

第二章教材习题与解析

1、判断下列表达式是否正确：

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_i, & i = 1, 2, \dots, n \\ \hat{y}_i &= \hat{\beta}_0 + \hat{\beta}_1 x_i, & i = 1, 2, \dots, n \\ E(y_i | x_i) &= \beta_0 + \beta_1 x_i + u_i, & i = 1, 2, \dots, n \\ E(y_i | x_i) &= \beta_0 + \beta_1 x_i, & i = 1, 2, \dots, n \\ E(y_i | x_i) &= \hat{\beta}_0 + \hat{\beta}_1 x_i, & i = 1, 2, \dots, n \\ y_i &= \beta_0 + \beta_1 x_i + u_i, & i = 1, 2, \dots, n \\ \hat{y}_i &= \hat{\beta}_0 + \hat{\beta}_1 x_i + u_i, & i = 1, 2, \dots, n \\ y_i &= \hat{\beta}_0 + \hat{\beta}_1 x_i + u_i, & i = 1, 2, \dots, n \\ y_i &= \hat{\beta}_0 + \hat{\beta}_1 x_i + \hat{u}_i, & i = 1, 2, \dots, n \\ \hat{y}_i &= \hat{\beta}_0 + \hat{\beta}_1 x_i + \hat{u}_i, & i = 1, 2, \dots, n \end{aligned}$$

答案：对于计量经济学模型有两种类型，一是总体回归模型，另一是样本回归模型。两类回归模型都具有确定形式与随机形式两种表达方式：

总体回归模型的确定形式： $E(Y | X) = \beta_0 + \beta_1 X$

总体回归模型的随机形式： $Y = \beta_0 + \beta_1 X + \mu$

样本回归模型的确定形式： $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$

样本回归模型的随机形式： $Y = \hat{\beta}_0 + \hat{\beta}_1 X + e$

除此之外，其他的表达形式均是错误的

2、给定一元线性回归模型： $y = \beta_0 + \beta_1 x + u$

- (1) 叙述模型的基本假定；
- (2) 写出参数 β_0 和 β_1 的最小二乘估计公式；
- (3) 说明满足基本假定的最小二乘估计量的统计性质；
- (4) 写出随机扰动项方差的无偏估计公式。

答案：(1) 线性回归模型的基本假设有两大类，一类是关于随机误差项的，包括零均值、同方差、不序列相关、满足正态分布等假设；另一类是关于解释变量的，主要是解释变量是非随机的，如果是随机变量，则与随机误差项不相关。

$$(2) \quad \hat{\beta}_1 = \frac{\sum x_i y_i}{\sum x_i^2}, \quad \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

(3) 考察总体的估计量，可从如下几个方面考察其优劣性：

- 1) 线性性，即它是否是另一个随机变量的线性函数；
- 2) 无偏性，即它的均值或期望是否等于总体的真实值；
- 3) 有效值，即它是否在所有线性无偏估计量中具有最小方差；
- 4) 渐进无偏性，即样本容量趋于无穷大时，它的均值序列是否趋于总体真值；
- 5) 一致性，即样本容量趋于无穷大时，它是否依概率收敛于总体的真值；

6) 渐进有效性, 即样本容量趋于无穷大时, 它在所有的一致估计量中是否具有最小的渐进方差。

这里, 前三个准则也称作估计量的有限样本性质或小样本性质, 因为一旦某估计量具有该性质, 它是不以样本的大小而改变的。拥有这类性质的估计量称为最佳线性无偏估计量。当然, 在有限样本情形下, 有时很难找到最佳线性无偏估计量, 这时就需要考察样本容量无限增大时估计量的渐进性质。

后三个准则称为估计量的无限样本性质或大样本渐进性质。如果有限样本情况下不能满足估计的准则, 则应扩大样本容量, 考虑参数估计量的大样本性质。

(4) 随机干扰项方差的估计式为: $\hat{\sigma}^2 = \frac{\sum e_i^2}{n-2}$

3、假定有如下回归结果,

$$\hat{Y}_t = 2.9611 - 0.4795 X_t$$

其中 Y = 我国的茶消费量 (每天每人消费的杯数)

X = 茶的零售价格 (元/公斤), t 表示时间

(1) 这是一个时间数列回归还是横截面序列回归?

(2) 画出回归线。

(3) 如何解释截距项的意义, 它有经济含义吗?

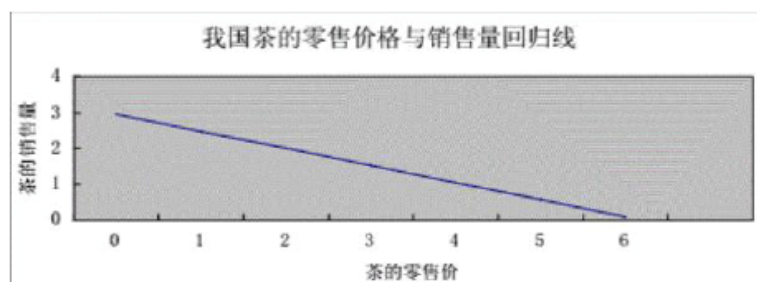
(4) 如何解释斜率?

(5) 你能求出真实的总体回归函数吗?

答案:

(1) 这是时间数列回归

(2)



(3) 截距 2.9611 表示茶的零售价在每公斤 0 元时, 我国平均茶消费量为每人每天 2.6911 杯, 这个数字没有明显的经济意义

(4) 茶的零售价格每公斤变动一元, 平均而言, 茶消费量向反方向变动 0.4795 个杯数。

(5) 不能, 因为要了解我国所有人的茶消费情况几乎是不可能的。

4、使用本章应用 CAPM 模型实证分析的数据, 估计不同年份贵州茅台的 β 系数, 考察 β 是否有变化。可以使用 2010-2019 这十年的日收益率数据, 看看 β 在不同年份是如何变化的?

答案:

在命令窗口输入:

```
regress RI RM if year==2010
```

输出结果如下:

Source	SS	df	MS	Number of obs	=	240
Model	.01120743	1	.01120743	F(1, 238)	=	33.72
Residual	.079103244	238	.000332367	Prob > F	=	0.0000
				R-squared	=	0.1241
				Adj R-squared	=	0.1204
Total	.090310675	239	.000377869	Root MSE	=	.01823

RI	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
RM	.4805769	.0827596	5.81	0.000	.317542	.6436119
_cons	.0003224	.0011841	0.27	0.786	-.0020103	.0026552

类似的，我们可以求出 2011 到 2019 的回归结果。β 系数如下表所示：

年份	β 系数	年份	β 系数
2010	0.481	2015	0.626
2011	0.555	2016	0.611
2012	0.702	2017	0.771
2013	0.401	2018	1.169
2014	0.686	2019	0.969

如上表所示，β 系数的整体变化趋势逐渐增大的。由回归结果我们可以看到，t 统计量通过了显著性检验，截距项的估计值也通过显著性检验，但值比较小，接近于 0。CAPM 可以认为基本成立。

第三章教材习题与解析

1、给定二元回归模型： $y_t = b_0 + b_1x_{1t} + b_2x_{2t} + u_t$ ($t=1, 2, \dots, n$)

- (1) 叙述模型的古典假定；
- (2) 写出总体回归方程、样本回归方程和样本回归模型；
- (3) 写出回归模型的矩阵表示；
- (4) 写出回归系数及随机误差项的最小二乘估计量，并叙述参数估计量的性质；
- (5) 试述总离差平方和、回归平方和、残差平方和之间的关系及其自由度之间的关系。

答案：

- (1) 模型的古典假定包括：

假设 1：模型设定正确/线性于参数

假设 2：X 固定，即 X 在重复抽样中是固定的，也就是说 X 是非随机的。

假设 3：X 有变异，解释变量 X 在所抽取的样本中具有变异性，而且随着样本容量的无限增加，解释变量 X 的样本协方差阵趋于一个非零的有限正定矩阵。

假设 4：不存在完全共线性，指在样本中（从而也在总体中），没有一个自变量是常数，自变量之间不存在严格（完全）的线性关系。

假设 5：零条件数学期望，即给定自变量的任何值，误差 u 的期望值为零。

假设 6：同方差，对于解释变量的所有观测值，随机误差项有相同的方差

假设 7：无序列相关，即

$$Cov(\mu_i, \mu_j) = E(\mu_i \mu_j) = 0, (i \neq j, i, j = 1, 2, \dots, n)$$

假设 8：正态性假设，即随机误差项 μ 服从均值为零，方差为 σ^2 的正态分布。

- (2) 总体回归方程： $E(y_t | x_t) = b_0 + b_1x_{1t} + b_2x_{2t}$

样本回归方程： $\hat{y}_i = \hat{b}_0 + \hat{b}_1 x_{1i} + \hat{b}_2 x_{2i}$

样本回归模型： $y_i = \hat{b}_0 + \hat{b}_1 x_{1i} + \hat{b}_2 x_{2i} + e$

(3) 模型回归的矩阵表示为： $Y_i = X_i b + \mu_i$

其中 $X_i = (1, x_{1i}, x_{2i})'$, $b = (b_0, b_1, b_2)'$,

(4) 回归系数的最小二乘估计式为 $\hat{b} = (X_i' X_i)^{-1} X_i' Y$, 随机扰动项方差的最小二乘估计

为： $\hat{\sigma}^2 = \frac{\sum e_i^2}{n-k-1}$

参数估计式的性质：具有线性性、无偏性和最小方差性

(5) 总平方和等于回归平方和加上残差平方和，即 $TSS=ESS+RSS$ 。残差平方和自由度为 $n-k-1$ ，总离差平方和的自由度为 $n-1$ 。

2、为研究中国各地区入境旅游状况，建立了各省市旅游外汇收入（Y，百万美元）、旅行社职工人数（ X_1 ，人）、国际旅游人数（ X_2 ，万人次）的模型，用某年 31 个省市的截面数据估计结果如下：

$$\begin{aligned}\hat{Y}_i &= -151.0263 + 0.1179 X_{1i} + 1.5452 X_{2i} \\ t &= (-3.066806) \quad (6.652983) \quad (3.378064) \\ R^2 &= 0.934331 \quad \bar{R}^2 = 0.92964 \quad F = 191.1894 \quad n = 31\end{aligned}$$

(1) 从经济意义上考察估计模型的合理性。

(2) 在 5% 显著性水平上，分别检验参数 β_1, β_2 的显著性。

(3) 在 5% 显著性水平上，检验模型的整体显著性。

答案：

有模型估计结果可看出：旅行社职工人数和国际旅游人数均与旅游外汇收入正相关。平均说来，旅行社职工人数增加 1 人，旅游外汇收入将增加 0.1179 百万美元；国际旅游人数增加 1 万人次，旅游外汇收入增加 1.5452 百万美元。

取 $\alpha = 0.05$ ，查表得 $t_{0.025}(31-3) = 2.048$

因为 3 个参数 t 统计量的绝对值均大于 $t_{0.025}(31-3) = 2.048$ ，说明经 t 检验 3 个参数均显著不为 0，即旅行社职工人数和国际旅游人数分别对旅游外汇收入都有显著影响。

取 $\alpha = 0.05$ ，查表得 $F_{0.05}(2, 28) = 3.34$ ，由于 $F = 191.1894 > F_{0.05}(2, 28) = 3.34$ ，说明旅行社职工人数和国际旅游人数联合起来对旅游外汇收入有显著影响，线性回归方程显著成立。

3、考虑下列利率和美国联邦预算赤字关系的最小二乘估计：

模型 A： $\hat{Y}_1 = 0.103 - 0.079 X_1$ $R^2 = 0.00$

其中： Y_1 ——Aaa 级公司债券的利率

X_1 ——联邦赤字占 GNP 的百分比

(季度模型：1970——1983)

模型 T： $\hat{Y}_2 = 0.089 + 0.369 X_2 + 0.887 X_3$ $R^2 = 0.40$

其中： Y_2 ——三个月国库券的利率

X_2 ——联邦预算赤字（以 10 亿美元为单位）

X_3 ——通货膨胀率（按百分比计）
（季度模型：1970年4月——1979年9月）

请回答以下问题：

- （1）“最小二乘估计”是什么意思？什么被估计，什么被平方？在什么意义下平方“最小”？
- （2） R^2 为 0.00 是什么意思？它可能为负吗？
- （3）计算两个方程的 $\bar{R}^2$ 值。
- （4）比较两个方程，哪个模型的估计值符号与你的预期一致？模型 T 是否自动的优于模型 A，因为它的 R^2 值更高？若不是，你认为哪个模型更好，为什么？

答案：（1）“最小二乘”估计是求出使 Y 的实际值和预测值之差的平方和最小的参数估计值。
平方“最小”是指它们的和最小。

（2）若 $R^2 = 0$ ，则误差平方和（ESS）等于总平方和（TSS），即回归平方和（RSS）等于 0。

因为 $R^2 = RSS / TSS$ ，故不可能为负。

（3）模型 A： $\bar{R}^2 = 1 - (56 - 1) / (56 - 1 - 1) = -0.02$

模型 T： $\bar{R}^2 = 1 - (1 - 0.4)[(38 - 1) / (38 - 2 - 1)] = 0.37$

（4）模型 T 与预期的估计符号一致，并且包含了一个重要变量（假定利率为名义利率），因此它优于模型 A，模型 A 的符号与预期不一致。较高的 $\bar{R}^2$ 并不意味着方程就自动地更优。

第四章教材习题与解析

1、回归模型中引入虚拟变量的作用是什么？有哪几种基本的引入方式，它们各适用于什么情况？

解析：作用：在模型中引入虚拟变量，主要是为了寻找某（些）定性因素对解释变量的影响。加法方式与乘法方式是最主要的引入方式，前者主要适用于定性因素对截距项产生影响的情况，后者主要适用于定性因素对斜率项产生影响的情况。除此外，还可以加法与乘法组合的方式引入虚拟变量，这时可测度定性因素对截距项与斜率项同时产生影响的情况。

2、什么是“虚拟变量陷阱”？

解析：根据虚拟变量的设置原则，一般情况下，如果定性变量有 m 个类别，则需在模型中引入 $m-1$ 个变量。如果引入了 m 个变量，就会导致模型解释变量出现完全的共线性问题，从而导致模型无法估计。这种由于引入虚拟变量个数与类别个数相等导致的模型无法估计的问题，称为“虚拟变量陷阱”。

3、假设已经得到关系式 $Y = \beta_0 + \beta_1 X$ 的最小二乘估计，请回答：

（1）如果把 X 变量的单位扩大 10 倍，对原回归的斜率和截距的影响如何？如果把 Y 变量的单位扩大 10 倍，又会怎样？

（2）如果给 X 的每个观测值都增加 1，对原回归的斜率和截距的影响如何？如果给 Y 的每个观测值都增加 1，又会怎样？

解析：（1）记 X^* 为原变量 X 单位扩大 10 倍的变量，则 $X = \frac{X^*}{10}$ ，于是

$$\begin{aligned}
Y &= \beta_0 + \beta_1 X \\
&= \beta_0 + \beta_1 \frac{X^*}{10} \\
&= \beta_0 + \frac{\beta_1}{10} X^*
\end{aligned}$$

可见，解释变量的单位扩大 10 倍时，回归的截距项不变，而斜率项将会变为原回归系数的 $\frac{1}{10}$ 。

同样地，记 Y^* 为原变量 Y 单位扩大 10 倍时的变量，则 $Y = \frac{Y^*}{10}$ ，于是

$$\frac{Y^*}{10} = \beta_0 + \beta_1 X$$

即
$$Y^* = 10\beta_0 + 10\beta_1 X$$

可见，被解释变量的单位扩大 10 倍时，截距项与斜率项都会比原回归系数扩大 10 倍。

记 $X^* = X + 1$ ，则原回归模型变为

$$\begin{aligned}
Y &= \beta_0 + \beta_1 X \\
&= \beta_0 + \beta_1 (X^* - 1) \\
&= (\beta_0 - \beta_1) + \beta_1 X^*
\end{aligned}$$

记 $Y^* = Y + 1$ ，则原回归模型变为

$$Y^* - 1 = \beta_0 + \beta_1 X$$

即
$$Y^* = (\beta_0 + 1) + \beta_1 X$$

可见，无论解释变量还是被解释变量以加法的形式变化，都会造成原回归模型的截距项变化，而斜率项不变。

4、根据搜集到的季度数据，我们有如下的咖啡需求函数的回归方程：

$$\ln \hat{Q}_t = 1.28 - 0.16P_t + 0.51 \ln I_t + 0.15 \ln P'_t - 0.01T - 0.10D_{1t} - 0.16D_{2t} - 0.01D_{3t}$$

$(-2.14) \quad (1.23) \quad (0.55) \quad (-3.36) \quad (-3.74) \quad (-6.03) \quad (-0.37)$

其中，括号内为各估计系数的 t 值。 Q 为人均咖啡消费量（单位：磅）， I 为人均收入， P 为咖啡的价格， P' 为茶的价格， T 为时间趋势变量。

$$D_{1t} = \begin{cases} 1, & \text{第一季度} \\ 0, & \text{其他季度} \end{cases}$$

$$D_{2t} = \begin{cases} 1, & \text{第二季度} \\ 0, & \text{其他季度} \end{cases}$$

$$D_{3t} = \begin{cases} 1, & \text{第三季度} \\ 0, & \text{其他季度} \end{cases}$$

回答下列问题：

- (1) 模型中 P 、 I 和 P' 系数的经济含义是什么？
- (2) 咖啡的价格需求是否很有弹性？
- (3) 咖啡和茶是互补品还是替代品？
- (4) 如何解释时间变量 T 的系数？
- (5) 如何解释模型中虚拟变量的作用？
- (6) 哪一个虚拟变量在统计上是显著的？

解析：(1) 从回归模型来看， P 的系数 -0.16，表示当咖啡的价格增加 1% 时，咖啡的需求量减少 0.16%，是咖啡需求的价格弹性系数； I 的系数 0.51，表示的是咖啡需求量对收入的弹性，即当收入增加 1% 时，咖啡需求量将增加 0.51%； P' 的系数 0.15，表示的是咖啡需求量对茶叶的交叉价格弹性系数，即当茶叶的价格增加 1% 时，咖啡需求量将增加 0.15%。

(2) 咖啡需求的价格弹性为 0.16 小于 1，属于缺乏弹性。

(3) 由于交叉价格弹性为正，表明两者是替代品。

(4) 时间 T 的系数为 -0.01，表明咖啡的需求量在逐年递减。

(5) 虚拟变量的引入反映了季节因素对咖啡需求量的影响。

(6) 在 5% 的显著性水平下， t 统计量的临界值为 $t_{0.025}(70-8) = 1.99$ ， D_1 与 D_2 系数的 t 统计量绝对值大于临界值，在统计上是显著的。

(7) 咖啡需求量存在季节效应，第一季度和第二季度的销售量要少于其他季度。

5、一个解释 CEO 薪水的估计方程如下：

$$\ln \hat{Y} = 4.59 + 0.257 \ln X_1 + 0.011 X_2 + 0.158 D_1 + 0.181 D_2 - 0.283 D_3$$

(15.30) (8.03) (2.75) (1.78) (2.13) (-2.895)

其中， Y 表示年薪水平（单位：万元）， X_1 表示年销售收入（单位：万元）， X_2 表示公司股票收益（单位：万元）， D_1, D_2, D_3 均为虚拟变量，分别表示金融业、消费品工业和公用事业。假设对比产业为交通运输业。

(1) 解释三个虚拟变量参数的经济含义。

(2) 保持 X_1 和 X_2 不变，计算公用事业和交通运输业之间估计薪水的近似百分比差异。这个差异在 1% 的显著性水平下是统计显著的吗？

(3) 消费品工业和金融业之间估计薪水的近似百分比差异是多少？写出一个使你能直接检验这个差异是否统计显著的方程。

解析：(1) D_1 的参数经济含义为：当销售收入与公司股票收益保持不变时，金融业的 CEO 与交通运输业的 CEO 所获薪水的对数的差异。由于某变量对数的差可近似地看成该变量变动的百分率，因此， D_1 的参数经济意义还可解释为：当销售收入与公司股票收益保持不变时，金融业的 CEO 要比交通运输业的 CEO 多获 15.8% 的薪水。其他两个可类似解释。

(2) 公共事业和交通运输业之间估计薪水的近似百分比差异就是以百分数解释的 D_3 的参数，即为 28.3%。由于参数的 t 统计量为 -2.895，其绝对值大于 1% 显著性水平下自由度为 203 的 t 分布的临界值为 2.326，

因此这种差异是统计显著的。

(3) 由于消费品工业和金融业相对于交通运输业的薪水百分比差异分别为 15.8%与 18.1%，因此它们之间的差异为 18.1%-15.8%=2.3%。一个能直接检验这一差异是否显著地方程为

$$\ln Y = \beta_0 + \beta_1 \ln X_1 + \beta_2 X_2 + \alpha_1 D_2 + \alpha_2 D_3 + \alpha_3 \bullet trans + \mu$$

其中，trans 为交通运输业虚拟变量。这里对比基准为金融业，因此 α_1 表示了消费品工业与金融业薪水的百分比差异，其 t 统计值可用来进行显著性检验。

6、在申请出国读学位的 16 名学生中有如下 GRE 数量与词汇成绩，全职 9 名学生获得入学批准。请根据下表中的资料估计 Probit 模型和 Logit 模型。

学生 编号	数量 成绩 Q	词汇 成绩 V	是否准 入 Y(1= 准, 0= 不准)	学生 编号	数量 成绩 Q	词汇 成绩 V	是否准 入 Y(1= 准, 0= 不准)
1	760	550	1	9	520	660	1
2	600	350	0	10	800	250	0
3	720	320	0	11	670	480	0
4	710	630	1	12	670	520	1
5	530	430	1	13	780	710	1
6	650	570	0	14	520	450	0
7	800	500	1	15	680	590	1
8	650	680	1	16	500	380	0

解析：根据 Eviews 软件得如下表：

Dependent Variable: Y
Method: ML - Binary Logit (Quadratic hill climbing)
Date: 05/22/11 Time: 22:19
Sample: 1 16
Included observations: 16
Convergence achieved after 5 iterations
Covariance matrix computed using second derivatives

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-11.10741	6.124290	-1.813665	0.0697
Q	0.003968	0.008008	0.495515	0.6202
V	0.017696	0.008752	2.021914	0.0432
McFadden R-squared	0.468521	Mean dependent var	0.562500	
S.D. dependent var	0.512348	S.E. of regression	0.382391	
Akaike info criterion	1.103460	Sum squared resid	1.900896	
Schwarz criterion	1.248321	Log likelihood	-5.827681	
Hannan-Quinn criter.	1.110878	Restr. log likelihood	-10.96503	
LR statistic	10.27469	Avg. log likelihood	-0.364230	
Prob(LR statistic)	0.005873			
Obs with Dep=0	7	Total obs	16	
Obs with Dep=1	9			

于是，我们可得到 Logit 模型为：

$$\hat{Y}_i = -11.107 + 0.004Q + 0.0177V$$

(-1.81) (0.49) (2.02)

$$R_{MCF}^2 = 0.4685, \quad LR(2)=10.27$$

如果在 Binary estimation 这一栏中选择 Probit 估计方法，可得到如下表：

Dependent Variable: Y
Method: ML - Binary Probit (Quadratic hill climbing)
Date: 05/22/11 Time: 22:25
Sample: 1 16
Included observations: 16
Convergence achieved after 5 iterations
Covariance matrix computed using second derivatives

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-6.634542	3.396882	-1.953127	0.0508
Q	0.002403	0.004585	0.524121	0.6002
V	0.010532	0.004693	2.244299	0.0248
McFadden R-squared	0.476272	Mean dependent var	0.562500	
S.D. dependent var	0.512348	S.E. of regression	0.381655	
Akaike info criterion	1.092836	Sum squared resid	1.893588	
Schwarz criterion	1.237696	Log likelihood	-5.742687	
Hannan-Quinn criter.	1.100254	Restr. log likelihood	-10.96503	
LR statistic	10.44468	Avg. log likelihood	-0.358918	
Prob(LR statistic)	0.005395			
Obs with Dep=0	7	Total obs	16	
Obs with Dep=1	9			

于是，我们可得到 Probit 模型为：

$$\hat{Y}_i = -6.635 + 0.0024Q + 0.0105V$$

$$(-1.95) \quad (0.52) \quad (2.24)$$

$$R^2_{MCF} = 0.4763, \quad LR(2)=10.44$$

第五章教材习题与解析

1、下列哪些因素会导致 OLS 估计量偏误？

- (1) 异方差性
- (2) 遗漏一个重要变量
- (3) 模型中同时包含的两个自变量之间的样本相关系数达到 0.95

解析：只有 (2) 才会导致 OLS 出现偏误，且所遗漏的变量必须与模型中包含的变量是相关的。模型中出现异方差并不影响 OLS 估计量的无偏性，仅影响估计量的方差。样本解释变量之间共线性的程度并不影响高斯-马尔科夫假定。只有当自变量之间完全共线时，才会违背 MLR.3。

2、多重共线性的原因有哪些？其造成的后果有哪些？该如何修正？

解析：原因有以下三点：

- (1) 经济变量存在相关的共同趋势。例如，在西方经济学里我们已经知道投资和利率之间存在负相关关系，但如果我们研究房地产价格波动将投资和利率同时纳入影响因素中，可能会造成多重共线性。
- (2) 滞后变量的引入。在计量经济学中，研究模型时经常需要加入滞后变量来反映真实的经济关系。例如，在相对收入假说中，居民当期消费不仅受到当期收入 Y_t 的影响，还会受到前期消费 C_{t-1} 的影响。

(3) 样本资料的限制。由于完全符合模型的数据较难收集,在现有的特定样本可能存在某种程度的多重共线性。如时间序列数据以简单形式建立的计量经济学模型,往往存在多重共线性。

计量经济学模型一旦出现多重共线性,如果仍采用普通最小二乘法估计模型参数,会产生下列不良后果:

- (1) 完全共线性下参数估计量不存在。
- (2) 近似共线性下普通最小二乘法参数估计量的方差变大。
- (3) 参数估计量的含义不合理。
- (4) 变量的显著性检验和模型的预测功能失去意义。

根据教材理论部分,主要介绍两种克服多重共线性的方法。

(1) 第一类方法:排除引起共线性的变量

找出引起多重共线性的解释变量,将它排除出去,是最为有效地克服多重共线性问题的方法,所以逐步回归法得到最为广泛的应用。但是,需要特别注意的是,当排除了某个或某些变量后,保留在模型中的变量的系数的经济意思将发生变化

(2) 第二类方法:岭回归法

岭回归法是在 20 世纪 70 年代提出的能够处理多重共线性问题的一种方法,该方法通过减小参数估计量的方差的方法消除多重共线性带来的显著性检验,预测失效等问题。

对于多元线性回归模型:

$$Y = X\beta + u$$

通过普通最小二乘法得到的参数估计量为:

$$\hat{\beta} = (X'X)^{-1} X'Y$$

岭回归法得到的系数估计量为:

$$\hat{\beta} = (X'X + I \cdot E)^{-1} X'Y$$

其中, I 为大于 0 的常数, E 为单位矩阵。岭回归法关键要得到 I 估计量。(有关 I 统计量的估计详见教材 5.1 节)

3、异方差性造成的后果有哪些?该如何检验以及修正?

解析:计量经济学模型一旦出现异方差性,如果仍采用普通最小二乘法估计模型参数,会产生一系列不良的后果。

- (1) 参数估计量是无偏的、一致的,但不具有有效性。
- (2) 参数估计量的方差出现偏误,变量的 t 检验和 F 检验失效。
- (3) 异方差性将导致预测区间偏大或偏小,预测失效。

异方差的检验主要有以下四种方法:

1.图示检验法; 2.布罗施—帕甘(Breusch-Pagan)检验; 3. $G-Q$ (Goldfeld-Quandt) 检验; 4.怀特(White)检验。

异方差的修正主要有以下两种方法:

1.加权最小二乘法(WLS) 2.异方差稳健标准误差法(异方差检验与修正方法的具体介绍参见教材 5.2 节)

4、序列相关性的原因有哪些?其造成的后果有哪些?该如何检验以及修正?

解析:对于时间序列数据,由于其它影响因素对被解释变量的影响都归到随机误差项,其它因素由于经济发展的惯性,存在序列相关,例如当期收入与前期收入相关,造成随机误差项存在序列相关。

如果存在序列相关性,会造成以下后果:

- (1)通过普通最小二乘法估计得到的估计量 $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k)$ 是无偏的和一致的,但是非有效的。
- (2)通过 OLS 估计得到的估计量 $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k)$ 的显著性检验失效。
- (3)模型的预测失效。

判断序列相关性的方法多种多样,如:图示法,回归检验法, D.W.检验法,拉格朗日乘数(LM)检验法, Q 统计量检验法等。

如果模型存在序列相关性,则需要用新的方法估计模型。我们有两种思路对模型进行修正。两种思路

为：一是将存在序列相关性的模型转换成不存在序列相关的模型，再用普通最小二乘估计，该思路有两种方法分别为广义最小二乘法法和广义差分法。二是先用普通最小二乘法估计模型，再对参数估计量的方差或标准差进行修正，这就是序列相关稳健估计法。(具体的检验与修正的方法请参见教材 5.3 节)

5、下表列出了某年中国部分省市城镇居民每个家庭平均全年可支配收入 X 与消费性支出 Y 的统计数据。

地区	可支配收入 (X)	消费性支出 (Y)	地区	可支配收入 (X)	消费性支出 (Y)
北 京	10349.69	8493.49	浙 江	9279.16	7020.22
天 津	8140.50	6121.04	山 东	6489.97	5022.00
河 北	5661.16	4348.47	河 南	4766.26	3830.71
山 西	4724.11	3941.87	湖 北	5524.54	4644.5
内蒙古	5129.05	3927.75	湖 南	6218.73	5218.79
辽 宁	5357.79	4356.06	广 东	9761.57	8016.91
吉 林	4810.00	4020.87	陕 西	5124.24	4276.67
黑龙江	4912.88	3824.44	甘 肃	4916.25	4126.47
上 海	11718.01	8868.19	青 海	5169.96	4185.73
江 苏	6800.23	5323.18	新 疆	5644.86	4422.93

- (1) 试用普通最小二乘法建立居民人均消费支出与可支配收入的线性模型；
- (2) 检验模型是否存在异方差性；
- (3) 如果存在异方差性，试采用适当的方法估计模型参数。
- 解答：城镇居民人均消费支出主要由人均收入决定。为了考察可支配收入对中国城镇居民消费支出增长的影响，设定如下线性模型：

$$Y = \beta_0 + \beta_1 X + \mu$$

其中，Y 表示城镇居民家庭消费性支出，X 表示平均每个全年可支配收入。
通过 stata16 软件估计结果如下表所示。

Source SS df MS Number of obs = 20									
F(1, 18) = 1048.91									
Model 49342144 1 49342144 Prob > F = 0.0000									
Residual 846742.352 18 47041.2418 R-squared = 0.9831									
Adj R-squared = 0.9822									
Total 50188886.4 19 2641520.34 Root MSE = 216.89									
Y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]				
X	.755125	.0233157	32.39	0.000	.7061404	.8041095			
_cons	272.3635	159.6772	1.71	0.105	-63.1059	607.8329			

图 5.3 OLS 估计结果

普通最小二乘法的估计结果为：

$$Y = 272.36 + 0.755X$$

(1.71) (32.39)

$$R^2 = 0.9831 F = 1048.91$$

(2) 利用 White 检验，具体命令如下

estat imtest,white

得到的分析结果如下：

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/048010077052006055>