时间序列分析笔记

卓功亦

2025 年 10 月 1 日

前一半的课程与计量经济学有着密切的联系，后一半课程则主要是时间序列分析的内容。

1 多元线性回归模型

我们已经很熟悉二元线性回归模型了，即通过最小二乘法拟合直线，但这种模型在多元情况下表现如何，最小二乘法本身又有什么特殊的性质，这是本节研究的内容。

1.1 模型的建立与求解

具体来说，我们研究的是这样的问题：给定一个被解释变量 y，和若干个解释变量 $x_1, \cdots, x_k$，希望得到尽可能好的预测和拟合，多元线性回归模型如下所示：

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \epsilon$$

其中 $\beta_0, \cdots \beta_k$ 为参数，ϵ 为扰动项（即被认为与解释变量无关）

模型给出四条假设：

- $Y = X\beta + \epsilon$
- $\mathbb{E}(\epsilon \mid X) = 0$（外生性）
- $\mathrm{var}(\epsilon \mid X) = \sigma^2 I$（球面方差）

- X 以概率 1 列满秩

注 1.1.1. 这里 2、3 的结果都是向量/矩阵，计算方法与一一计算相同，第三点中

$$\operatorname{var}(\epsilon \mid X) = \mathbb{E}(\epsilon\epsilon^T \mid X) - \mathbb{E}(\epsilon \mid X)\mathbb{E}(\epsilon \mid X)^T = \mathbb{E}(\epsilon\epsilon^T \mid X)$$

我们的目标就是在已知一些 $(x_1, x_2, \cdots, x_k)$ 及其对应的 y 以后，找到最好的对参数 $\beta_0, \cdots, \beta_k$ 的估计。

注 1.1.2. 我们在这里引入经济学上的两个概念：考虑 $\frac{\partial y}{\partial x_i}$，这即是在其他参数不变的情况下 y 关于 x_i 的变化率，在经济学中称为**边际**。而若考虑 $\frac{\partial(\ln y)}{\partial(\ln x_i)} = \frac{\frac{\partial y}{y}}{\frac{\partial x_i}{x_i}}$ 为两者变化幅度的比值，在经济学中称为**弹性**。

在上述模型中，β_i 即为 x_i 的边际数据，而若将模型修改为

$$\ln y = \beta_0 + \beta_1 \ln x_1 + \cdots + \beta_k \ln x_k + \epsilon$$

则此时 β_i 即为 x_i 的弹性数据。

在估计过程中，我们面对的是 m 数据 $x_i = (x_{i1}, \cdots, x_{ik})(1 \leq i \leq m)$，若干个行向量拼在一起，再在左侧添上一个全为 1 的列向量得到矩阵 X，再记 $\beta = (\beta_0, \cdots, \beta_k)^T, Y = (y_1, \cdots, y_m)^T$，则应该有 $Y = X\beta + \epsilon$ 这里 ϵ 为列向量。

注 1.1.3. 这是频率派的观点，即 β 是未知确定的，我们的目标是对这个 β 做恰当的估计

我们假设得到的估计结果是 b，一个重要的工作是讨论哪些估计更加“接近”真实值。一个经典的想法是考虑真实值和预测值的差，但考虑到其有正负差异，我们使用平方来作为判断标准。具体来说，令 $e = Y - \hat{Y}$ 为这一估计的**残差**，其中 $\hat{Y}$ 为估计值，我们用 e^Te 来评估估计值与真实值间的距离，我们接下来的目标就是找到使 e^Te 最小的 b

而注意到 $\hat{Y} = Xb$，故我们有

$$L = e^Te = (Y - Xb)^T(Y - Xb) = Y^TY - Y^TXb - b^TX^TY + b^TX^TXb$$

为取使该表达式最小的 b，我们令其关于 b 偏导 $\frac{\partial L}{\partial b}$ 为 0：

$$0 - X^TY - X^TY + 2X^TXb = 0 \Rightarrow b = (X^TX)^{-1}X^TY$$

这说明该式在 $b=(X^TX)^{-1}X^TY$ 时取极值，再计算二阶偏导的 Hessiam 矩阵

$$\frac{\partial^2 L}{\partial b \partial b^T} = 2X^TX$$

该矩阵半正定，故 b 在该点取最小值。这样我们就求解了该模型，但该模型得到的结果 b 效果如何，我们在下一节中做分析

1.2 模型的评估

首先考虑 $\mathbb{E}(b \mid X)$，我们显然有

$$\mathbb{E}(b \mid X) = \mathbb{E}((X^TX)^{-1}X^TY \mid X) = (X^TX)^{-1}X^T\mathbb{E}(X\beta + \epsilon \mid X) = \beta$$

这说明多元线性回归模型的估计是无偏估计。

在都是无偏估计的情况下，若要比较两种估计的优劣，重要的方式便是研究它们的条件方差，这代表了它们估计的稳定度。事实上，注意到

$$b = \beta + (X^TX)^{-1}X\epsilon$$

我们有

$$\begin{aligned}
\mathrm{var}(b \mid X) &= \mathrm{var}((X^TX)^{-1}X\epsilon \mid X) \\
&= \mathbb{E}((X^TX)^{-1}X\epsilon\epsilon^TX^T(X^TX)^{-1} \mid X) \\
&= (X^TX)^{-1}X\mathbb{E}(\epsilon\epsilon^T \mid X)X^T(X^TX)^{-1} \\
&= (X^TX)^{-1}X\sigma^2IX^T(X^TX)^{-1} \\
&= \sigma^2(X^TX)^{-1}
\end{aligned}$$

而考虑线性估计 $\tilde{b} = CY$，如果它是无偏的，我们有 $\mathbb{E}(\tilde{b} \mid X) = \beta$，即

$$\mathbb{E}(C(X\beta + \epsilon) \mid X) = \mathbb{E}(CX\beta \mid X) = \beta$$

故 $CX = I, \tilde{b} = \beta + C\epsilon$ 还是和之前一样，我们来计算 $\mathrm{var}(\tilde{b} \mid X)$

$$\begin{aligned}
\mathrm{var}(\tilde{b} \mid X) &= \mathrm{var}(C\epsilon \mid X) \\
&= \mathbb{E}(C\epsilon\epsilon^TC^T \mid X) \\
&= \sigma^2CC^T
\end{aligned}$$

注 1.2.1. 注意到 C 是由 X 给出的估计，故可以直接在条件概率中被提出

我们要比较 $\mathrm{var}(\tilde{b} \mid X)$ 和 $\mathrm{var}(b \mid X)$，只需将其做差考虑矩阵的正定性即可。令 $D = (C - (X^TX)^{-1}X^T)$，则

$$\begin{aligned}\mathrm{var}(\tilde{b} \mid X) - \mathrm{var}(b \mid X) &= \sigma^2(CC^T - (X^TX)^{-1}) \\ &= \sigma^2((D + (X^TX)^{-1}X^T)(D^T + X(X^TX)^{-1}) - (X^TX)^{-1}) \\ &= \sigma^2(DD^T + (X^TX)^{-1}X^TD^T + DX(X^TX)^{-1}) \\ &= \sigma^2 DD^T\end{aligned}$$

倒数第二步来自于 $DX = 0$ 这一观察。这样我们发现它们的差为半正定矩阵，于是我们得到

$$\mathrm{var}(\tilde{b} \mid X) \succeq \mathrm{var}(b \mid X)$$

这就说明了**简单最小二乘法是所有线性无偏估计中有效性最好的**。

定义 1.2.1. 多元线性回归模型的非中心化拟合优度 $R^2_{uc} := \frac{\hat{Y}^T\hat{Y}}{Y^TY} = 1 - \frac{e^Te}{Y^TY}$

非中心化拟合优度在实际建模中应用不多

一般来说，在建立模型时要满足以下几点：

- 数据量大于参数的三倍
- 数值计算调整量纲使得数据量级一致，减小误差
- 对于宏观统计量如 GDP, 要做平减，即以某一固定的物价水平换算
- 回归参数应该符合理论

1.3 中心化

中心化是指有 1 作为解释变量，这会进一步导出一些良好的性质，首先是

$$X^Te = 0 \Rightarrow \sum e_i = 0$$

同时我们引入 $M_x := I - X(X^TX)^{-1}X^T$，易见 $e = M_xX$，M_x 是对角等幂阵，即

$$M_x^T = M_x M_x^2 = M_x$$

善用这一点可以使我们后面的计算变得简洁。接下来我们考察拟合优度，我们希望拟合优度考察的是变动项的拟合结果，因此自然要通过减去 $\overline{Y}$ 得到

定义 1.3.1. 中心化的拟合优度为 $\frac{(\hat{Y}-\overline{Y})^T(\hat{Y}-\overline{Y})}{(Y-\overline{Y})^T(Y-\overline{Y})}$

我们自然希望这一结果与之前有相同的形式，即

$$R^2 = 1 - \frac{e^Te}{(Y-\overline{Y})^T(Y-\overline{Y})}$$

$$Y - \overline{Y} = Y - \overline{y}(1, \cdots 1)^T = M_x \mid_{x=(1,\cdots,1)^T} y = M_0 y$$

$$(Y-\overline{Y})^T(Y-\overline{Y}) = y^TM_0y = \hat{y}^TM_0\hat{y} + e^TM_0e + e^TM_0\hat{y} + \hat{y}^TM_0e = (\hat{Y}-\overline{Y})^T(\hat{Y}-\overline{Y}) + e^Te$$

$$M_0e = e$$

离差 $y_i - \hat{y}$ 考察变动项

1.4 参数间的关系：偏回归和偏相关

我们来考察各个参数间的相互影响，将参数分为两组，数据分别组成两个矩阵 X_1, X_2

$$X = (X_1, X_2), Y \sim X$$

则我们同样将 b 分块，得到：

$$Y = \begin{pmatrix} X_1 & X_2 \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} + \epsilon$$

则在原来的回归计算中，我们有：$b = (X^TX)^{-1}X^TY$，故得到分块矩阵间的关系

$$X_1^TX_1b_1 + X_1^TX_2b_2 = X_1^TY$$

$$X_2^TX_1b_1 + X_2^TX_2b_2 = X_2^TY$$

第一式得到 $b_1 = (X_1^TX_1)^{-1}(X_1^TY - X_1^TX_2b_2)$ 代入第二式得到

$$X_2^TM_{X_1}X_2b_2 = X_2^TM_{X_1}Y$$

从而变形为

$$(M_{X_1}X_2)^T(M_{X_1}X_2)b_2 = (M_{X_1}X_2)M_{X_1}Y$$

不难发现这个形式就是原先回归形式中全部左乘 M_{X_1} 的结果，从某种意义上来说是一种相关性的体现。

我们接下来进一步考察加入一个变量对拟合精准度的影响：考虑模型 $y \sim X_1, z$，记 u 为 $y \sim (X_1, z)$ 残差，e 为 $y \sim X_1$ 残差 $e = M_{X_1}Y$，则我们希望考察 u^Tu, e^Te 间的关系。记 $(c^T, d^T)^T$ 为 $y \sim X_1, z$ 回归结果的分块，b 为 $y \sim X_1$ 回归结果，则自然有 $Y - X_1b = Y - X_1d - zc$

$$y^* = M_{X_1}Y, z^* = M_{X_1}Z$$

两者的简单相关系数称为 y, z 间的偏相关系数，记作 r^*_{yz}。则根据前文对偏回归的讨论我们有 $c = (z^{*T}z^*)^{-1}(z^{*T}y^*)$

$$\begin{aligned}
u &= Y - X_1d - zc \\
&= Y - X_1(b - (X_1^TX_1)^{-1}X_1^Tzc) - zc \\
&= Y - X_1b + X_1(X_1^TX_1)^{-1}X_1^Tzc - zc \\
&= e - c(I - X_1(X_1^TX_1)^{-1}X_1^T)z \\
&= e - cz^*
\end{aligned}$$

于是我们得到

$$u^Tu = e^Te - c^2(z^{*T}z^*)$$

从而引进新变量后模型拟合优度总是提高

1.5 模型参数准入

我们应该如何判断一个参数应不应该纳入？理想状况下自然是若有 $\beta_j = 0$ 则 x_j 不应该纳入模型中，但计算过程中不会出现正好为 0 的情况，于是我们采用假设检验的方法来验证。

$$b = \beta + (X^TX)^{-1}X^T\epsilon$$

引入关于方差的新假设：

$$\epsilon \mid X \sim N(0, \sigma^2 I)$$

于是自然有

$$b \mid X \sim N(\beta, \sigma^2(X^TX)^{-1})$$

作出原假设 $H_0 : \beta_j = 0$，我们希望把统计量凑成某种特定分布的形式以进行检验，我们引入以下几个分布

$$\chi^2(n) = \sum_{i=1}^{n} x_i^2 x_i \overset{i.i.d}{\sim} N(0,1)$$

$$t(k) = \frac{N(0,1)}{\sqrt{\chi^2(k)/k}}$$

$$F(m,n) = \frac{\chi^2(m)/m}{\chi^2(n)/n}$$

这些是在做统计检验时常用的分布，我们尝试配凑出这些分布从而进行检验，由之前的假设我们有

$$b_j \mid X \sim N(\beta_j, \sigma^2(X^TX)^{-1}_{j,j})$$

从而

$$u_j = \frac{b_j - \beta_j}{\sqrt{\sigma^2(X^TX)^{-1}_{j,j}}} \sim N(0,1)$$

但需要注意的是 σ^2 是真实方差，

无法得到，因此要通过其无偏估计 S^2 去拟合，而 $S^2 = \frac{e^Te}{n-k}$，故只需计算 e^Te 即可

$$e = M_X\epsilon \Rightarrow e^Te = \epsilon^T M_X \epsilon = \sigma^2\chi^2(Tr(M_X)) = \sigma^2\chi^2(n-k)$$

于是我们有

$$\frac{b_k - \beta_k}{\sqrt{S^2(X^TX)^{-1}_{k,k}}} \sim t(n-k)$$

代入显著性水平 $\alpha = 5\%$，即可做参数的显著性检验。

注 1.5.1. 需要注意的是，若不拒绝 $\beta_j = 0$ 但经验中有 $\beta_j \neq 0$，则 x_j 应当引入模型

1.6 参数线性约束

我们希望能够简化模型的计算，优化拟合效果，一个想法是找到 β 中的线性关系，从而通过线性变换减少变量数量，优化拟合结果。但这就要求我

们对其中的线性关系做显著性检验。我们来尝试对一般的线性关系 $R\beta = 0$ 做显著性检验（因为我们之前假设中 β 含有 1 作为解释变量，故所有的线性约束都是这种形式）

我们不妨 R 行满秩，$J = rank(R)$，基本的思路是对原来的变量在线性约束下做优化，计算带约束模型的拟合结果，如果得到的结果和原结果的差很接近 0，则可以认为它们存在线性关系。具体来说，我们要最小化 $(Y - Xb)^T(Y - Xb)$ 使得 $Rb = q$，我们利用拉格朗日乘子法来完成最小化，记

$$L(b, \lambda) = (Y - Xb)^T(Y - Xb) + 2\lambda^T(Rb - q)$$

则最小值点应有 $\frac{\partial L}{\partial \lambda} = 0, \frac{\partial L}{\partial b} = 0$，即

$$-2X^T(Y - Xb) + 2R^T\lambda = 0, 2(Rb - q) = 0$$

故得到

$$b = (X^TX)^{-1}X^TY - (X^TX)^{-1}R^T\lambda$$

代入第二式得到

$$\lambda = (R(X^TX)^{-1}R^T)^{-1}(R((X^TX)^{-1}X^TY) - q)$$

从而

$$b = \tilde{b} - (X^TX)^{-1}R^T(R(X^TX)^{-1}R^T)^{-1}(R\tilde{b} - q)$$

其中 $\tilde{b}$ 为不带约束的回归结果。

我们来计算 $e_*^Te_* - e^Te$

$$\begin{aligned} e_*^Te_* - e^Te &= (Y - Xb)^T(Y - Xb) - (Y - X\tilde{b})^T(Y - X\tilde{b}) \\ &= (X(b - \tilde{b}))^T(X(b - \tilde{b})) - 2e^T(X(b - \tilde{b})) \\ &= (R\tilde{b} - q)^T(R(X^TX)^{-1}R^T)^{-1}(R\tilde{b} - q) \end{aligned}$$

于是根据之前的讨论，我们有

$$(Rb - q)^T(R\sigma^2(X^TX)^{-1}R^T)^{-1}(Rb - q) \sim \chi^2(J)$$

其中 σ^2 为真实方差，替换为估计值 S^2 即有

$$(Rb - q)^T(S^2R(X^TX)^{-1}R^T)^{-1}(Rb - q)/J \sim F(J, n - k)$$

也即

$$\frac{e_*^T e_* - e^T e/J}{e^T e/n-k} \sim F(J, n-k)$$

于是我们通过统计量 F 来完成对参数线性约束的检验，这两个统计量叫做 Wald 统计量。

1.7 补充：非线性约束的检验

对于非线性约束 $f(\beta) = q$ 做检验，其中 $\frac{\partial f}{\partial \beta^T} = (\frac{\partial f(\beta)}{\beta})^T = J$ 行满秩。我们应该如何做？考虑跟前述 Wald 统计量一样的想法，即统计量

$$(f(b)-q)^T(\mathrm{var}(f(b)-q))^{-1}(f(b)-q) \sim \chi^2$$

将 $f(b)$ 展开得到

$$f(b) = f(\beta) + \frac{\partial f}{\partial \beta^T}(\beta - b) + \cdots$$

从而 $\mathrm{var}(f(b)-q) = J(\beta - b)$，从而转化为线性情况解决

1.8 预测与置信区间

在得到对 β 的估计 b 后，我们会代入新的 x_0 做预测，当 x_0 为观测值时预测值 $\hat{y_0} = x_0^T b$，此时

$$e_0 = x_0^T(-(X^T X)^{-1}\epsilon) + \epsilon_0$$

$$\mathrm{var}(e_0|X, x_0) = \sigma^2(1 + x_0^T(X^T X)^{-1}x_0)$$

预测方差可由 s^2 代替 σ^2 做估计，从而得到置信区间

$$\hat{y_0} \pm t_{\frac{\lambda}{2}}\mathrm{se}(\hat{y_0})$$

1.9 模型诊断

我们之前所做的种种计算都是在几条假设成立的情况下完成的，那如果我们的假设并不成立又会怎么样？模型诊断就是要检验并完善这一部分内容

1.9.1 完全共线性

完全共线性指的是在建立模型时错误地引入了等价的解释变量，比如以四个季节作为解释变量的同时还引入了 1，这样就会产生虚拟变量陷阱，进而影响模型。

1.9.2 近似共线性

与完全共线性不同，近似共线性是指在数据中有共线的情况出现，从而导致计算过程不稳定，即方程 $(X^TX)b = X^TY$ 求解不稳定。我们引入矩阵条件数

$$cond(A) := \parallel A \parallel \cdots \parallel A^{-1} \parallel$$

来描述这一点，对线性方程组 $AX = b$，当 $cond(A)$ 较大时该线性方程组不稳定，而 X^TX 为实对称阵，从而它的条件数小其实就是说它的特征值比较小。

矩阵条件数之外，我们也可以通过统计检验的方法更简便地判断，我们引入方差膨胀因子

$$VIF_j := \frac{1}{1 - R_j^2}$$

其中 R_j 为 x_j 关于余下解释变量回归得到的可决系数。认为 $VIF_j > 10$ 则数据有近似共线性问题。

为解决近似共线性这一问题我们向原损失函数中添加一正则项，即令

$$L(b) = (Y - Xb)^T(Y - Xb) + \lambda \parallel b \parallel^2$$

其中 λ 为给定常数，我们来最小化 $L(b)$，直接计算 $\frac{\partial L(b)}{\partial b} = 0$ 即得

$$b = (X^TX + \lambda I)^{-1}X^TY$$

实际上是将 X^TX 沿特征方向做缩放，从而解决近似共线性的问题。

注 1.9.1. 这个方法叫做岭（Ridge）回归

1.9.3 方差假设的违背

$$\mathrm{var}(\epsilon \mid X) = \sigma^2\Sigma \neq \sigma^2 I$$

这里会有两种不同的情况，对角阵但对角元素不完全相同（即独立但方差不同），对角元素相同但不是对角阵（即方差相同但存在自相关性），基本的思路是将其变为普通最小二乘能解决的问题。希望做线性变换满足假设，即希望取矩阵 G 使得

$$GY = GX\beta + G\epsilon, \mathrm{var}(G\epsilon \mid X) = \sigma^2 I$$

这个要求即 $G^T G = \sigma^{-1}$，我们记普通最小二乘得到估计结果为 b_{ols}，广义最小二乘得到的为 b_{gls}，于是上面的变换下我们给出

$$b_{gls} = (X^T G^T G X)^{-1}(X^T G^T G Y) = (X^T \Sigma^{-1} X)^{-1}(X^T \Sigma^{-1} Y)$$

但这一切建立在 Σ 已知的基础上，那在一个原始模型下，我们怎么知道 ϵ_i 的自相关性和异方差性呢？

对于一阶自相关性，我们引入统计量

$$DW := 2(1 - \frac{\sum e_i e_{i-1}}{\sum e_i^2}) \approx 2(1-\rho)$$

来对 ρ 进行估计。同样的，对异方差性我们也应该从数据出发进行检验和估计，一个基本的思路是从 e_i^2 和 $|e_i|$ 出发进行检验，我们引入被称为怀特检验的方法

对模型 $y \sim x$，我们可考虑

$$e^2 \sim 1, x, x \otimes x$$

对其中变量去重后得到的模型，这里 e^2 看作是 e_i^2 构成的列向量。如果有异方差，则非 1 的解释变量前系数不能同时为 0，用之前的 F 统计量检验显著性即可。特别地，当 n 过于大，即数据过多时，可以认为 $mF(m,n) \approx \chi^2(m)$，这有时可以大大减少计算量。

事实上，对异方差性的的检验不止这一种方法，另一种方法是将数据从小到大排序，去掉中间的一部分数据后两边分别回归，得到 χ^2 分布后通过做比得到 F 统计量检验，进而校验异方差性

1.9.4 外生性假设的违背

即考察 $\mathbb{E}(\epsilon|x) \neq 0$ 时的情况。事实上，对模型

$$y = X\beta + \epsilon$$

若 $\mathbb{E}(\epsilon|x)=0$，则自然 $\mathbb{E}(\epsilon)=0$，从而

$$cov(\epsilon, X)=cov(\mathbb{E}(\epsilon \mid X), X)=0 \Rightarrow \mathbb{E}(XY)=0$$

拆分开得到

$$\mathbb{E}(x_i y)=\mathbb{E}(\beta_1 x_1^2+\beta_2 x_1 x_2+\cdots+\beta_k x_1 x_k+\beta\epsilon)$$

从而通过矩估计得到

$$\frac{\sum_{i=1}^n x_{i1} y_i}{n}=\beta_1 \frac{\sum_{i=1}^n x_{i1}^2}{n}+\cdots+\beta_k \frac{\sum_{i=1}^n x_{i1} x_{ik}}{n}$$

将 n 个方程并在一起其实就得到了普通最小二乘的估计量，而若将前面乘的由 $x_1, \cdots, x_n$ 改为 Z，我们会得到矩估计量 $b_{IV}=(Z^T X)^{-1} Z^T Y$，我们现在在一个例子里面看这种方法在不满足外生性时的处理方法

例 1. 考虑模型

$$y \sim x_1, x_2, y=\beta_1 x_1+\beta_2 x_2+\epsilon$$

我们假设 x_1 满足外生性而 x_2 不满足，于是我们有

$$\mathbb{E}(x_1 \epsilon)=0 \Leftrightarrow cov(x_1, \epsilon)=0, \mathbb{E}(x_2 \epsilon) \neq 0 \Leftrightarrow cov(x_2, \epsilon) \neq 0$$

我们引入 x_2 的工具变量 z_2 满足

(1) $\mathbb{E}(z_2 \epsilon)=0$（与扰动项无关）

(2) $cov(z_2, \epsilon) \neq 0$（与原变量有关）

(3) (x_1, z_2) 列满秩（与其余变量无关）

此时用 z_2 替换 x_2，记 $Z=(X_1, Z_2)$ 为对应数据组成矩阵，对其用矩估计得到估计

$$b_{IV}=(Z^T X)^{-1} Z^T Y=\beta+(Z^T X)^{-1} Z^T \epsilon$$

于是自然有

$$\mathbb{E}(b_{IV} \mid X)=\beta$$

从而 b_{IV} 为 β 的无偏估计

1.10 模型的实际处理

1.10.1 嵌套模型与非嵌套模型

对两个不同的模型，我们应该选择哪个接受呢？如果两者解释变量存在包含关系，根据前面的讨论，我们只需要对多出来的变量做假设检验即可，但如果模型是非嵌套的呢？

例 2. 我们来看两个模型

$$y = x_1\beta + z_1\gamma_1 + \epsilon, y = x_1\beta_1 + z_2\gamma_2 + \epsilon$$

要判断是否要接受第二个模型，我们先做 $y \sim x_1, z_2$ 的最小二乘回归，得到拟合项 $\hat{y}$，再将拟合项代入新模型

$$y = (1-\alpha)(x_1\beta + z_1\gamma_1) + \alpha\hat{y} + \epsilon$$

做假设检验 $\alpha = 0$ 即可。

1.10.2 断点与 chow 预测检验

有时模型可能并不如我们所想拟合出来是一条直线，因此我们会想要判断两组数据回归是否一致，还是不一致从而会产生断点。一个基础的思路是通过约束数据进行 F 统计量假设检验。

但有时分组会导致一方数据量过小，从而无法回归，为此我们把长数据看作不带约束的问题，原模型看作带约束的问题，从而利用 F 统计量做检验，即假设 n 个数据，k 个参数，短数据包含 l 个，于是

$$\frac{e_*^T e_* - e^T e/l}{e^T e/(n-l-k)} \sim F(l, n-l-k)$$

这种方法叫做 chow 检验

1.11 极大似然估计

我们前面介绍了两种估计方法：简单最小二乘和矩估计，我们在这里介绍一种新的方法，极大似然估计。极大似然估计需要所有模型中的性质，

包括 ϵ 的正态分布假设。对独立同分布的数据 (x_i, y_i)，我们考察他们的联合分布的密度函数

$$f(y_1, \cdots, y_n \mid x_1, \cdots, x_n) = f(y_1 \mid x_1) f(y_2 \mid x_2) \cdots f(y_n \mid x_n)$$

而由模型假设 $\epsilon \sim N(0, \sigma^2 I)$，于是 $y_i \mid x_i \sim (x_i\beta, \sigma^2)$，从而

$$f(y_i \mid x_i) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_i - x_i\beta)^2}{2\sigma^2}}$$

于是代入之前的计算中得到

$$f(y_1 \mid x_1) f(y_2 \mid x_2) \cdots f(y_n \mid x_n) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_i - x_i\beta)^2}{2\sigma^2}}$$

取 ln 后关于 β, σ^2 分别求偏导，得到

$$b_{ML} = (X^T X)^{-1} X^T Y = b_{ols}, \hat{\sigma^2} = \frac{\sum_{i=1}^{n} (y_i - x_i b_{ML})^2}{n} = \frac{e^T e}{n}$$

需要注意这里给出的 $\hat{\sigma^2}$ 并非无偏估计。

极大似然的最大好处是易推广，我们接下来介绍离散因变量模型（分类）来展示极大似然估计的应用。我们以二分类模型为例

$$(x_i, y_i) y_i \in \{0, 1\}$$

我们的基本思路是将离散的因变量转化为实际的“近似度”，以“近似度”的大小来判断其归属，为此我们引入函数 $\sigma(z) = \frac{1}{1+e^{-z}}$，从而记

$$P(y_i = 1 | x_i) = \sigma(\beta^T x_i) \Rightarrow f(y_i \mid x_i) = (\sigma(\beta^T x_i))^{y_i} (1 - \sigma(\beta^T x_i))^{1 - y_i}$$

故通过最大似然估计

$$L(\beta) = f(y_1, \cdots, y_n \mid x_1, \cdots, x_n) = f(y_1 \mid x_1) f(y_2 \mid x_2) \cdots f(y_n \mid x_n) = \prod_{i=1}^{n} (\sigma(\beta^T x_i))^{y_i} (1 - \sigma(\beta^T x_i$$

在这个式子中取 $\frac{\partial L(\beta)}{\partial \beta} = 0$，通过迭代方法求解。

注 1.11.1. 这个模型叫做 Logistic 回归，如果将我们选择的概率函数替换，令

$$P(y_i = 1 \mid x_i) = \int_{-\infty}^{\beta^T x_i} \phi(x) dx$$

其中 $\phi(x)$ 为 $N(0, 1)$ 概率密度函数，则得到的新模型称为 probit

那这个模型能通过回归的方式求解吗？注意 $\mathrm{var}(\epsilon_i \mid x_i) = \sigma(\beta^T x_i)(1 - \sigma * \beta^T x_i)$ 有异方差性。