22920212204392 黄勖 第八次作业

chapter 8

作业

- 作业:P199
- 8.9-12,8.14 8.20 8.27

【8.9】填空。

a. 在接近多重共线性的情况下，回归系数的标准误趋于________，t 值趋于________。

b. 在完全多重共线性的情况下，普通最小二乘估计量是________，其方差是________。

c. 在其他情况不变的条件下，VIF 越高，则普通最小二乘估计量的________越高。

解答：（a）大，小。

（b）没有定义的，没有定义的。

（c）方差。

【8.10】判断正误并说明理由。

a. 尽管存在完全多重共线性，但普通最小二乘估计量仍然是最优线性无偏估计量(BLUE)。

b. 在高度多重共线性的情况下，无法评估一个或多个偏回归系数的显著性。

c. 如果辅助回归表明某一 R_i^2 较高，则表明一定存在高度共线性。

d. 较高的相关系数并不一定表明存在高度多重共线性。

e. 如果分析的目的仅仅是为了预测，则多重共线性并无大碍。

解答：（a）错误，当模型存在完全的多重共线性时，OLS 估计量是没有定义的。

（b）正确。

（c）不一定，较高的 R^2 可以通过较小的 σ^2、相关解释变量较高的方差，或两者的共同作用来弥补。

（d）正确，有时两解释变量之间的相关系数较高，但是当模型中其他变量的影响保持恒定时，这两个解释变量的偏相关系数就有可能会变小。

（e）不一定，如果数据中观察到的共线性会在未来持续下去，那么这个论断就是正确的，否则就是错误的。

【8.11】在用失业、货币供给、利率、消费支出等经济时间序列数据进行回归分析时，常常怀疑存在多重共线性，为什么？

解答：因为大多数时间序列变量都会同经济周期及（或）经济增长趋势（反映了经济增长率）的发展方向，因此在消费者价格指数（CPI）对货币供给和失业率的回归中，解释变量之间很可能出现共线性问题。

【8.12】考虑下面模型：

$$Y_t = B_1 + B_2X_t + B_3X_{t-1} + B_4X_{t-2} + B_3X_{t-3} + u_t$$

其中，Y——消费；X——收入；t——时间。

模型表明：t 期的消费支出是同期收入以及前两期收入的线性函数。这类模型称为分布滞后模型，也称为动态模型（即模型涉及时间变化）。

a. 是否预期这类模型中存在多重共线性，为什么？

b. 如果怀疑存在多重共线性，那么如何“消除”？

解答：（a）是的。在经济周期中，许多变量（如收入）在 t 期和 $t-1$ 期时通常会向同一个方向变动，因此在经济周期的上扬阶段，收入的本期值通常会比滞后一期值高。

（b）有很多方法可以解决此问题，如柯依克变换、取一阶差分等。有些方法会在第 16 章进行详细介绍。

【8.14】表 8-5 给出了美国 1971～1986 年的年度数据。

表 8-5　美国对新轿车的需求

年　份	Y	X_2	X_3	X_4	X_5	X_6
1971	10 227	112.0	121.3	776.8	4.89	79 367
1972	10 872	111.0	125.3	839.6	4.55	82 153
1973	11 350	111.1	133.1	949.8	7.38	85 064
1974	8 775	117.5	147.7	1 038.4	8.61	86 794
1975	8 539	127.6	161.2	1 142.8	6.16	85 846
1976	9 994	135.7	170.5	1 252.6	5.22	88 752
1977	11 046	142.9	181.5	1 379.3	5.50	92 017
1978	11 164	153.8	195.4	1 551.2	7.78	96 048
1979	10 559	166.0	217.4	1 729.3	10.25	98 824
1980	8 979	179.3	246.8	1 918.0	11.28	99 303
1981	8 535	190.2	272.4	2 127.6	13.73	100 397
1982	7 980	197.6	289.1	2 261.4	11.20	99 526
1983	9 179	202.6	298.4	2 428.1	8.69	100 834
1984	10 394	208.5	311.1	2 670.6	9.65	105 005
1985	11 039	215.2	322.2	2 841.1	7.75	107 150
1986	11 450	224.4	328.4	3 022.1	6.31	109 597

注：Y——新轿车的销售量（千辆），未做季节调整；
X_2——新车消费者价格指数，1967 年 = 100，未做季节调整；
X_3——城市居民消费者价格指数，1967 年 = 100，未做季节调整；
X_4——个人可支配收入（PDI）（10 亿美元），未做季节调整；
X_5——利率（%），金融公司直接支付的票据利率；
X_6——城市就业劳动力（千人），未做季节调整。
资料来源：*Business Statistics*，1986，a Supplement to the *Current Survey of Business*，U. S. Department of Commerce.

考虑下面的轿车总需求函数：

$$\ln Y_t = B_1 + B_2 \ln X_{2t} + B_3 \ln X_{3t} + B_4 \ln X_{4t} + B_5 \ln X_{5t} + B_6 \ln X_{6t} + u_t$$

其中，ln 表示自然对数。

a. 同时引入价格指数 X_2 和 X_3 的理论根据是什么？

b. 为什么在需求函数中引入“城市就业劳动力”？

c. 如何解释各偏斜率系数的经济意义？

d. 求上述模型的 OLS 估计值。

解答：（a）X_2 为新车消费者价格指数，而 X_3 为城市居民消费者价格指数。如果两者之间服从一种领先－滞后机制，则两者可能就不是同步变动的。

（b）其反映了劳动力市场的就业情况。在其他条件保持不变的前提下，就业水平越高，对汽车的需求量越大。

（c）本例为双对数模型，因此偏回归系数代表被解释变量对解释变量的偏弹性系数。

（d）用 $\ln Y_t$ 对所有的解释变量的自然对数进行回归，得到汽车的总需求函数的估计结果如下：

变量	系数	t 值
截距	11.058 2	0.508 6②
$\ln X_{2t}$	1.940 9	2.109 9②
$\ln X_{3t}$	−4.681 5	−2.547 5①
$\ln X_{4t}$	2.716 4	1.843 8②
$\ln X_{5t}$	−0.025 9	−0.210 6②
$\ln X_{6t}$	−0.582 1	−0.249 6②
$R^2 = 0.855\ 1$		

① 表示在 5% 的显著性水平下显著（双边检验）。
② 表示在 5% 的显著性水平下不显著（双边检验）。

【8.20】莱顿·托马斯（R. Leighton Thomas）在研究英国 1961 ~ 1981 年间砖、瓷、玻璃和水泥行业的生产函数时，得到如下结果：

1. $\log Q = -5.04 + 0.887 \log K + 0.893 \log H$

   se = (1.40)　(0.087)　(0.137)　$R^2 = 0.878$

2. $\log Q = -8.57 + 0.027\ 2t + 0.460 \log K + 1.285 \log H$

   se = (2.99)　(0.020 4)　(0.333)　(0.324)　$R^2 = 0.889$

其中，Q——生产指数（不变要素成本）；K——总资本存量（1975 年重置成本）；H——工作小时数；t——时间趋势（技术进步的替代量）。

其中，Q——生产指数（不变要素成本）；K——总资本存量（1975 年重置成本）；H——工作小时数；t——时间趋势（技术进步的替代量）。

括号中的数字是估计的标准差。

a. 解释这两个回归方程。

b. 在回归方程 1 中验证在 5% 的显著性水平下，偏斜率系数是统计显著的。

c. 在回归方程 2 中验证在 5% 的显著性水平下，t 和 $\log K$ 的系数是统计不显著的。

d. 如何解释模型 2 中变量 $\log K$ 的不显著性？

e. 如果得知 t 和 K 之间的相关系数为 0.980，那么能够得出什么结论？

f. 如果在模型 2 中 t 和 K 都是不显著的，那么，是接受还是拒绝假设：模型 2 中所有偏斜率系数同时为零，使用哪种检验？

g. 在模型 1 中，规模收益是多少？

解答：(a) 第一个模型中的回归系数为偏弹性。而在第二个模型中，$\log K$，$\log H$ 前的回归系数为弹性。趋势项 t 前的系数意味着在其他条件不变时，（指数化的）产出水平的年增长率为 2.7%。

(b) 回归系数的 t 值分别为 -3.600、10.195 和 6.518。它们在 5% 的显著性水平下是显著的，因为自由度为 18 的临界值为 2.101。

(c) 趋势项和 $\log K$ 回归系数的 t 值分别为 1.333 和 1.381，其在统计上是不显著的。

(d) 这可能意味着趋势项（在一定程度上反映了科技水平）和 $\log K$ 存在共线性。

(e) 虽然较高的相关系数不一定意味着共线性，但至少可以怀疑两者之间存在共线性。

(f) 可以拒绝该假设，因为 F 统计量（运用 R^2 进行计算）为 45.384 4，其在 1% 的显著性水平上是显著的，因为 $F_{3,17}=5.18$。

(g) 回归结果显示规模收益为 $0.887+0.893=1.780$，这意味着规模收益递增。

【8.27】艾斯特里欧（Asteriou）和霍尔（Hall）根据英国 1990 年第一季度至 1998 年第二季度的季度数据得到如下回归结果。应变量是 log（IM）= 出口的对数（括号内的是 t 值）。

解释变量	模型 1	模型 2	模型 3
Intercept	0.631 8	0.213 9	0.685 7
	(1.834 8)	(0.596 7)	(1.850 0)
Log（GDP）	1.926 9	1.969 7	2.093 8
	(11.411 7)	(12.561 9)	(12.132 2)
Log（CPI）	0.274 2	1.025 4	—
	(1.996 1)	(3.170 6)	0.119 5
Log（PPI）	—	−0.770 6	0.119 5
		(−2.524 8)	(0.878 7)
Adjusted − R^2	0.963 8	0.969 2	0.960 2

a. 解释每个方程。

b. 在模型 1 中（去掉变量 log（PPI）），在 5% 的显著水平下，log（CPI）的系数为正，并且是统计显著的，这是否有经济意义？

c. 在模型 3 中（去掉变量 log（CPI）），log（PPI）的系数为正，但不是统计显著

的，这是否有经济意义？

d. 在模型 2 中，两个价格变量的系数各自都是统计显著的，但是 log（CPI）的系数为正，log（PPI）的系数为负。如何解释这个结果？

e. 出现这样矛盾的结果是不是因为多重共线性？证明你的结论。

f. 如果 CPI 和 PPI 的相关系数为 0.981 9，那么是否表明存在多重共线性问题？

g. 你将选择上述三个模型中的哪一个？为什么？

解答：（a） 模型 1：英国的 GDP 每增长 1%，进口会增长 193%；而 CPI 每上升 1%，进口会上升 27%。但回归结果中只有 logGDP 是显著的。

模型 2：英国的 GDP 每增长 1%，进口会增长 197%；而 CPI 每上升 1%，进口会上升 103%；PPI 每上升 1%，进口将下降 77%。此时三个解释变量都是显著的。

模型 3：英国的 GDP 每增长 1%，进口会增长 209%；PPI 每上升 1%，进口将上升 12%。但此时只有 log GDP 是显著的。

（b） 这是不符合经济学常识的。

（c） 这也是不符合经济学常识的。

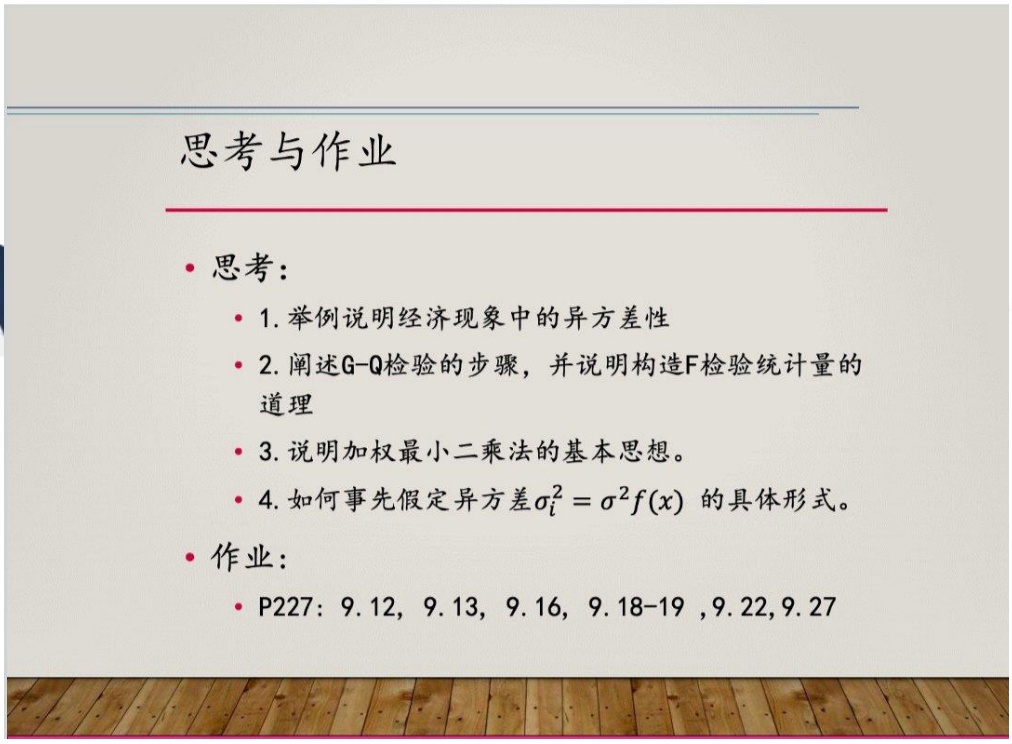
（d） 这可能是因为模型中存在多重共线性。

（e） 是的，正如之前所述一样。

（f） 是的，因为相关系数比 log *Imports* 同所有解释变量之间的总相关系数高。

（g） 没有必要将 PPI 和 CPI 都留在模型中，因为相关系数表明模型中可能存在多重共线性问题，包含了冗余信息。因此我们应在模型 1 和模型 3 中进行选择，由于模型 1 的判定系数较高且变量都是显著的，因此模型 1 更好。

chapter 9



1. 思考：异方差性是指随机变量的方差在不同取值或不同条件下不等的现象。它表示随机变量的方差不是恒定的，而是与某个因素或条件相关。下面是一个关于收入和消费之间关系的例子：

假设我们研究了一组人的收入和消费数据。在低收入水平下，人们的消费可能相对较低，方差也较小；而在高收入水平下，人们的消费可能相对较高，方差也较大。这就展现了收入和消费之间的异方差性，即收入水平的不同导致了消费方差的变化。

2. G-Q 检验是一种检验线性回归模型中误差项的异方差性的统计方法。它的步骤如下：
    - 第一步：先拟合一个普通的线性回归模型，得到残差项。
    - 第二步：将残差项的平方与自变量之间进行回归，得到一个新的回归模型。
    - 第三步：检验新模型中自变量的系数是否显着不为零，若显着，则表明存在异方差性。

构造 F 检验统计量的原理是比较新模型中自变量的系数与零之间的偏离程度。如果偏离程度较大且显着，则说明存在异方差性。

3. 加权最小二乘法是一种用于处理异方差性的回归方法。其基本思想是对每个样本点进行加权，使得具有较小方差的点在拟合过程中起较大的作用，而具有较大方差的点则起较小的作用。

加权最小二乘法的步骤如下：

- 第一步：确定加权函数，即根据样本点的方差大小为每个样本点分配权重。
- 第二步：根据加权函数对样本进行加权，得到加权样本。
- 第三步：利用加权样本进行最小二乘拟合，得到回归方程。
- 第四步：对回归方程进行统计推断和模型评估。

通过加权最小二乘法，可以更准确地估计回归方程的系数，并减小异方差对拟合结果的影响。

4. 在实际应用中，事先假定异方差 $\sigma^2 = \sigma^2 f(x)$ 的具体形式可以根据经验或理论知识来确定。常见的假定形式包括：
    - 线性假定：假设异方差与自变量之间存在线性关系，即 $\sigma^2 = \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k$。
    - 幂函数假定：假设异方差与自变量之间存在幂函数关系，即 $\sigma^2 = \alpha x_{\beta}$，其中 α 和 β 是待估计的参数。
    - 指数函数假定：假设异方差与自变量之间存在指数函数关系，即 $\sigma^2 = \exp(\alpha + \beta x)$。

根据具体问题和数据特点，可以选择适合的异方差形式。这种事先假定异方差形式的目的是为了在建模过程中更好地处理异方差性，并得到更准确的统计推断和拟合结果。

【9.12】1964 年，曾经对 9 966 名经济学家进行了调查，数据如下：

年龄	工资中值（美元）	年龄	工资中值（美元）
20 ~ 24	7 800	50 ~ 54	15 000
25 ~ 29	8 400	55 ~ 59	15 000
30 ~ 34	9 700	60 ~ 64	15 000
35 ~ 39	11 500	65 ~ 69	14 500
40 ~ 44	13 000	70 +	12 000
45 ~ 49	14 800		

资料来源："The Structure of Economists' Employment and Salaries," Committee on the National Science Foundation Report on the Economics Profession, *American Economics Review*, vol. 55, no. 4, December 1965, p. 36.

a. 建立适当的模型解释工资与年龄的关系。为了进行回归，假设工资中值对应于年龄区间的中点。

b. 假设误差方差与年龄成比例变动，变换数据，求 WLS 回归。

c. 假设误差方差与年龄的平方成比例变动，求 WLS 回归。

d. 哪一个假设看来更可行？

解答：令 Y = 工资中位数，X = 年龄，（假设最后一组的值为 $X = 72$）。

(a) $\hat{Y}=6\ 419.818\ 2+127.818\ 2X$

$t=(3.640\ 8) \quad (3.594\ 6) \qquad r^2=0.589\ 4$

(b) $\frac{Y}{\sqrt{X}}=5\ 133.854\ 8\left(\frac{1}{\sqrt{X}}\right)+155.179\ 1\sqrt{X}$

$t=(3.670\ 2) \qquad (4.876\ 4) \qquad r^2=0.960\ 8$

注：这是一个没有截距项的回归，这里的 r^2 是根据原始公式计算得出的。当没有截距项的时候，Eviews 得到的 r^2 经常是负的。

(c) $\frac{Y}{X}=4\ 216.910\ 5\left(\frac{1}{X}\right)+177.483\ 6$

$t=(3.859\ 6) \qquad (6.213\ 8) \qquad r^2=0.623\ 4$

(d) 表面上看（b）和（c）的转换都降低了系数的标准误，减弱了数据中异方差程度。做出（b）和（c）残差的散点图看看是否存在系统模式。如果有的话，用帕克检验或格莱泽检验进一步验证数据到底存不存在异方差。

【9.13】异方差的 Spearman 秩相关检验。我们用工资回归方程（9-3）说明该检验步骤。

a. 从回归方程（9-3）中求得残差 e_i。

b. 求残差的绝对值 $|e_i|$。

c. 将教育 X_i 和 $|e_i|$ 按降序（从高到低）或升序（从低到高）排列。

d. 对于每个观察值，取两列间的差，称为 d_i。

e. 计算 Spearman 秩相关系数 r_s，定义为：

$$r_s=1-6\left[\frac{\sum d_i^2}{n(n^2-1)}\right]$$

其中，n 表示样本观察值的个数。

如果在 $|e_i|$ 和 X_i 之间存在系统关系，则二者之间的秩相关系数是统计显著的，这种情况下可能存在异方差。

给定零假设：真实总体秩相关系数为零，且 $n>8$，则可以证明

$$\frac{r_s\sqrt{(n-2)}}{\sqrt{1-r_s^2}}\sim t_{n-2}$$

服从自由度为 $n-2$ 的学生 t 分布。

因此，在实际应用中，根据 t 检验，若秩相关系数是显著的，则不能拒绝零假设：存在异方差问题。用这个方法检验本章的工资数据，确认数据是否存在异方差。

解答： 斯皮尔曼秩相关系数为 0.440 7。将这个值代入给定公式，得到 t 值为 1.963 6。自由度为 16，显著性为 5%（单边检验），临界 t 值为 1.746。因此，所得的 t 值在 5% 的水平下是显著的。说明数据存在异方差。

【9.16】在平均工资对就业人数的回归分析中（包括 30 个公司的随机样本），得到如下回归结果：

$$\hat{W}=7.5+0.009N$$

$$t=\text{N. A.}\ (16.10) \quad R^2=0.90 \tag{1}$$

$$\frac{\hat{W}}{N}=0.008+7.8\frac{1}{N}$$

$$t=(14.43)(76.58)\quad R^2=0.99 \tag{2}$$

a. 如何解释这两个回归？

b. 从方程（1）到（2）做了哪些假设？是否担心存在异方差？

c. 能否把两个回归方程中的斜率和截距联系起来？

d. 能否比较两个回归方程中的 R^2？为什么？

解答：（a）回归方程（1）中的斜率说明了如果雇员增加 1 人，那么平均工资上涨 0.009 美元。乘以 N 后，模型（2）的斜率同模型（1）相等。

（b）作者认为模型中存在异方差，且假设误差项的方差同 N 的平方成比例。

（c）如（a）中提到的，两个斜率和两个截距相等。

（d）因为两个因变量不同，所以两个 R^2 不能进行直接比较。

【9.18】表 9-5（参见网上教材）给出了 20 个国家五项社会经济指标的有关数据，样本分为四个收入等级：低收入（人均年收入 500 美元以下），中等偏低收入（人均年收入在 500～2 200美元之间），中等偏上收入（人均年收入在 2 200～5 500 美元之间），以及高收入（人均年收入超过 5 500 美元）。表中前 5 个国家属于低收入国家，接下来的五个国家属于中等偏低收入国家，以此类推。

a. 建立一个包括所有 5 个解释变量的回归模型。先验地，你认为人口增长率 X_4 和每日卡路里吸收量 X_5 对婴儿死亡率有什么样的影响？

b. 对回归方程进行估计，并检验你的预期是否正确。

c. 如果在上述方程中遇到了多重共线性问题，该怎么办？可以采取任何你认为正确的措施。

解答：（a）先验的结论为，卡路里吸收量对婴儿死亡率存在负效应，对人口增长率存在正效应。

（b）Eviews 回归结果如下：

Dependent Variable: IMOR
Sample: 1 20

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	172.619 5	52.455 98	3.290 749	0.005 0
PCGNP	−0.002 502	0.001 535	−1.629 641	0.124 0
PEDU	−1.279 618	0.316 722	−4.040 198	0.001 1
POPGROWTH	6.379 603	7.045 706	0.905 460	0.379 5
CSPC	−0.001 363	0.018 708	−0.072 873	0.942 9
R-squared	0.815 002	F-statistic		16.520 53
Adjusted R-squared	0.765 670	Prob(F-statistic)		0.000 023

注：我们在回归结果中给出了 F 统计量的值和相应的 p 值。

人口增长率和卡路里吸收量的关系和预期相一致。

（c）之前的回归中只有一个系数是统计显著的，但 F 值很大，且是高度显著的，这是存在多重共线性的典型特征。去掉变量人口增长（POPGROWTH）和人均 GNP（PCGNP），得到结果如下：

Dependent Variable: IMOR
Sample: 1 20

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	250.099 6	28.191 64	8.871 411	0.000 0
PEDU	−1.210 991	0.329 756	−3.672 385	0.001 9
CSPC	−0.030 999	0.013 099	−2.366 448	0.030 1
R-squared	0.760 462	F-statistic		26.984 94
Adjusted R-squared	0.732 281	Prob(F-statistic)		0.000 005

现在，两个自变量都显著了。

【9.19】如果在习题9.18的模型中不包括X_4和X_5两个解释变量，对回归结果进行异方差检验，按照怀特异方差检验方法，得到如下回归结果：（注：为了节省篇幅，只给出了t统计量和它们的p值。这些结果通过EViews软件实现）

$$e_i^2 = -15.76 + 0.3810X_{2i} - 4.5641X_{3i} + 0.000005X_{2i}^2 + 0.1328X_{3i}^2 - 0.0050X_{2i}X_{3i}$$

$$t = (-0.01)\quad (0.60)\quad (-0.13)\quad (0.87)\quad (0.56)\quad (-0.85)$$

$$pvalue = (0.989)\quad (0.556)\quad (0.895)\quad (0.394)\quad (0.581)\quad (0.400)$$

$$R^2 = 0.23$$

a. 如何解释上述回归方程？

b. 回归方程是否表明存在异方差问题。

c. 如果方程存在异方差，如何消除异方差？

解答：（a）回归结果显示辅助回归中所有变量的回归系数均不显著。

（b）由于系数不显著且R^2同样本容量之间的乘积小于自由度为5的χ^2分布的临界值，因此模型中不存在异方差。

（c）可以通过变换后模型的残差图来进行检验。可以通过怀特修正法来修正变换后的回归模型，以保证转换后的模型不存在异方差。

【9.22】利用Exper和Wagef作为缩减因子估计方程（9-10）到方程（9-12）。

解答：对于经验（Experience）：

$$|e_i| = 3.165 + 0.0108Exper$$
$$t = (12.77)\quad (0.95)\qquad r^2 = 0.002$$

$$|e_i| = 3.0989 + 0.0655\sqrt{Exper}$$
$$t = (7.79)\quad (0.70)\qquad r^2 = 0.001$$

$$|e_i| = 3.2480 + 0.9528\frac{1}{\sqrt{Exper}}$$
$$t = (19.19)\quad (1.15)\qquad r^2 = 0.003$$

对于工资（Wage）：

$$|e_i| = 0.3390 + 0.3314Wage$$
$$t = (1.43)\quad (14.63)\qquad r^2 = 0.291$$

$$|e_i| = 1.5942 + 1.6988\sqrt{Wage}$$
$$t = (-3.29)\quad (10.59)\qquad r^2 = 0.177$$

$$|e_i| = 3.7773 + 2.894\frac{1}{\sqrt{Wage}}$$
$$t = (13.38)\quad (-1.69)\qquad r^2 = 0.005$$

【9.27】表 9-8（参见网上教材）给出了《财富》500 强企业中的 447 个高管薪水数据。Salary 表示 1999 年薪水和奖金；totcomp 表示 1999 年 CEO 总薪水；tenure 表示任职 CEO 的年数；age 表示 CEO 的年龄；sale 表示 1998 年公司销售收入；profit 表示 1998 年公司利润；assets 表示 1998 年公司总资产。

a. 利用表中提供的数据估计下面的方程，并利用布鲁尔什－培甘检验是否存在异方差。

$$Salary_i = B_1 + B_2 tenure_i + B_3 age_i + B_4 sales_i + B_5 profits + B_6 assets_i + u_i$$

异方差是一个严重的问题吗？

b. 利用 ln*Salary* 作为应变量建立回归模型。异方差有所改善吗？

c. 做薪水对各个解释变量的散点图。能否看出哪些变量导致了异方差问题？你将采取什么样的方法？最终选择什么样的模型？

d. 求怀特稳健标准误。有什么明显不同吗？

解答：（a）Eviews 的回归结果如下：

Dependent Variable: SALARY
Sample: 1 447
Included observations: 447

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	998.709 5	623.695 4	1.601 277	0.110 0
TENURE	31.672 79	9.465 097	3.346 272	0.000 9
AGE	5.492 393	11.460 86	0.479 230	0.632 0
SALES	0.014 287	0.006 614	2.160 040	0.031 3
PROFITS	0.141 302	0.068 845	2.052 471	0.040 7
ASSETS	0.007 630	0.001 326	5.754 849	0.000 0

R-squared	0.248 829	Mean dependent var	2 027.517
Adjusted R-squared	0.240 312	S.D. dependent var	1 722.566
S.E. of regression	1 501.390	Akaike info criterion	17.479 50
Sum squared resid	9.94E+08	Schwarz criterion	17.534 57
Log likelihood	-3 900.669	F-statistic	29.216 60
Durbin-Watson stat	2.014 806	Prob(F-statistic)	0.000 000

怀特检验结果如下：

White Heteroskedasticity Test:

F-statistic	2.089 313	Probability	0.024 109
Obs*R-squared	20.440 73	Probability	0.025 349

p 值为 0. 025 3，所以数据中存在明显的异方差。请读者用布鲁尔什－培甘（Breusch-Pagan）检验再次验证一下，当做练习。

（b）对数－线性（log-lin）模型和怀特异方差检验的结果如下：

Dependent Variable: LN_SAL
Sample: 1 447
Included observations: 447

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	6.753 659	0.236 823	28.517 78	0.000 0
TENURE	0.008 251	0.003 594	2.295 836	0.022 2
AGE	0.007 228	0.004 352	1.661 010	0.097 4
SALES	6.09E-06	2.51E-06	2.425 693	0.015 7
PROFITS	5.72E-05	2.61E-05	2.186 738	0.029 3
ASSETS	2.03E-06	5.03E-07	4.035 537	0.000 1

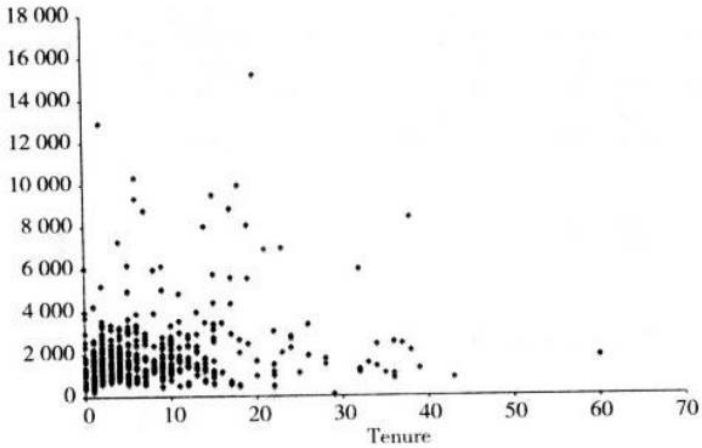
R-squared	0.208 984	Mean dependent var	7.391 898
Adjusted R-squared	0.200 016	S.D. dependent var	0.637 388
S.E. of regression	0.570 091	Akaike info criterion	1.727 292
Sum squared resid	143.326 8	Schwarz criterion	1.782 359
Log likelihood	−380.049 7	F-statistic	23.302 17
Durbin-Watson stat	1.920 217	Prob(F-statistic)	0.000 000

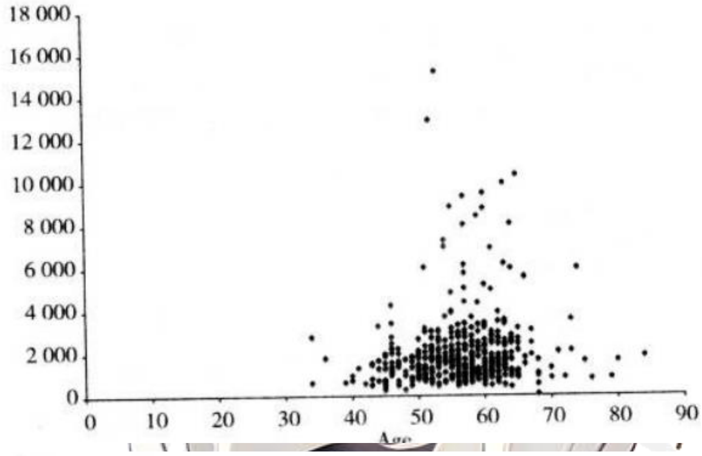
White Heteroskedasticity Test:

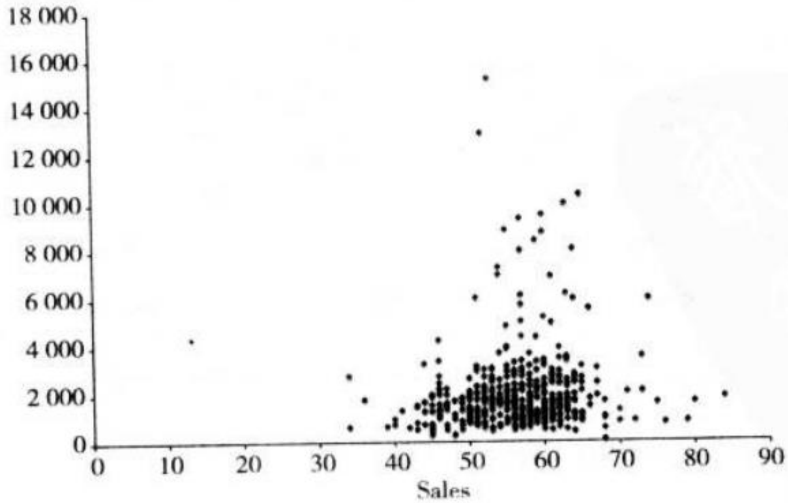
F-statistic	2.581 930	Probability	0.004 784
Obs*R-squared	24.990 78	Probability	0.005 363

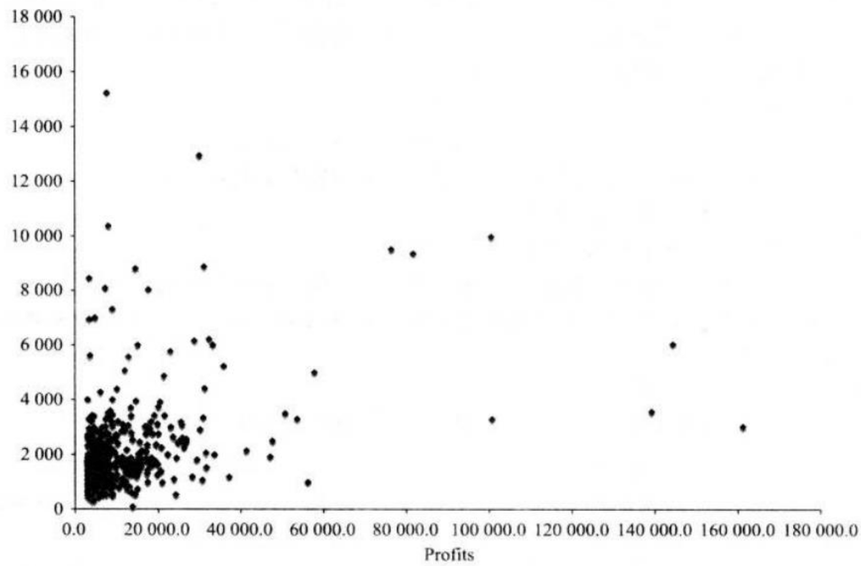
很明显，数据中依旧存在异方差。

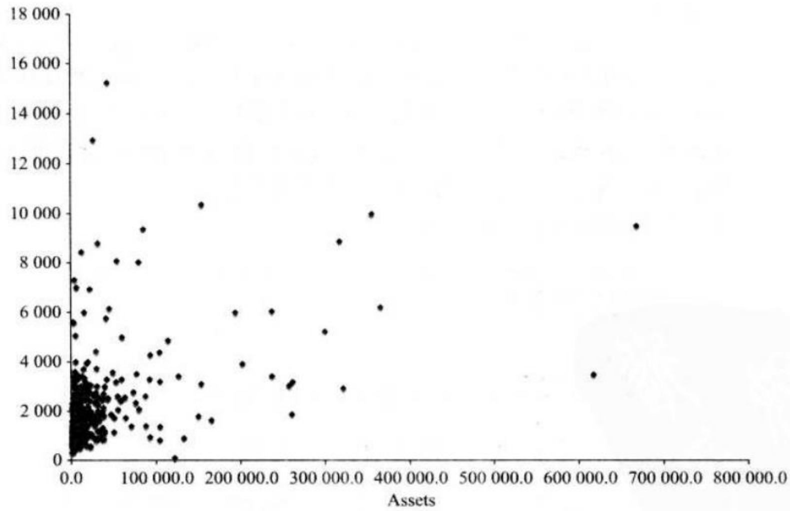
（c）











根据散点图，本例中的异方差可分别用多个变量来解释，读者可以在相应的修正方法中找出最有效的方法来降低模型的异方差程度。

chapter 10

作业

- 第233页：10.15,10.17,10.21
- 思考:10.14,10.16,10.20
- 阅读:附录10A-B

【10.15】考虑如下回归模型：

$$\hat{Y}_t = -49.4664 + 0.88544X_{2t} + 0.09253X_{3t};\quad R^2 = 0.9979; d = 0.8755$$
$$t = (-2.2392)\quad (70.2936)\quad (2.6933)$$

式中，Y——个人消费支出（1982年美元价，10亿美元）；X_2——个人可支配收入（1982年美元价，10亿美元）（PDI）；X_3——道·琼斯工业平均股票指数。回归利用了1961～1985年间美国数据。

a. 回归残差存在一阶自相关吗？你是如何知道的？

b. 利用德宾两阶段方法，将上述回归转换成式（10-15），结果如下：

$$\hat{Y}_t^* = -17.97 + 0.89X_{2t}^* + 0.09X_{3t}^*;\quad R^2 = 0.9816;\quad d = 2.28$$
$$t = \qquad (30.72)\quad (2.66)$$

自相关的问题解决了吗？你是如何知道的？

c. 比较初始回归和变换后的回归，PDI的t值急剧下降，这一变化表明了什么？

d. 根据变换后模型获得的d值能否确定变换后的数据存在自相关？

解答：(a) 当$n=25$，$k'=2$时，显著性水平为5%的d统计量临界值为1.206和1.550。因为0.8755小于1.206，因此模型中存在正（一阶）自相关。

(b) 因为本例并不适宜进行杜宾检验，因此杜宾检验的结果并不可信。如果可以得到原始数据的话，可以通过游程检验来判断是否存在自相关。

(c) 由于自相关的存在，造成传统的标准误估计量有偏，在本例中标准误可能被低估，从而造成t统计量被高估，这一点可以从修正后模型的回归结果中看出。

(d) 答案见（b）

【10.17】考虑表 10-7（参见网上教材）给出的 1980 ~ 2006 年间股票价格和 GDP 的数据。

a. 估计 OLS 回归：

$$Y_t = B_1 + B_2 X_t + u_t$$

b. 根据 d 统计量判定数据中是否存在一阶自相关。

c. 如果存在，用 d 值估计自相关参数 ρ。

d. 利用估计的 ρ 对数据变换，用 OLS 法估计广义差分方程（10-14）：①舍去第一个观察值；②包括第一个观察值。

e. 重复（b），根据形如式（10-20）的残差估计 ρ 值。利用估计的 ρ 值，估计广义差分方程（10-14）。

f. 利用一阶差分方法将模型变换成方程（10-17）的形式，并对变换后的模型进行估计。

g. 比较（d）、（e）和（f）的回归结果。你能得出什么结论？在变换后模型中还存在自相关吗？你是如何知道的？

解答：（a）回归结果为：

$$\hat{Y}_t = -2\,015.2 + 0.772\,3X_t$$

$$t = (-6.58) \qquad (19.52) \qquad R^2 = 0.938\,0;\quad d = 0.428\,5$$

（b）当 $n=27$，$k'=1$ 时，显著性水平为 5% 的 d 统计量临界值为 1.089 和 1.233。因为 0.428 5 小于 1.089，因此模型中存在正（一阶）自相关。

（c）$\hat{\rho} = 1 - d/2 = (1 - 0.428\,5/2) = 0.785\,8$

（d）不包含首个观测值的回归结果为：

$$\hat{Y}_t^* = -617.9 + 0.862\,4X_t^*$$

$$t = (-3.01) \qquad (8.47) \qquad R^2 = 0.749;\quad d = 0.911\,8$$

包含首个观测值的回归结果为：

$$\hat{Y}_t^* = -642.1 + 0.867X_t^*$$

$$t = (-3.13) \qquad (8.48) \qquad R^2 = 0.742;\quad d = 0.924\,8$$

（e）$\hat{e}_t = 0.768\,9e_{t-1}$

$t =$ （6.26）

注：该模型中没有截距项（为什么），因此 $\hat{\rho} = 0.768\,9$

不包含首个观测值的回归结果为：

$$\hat{Y}_t^* = -927.8 + 0.816\,6X_t^*$$

$$t = (-4.42) \qquad (12.76) \qquad R^2 = 0.872;\quad d = 0.793\,1$$

包含首个观测值的回归结果为：

$$\hat{Y}_t^* = -958.5 + 0.823X_t^*$$

$$t = (-4.70) \qquad (13.10) \qquad R^2 = 0.873;\quad d = 0.820\,9$$

（f）取一阶差分后有：

$$\Delta\hat{Y}_t = 0.868\,4\Delta X_t$$

$$t = (4.75) \qquad d = 0.931\,5$$

注：修正后模型截距所对应的样本值应为 $(1-\rho)$，而不再是 1。

（g）从上述结果中可以看出，不论是否包含首个观测值，修正后模型同原始模型的结果都有一定差距。读者可以通过第 10 章附录 A 中介绍的游程检验来对上述模型进行评价。

【10.21】在例 7-3 进口支出（Y）对个人可支配收入（X）回归中，考虑如下模型：

	模型 1	模型 2	模型 3
截距	-136.16	22.69	12.18
X	0.208 2	0.297 5	0.038 2
Time	—	-18.525	-3.045
Time-squared	—	—	0.965 9
校正的 R^2	0.969	0.984	0.994
d	0.216	0.341	1.611

a. 这些结果表明是否存在自回归？

b. 如何解释模型 3 的时间项和时间平方项？

注：在 5% 或更低的显著水平下，上面三个模型的系数都是统计显著的。

解答： 在 5% 的显著性水平下，通过杜宾检验可知，模型 1 存在正自相关，因为模型的 d 统计量小于 d 值的下限 1.288（$d_L = 1.288$）。模型 2 也存在正自相关，因为模型的 d 统计量小于 d 值的下限 0.341 1（$d_L = 1.245$）。模型 3 不存在（一阶）自相关，因为模型的 d 统计量 1.611 大于 d 值的上限（$d_U = 1.474$）。

从本例中可知，如果模型存在设定误差的话，d 检验则检验的是模型设定误差而不是纯的自相关。

【10.14】德宾两阶段法估计 ρ。广义差分方程（10-14）写成如下等价形式：

$$Y_t = B_1(1-\rho) + B_2X_t - \rho B_2X_{t-1} + \rho Y_{t-1} + v_t$$

第一阶段，德宾建议以 Y 作为应变量，X_t、X_{t-1} 和 Y_{t-1} 作为解释变量进行回归。Y_{t-1} 的系数提供了 ρ 的一个估计量，因此得到的 ρ 是一致估计量；也就是说，对大样本，它是真实 ρ 的一个好的估计值。

第二阶段，利用从第一阶段中获得的 ρ 对数据变换，并估计广义差分方程（10-14）。

利用德宾两阶段法估计第 7 章讨论的美国进口支出数据，并将得到的结果与初始回归结果做比较。

解答： $\hat{Y}_t = -117.801\,4 + 0.260\,8X_t - 0.629\,X_{t-1} + 0.656\,2Y_{t-1}$

$t = (-1.879\,6)\quad(2.621\,9)\quad(-1.421\,0)\quad(2.809\,6)\qquad R^2 = 0.954\,7$

ρ 的估计值为 0.656 2。

运用经过修正的 X，Y 进行回归的结果为：

$$\hat{Y}_t^* = -120.328\,8 + 0.179\,0X_t^*$$

$$t = (-1.238\,3)\qquad(4.293\,6)\qquad R^2 = 0.506\,0$$

注：模型中的首个观测值经过了 Prais-Winsten 变换，且修正后模型截距的所对应的样本值应为（$1-\rho$），而不再是 1。

【10.16】德宾 h 统计量。在形如式（10-7）的自回归模型中：

$$Y_t = B_1 + B_2X_t + B_3Y_{t-1} + v_t$$

通常的 d 统计量不适合用于诊断自相关。对于这类模型，德宾建议用 h 统计量来代替 d 统计量，h 统计量定义为：

$$h \approx \hat{\rho}\sqrt{\frac{n}{1 - n \cdot \mathrm{var}(b_3)}}$$

式中，n——样本容量；$\hat{\rho}$——自相关系数 ρ 的估计量；var（b_3）——滞后变量 Y 的系数 B_3 的估计量的方差。

德宾证明了对于大样本而言，若零假设 $\rho = 0$，h 统计量服从均值为 0、方差为 1 的标准正态分布：

$$h \sim N(0,1)$$

因此，如果计算的 h 统计量超过了 h 的临界值，则拒绝 $\rho = 0$ 这一零假设；如果没有超过临界的 h 值，则不能拒绝无（一阶）自相关的零假设。此外，h 公式中的 $\hat{\rho}$ 可以利用本章讨论的任意一种方法获得。

现在考虑 1948～1949 年到 1964～1965 年间印度的货币需求函数：

$$\widehat{\ln M_t} = 1.6027 - 0.1024\ln R_t + 0.6869\ln Y_t + 0.5284\ln M_{t-1}$$

$$\mathrm{se} = (1.2404) \quad (0.3678) \quad (0.3427) \quad (0.2007) \quad R^2 = 0.9227$$

$$d = 1.8624$$

式中，M——实际现金余额；R——长期利率；Y——实际国民收入。

a. 求 h 统计量，并检验假设：上述回归中不存在一阶自相关。

b. 德宾－沃森 d 统计量为 1.8624。说明为什么本例不适合用 d 统计量，但可以用 d 统计量估计 ρ（$\hat{\rho} = 1 - d/2$）。

解答：（a）运用本例中所给出的 d 统计量可知，ρ 的估计值为 $\left(1 - \frac{1.8624}{2}\right) = 0.0688$。根据 h 统计量的计算公式可知：

$$h \approx \left(0.0688\sqrt{\frac{17}{1 - 17(0.0403)}}\right) = 0.5055$$

显然，h 统计量是不显著的，这意味着数据中并不存在自相关。但应注意本例中样本容量不大，所以得到的结论也还是值得商榷的。

（b）自回归模型（本例中所涉及的模型）的 d 统计量一般都在 2 左右，即意味着数据中不存在自相关问题。所以通过 d 统计量来检验这类模型自相关问题本身就是不合理的。

【10.20】根据 d 统计量估计的泰尔－纳加尔（Theil-Nagar）ρ。泰尔和纳加尔认为，在小样本中不应用（$1 - d/2$）来估计 ρ，而是用：

$$\hat{\rho} = \frac{n^2(1 - d/2) + k^2}{n^2 - k^2}$$

式中，n——样本数；d——德宾－沃森 d 统计量；k——待估系数的个数（包括截距项）。

证明：当 n 充分大时，ρ 的估计值等于用简单公式（$1 - d/2$）估计的 ρ 值。

解答：分子分母同时除以 n^2 有，

$$\hat{\rho} = \frac{[(1 - d/2) + k^2/n^2]}{1 - \frac{k^2}{n^2}}$$

当 n 趋于无穷时，上式趋向于（$1 - d/2$）。