第四章

经典单方程计量经济学模型

放宽基本假定的模型

第四章 放宽基本假定的模型

4.1 多重共线性

4.2 异方差性

4.3 内生解释变量问题

4.4 模型设定偏误问题

4.5 序列相关性

4.4 模型设定偏误的概念与类型

定义：如果设定了一个“错误的”或者说是“有偏误的”模型，即使其他经典假设满足，得到的估计结果也会与“实际”有偏误，这种偏误称为模型设定偏误。

模型设定偏误的类型：

1. 相关变量的遗漏
2. 无关变量的误选
3. 错误的函数形式

1.相关变量的遗漏

产生原因：由于认识上的偏差、理论分析的缺陷，或者是有关统计数据的限制，导致有意或无意的忽略了某些重要变量。

动态设定偏误：遗漏相关变量表现为对Y或X滞后项的遗漏。

【例】

正确模型为

$$Y=\beta_0+\beta_1X_1+\beta_2X_2+\mu$$

误设定模型为

$$Y=\alpha_0+\alpha_1X_1+v$$

即设定模型时漏掉了相关的解释变量，这类错误称为遗漏相关变量。

2.无关变量的误选

无关变量的误选是指在设定模型时，包括了无关的解释变量。

【例】

正确模型为

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \mu$$

误设定模型为

$$Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 + v$$

即设定模型时，多选了一个无关解释变量。

3.错误的函数形式

错误的函数形式是指在设定模型时，选取了不正确的函数形式。最常见的是当真是函数形式为非线性时，却选取了线性的函数形式。

【例】

正确模型为

$$Y = A{X_1}^{\beta_1}{X_2}^{\beta_2}e^{\mu}$$

误设定模型为

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + v$$

二、模型设定偏误的后果

（一）遗漏相关变量偏误带来的后果

采用遗漏相关变量的模型进行估计而带来的偏误称为遗漏相关变量偏误。

如正确模型形式为

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \mu$$

而我们却对

$$Y = \alpha_0 + \alpha_1 X_1 + v$$

进行回归，得到X_1参数的估计值为

$$\hat{\alpha}_1 = \frac{\sum x_{i1} y_i}{\sum {x_{i1}}^2}$$

正确模型离差形式

$$
\begin{aligned}
y_i &= Y - \bar{Y} \\
&= \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \mu - (\beta_0 + \beta_1 \bar{X}_1 + \beta_2 \bar{X}_2 + \bar{\mu}) \\
&= \beta_1 x_{i1} + \beta_2 x_{i2} + \mu_i - \bar{\mu}
\end{aligned}
$$

代入$\hat{\alpha}_1 = \frac{\sum x_{i1} y_i}{\sum x_{i1}{}^2}$中，得到

$$
\begin{aligned}
\hat{\alpha}_1 = \frac{\sum x_{i1} y_i}{\sum x_{i1}{}^2} &= \frac{\sum x_{i1}(\beta_1 x_{i1} + \beta_2 x_{i2} + \mu_i - \bar{\mu})}{\sum x_{i1}{}^2} \\
&= \beta_1 + \beta_2 \frac{\sum x_{i1} x_{i2}}{\sum x_{i1}{}^2} + \frac{\sum x_{i1}(\mu_i - \bar{\mu})}{\sum x_{i1}{}^2} \\
&= \beta_1 + \beta_2 \frac{\sum x_{i1} x_{i2}}{\sum x_{i1}{}^2} + \frac{\sum x_{i1} \mu_i}{\sum x_{i1}{}^2}
\end{aligned}
$$

● 遗漏相关变量的后果

1. 如果漏掉的X_2与X_1相关，则上式

$$\hat{\alpha}_1 = \beta_1 + \boxed{\beta_2 \frac{\sum x_{i1}x_{i2}}{\sum {x_{i1}}^2}} + \frac{\sum x_{i1}\mu_i}{\sum {x_{i1}}^2}$$

中$\beta_2 \frac{\sum x_{i1}x_{i2}}{\sum {x_{i1}}^2}$在小样本下求期望和大样本下求概率极限都不为0（$\frac{\sum x_{i1}\mu_i}{\sum {x_{i1}}^2}$的条件期望和概率均为0），从而使模型估计量在小样本下有偏，在大样本下非一致；

2. 如果X_2与X_1在给定的样本下不相关，α_1的估计满足无偏性和一致性，但是α_0的估计却是有偏且非一致的；

3. 随机干扰项的方差估计σ^2是有偏的。（$\sigma^2=\frac{\sum e_i^2}{n-k-1}$，样本相同情况下，残差不同）

4. $\hat{\alpha}_1$的方差是正确估计量$\hat{\beta}_1$的方差的有偏估计

$$\hat{\alpha}_1=\frac{\sum x_{i1}y_i}{\sum {x_{i1}}^2} \Longrightarrow Var(\hat{\alpha}_1)=\frac{\sigma^2}{\sum {x_{i1}}^2}$$

$$Var(\hat{\beta}_1)=\sigma^2\frac{\sum {x_{i2}}^2}{\sum {x_{i1}}^2\sum {x_{i2}}^2-(\sum x_{i1}x_{i2})^2}=\frac{\sigma^2}{\sum {x_{i1}}^2(1-r_{x_1x_2}^2)}$$

无论X_2与X_1是否相关，$\mathrm{Var}(\hat{\alpha}_1)\neq\mathrm{Var}(\hat{\beta}_1)$。（$\sigma^2$不同）

（二）包含无关变量偏误带来的后果

正确模型为

$$Y = \alpha_0 + \alpha_1 X_1 + v$$

如果误设定模型为

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \mu$$

- 其估计量仍然具有无偏性（ $\beta_2 = 0$， $E(\beta_2) = 0$ ）。
- X_2与X_1 无关时，方差也相同。
- X_2与X_1相关时，错误模型中的方差增大，不具有最小方差性。

$$Var(\hat{\beta}_1) = \frac{\sigma^2}{\sum x_{i1}{}^2 (1 - r_{x_1 x_2}^2)} > Var(\hat{\alpha}_1)$$

（三）错误函数形式的偏误带来的后果

- 当选取了错误的函数形式并对其进行估计时，带来的偏误称为错误函数形式偏误。
- 容易判断，这种偏误是全方位的。
- 参数经济意义完全不同，其结果也不相同。

三、模型设定偏误问题的检验

（一）检验是否含有无关变量

基本思想：如果模型中误选了无关变量，则其系数的真值应为零。因此，只须对无关变量系数的显著性进行检验。

t检验：检验某1个变量是否应包括在模型中。

F检验：检验若干个变量是否应同时包括在模型中。

F检验过程：

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \mu \text{.......................................}①$$

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k + \beta_{k+1} X_{k+1} + \cdots + \beta_{k+q} X_{k+q} + \mu \text{ .....} ②$$

可以把①看做是②施加了一组约束条件的受约束回归

F检验目的：检验q个变量是否应同时包含在模型之中

$$H_0: \beta_{k+1} = \beta_{k+2} = \cdots = \beta_{k+q} = 0$$

$$H_1: \beta_{k+1}, \beta_{k+2}, \ldots, \beta_{k+q}\text{不全为}0$$

- 如果受约束条件为真，原模型和受约束回归模型具有相同的解释能力，即额外的变量$X_{k+1}\ldots\ldots X_{k+q}$对$Y$没有解释能力，则$F$统计量较小。
- 如果约束条件为假，即额外的变量$X_{k+1}\ldots\ldots X_{k+q}$对$Y$有较强的解释能力，则$F$统计量较大。
- 因此，可以通过F的计算值与临界值的比较，来判断额外变量是否应该包含在模型中。

此时，F统计量的值为

$$F=\frac{(ESS_u-ESS_R)/q}{RSS_u/[n-(k+q+1)]}$$

$$=\frac{(r_u^2-r_R^2)/q}{(1-r_u^2)/[n-(k+q+1)]}$$

$$F\sim F_\alpha(q\ ,\ n-(k+q+1))$$

其中：r_u^2—新模型（无约束回归方程②）的可决系数，

r_R^2—原模型（受约束回归①）的可决系数。

（二）检验是否有遗漏变量或函数形式设定偏误

遗漏相关变量与设定错误的函数形式的后果非常严重，不仅估计量有偏且不一致，而且随机干扰项的方差也往往会被高估，从而使通常的推断程序变的无效。

检验遗漏变量和函数形式设定偏误的方法：

1.残差图示法

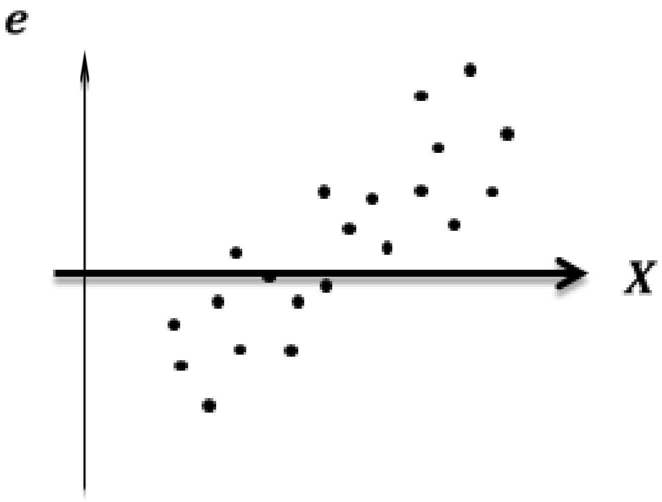
2.一般性偏误检验——RESTE检验

1.残差图示法

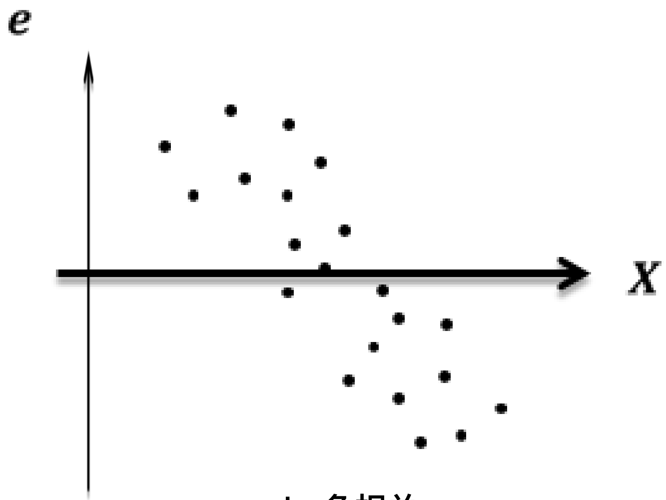
残差图示法的检验过程及原理：

对所设定的模型进行OLS回归，得到估计的残差序列e_i，做出e_i与某解释变量X_i的散点图，从图形考察估计的残差序列e_i是否有规律的变动，来判断是否遗漏了重要的解释变量或函数形式选取有误。

● 遗漏解释变量的残差图示



a. 正相关

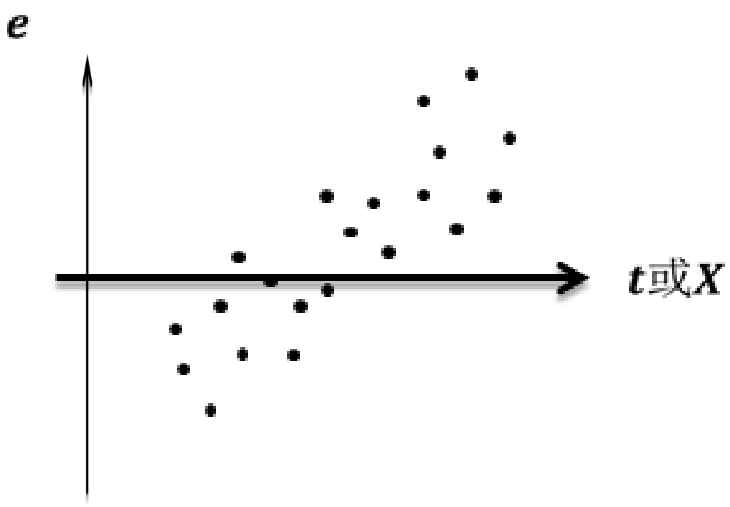


b. 负相关

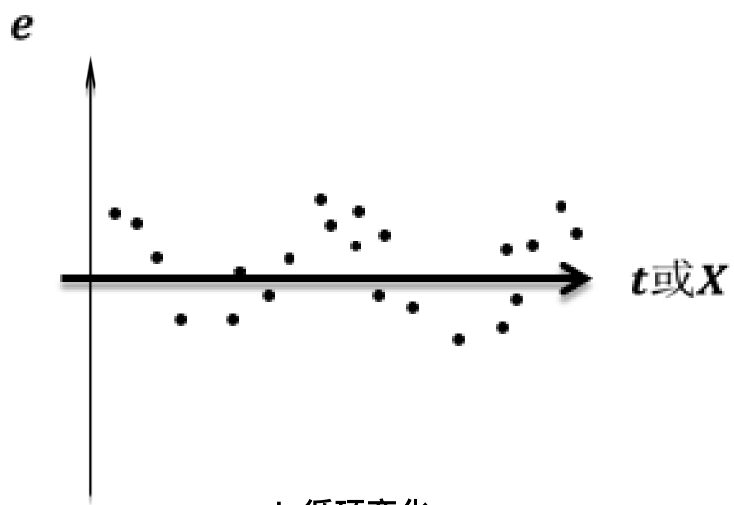
残差项与X正相关，意味着模型中遗漏了与某解释变量正相关的变量。

残差项与X负相关，意味着模型中遗漏了与某解释变量负相关的变量。

遗漏解释变量的残差图示



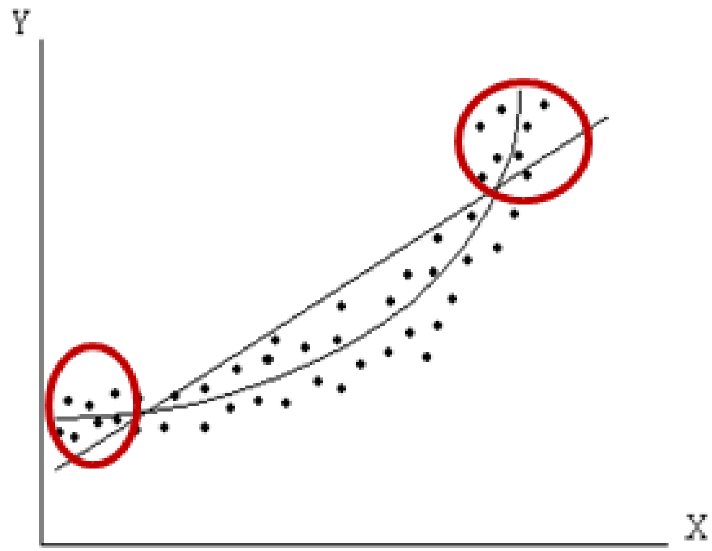
c. 趋势变化



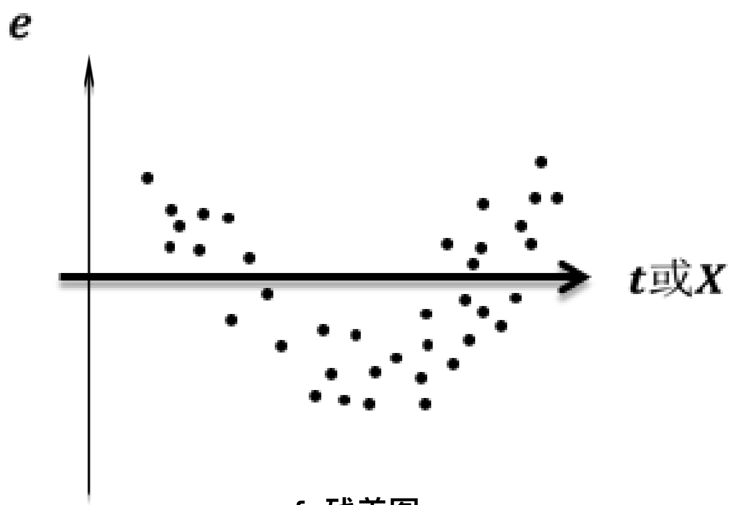
d. 循环变化

残差项随t或X增加而增加，意味着模型中可能遗漏了一个随时间推移而持续上升的变量。

残差项随t或X增加而出现周期性变化，意味着模型中可能遗漏了一个随时间推移而出现循环变化的变量。



e. 函数形式设定偏误



f. 残差图

当函数形式出现偏误时，残差序列也往往表现出某种有规律的变化特征。

对于图e所示模型，其真实函数成幂指数形式，但是选取了线性函数进行回归，这种情形下，残差序列往往呈现先正、后负、再正的变化特征。

2.一般性设定偏误检验——*RESET*检验

基本思想：若事先知道遗漏了哪个变量，只需将此变量引入模型，估计并检验其参数是否显著不为0即可；

若果不知道遗漏了哪个变量，需要寻找一个替代变量Z，来进行上述检验。

【注意】*RESET*检验中，采用所设定模型中被解释变量Y的估计值$\hat{Y}$的若干次幂来充当该“替代”变量。

- ***RESET*检验应用于遗漏变量的检验**

对于模型

$$Y=\alpha_0+\alpha_1X_1+v$$

怀疑其遗漏了某解释变量，进行*RESET*检验的步骤为：

第一步：先估计原模型，得到

$$\hat{Y}=\hat{\alpha}_0+\hat{\alpha}_1X_1$$

再通过残差项e_i与估计的$\hat{Y}$的图形判断引入$\hat{Y}$的若干次幂充当“替代”变量；

第二步：当e_i与Y的图形呈现曲线变化时，回归模型可选为

$$Y=\beta_0+\beta_1X_1+r_1\hat{Y}^2+r_2\hat{Y}^3+\mu$$

再根据增加解释变量的F检验来判断是否增加这些替代变量。若只增加一个变量，可以用t检验来判断。

- ***RESET*检验应用函数设定形式偏误的检验**

一元线性回归模型中：

真实函数形式为非线性，用泰勒定理可以近似表示为多项式

$$Y=\beta_0+\beta_1X_1+\beta_2X_1^2+\beta_3X_1^3+\cdots+\mu$$

如果设定为线性形式，就意味着遗漏了相关变量X_1^2，X_1^3……

因此，在一元回归中，可通过检验式中的各高次幂参数的显著性来判断是否将非线性模型误设成了线性模型。

多元模型中：

对多元回归，非线性函数可能是关于若干或全部X的非线性，这时可按遗漏变量的程序进行检验。

若怀疑回归

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \mu$$

是非线性的，需要估计出$\hat{Y}$的若干次幂为“替代”变量，进行类似于如下模型的估计

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + r_1 \hat{Y}^2 + r_2 \hat{Y}^3 + \mu$$

再判断各替代参数是否显著不为0即可。