

作业

作业P180

7. 12, 7. 14, 7. 15, 7. 16

【7. 12】考虑如下模型：

模型 I：消费_{*i*} = $B_1 + B_2$ 收入_{*i*} + u_i

模型 II：消费_{*i*} = $A_1 + A_2$ 财富_{*i*} + v_i

a. 如何确定哪个模型是“正确的”模型？

b. 假定同时做消费对收入和财富的回归。这有助于模型选择吗？

解答：(a) 虽然在大多数中级宏观经济学的教科书中将（凯恩斯）消费函数看成是收入的函数，但是有不少经济学家认为财富拥有量也是决定消费支出的重要因素之一。因此无法仅仅通过纯经济理论来对模型 I 和模型 II 做出选择。

(b) 令 $Consumption = C_1 + C_2 Income + C_3 Wealth + w$ ，其中 $Consumption$ 为消费支出， $Income$ 为收入水平， $Wealth$ 为财富拥有量。如果实证分析研究表明， C_2 、 C_3 在统计上都是显著的，则模型 I 和模型 II 都是不正确的。如果 C_2 是显著的，而 C_3 是不显著的，则选择模型 I 更为合适。如果 C_3 是显著的，而 C_2 是不显著的，则选择模型 II 是较为合适的。但在选择模型中也应注意在第 8 章讨论的共线性问题。

【7.14】表 7-6（参见网上教材）给出了 1954~1981 年美国普通股实际收益率（ Y ）、产出增长率（ X_2 ）和通货膨胀（ X_3 ）的数据。

- 做 Y 对 X_3 的回归。
- 做 Y 对 X_2 和 X_3 的回归。
- 尤金（Eugene Fama）教授指出：“实际股票收益和通货膨胀之间的简单负相关关系是虚假的（或者说是错误的），它是两个结构关系的结果：一个是当前实际股票收益和预期产出增长率之间的正相关关系，另一个是预期产出增长率和当前通货膨胀之间的负相关关系。”根据这个观点评论以上两个回归结果。
- 利用 1956~1976 年的数据进行（b）部分的回归，略去 1954 年和 1955 年的数据（这两年的股票收益行为异常）。将回归结果与（b）部分的回归结果比较，如果两者存在差异的话，简单分析产生差异的原因。
- 假定根据 1956~1981 年的数据进行回归，但区分为 1956~1976 年和 1977~1981 年两个阶段，如何进行这样的回归？（提示：考虑虚拟变量。）

解答：（a）回归结果如下

$$\hat{Y} = 23.9869 - 4.3756X_3$$

$$t = (4.5820) \quad (-4.2805) \quad R^2 = 0.4134$$

（b）回归结果如下：

$$\hat{Y} = 3.5318 + 3.9433X_2 + 2.4994X_3$$

$$t = (0.4354) \quad (3.0487) \quad (-2.3098) \quad R^2 = 0.5724$$

（c）从下列回归结果可知，法马的观点是正确的：

$$(i) \hat{Y} = -12.2815 + 5.6424X_2$$

$$t = (-2.6137) \quad (4.9099) \quad R^2 = 0.4811$$

$$(ii) \hat{X}_2 = 5.1873 - 0.4758X_3$$

$$t = (7.5055) \quad (-3.5256) \quad R^2 = 0.3234$$

（d）首先，将样本期设定为 1954~1976 年（即包括 1954 和 1955 年数据）时的回归结果如下：

$$\hat{Y} = -1.3462 + 5.3231X_2 - 2.6777X_3$$

$$t = (-0.1657) \quad (4.1037) \quad (-2.1202) \quad R^2 = 0.6911$$

省去 1954 年和 1955 年数据即样本期为 1956~1976 年时，回归结果为：

$$\hat{Y} = -11.3627 + 6.0120X_2 - 1.0744X_3$$

$$t = (-1.4726) \quad (5.1418) \quad (-0.9033) \quad R^2 = 0.7288$$

可见，去掉两个样本值后，回归结果较之前有很大变化。通货膨胀率对于普通股实际收益率的影响不再显著。

（e）在模型中引入一个虚拟变量 D ， $D=0$ 代表 1956~1976 年的数据， $D=1$ 代表 1977~1981 年的数据，之后我们得到如下结果：

$$\hat{Y} = -3.3591 + 4.2531X_2 - 1.6024X_3 + 1.5156D$$

$$t = (-0.3873) \quad (3.3337) \quad (-1.1646) \quad (0.1757) \quad R^2 = 0.5546$$

从结果可见，虚拟变量的回归系数并不显著，因此可以认为两个时期的股票实际收益率没有发生显著变化。当然，上述模型假设两个时期的截距有所不同，而认为斜率是相同的。可以通过向模型中加入虚拟变量同其他解释变量的交叉项来检验两个时期的斜率是否相同。

【7.15】表 7-7（参见网上教材）给出了 1960~1982 年美国、加拿大、德国、法国、英国、意大利和日本等 OECD 国家的最终能源总需求（ Y ），实际国内生产总值，GDP（ X_2 ）和实际能源价格（ X_3 ）的数据。（所有变量都是指数形式，以 1973 年为 100）

a. 估计以下模型：

$$\text{模型 A: } \ln Y_t = B_1 + B_2 \ln X_{2t} + B_3 \ln X_{3t} + u_{1t}$$

$$\text{模型 B: } \ln Y_t = A_1 + A_2 \ln X_{2t} + A_3 \ln X_{2(t-1)} + A_4 \ln X_{3t} + u_{2t}$$

$$\text{模型 C: } \ln Y_t = C_1 + C_2 \ln X_{2t} + C_3 \ln X_{3t} + C_4 \ln X_{3(t-1)} + u_{3t}$$

$$\text{模型 D: } \ln Y_t = D_1 + D_2 \ln X_{2t} + D_3 \ln X_{3t} + D_4 \ln Y_{(t-1)} + u_{4t}$$

其中 u 为误差项。模型 B 和 C 称为动态模型——考虑了变量的跨期变化。模型 B 和 C 称为分布滞后模型，因为解释变量对于应变量的影响扩散到了多个时期，这里是两个时期。模型 D 称为自回归模型，因为其中的一个解释变量是应变量的上一期滞后。

- b. 如果只估计了模型 A，而真实模型是 B、C 或 D，则会犯什么样的设定误差？
- c. 由于以上模型都是线性对数形式的，因此斜率系数代表了弹性系数。模型 A 的收入弹性和价格弹性是多少？如何估计其他三个模型的弹性？
- d. 由于在模型 D 中滞后的 Y 变量以解释变量的形式出现，在对模型 D 进行 OLS 估计时预计出现什么问题？（提示：回顾 CLRM 的假定条件）

解答：（a）4 个模型的回归结果如下：

A:	$\ln Y_t = 1.553\ 6 + 0.997\ 6 \ln X_{2t} - 0.332\ 8 \ln X_{3t}$ $t = (17.370) \quad (52.606) \quad (-13.795) \quad R^2 = 0.994\ 2$			
B:	$\ln Y_t = 1.593\ 2 + 0.835\ 3 \ln X_{2t} + 0.175\ 8 \ln X_{2(t-1)} - 0.352\ 6 \ln X_{3t}$ $t = (12.219) \quad (3.045) \quad (0.652) \quad (-12.511) \quad R^2 = 0.994\ 2$			
C:	$\ln Y_t = 1.629\ 5 + 1.005\ 8 \ln X_{2t} - 0.236\ 3 \ln X_{3t} - 0.120\ 8 \ln X_{3(t-1)}$ $t = (17.008) \quad (52.027) \quad (-3.951) \quad (-1.920) \quad R^2 = 0.995\ 0$			
D:	$\ln Y_t = 1.249\ 0 + 0.671\ 3 \ln X_{2t} - 0.270\ 4 \ln X_{3t} + 0.333\ 2 \ln Y_{(t-1)}$ $t = (11.599) \quad (6.593) \quad (-9.469) \quad (3.404) \quad R^2 = 0.996\ 4$			

（b）遗漏了相关解释变量。

（c）收入弹性和价格弹性如下：

模型	收入弹性	价格弹性
A	0.997 6	-0.332 8
B	$(0.835\ 3 + 0.175\ 8) = 1.011\ 1$	-0.352 6
C	1.005 8	$(-0.236\ 3) + (-0.120\ 8) = -0.357\ 1$
D	$0.671\ 3 / 0.666\ 8 = 1.006\ 7$	$(-0.270\ 4) / 0.666\ 8 = -0.405\ 5$

（d）经典线性回归模型假设解释变量都为非随机变量。也就是说，在反复多次抽样中，它们的取值是固定的。但如果被解释变量的滞后期当做解释变量出现在模型中，就违反了该假设。第 12 章将介绍，这种情况会导致 OLS 失效。

【7.16】参考习题 7.11。假定模型中引入代表技术的趋势变量 X_4 ，得到扩展的柯布-道格拉斯函数。进一步假定 X_4 是统计显著的。在这种情形下，犯了哪种设定误差？如果 X_4 是统计不显著的，则又犯了哪种设定误差？给出必要的计算。

解答：包含时间趋势项的柯布-道格拉斯生产函数的回归结果如下：

$$\ln Y_t = 4.944\ 3 + 0.121\ 8 \ln X_{2t} + 0.403\ 4 \ln X_{3t} + 0.118\ 1 \ln X_{4t}$$

$$t = (1.228\ 5) \quad (-0.475\ 3) \quad (1.394\ 7) \quad (3.602\ 3) \quad R^2 = 0.992\ 5$$

时间趋势项在 5% 的显著性水平下是显著的。如果不包含时间趋势项，模型将会犯有遗漏相关变量所造成的设定误差问题。本例中遗漏相关变量的后果是很明显的，当模型中包含时间趋势项（大致可以代表技术水平）时，资本投入和劳动投入对产出的贡献并不明显。

在本例中， Y 、 X_2 、 X_3 有明显的时间趋势性，所以上述回归结果显示的是去（正常）趋势化后的产出和资本与劳动之间的关系。换句话说，上述回归模型给出了短期中，劳动与资本投入同产出之间的关系。本例中这种关系在统计上是不显著的。

注：当前，在模型中引入时间趋势变量的方法已经受到了较为广泛的关注和审查。回归结果表明，本例中的趋势项为确定性趋势，而不是随机性趋势。对于这方面的知识，第 12 章将用一定的篇幅进行介绍。

