

Econometrics

数学回顾

1. 形如 $\mathbf{x}^T [\text{Var}(\mathbf{x})]^{-1} \mathbf{x}$ 的二次型：直观含义是以 $[\text{Var}(\mathbf{x})]^{-1}$ 为权重，将 \mathbf{x} 到零向量 $\mathbf{0}$ (原点) 的距离标准化。一维情形下，化为 $(\frac{x}{\sqrt{\text{Var}(x)}})^2$ ，即 x 离原点有多少个标准差距离（马氏距离）的平方。
2. 夹心估计量， $\text{Var}(AX) = A\text{Var}(X)A^T$ 。其中 A 为对称矩阵
3. $\text{Var}(X) = E(XX^T) - E(X)[E(X)]^T$
4. 正态分布峰度为3，峰度越大越厚尾
5. 期望迭代

$$E_X[E(Y|X)] = \int_{-\infty}^{+\infty} f(x)E(Y|X)dx \quad (1)$$

$$= \int_{-\infty}^{+\infty} f(x) \int_{-\infty}^{+\infty} y \frac{f(x,y)}{f(x)} dy dx \quad (2)$$

$$= \int_{-\infty}^{+\infty} y dy \int_{-\infty}^{+\infty} f(x,y) dx \quad (3)$$

$$= \int_{-\infty}^{+\infty} y f(y) dy \quad (4)$$

$$= EY \quad (5)$$

6. 相互独立 \Rightarrow 均值独立 \Rightarrow 线性不相关。且均值独立不是对两个变量对称的关系
7. 多维正态分布

1. 二维正态

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - \frac{2(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]\right\} \quad (6)$$

2. 多维矩阵表达

$$f(x_1, \dots, x_n) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(X-\mu)^T \Sigma^{-1} (X-\mu)\right\} \quad (7)$$

$$\text{其中, } \Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_2\sigma_1 & \sigma_2^2 \end{pmatrix}$$

$$|\Sigma| = (1-\rho^2)\sigma_1^2\sigma_2^2$$

$$\Sigma^{-1} = \begin{pmatrix} \frac{1}{(1-\rho^2)\sigma_1^2} & -\frac{\rho}{(1-\rho^2)\sigma_1\sigma_2} \\ -\frac{\rho}{(1-\rho^2)\sigma_2\sigma_1} & \frac{1}{(1-\rho^2)\sigma_2^2} \end{pmatrix}$$

小样本OLS

整体模型，线性假定 $\mathbf{y} = X\beta + \epsilon$ ，注意矩阵 X 的第一列元素 $x_{i1} = 1$ ，原因是常数项

最小化残差平方和，注意 ϵ 为误差， \mathbf{e} 为残差

$$\min_{\tilde{\beta}} SSR(\tilde{\beta}) = \sum_{i=1}^n e_i^2 = \mathbf{e}^T \mathbf{e} \quad (8)$$

$$= (\mathbf{y} - X\tilde{\beta})^T (\mathbf{y} - X\tilde{\beta}) \quad (9)$$

$$= \mathbf{y}^T \mathbf{y} - \mathbf{y}^T X\tilde{\beta} - \tilde{\beta}^T X^T \mathbf{y} + \tilde{\beta}^T X^T X\tilde{\beta} \quad (10)$$

$$= \mathbf{y}^T \mathbf{y} - 2\mathbf{y}^T X\tilde{\beta} + \tilde{\beta}^T X^T X\tilde{\beta} \quad (11)$$

其中， \mathbf{y} 为 $n \times 1$ 向量， X 为 $n \times K$ 矩阵， β 为 $K \times 1$ 向量 可以看到各项均为内积形式的 1×1 常数

一些基本矩阵微分规则

这里都是按列向量求导的梯度微分 ∇

$$\text{内积微分, } \frac{\partial(\alpha^T \beta)}{\partial \beta} = \alpha \quad (12)$$

$$\text{二次型微分, } \frac{\partial(\beta^T A \beta)}{\partial \beta} = 2A\beta, \text{ 其中 } A \text{ 为对称矩阵}$$

使用上面的微分规则，对残差平方和目标函数求一阶导

$$\frac{\partial(SSR)}{\partial \tilde{\beta}} = -2X^T \mathbf{y} + 2X^T X\tilde{\beta} = 0 \quad (13)$$

从而有正规方程组：

$$(X^T X)_{K \times K} \tilde{\beta}_{K \times 1} = X_{K \times n}^T \mathbf{y}_{n \times 1} \quad (14)$$

$$\text{移项有 } X^T \mathbf{y} - (X^T X)\tilde{\beta} = \mathbf{0} \quad (15)$$

$$\text{提公因式, } X^T (\mathbf{y} - X\tilde{\beta}) = \mathbf{0}$$

$$\text{亦即 } X^T \mathbf{e} = \mathbf{0}, \text{ 残差和解释变量正交}$$

因 $x_{i1} = 1$, 转置后 X^T 第一行元素全为 1，从而上式也意味着 $\sum_{i=1}^n e_i = 0$

最终得到

$$\tilde{\beta} = (X^T X)^{-1} X^T \mathbf{y} \quad (16)$$

$$\hat{\mathbf{y}} = X\tilde{\beta} = X(X^T X)^{-1} X^T \mathbf{y} = \mathbf{P}\mathbf{y}$$

其中 $\mathbf{P} = X(X^T X)^{-1} X^T$ 称为投影矩阵，用 \mathbf{P} 左乘任何向量就得该向量在超平面 X 上的投影 另一方面 $\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{P}\mathbf{y} = (\mathbf{I}_n - \mathbf{P})\mathbf{y} = \mathbf{M}\mathbf{y}$ 其中 $\mathbf{M} = \mathbf{I}_n - \mathbf{P}$ 为消灭矩阵，用它左乘任何向量就是该向量对超平面 X 投影后的残差向量

平方和分解公式

离差平方和 Total Sum of Squares (TSS)

$$\sum_{i=1}^n (y_i - \bar{y})^2 \quad (17)$$

Explained Sum of Squares (ESS)

$$\sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \quad (18)$$

残差平方和 Residual Sum of Squares (RSS)

$$\sum_{i=1}^n e_i^2 \quad (19)$$

分解有

$$TSS = ESS + RSS \quad (20)$$

拟合优度

$$R^2 = \frac{ESS}{TSS} \quad (21)$$

校正拟合优度

$$\bar{R}^2 = 1 - \frac{\sum_{i=1}^n e_i^2 / (n - K)}{\sum_{i=1}^n (y_i - \bar{y})^2 / (n - 1)} \quad (22)$$

通过自由度来惩罚模型复杂度

古典线性回归模型关键假定

1. 线性假定，真实DGP线性于参数
2. 严外生性 $E(\varepsilon_i | X) = 0$ ，即扰动项和所有解释变量均不相关， $Cov(\varepsilon_i, x_{jk}) = 0, \forall j, k$ 。用期望迭代即可有 $E(\varepsilon_i) = 0$
3. 不存在严格多重共线性，亦即 X 满列秩。进而 $X^T X$ 正定、可逆，因为 $r(X^T X) = r(X)$ ，对称且满秩（这里 X 不一定是方阵，所以谈不上 X 可逆，实际上不能用分解证明其正定）。此性质是回归能够算得出来的必要条件
4. 球形扰动：同方差、无自相关，亦即 $Var(\varepsilon | X) = \sigma^2 \mathbf{I}_n$
5. 扰动正态： $\varepsilon | X \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n)$

可以这样理解这些古典假定：手上有了 \mathbf{y} 和 X ，希望把它们投到超平面上 $\hat{\mathbf{y}} = X\hat{\beta} + \mathbf{e}$

首先要让这个超平面存在，不只是像 SVM 那样插入一块作为分隔。所以要有线性假定，真实 DGP 线性于参数，然后希望有如下的合意性质。

合意性质

1. 线性性， $\hat{\beta}$ 可以视为 \mathbf{y} 的线性组合。由线性假定所得
2. 无偏性。由严外生性所得
3. 协方差矩阵形式简洁， $Var(\hat{\beta}|X) = (X^T X)^{-1} X^T Var(\varepsilon|X) X (X^T X)^{-1} = \sigma^2 (X^T X)^{-1}$ 。由球形扰动所得
4. BLUE 中的方差最小。同样有球形扰动所得
5. 假设检验和统计推断。由扰动正态支持

假设检验/分类标准的术语

			标准集	
		正样本	负样本	合计
	正样本	TP	FP	P
测试集	负样本	FN	TN	N
	合计	T	F	$P + N = T + F$

1. Precision 查准率 = TP/P
2. Recall 召回率（真阳性率/敏感性）= TP/T
3. 假阳性率/特异性 = FP/F
4. Accuracy 准确率 = $(TP + TN)/(T + F)$
5. F 分数 = $\frac{2}{\frac{1}{Recall} + \frac{1}{Precision}}$ ，即查准率和召回率的调和平均数
6. ROC Curve：假阳性率为横轴，真阳性率为纵轴
7. PR Curve：Recall 为横轴，Precision 为纵轴
8. 显著性水平 = $FN/T = \alpha$
9. Power 检验的势 = $TN/F = 1 - \frac{FP}{F} = 1 - \beta$

大样本 OLS

小样本需要严外生和正态扰动的假设太强；即便有假设成立都还需要推导精确分布。所以放松要求，换用大样本（ $n \geq 30$ ）

随机收敛

依均方收敛 $\xrightarrow{\text{Chebyshev不等式}}$ 依概率收敛 \Rightarrow 依分布收敛

Khinchin 大数定律，由依均方收敛推出的依概率收敛

Levy-Lindberg 中心极限定理，依分布收敛

随机过程

严平稳：要求相对距离下有限维分布不变，而不取决于实际的绝对距离，自然也就有单个同分布。但不
对跨期是否存在序列相关做要求。此性质主要用来代替同分布的要求

弱平稳：相比严平稳不要求同分布，只要求到二阶矩（期望、方差、协方差）相同即可

渐进独立性/遍历性/弱相依/各态历经性：允许存在序列相关，只要这种相关关系在极限处消失即可。只
要相距够远，就可以视为近似独立（严格来说是不相关）。主要用来代替独立的要求

渐近独立定理

渐近独立的严平稳过程的样本均值是总体均值的一致估计。

配合连续函数变换，也就推广了矩法估计。只要是渐近独立严平稳，样本矩都是总体矩的一致估计

仅要求：

- 严平稳
- 渐近独立

也就是说解除了独立的要求，但还是要求同分布

推广了：

- 大数定律
- 中心极限定理
- 矩估计法

大样本假定

1. 线性假定，和小样本一样
2. $(K+1)$ 维随机过程 $y_i, x_{i1}, \dots, x_{iK}$ 为渐近独立严平稳过程，从而适用大数定律和中心极限定理
3. 前定解释变量，同期外生。 $E(x_{ik}\varepsilon_i) = 0, \forall i, k$ ，所谓“前定”说的是仿佛在扰动项产生前解释变量
就已产生，只是要求同期外生，不是说跨期也要不相关。放松了严外生
4. 秩条件，不存在严格多重共线性。和小样本一样
5. 不再要求球形扰动，大样本下渐近方差支持异方差稳健（自相关要另用其他办法）
6. 不再要求扰动正态，因为已有渐近正态性质

合意性质

1. 一致性，由前定解释变量（同期外生）所得
2. 渐近正态性，由渐近独立严平稳所得
3. 统计推断使用标准正态分布代替 t 分布， χ^2 分布代替 F 分布，由渐进正态推得

时间序列典例

AR (1)

$$y_t = \rho y_{t-1} + \varepsilon_t \quad (23)$$

其中 $\{\varepsilon_t\}$ 为 i.i.d, 且 $cov(y_{t-1}, \varepsilon_t) = 0$

- 当 $\rho = 1$ 时， $\{y_t\}$ 不平稳，随机游走
- 当 $\rho < 1$ 时， $\{y_t\}$ 严平稳

白噪音

对于弱平稳过程 $\{x_t\}$ ，如果对 $\forall t$ 都有 $E(x_t) = 0$ 且 $cov(x_t, x_{t+k}) = 0, (\forall k \neq 0)$ 则为白噪音。直观即每一期期望都保持为零，不同期序列不相关，且各期直到二阶矩都保持相同（当然也就有方差不变 $var(x_t) = \sigma^2$ ）

异方差

1. 球形扰动的主对角线元素不完全相等
2. 后果
 1. OLS仍然无偏、一致且渐进正态
 2. 估计量方差不再是同方差假设下的，t检验、F检验失效
 3. 高斯马尔科夫定理不再成立
3. 检验手段
 1. BP 检验
 2. White 检验
4. 处理方法
 1. OLS + 稳健标准误
 2. 加权最小二乘法 (WLS)
 3. 可行加权最小二乘法 (FWLS)

自相关

1. 球形扰动的非主对角线元素不全为0
2. 后果
 1. OLS仍然无偏、一致且渐进正态
 2. 估计量方差不再是同方差假设下的，t检验、F检验失效
 3. 高斯马尔科夫定理不再成立
3. 检验手段
 1. BG检验
 2. Q检验
 3. DW检验
4. 处理方法
 1. OLS+异方差自相关稳健标准误
 2. 准差分法
 3. 广义最小二乘法（GLS）
 4. 修改模型设定

模型设定和数据问题

遗漏变量

1. 后果
 1. 如果被遗漏变量和解释变量不相关，那么扰动项还是和解释变量不相关。从而OLS没什么问题，只是扰动项方差增大了，准确性有所降低
 2. 如果被遗漏变量和解释变量相关，那么扰动项也和解释变量相关。大样本下OLS不一致
2. 处理方法
 1. 加入尽可能多的控制变量
 2. 随机试验和自然实验
 3. 工具变量法
 4. 使用面板数据

加入无关变量

1. 后果：OLS没什么问题，但是解释变量系数 $\hat{\beta}$ 的方差一般会增大
2. 处理方法：别乱加变量，最好遵循理论指导。但这里就有建模策略是从小到大还是从大到小的问题，结合ML有lasso选择变量方法

解释变量个数的选择

1. 奥卡姆剃刀原则
2. 可用的权衡准则
 1. 校正可决系数 \bar{R}^2
 2. 赤池 (Akaike) 信息准则
 3. 贝叶斯信息准则
 4. 从大到小的序贯t规则

对函数形式（模型设定）的检验

1. Ramsey的RESET检验

（非严格）多重共线性

1. 后果：
 1. 高斯马尔科夫定理仍然成立，OLS仍然BLUE。但这个BLUE只是说线性无偏估计量里方差最小，可能其他估计量可以更小
 2. 矩阵 $(X^T X)$ 变得“几乎”不可逆， $(X^T X)^{-1}$ 变得很大，使得方差 $Var(\hat{\beta}|X) = \sigma^2 (X^T X)^{-1}$ 增大，系数估计变得不准确。 X 中的元素轻微的变化都会引起很大的变化
 3. 典型症状是整个方程的 R^2 较大， F 检验也显著，但是单个系数的 t 检验却不显著。而且增减解释变量会使原有系数估计值产生较大变化（增加的变量和原来的变量多重共线性）。直观上就是无法区分高度相关的解释变量对被解释变量的单独影响力
2. 检验手段：
 1. R_k^2 与方差膨胀因子 VIF
3. 处理方法：
 1. 如果不关心具体的回归系数，只关心整个方程的预测能力（ML倾向），那大可不必理会多重共线性
 2. 如果重点讨论的变量已经显著，同样可以不理睬。因为方差膨胀下都显著，那么不膨胀（没有多重共线性）就会更显著
 3. 如果重点讨论变量不显著，就要尝试处理：
 1. 加大样本容量
 2. 剔除多重共线性变量
 3. 将变量标准化，减均值除标准差

极端数据离群值

1. 检验手段：遍历去除第 i 个观测值，比较回归系数。计算杠杆作用 lev
2. 处理方法：截尾、缩尾

虚拟变量

1. 注意虚拟变量陷阱：不能放入所有类别，否则将出现严格多重共线性

结构变动检验

1. 结构变动日期已知：邹检验。全样本、前后总共分三次回归
2. 结构变动日期未知：匡特似然比（QLR）

缺失数据与线性插值

变量单位选择

1. 变量间数量级不要过于悬殊，影响到 $(X^T X)^{-1}$ 的计算

工具变量法

1. 内生性：
 1. 解释变量和扰动项相关。违背前定变量或者叫同期外生的假设
 2. 来源可以是
 1. 遗漏变量
 2. 联立方程双向因果
 3. 测量误差
2. 工具变量条件：
 1. 相关性：与内生解释变量相关
 2. 外生性：与扰动项不相关
3. 二阶段最小二乘法（2SLS/TSLS）：
 1. 第一阶段，内生解释变量对工具变量回归
 2. 第二阶段，被解释变量对第一阶段回归的拟合值再做回归
 3. 阶条件：工具变量个数不少于内生解释变量个数
 1. 不可识别：工具变量个数少于内生解释变量个数
 2. 恰足识别：工具变量个数等于内生解释变量个数

3. 过度识别：工具变量个数大于内生解释变量个数
4. 弱工具变量
5. 过度识别检验：Sargan统计量
6. 内生性豪斯曼检验：Hausman test

二值选择模型

1. logit
2. probit

面板数据

1. 固定效应
2. 随机效应

平稳时间序列

1. 自回归
2. ARMA
3. VAR

单位根与协整

- 1.

具体的统计量

直观上，如果未知参数的估计值离预期假想值比较“远”，则应倾向于拒绝原假设（ H_0 ：未知参数等于预期假想值）。使用此原理的这类统计检验成为沃尔德检验（Wald test）

t 统计量

小样本

满足全部五个小样本 OLS 古典假定，并且原假设 $H_0 : \beta_k = c$ 也成立的情况下

$$t_k = \frac{\hat{\beta}_k - c}{SE(\hat{\beta}_k)} \sim t(n - K) \quad (24)$$

一般的通用公式为

$$t \equiv \frac{\text{估计量} - \text{假想值}}{\text{估计量的标准误}} \quad (25)$$

大样本

$$t_k = \frac{\hat{\beta}_k - c}{SE_1(\hat{\beta}_k)} \xrightarrow{d} N(0, 1) \quad (26)$$

$$\text{其中 } SE_1(\hat{\beta}_k) = \sqrt{\frac{1}{n} Avar(\hat{\beta}_k)}$$

服从正态分布而不再是 t 分布

F 统计量

小样本

原假设为检验多个线性约束条件，矩阵表示为 $H_0 = \mathbf{R}_{m \times K} \beta_{K \times 1} = \mathbf{r}_{m \times 1}$ ；其中 \mathbf{R} 要满行秩，也就是 $rank(\mathbf{R}) = m$ ，待检验线性约束条件不能有多余或者自相矛盾

根据 Wald test 原理，由于 $\hat{\beta}$ 是 β 的估计量，故如果 H_0 成立，那么 $\mathbf{R}\hat{\beta} - \mathbf{r}$ 应该比较接近 $\mathbf{0}$ 。这种接近程度用其二次型来衡量，比如说

$$(\mathbf{R}\hat{\beta} - \mathbf{r})^T [Var((\mathbf{R}\hat{\beta} - \mathbf{r}))]^{-1} (\mathbf{R}\hat{\beta} - \mathbf{r}) \quad (27)$$

其中，

$$Var(\mathbf{R}\hat{\beta} - \mathbf{r}) = Var(\mathbf{R}\hat{\beta}) \quad (28)$$

$$= \mathbf{R} Var(\hat{\beta}) \mathbf{R}^T \quad (\text{夹心估计量}) \quad (29)$$

$$= \sigma^2 \mathbf{R} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{R}^T \quad (\text{因为 } Var(\hat{\beta}) = \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1}) \quad (30)$$

从而在满足全部五个小样本 OLS 古典假定，并且原假设 $H_0 = \mathbf{R}\beta = \mathbf{r}$ 的情况下，有

$$F \equiv \frac{(\mathbf{R}\hat{\beta} - \mathbf{r})^T [\mathbf{R} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{R}^T]^{-1} (\mathbf{R}\hat{\beta} - \mathbf{r}) / m}{s^2} \sim F(m, n - K) \quad (31)$$

另外，在做假设检验时，如果接受原假设，则可将原假设作为约束条件，代入最小二乘法的最优化问题。

从而 F 检验可以变为

$$\begin{aligned} \min_{\hat{\beta}} \quad & SSR(\hat{\beta}) \\ \text{s.t.} \quad & \mathbf{R}\hat{\beta} = \mathbf{r} \end{aligned} \quad (32)$$

如果原假设 $H_0 : \mathbf{R}\beta = \mathbf{r}$ 正确，那么加上约束后不应使得残差平方和增大很多

记无约束回归的残差平方和为 SSR_0 ，而有约束回归的残差平方和为 SSR_1 。这意味着，在 H_0 正确的情况下， $(SSR_0 - SSR_1)$ 不应很大。从而构造如下的 F 统计量

$$F = \frac{(SSR_1 - SSR_0)/m}{SSR_0/(n - K)} \quad (33)$$

$$= \frac{(R_0^2 - R_1^2)/m}{(1 - R_0^2)/(n - K)} \quad (34)$$

其中用了 $R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$ 来替换，注意第二个等号后的 R^2 都加了负号

大样本

假设统计量 $F \sim F(m, n)$ ，则当 $n \rightarrow \infty$ 时， $mF \xrightarrow{d} \chi^2(m)$