delta_1 + \alpha_1 y_2 + u_1 > 0] \tag{670}$$

$$y_2 = 1[z\delta_2 + v_2 > 0] \quad \text{离散内生} \tag{671}$$

$$u_1, v_2 \sim N(0, 1) \tag{672}$$

two stage 不能使用 $\hat{v}_2$ 算不出来，属于一个区间

MLE:$P(y_1 = i, y_2 = j) = P(y_1 = i|y_2 = j, z)P(y_2 = j|z), \quad i, j = 0, 1$,e.g.

$$P(y_1 = 1, y_2 = 1) = P(y_1 = 1|y_2 = 1, z)P(y_2 = 1|z) \tag{673}$$

$$P(y_1 = 1|v_2, z) = \Phi[(z_1\delta_1 + \alpha_1 y_2 + \rho_1 v_2)/(1-\rho_1^2)^{\frac{1}{2}}] \tag{674}$$

$$E[P(y_1 = 1|v_2, z)|y_2 = 1, z] \tag{675}$$

$$\text{对 } v_2 \text{ 积分} = \int_{-z\delta_2}^{\infty} P(y_1 = 1|v_2, z)f(v_2|y_2 = 1, z)dv_2 \tag{676}$$

$$= \int_{-z\delta_2}^{\infty} f(y_1 = 1, v_2|y_2 = 1, z)dv_2 \tag{677}$$

$$= P(y_1 = 1|y_2 = 1, z) \tag{678}$$

$$(676) = \int_{-z\delta_2}^{\infty} \Phi[(z_1\delta_1 + \alpha_1 y_2 + \rho_1 v_2)/(1-\rho_1^2)^{\frac{1}{2}}]\frac{\phi(v_2)}{P(v_2 > -z\delta_2)}dv_2 \tag{679}$$

$$= \frac{1}{\Phi(z\delta_2)}\int_{-z\delta_2}^{\infty} \Phi[(z_1\delta_1 + \alpha_1 y_2 + \rho_1 v_2)/(1-\rho_1^2)^{\frac{1}{2}}]\phi(v_2)dv_2 \tag{680}$$

$$P(y_2 = 1|z) = \Phi(z\delta_2) \tag{681}$$

$$P(y_1 = 1, y_2 = 1|z) = P(y_1 = 1|y_2 = 1, z)P(y_2 = 1|z) = \int_{-z\delta_2}^{\infty}\int_{-z_1\delta_1-\alpha_1 y_2}^{\infty} \frac{1}{2\pi\sqrt{1-\rho_1^2}}e^{-\frac{1}{2(1-\rho_1^2)}[u_1^2 - 2\rho_1 u_1 u_2 + v_2^2]} \tag{682}$$

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

$$y_1 = 1[x_1\beta_1 + e_1 > 0] \tag{683}$$

$$y_2 = 1[x_2\beta_2 + e_2 > 0] \tag{684}$$

$$\begin{pmatrix} e_1 \\ e_2 \end{pmatrix} \sim N(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}) \tag{685}$$

if $\rho = 0$,separate probit, joint MLE 都一致有效

if $\rho \neq 0$ 参考式 (682)

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

6.3 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

$$E(y_i|x_i, y \leq a) \neq x_i'\beta \tag{686}$$

$$f(x|x>a) = \frac{f(x)}{P(x>a)} \tag{687}$$

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)

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma}e^{\frac{1}{2}(\frac{x-\mu}{\sigma})^2} \quad x \sim N(\mu, \sigma^2) \tag{688}$$

$$P(X>a) = 1 - \Phi(\frac{a-\mu}{\sigma}) \tag{689}$$

$$f(x|x>a) = \frac{\frac{1}{\sigma}\phi(\frac{x-\mu}{\sigma})}{1-\Phi(\frac{a-\mu}{\sigma})} \tag{690}$$

$$E[x|x>a] = \int_a^\infty xf(x|x>a)dx \tag{691}$$

$$= \int_a^\infty x\frac{\frac{1}{\sigma}\phi(\frac{x-\mu}{\sigma})}{1-\Phi(\frac{a-\mu}{\sigma})}dx \tag{692}$$

$$\frac{x-\mu}{\sigma} \overset{def}{==} v \quad \frac{a-\mu}{\sigma} \overset{def}{=} \alpha \tag{693}$$

$$= \int_\alpha^\infty (\sigma v + \mu)\frac{\cancel{\frac{1}{\sigma}}\phi(v)}{1-\Phi(\alpha)}\cancel{\sigma}dv \tag{694}$$

$$\phi'(x) = -x\phi(x) \tag{695}$$

$$= \frac{\sigma}{1-\Phi(\alpha)}\int_\alpha^\infty (-\phi'(x))dv + \frac{\mu}{1-\Phi(\alpha)}\int_\alpha^\infty \phi(v)dv \tag{696}$$

$$= \mu + \sigma\frac{\phi(\frac{a-\mu}{\sigma})}{1-\Phi(\alpha)} \tag{697}$$

$$y_i = x_i'\beta + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2) \tag{698}$$

$$y_i|x_i \sim N(x_i'\beta, \sigma^2) \tag{699}$$

$$E[y_i|y_i>a] = x_i'\beta + \sigma\frac{\phi(\frac{a-x_i'\beta}{\sigma})}{1-\Phi(\frac{a-x_i'\beta}{\sigma})} \tag{700}$$

$$\text{MLE: } L(\beta, \sigma; x, y) = \prod_{i=1}^n \frac{\frac{1}{\sigma}\phi(\frac{y_i-x_i'\beta}{\sigma})}{1-\Phi(\frac{a-x_i'\beta}{\sigma})} \tag{701}$$

7 Non Paramtric Model

参数模型 $y = m(x) + u \quad E(y|x)$, $\hat{\theta} \to \theta$, $\sqrt{n}$ 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

$$f(x) = \frac{dF(x)}{dx} \tag{702}$$

$$= \lim_{h\to 0} \frac{F(x+h) - F(x-h)}{2h} \tag{703}$$

$$= \lim_{h\to 0} \frac{x-h \le rv \le x+h}{2h} \tag{704}$$

$$= \lim_{h\to 0} \frac{\{\text{\# of } x_i' \text{s falling in the interval [x-h,x+h]}\}}{2h\dot{n}} \tag{705}$$

$$\text{Def } k(\frac{x_i - x}{h}) = k(z) = \{ \begin{matrix} \frac{1}{2} & if|z| < 1 \\ 0 & \text{otherwise} \end{matrix} \tag{706}$$

$$= \lim_{h\to 0} \frac{1}{nh} \sum_{i=1}^{n} k(\frac{x_i - x}{h}) \tag{707}$$

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{h} k(\frac{x_i - x}{h}) \tag{708}$$

k Kernel function 距离中心点越远越小

2nd order kernel function

$$\begin{cases} \int k(v)dv = 1 \\ \int vk(v)dv = 0 \quad \text{奇函数} \\ \int v^2 k(v)dv = k_2 > 0 \end{cases} \tag{709}$$

4th order kernel function

$$\begin{cases} \int k(v)dv = 1 \\ \int vk(v)dv = 0 \\ \int v^2 k(v)dv = 0 \\ \int v^3 k(v)dv = 0 \\ \int v^4 k(v)dv = k_4 \neq 0 \end{cases} \tag{710}$$

h is bandwith $n, h, X, k(\dot{})$ 已知

$\overset{MSE}{\rightarrow}$

$$E\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} Ek(\frac{x_i - x}{h}) \quad \text{k is 2nd kernel} \tag{711}$$
$$= \frac{1}{h}Ek(\frac{x_i - x}{h}) \tag{712}$$
$$Ek(\frac{x_i - x}{h}) = \int f(x_i)k(\frac{x_i - x}{h})dx_i \quad \frac{x_i - x}{h} = v, x_i = x + hv \tag{713}$$
$$= \int f(x + hv)k(v)hdv \tag{714}$$
$$\text{Taylor} = \int [f(x) + f'(x)hv + \frac{1}{2}f''(x)h^2x^2 + \dots]k(v)hdv \tag{715}$$
$$= hf(x) + \frac{h^3}{2}f''(x)\int v^2k(v)dv + o(h^3) \tag{716}$$
$$E\hat{f}(x) = f(x) + \frac{h^2}{2}f''(x)K_2 + o(h^2) \tag{717}$$
$$\text{bias: } E\hat{f} - f = \frac{h^2}{2}f''(x)K_2 + o(h^2) \tag{718}$$

$$var(\hat{f}) = \frac{1}{nh^2}var(k(\frac{x_i - x}{h})) \tag{719}$$
$$var(k(\frac{x_i - x}{h})) = Ek^2(\frac{x_i - x}{h}) - [Ek(\frac{x_i - x}{h})]^2 \tag{720}$$
$$Ek^2(\frac{x_i - x}{h}) = \int f(x_i)k^2(\frac{x_i - x}{h})dx \tag{721}$$
$$= \int f(x + hv)k^2(v)hdv \tag{722}$$
$$= \int [f(x) + f'(x)hv + \frac{f''(x)}{2}h^2v^2 + o(h^2)]k^2(v)hdv \tag{723}$$
$$= hf(x)\int k^2(v)dv + o(h) \tag{724}$$
$$var(\hat{f}(x)) = \frac{1}{nh}f(x)\int k^2(v)dv + o(\frac{1}{nh}) = \frac{1}{nh}fK + o(\frac{1}{nh}) \tag{725}$$
$$bias^2 + var = \overset{MSE}{\rightarrow} 0 \Rightarrow \overset{p}{\rightarrow} 0 \tag{726}$$
$$bias^2 + var \overset{?}{\rightarrow} 0 \tag{727}$$
$$MSE\hat{f} = \frac{h^4}{4}(K_2f'')^2 + \frac{fk}{nh} + o(h^4 + \frac{1}{nh}) \tag{728}$$
$$\frac{\partial MSE}{\partial h}\hat{f} = 0 \Rightarrow h_{opt} = [\frac{kf}{(K_2f'')^2}]^{\frac{1}{5}}n^{-\frac{1}{5}} \overset{def}{=} C(x)n^{-\frac{1}{5}} \tag{729}$$
$$C(x) \quad \text{finite} \tag{730}$$

assume $x \sim N(\mu, \sigma^2)$, $\quad v \sim N(0, 1)$, $\quad h \propto n^{-\frac{1}{5}}$

$$\frac{fk}{nh} = O(n^{-\frac{4}{5}}), \frac{h^4}{4}(k_2f'')^2 = O(n^{-\frac{4}{5}}) \tag{731}$$

$$g(x) = E(y|x) \tag{732}$$
$$= \int yf(y|x)dy \tag{733}$$
$$= \int y\frac{f(y,x)}{f(x)}dy \tag{734}$$
$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} k(\frac{x_i - x}{h}) \tag{735}$$
$$\hat{f}(y,x) = \frac{1}{nh_xh_y}\sum_{i=1}^{n} k(\frac{x_i - x}{h_x})k(\frac{y_i - y}{h_y}) \tag{736}$$

$$\hat{g}(x) = \frac{\int y\hat{f}(y,x)dy}{\hat{f}(x)} \overset{def}{=} \frac{\hat{m}(x)}{\hat{f}(x)} \tag{737}$$

$$\hat{m}(x) = \int y \frac{1}{nh_x h_y} \sum_{i=1}^{n} k(\frac{x_i - x}{h_x})k(\frac{y_i - y}{h_y})dy \tag{738}$$

$$\int yk(\underset{def\ v}{\frac{y_i - y}{h_y}})dy, \quad y_i = y + h_y v \tag{739}$$

$$= \int_{+\infty}^{-\infty} (y_i - h_y v)k(v)(-h_y)dv \tag{740}$$

$$= \int_{-\infty}^{+\infty} (y_i - h_y v)k(v)(h_y)dv \tag{741}$$

$$= h_y \int_{-\infty}^{+\infty} (y_i - h_y v)k(v)dv \tag{742}$$

$$= h_y y_i \tag{743}$$

$$\hat{m}(x) = \frac{1}{nh_x} \sum_{i=1}^{n} y_i k(\frac{x_i - x}{h_x}) \tag{744}$$

$$\hat{g}(x) = \frac{\cancel{\frac{1}{nh}} \sum_{i=1}^{n} y_i k(\frac{x_i - x}{h_x})}{\cancel{\frac{1}{nh}} \sum_{i=1}^{n} k(\frac{x_i - x}{h})} = \frac{\hat{m}(x)}{\hat{f}(x)} \tag{745}$$

$$= \frac{E\hat{m} + \hat{m} - E\hat{m}}{E\hat{f} + \hat{f} - E\hat{f}} = \frac{E\hat{m} + \hat{m} - E\hat{m}}{E\hat{f}[1 + \frac{\hat{f} - E\hat{f}}{E\hat{f}}]} \tag{746}$$

$$= \frac{E\hat{m} + \hat{m} - E\hat{m}}{E\hat{f}}[1 - \frac{\hat{f} - E\hat{f}}{E\hat{f}} + ()^2 - ()^3 + \dots] \tag{747}$$

$$= \frac{E\hat{m}}{E\hat{f}} + \frac{\hat{m} - E\hat{m}}{E\hat{f}} - \frac{E\hat{m}(\hat{f} - E\hat{f})}{(E\hat{f})^2} - \frac{(\hat{m} - E\hat{m})(\hat{f} - E\hat{f})}{(E\hat{f})^2} + \frac{E\hat{m}(\hat{f} - E\hat{f})^2}{E(\hat{f})^3} + \dots \tag{748}$$

$$E\hat{m} = E\{E(\hat{m}(x)|x_i)\} \tag{749}$$

$$= E\{E(\frac{1}{nh}\sum y_i k(\frac{x_i - x}{h})|x_i)\} \tag{750}$$

$$= E\{\frac{1}{nh}\sum E(y_i|x_i)k(\frac{x_i - x}{h})\} \tag{751}$$

$$= E\{\frac{1}{nh}\sum g(x_i)k(\frac{x_i - x}{h})\} \tag{752}$$

$$= \frac{1}{h}Eg(x_i)k(\frac{x_i - x}{h}) \tag{753}$$

$$= \frac{1}{h}\int g(x_i)k(\frac{x_i - x}{h})f(x_i)dx_i \tag{754}$$

$$= \frac{1}{\cancel{h}}\int g(x + hv)k(v)f(x + hv)\cancel{h}dv \tag{755}$$

$$= \int [g + g'hv + \frac{g''}{2}h^2v^2 + o(h^2)]k(v)[f + f'hv + \frac{f''}{2}h^2v^2 + o(h^2)]dv \tag{756}$$

$$= gf + \frac{h^2}{2}[2g'f' + gf'' + g''f]K_2 + o(h^2) \tag{757}$$

$$Var\hat{m} = E\{\underset{A}{Var(\hat{m}|x_i)}\} + Var\{\underset{B}{E(\hat{m}|x_i)}\} \tag{758}$$

law of total variance/variance decomposion/conditional variance /law of iterated variance, 也可以用回归来理解

$$A = E\{Var[\frac{1}{nh}\sum y_i k(\frac{x_i - x}{h})|x_i]\} \tag{759}$$
$$= E\{Var[\frac{1}{nh}\sum y_i k(\frac{x_i - x}{h})|x_i]\} \tag{760}$$
$$= E\{\frac{1}{n^2h^2} n \underbrace{Var(y_i|x_i)}_{=\sigma^2} k^2(\frac{x_i - x}{h}) + \underbrace{Cov}_{=0}\} \tag{761}$$
$$= \frac{\sigma^2}{nh^2} Ek^2(\frac{x_i - x}{h}) = \frac{\sigma^2}{nh} f \int k^2(v)dv \tag{762}$$
$$= \frac{\sigma^2}{nh} fK + o(\frac{1}{nh}) \tag{763}$$

$$B = Var\{E[\frac{1}{nh}\sum y_i k(\frac{x_i - x}{h})|x_i]\} \tag{764}$$
$$= Var\{\frac{1}{nh}\sum g(x_i) k(\frac{x_i - x}{h})\} \tag{765}$$
$$= \frac{1}{n^{\cancel{2}}h^2} \cancel{n} Var(g(x_i)k(\frac{x_i - x}{h})) \tag{766}$$
$$= \frac{1}{nh^2}\{Eg^2k^2(\frac{x_i - x}{h}) - [\underbrace{Egk(\frac{x_i - x}{h}}_{o_p(1)})]^2\} \tag{767}$$
$$= \frac{1}{nh^2} g^2(f + o(\frac{1}{nh})) \int k^2(v)dv \tag{768}$$
$$= \frac{1}{nh} g^2 fK + o(\frac{1}{nh}) \tag{769}$$

$$Var(\hat{m}(x)) = \frac{1}{nh}[\sigma^2 f + g^2 f]K + o(\frac{1}{nh}) \tag{770}$$
$$Cov(\hat{m}, \hat{f}) = E\{(\hat{m} - E\hat{m})(\hat{f} - E\hat{f})\} \tag{771}$$
$$= E\{E(\hat{m}|x_i)(\hat{f} - E\hat{f})\} \tag{772}$$
$$= E\{\frac{1}{nh}\sum g(x_i) k(\frac{x_i - x}{h})\} \tag{773}$$
$$= E\{\frac{1}{nh}\sum g(x_i) k(\frac{x_i - x}{h})[\frac{1}{nh}]\} \tag{774}$$
$$= \frac{1}{nh^2}\{Eg(x_i)k^2(\frac{x_i - x}{h}) - Eg(x_i)k(\frac{x_i - x}{h})Ek(\frac{x_i - x}{h})\} \tag{775}$$
$$= \frac{1}{nh} gfK + o(\frac{1}{nh}) \tag{776}$$

$$E\hat{g}(x) = \frac{E\hat{m}}{E\hat{f}} - \underbrace{\frac{Cov(\hat{m}, \hat{f})}{(Ef)^2}}_{O(\frac{1}{nh})} + \underbrace{\frac{E\hat{m}Var\hat{f}}{(E\hat{f})^3}}_{O(\frac{1}{nh})} \tag{777}$$
$$\approx \frac{\{gh + \frac{h^2}{2}[g'' + gf'' + 2g'f'] \int v^2 k(v)dv\}}{f + \frac{h^2 f''}{2} \int v^2 k(v)dv} \tag{778}$$
$$= \frac{\{\}}{f[1 + \frac{h^2 f''}{2f} K_2]} \tag{779}$$
$$= \frac{\{\}}{f}[1 - (\frac{h^2 f''}{2f} K_2) + ()^2 - ()^3 + \dots] \tag{780}$$
$$= g + \frac{h^2}{2f}[g''f + gf'' + 2g'f']K^2 - \frac{h^2 g f''}{2f} K_2 + o(h^2) \tag{781}$$
$$= g + \frac{h^2}{2f}[g''f + 2g'f']K_2 + o(h^2) \tag{782}$$

$$Var\hat{g} = E[\hat{g} - E\hat{g}]^2 \tag{783}$$

$$\hat{g} - E\hat{g} \approx \frac{\hat{m} - E\hat{m}}{E\hat{f}} - \frac{E\hat{m}(\hat{f} - E\hat{f})}{(E\hat{f})^2} \tag{784}$$

$$Var\hat{g} = \frac{Var(\hat{m})}{(E\hat{f})^2} + \frac{(E\hat{m})^2 Var\hat{f}}{(E\hat{f})^4} - \frac{2E\hat{m}Cov(\hat{m}, \hat{f})}{(E\hat{f})^3} \stackrel{def}{=} A + B - C \tag{785}$$

$$Var\hat{g}(x) = A + B - C \tag{786}$$

$$= \frac{1}{nhf}[\sigma^2 + g^2 + g^2 - 2g^2] \int k^2(v)dv + o(\frac{1}{nh}) \tag{787}$$

$$= \frac{\sigma^2}{nhf} \int k^2(v)dv + o(\frac{1}{nh}) \tag{788}$$

$$MSE\hat{g} = \frac{h^4}{4f^2(x)}[g''f + 2g'f']^2 K_2^2 + \frac{\sigma^2 K}{nhf} + o(h^4 + \frac{1}{nh}) \tag{789}$$

$$h^4 \propto \frac{1}{nh} \Rightarrow h \propto n^{-\frac{1}{5}} \tag{790}$$

$$MSE\hat{g} \propto n^{-\frac{4}{5}} \tag{791}$$

$$\hat{f} = f + o_p(n^{-\frac{2}{5}}) \tag{792}$$

$$\hat{g} = g + o_p(n^{-\frac{2}{5}}) \tag{793}$$

多维 Kernel

$$\hat{f}(x_1, x_2, \ldots, x_q) = \frac{1}{nh_1, h_2, \ldots, h_q} \sum_{i=1}^{n} k(\frac{x_{i1} - x_1}{h_1} \frac{x_{i2} - x_2}{h_2} \ldots \frac{x_{iq} - x_q}{h_q}) \tag{794}$$

2nd,4th,6th,...kernel 有区别，假设 2nd order kernel

$$E\hat{f} = \frac{1}{h_1 \ldots h_q} E \frac{x_{i1} - x_1}{h_1} \ldots \frac{x_{iq} - x_q}{h_q} \tag{795}$$

$$\int \frac{x_{i1} - x_1}{h_1} \ldots \frac{x_{iq} - x_q}{h_q} f(x_{i1}, \ldots, x_{iq}) dx_{i1}, \ldots, dx_{iq} \tag{796}$$

$$= \int k(v_1) \ldots k(v_q) f(x_1 + h_1 v_1, \ldots, x_q + h_q v_q) h_1 \ldots h_q dv_1 \ldots dv_q \tag{797}$$

$$= \int k(v_1) \ldots k(v_q) \{f(x_1, \ldots, x_q) + \sum_{s=1}^{q} h_s v_s f_s + \frac{1}{2} \sum_s \sum_t h_s h_t v_s v_t f_{st} + (s.o.)\} h_1 \ldots h_q dv_1 \ldots dv_q \tag{798}$$

$$= h_1 \ldots h_q f(x_1, \ldots, x_q) + h_1 \ldots h_q \frac{\sum_s h_s^2}{2} f_{ss} \int v_s^2 k(vs) ds + (s.o.) \tag{799}$$

$$E\hat{f} = f + \frac{\sum_s h_s^2 f_{ss}}{2} K_2 + (s.o.) \tag{800}$$

$$Var\hat{f} = \frac{1}{nh_1^2 \ldots h_q^2} \{Ek^2(\frac{x_{i1} - x_1}{h_1}) \ldots k^2(\frac{x_{iq} - x_q}{h_q})[Ek(\frac{x_{i1} - x_1}{h_1}) \ldots k(\frac{x_{iq} - x_q}{h_q})]\} \tag{801}$$

$$= \frac{1}{nh_1 \ldots h_q} f(x_1, \ldots, x_q)(\int k^2(v_i)dv)^q + (s.o.) \tag{802}$$

$$E\hat{f}, Var\hat{f} \tag{803}$$

$$\sqrt{nh}(\hat{f} - f - \frac{h^2}{2} f'' K_2) \sim N(0, fK) \tag{804}$$