

- 3.1 如何理解“经济变量之间的关系常常是一种随机关系”？
- 3.2 古典线性回归模型的四个假设分别起什么作用？
- 3.3 举出线性回归模型误差项存在异方差的例子。
- 3.4 如果用于回归的样本量足够大，当模型误差项存在异方差和/序列相关时，OLS 估计结果仍然可用吗？不可用的是什么？需要做怎样的改变才可用？
- 3.5 什么是回归残差？残差有哪些性质？为什么可以将残差看做误差项的估计？回归残差都用在了哪里？
- 3.6 假设要建立一个一元线性回归模型，解释变量为利率水平，被解释变量为银行储蓄总额，用来研究利率变动对银行储蓄的影响。在搜集数据对模型进行估计时，你是希望采用利率变化较小的时期的数据还是利率变化较大时期的数据？为什么？
- 3.7 设回归模型 $Y = \beta_0 + \beta_1 X + u$ 满足基本假设，OLS 估计结果为 $Y = 3.7 + 0.003X$ ，斜率系数估计的标准误为 $s_{\beta_1} = 0.001$ 。经计算用于检验假设 $H_0: \beta_1 = 0$ 的 t 值为 3，因此可以认为斜率系数 β_1 显著不为 0。请解释，为什么 β_1 的估计值 0.003 很小而 t 检验的结果却表明 β_1 显著不为 0。
- 3.8 推导出回归模型的 R^2 和 F 统计量之间的关系。用哪个量来衡量模型的拟合效果更为科学，为什么？
- 3.9 考虑解释变量只包含常数项的回归模型

$$Y_i = \beta_0 + u_i, \quad i = 1, 2, \dots, n$$

，误差项 u_i 满足基本假设。求出 β_0 的 OLS 估计。

- 3.10 以下采用一元线性回归模型的 EViews 输出结果，

White heteroskedasticity-consistent standard errors & covariance				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
X	0.493775		5.451075	0.0000
C	-0.550065	0.292620		0.0729
R-squared	0.588892	Mean dependent var	0.840800	
Adjusted R-squared	0.571018	S.D. dependent var	1.023885	
S.E. of regression		Akaike info criterion	2.115364	
Sum squared resid	10.34355	Schwarz criterion	2.212875	
Log likelihood	-24.44206	Hannan-Quinn criter.	2.142410	
F-statistic	32.94638	Durbin-Watson stat	1.697328	
Prob(F-statistic)	0.000008			

- (1) 输出结果的信息行“White heteroskedasticity-consistent standard errors”表明了什么？
- (2) 用有关数据计算出空白方格内缺失的值。

- 3.11 科技创新是企业发展的原动力。企业科技创新的投入用研发投入衡量，科技创新产出用企业申请的专利数衡量。从投入到产出需要一定的时间周期，设 y 表示当年企业申请专利的数量的自然对数， x 表示前五年前研发投入的自然对数。表 3.3 中给出了 30 个企业的有关数据

表 3.3

序号	y	x	序号	y	x	序号	y	x
1	3.806662	0.99361	11	3.73767	3.24168	21	3.135494	2.50803
2	3.433987	0.95654	12	4.110874	3.28292	22	2.564949	1.74407
3	3.433987	0.809	13	4.143135	3.31609	23	2.639057	1.76072
4	3.332205	0.71507	14	4.043051	3.34218	24	3.332205	1.94852
5	3.135494	0.6605	15	3.78419	2.51433	25	3.401197	1.98226

6	3.555348	0.59358	16	3.806662	2.65189	26	3.135494	1.88335
7	3.401197	1.06047	17	3.78419	2.73921	27	3.583519	2.02024
8	3.044522	2.87721	18	3.7612	2.78211	28	3.583519	2.04372
9	3.258097	3.04874	19	2.944439	2.81341	29	5.117994	3.64376
10	3.850148	3.17782	20	3.610918	2.69686	30	5.308268	3.74369

- (1) 用 OLS 估计出 y 和 x 之间的回归模型 $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ 。
- (2) 计算回归残差，并验证残差的两个性质。
- (3) 计算残差平方和和回归标准误 s ，并计算 R^2 。
- (4) 计算 $s_{\hat{\beta}_0}$ 和 $s_{\hat{\beta}_1}$ 。
- (5) 对假设 $H_0: \beta_1 = 0$, $H_1: \beta_1 \neq 0$ 进行 t 检验。
- (6) 解释回归模型中截距系数和斜率系数的含义，并说明模型中是否需要截距系数。

3.12 表 3.4 给出了英国一些大企业高管的薪水及其工作年限的数据。 y 表示薪水(千镑), x_1 表示在该企业工作的年限(年), x_2 : 表示在该企业任高管的年限(年),

表 3.4

y	x_1	x_2
106.7	7	7
94.5	35	10
...
62.2	35	4
99.9	28	17

数据来源: Dimitrios Asteriou (2006)

- (1) 将 y 对常数项和 x_1 进行回归, 研究工作年限对薪水的影响;
- (2) 将 y 对常数项和 x_2 进行回归, 研究任职年限对薪水的影响;
- (3) 对两个模型的回归结果进行比较, 在确定高管薪水时, 哪个模型更合理些?

◆ 参考答案

1. 经济活动的主题是人, 人的行为随机性导致了经济变量大多为随机变量。经济系统的复杂性导致了经济变量之间的关系是一种随机关系, 其主要原因在于影响一个经济变量的其他经济变量很多, 有些可以观测和计量, 有一些无法观测和计量。当给定影响一个经济变量的可观测变量时, 其他大量的不可观测的影响变量无法确定, 导致经济变量之间的关系时随机关系。
2. 假设 1 (或者假设 1') 保证了 OLS 估计的一致性和渐近正态性, 假设 2 和假设 3 一起保证了 OLS 估计的标准误计算具有简单的公式 (3.15), 假设 4 则保证了 OLS 估计具有精确的正态分布。假设 1~假设 3 还保证了 OLS 估计是最优估计, 即马尔科夫性。
3. 以公司年利润为因变量, 以公司规模为自变量, 形成一元线性回归模。如果抽取某一年若干家上市公司的财务报表数据为样本, 对模型进行回归分析, 则模型误差项存在异方差。原因在于大规模公司抵抗各种冲击的能力强, 利润稳定, 利润变量的方差小, 而小规模公司则相反。因此, 给定公司规模, 模型误差项的方差是公司规模的函数, 而不是常数。
4. OLS 估计结果仍然可用, 因为异方差和序列相关不改变 OLS 估计的一致性和渐近正态性。不可用的是计算标准误公式 (3.15), 需要用 White 计算方法计算 (异方差情况下) 和

Newey-West 方法计算（异方差和序列相关情况下）OLS 估计量的标准误，并据此计算参数检验的 t -统计量值对回归系数的显著性进行假设检验。

5. 回归残差是用因变量实际值减去因变量拟合值得到的差。残差形成的向量与自变量样本值形成的向量正交。如果回归模型中有截距项，则回归残差的和等于零。由于 OLS 估计以概率收敛到回归系数，残差以概率收敛到误差项。同方差假定下，估计回顾标准误要用到残差，计算拟合优度 R^2 时，计算残差平方和要用到残差，构造 t -检验统计量需要用到残差，构造 F -检验统计量是也要用到残差。
6. 用利率变化大的时候获得的数据估计模型。结论 6 中 OLS 估计的方差公式）来看，分布为自变量的总平方和，自变量变化越大，参数估计的方差越小，估计越精确。直观上看，建立模型研究的是利率变化如何影响储蓄率变化，只有利率变化较大时，影响才能充分体现。利率不变化，是不能看出对储蓄率变化的影响的。极端情况是利率为常数，此时 $X_i = c = \bar{X}$ ，自变量的总平方和为零，方差为无限大。
7. 因为 $\hat{\beta}_1$ 是随机变量。试想，如果 $\hat{\beta}_1$ 是非随机的，标准差为 0，即使 $\hat{\beta}_1$ 再小，只要不为 0，我们也不会怀疑 β_1 可能为 0（因为无偏，再加上 $\hat{\beta}_1$ 方差为 0，此时 $\hat{\beta}_1 = \beta_0$ ）。而如果 $\hat{\beta}_1$ 很小（不为 0），但其方差更小，说明它接近一个常数，以此推测 β_0 不等于 0 是合理的。
8. 由 $R^2 = ESS/TSS = 1 - SSR/TSS$ 得出 $ESS = R^2 \times TSS$ ， $SSR = (1 - R^2) \times TSS$ 。由此得出

$$F = (ESS/1) / [SSR/(n-2)] = (R^2 \times TSS) / [(1 - R^2) \times TSS / (n-2)] = [R^2 / (1 - R^2)] \times (n-2)$$
 。用 F 检验更为合理，它建立在严格的假设检验之上。 R^2 只是一个主观指标，没有严格的标准。
9. 此时只用到矩条件 $E(u) = E(Y - \beta_0) = 0$ 。根据类比原则得出样本矩条件

$$n^{-1} \sum_{i=1}^n (Y_i - \hat{\beta}_0) = 0$$

从中解出 OLS 估计 $\hat{\beta}_0 = \bar{Y}$ 。

10. (1) 表明采用了 White 方法计算 OLS 估计的标准误；
- (2)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
X	0.493775	0.090583	5.451075	0.0000
C	-0.550065	0.292620	-1.879794	0.0729
R-squared	0.588892	Mean dependent var		0.840800
Adjusted R-squared	0.571018	S.D. dependent var		1.023885
S.E. of regression	0.670611	Akaike info criterion		2.115364
Sum squared resid	10.34355	Schwarz criterion		2.212875

11. (1) ~ (5) 可采用 EViews7.2 进行计算
- (6) 截距系数表示当研发投入为 0 时，企业申请的专利数量。斜率系数则表示，研发投入每增加一个百分点带来的企业申请专利数量增加的百分比（由于是两个变量先取对数回归，因此回归系数为弹性，即增加百分不/增加百分比）。从理论上讲，没有投入就没有产出，因此截距系数应该为 0，但在实际操作中，模型要截距项，以保证拟合优度计算按 (3.18) 平方和分解公式进行（因为 EViews 对不带截距项的回归仍然采用 (3.18) 计算拟合优度，从而会出现负的或者大于 1 的拟合优度值）。
12. (1) ~ (2) 用 EViews 软件操作得到
- (3) 从 (1)、(2) 的回归结果看出，以工作年限为自变量的回归模型中对应工作年限

变量的斜率参数估计值为 0.254， t -统计量的值为 0.546，对应概率值为 0.5859，表明工作年限对工资没有明显的解释能力。整个模型的拟合优 0.0022，接近 0。以任 1 企业高管的年限为自变量的回归中，高管年限变量的斜率参数估计值为 1.695， t -统计量的值为 2.033，对应概率值为 0.044，在 5% 显著水平上拒绝原假设，表明任高管年限对对工资有明显的解释能力。在确定高管薪水时，以任高管年限为解释变量的模型更合理。