

- 5.1 什么是因素控制？如何采用多元回归分析进行因素控制？
- 5.2 线性回归残差是因变量剔除自变量影响后的剩余部分，为什么？
- 5.3 将变量取自然对数后建立模型有什么好处？
- 5.4 为什么要在模型中增加解释变量的高阶项和交叉相乘项？
- 5.5 用逻辑斯蒂曲线 $y = \alpha / (1 + \delta e^{\alpha x})$ 描述商品从上市到饱和的过程，求曲线的拐点。说明拐点的意义。
- 5.6 为什么带常数项的模型中引入虚拟变量时，要对引入的虚拟变量个数加以限制？对不带常数项的模型呢？在什么情况下需要以乘法形式引入虚拟变量？
- 5.7 在参数约束检验中，设无约束模型和约束模型回归的拟合优度分别为 R_{ur}^2 和 R_r^2 ，约束条件个数为 q ，用于回归的样本量为 n ，证明 (5.13) 中的参数约束检验统计量 T_r 可以表示为

$$T_r = \frac{(R_{ur}^2 - R_r^2) / q}{(1 - R_{ur}^2) / [n - (k + 1)]}$$

- 5.8 对工资收入建立如下回归模型

$$\ln(Wage) = \beta_0 + \beta_1 Edu + \beta_3 Age + \beta_4 Age^2 + \beta_5 Edu \cdot Age + u$$

，其中 $Wage$ 、 Edu 和 Age 分别表示工资、受教育水平和年龄。请分析年龄和教育水平对工资收入的影响。

- 5.9* 考虑如下的回归模型

$$C_t = \beta_0 + \beta_1 Inc_t + \beta_2 (Inc_t - Inc_{t_0}) D_t + \varepsilon_t$$

，其中 C_t 和 Inc_t 分别表示 t 年的消费和可支配收入， D_t 为年份虚拟变量，定义为

$$D_t = \begin{cases} 1 & t > t_0 \\ 0 & t \leq t_0 \end{cases}$$

。这样定义的回归模型称为分段回归，画出分段回归直线的示意图。以 t_0 为分界点将样本分为两个时间段分别进行回归分析与采用分段回归有什么区别？不对数据进行分段进行回归与分段回归有什么区别？分段回归有什么优点？如何检验是否需要采用分段回归？

- 5.10 从中国统计年鉴寻找国民经济年度数据，估计我国的柯布一道格拉斯生产函数。

- (1) 分别对 1959 年-1979 年和 1980 年-2010 年间的生产函数进行估计；
- (2) 检验 1959 年-1979 年间的生产函数是否规模报酬不变的 ($\alpha + \beta = 1$)；
- (3) 采用邹变点检验方法检验两个时段的生产函数是否相同。

◆ 参考答案

1. 通过线性回归模型刻画所关注的自变量对因变量的影响时，需要剔除相关变量对关注变量的影响，称为因素控制机。在多元线性回归模型的自变量中，加入与关注变量相关的变量，可以实现因素控制。
2. 由线性回归残差的性质知道，OLS 回归残差形成的向量与解释变量样本形成的向量正交，表明残差中已不再包含自变量的信息和影响，是因变量剔除自变量影响后的剩余部分

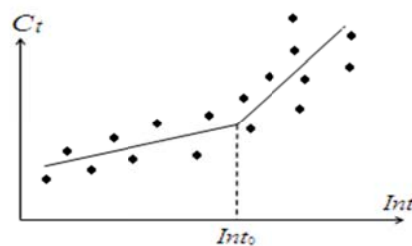
3. 将变量取自然对数后建立模型得出的回归系数估计，是因变量关于自变量的弹性或者半弹性，更能客观地计量自变量对应变量的影响。此外，取对数能够减轻变量之间的共线性，降低异方差程度。再者，取对数能够降低变量取值，减少计算误差。
4. 有时，经济变量之间的关系是线性关系，需要加入自变量的高阶项。例如，平均成本和产量的关系是一种二次函数(抛物线)的关系，存在使平均成本达到最小的最优产量。模型中加入自变量的交叉相乘项，反映的是一种变系数模型。例如在工资收入模型中加入教育(*educ*)和年龄的(*age*)交叉相乘项($age \times educ$)，模型变为 $\ln wage = c + \gamma age \times educ +$ 其他变量+误差项，表明教育对工资的边际影响与年龄有关：不同年龄阶段多受一年教育带来的工资收入增加并不相同。
5. 一种商品从上市初期到逐渐流行有一个被市场接受的过程，开始时销量和市场饱和度很小，随着被市场接受而迅速发展。当市场饱和度达到一定水平后，销售将放缓，当接近饱和后，销量不再发生变化。逻辑曲线的 S 型很好地描述了商品上市到饱和的过程。求二阶导数并令其为 0，可得出曲线拐点为 $t_0 = -(\ln \delta) / \gamma$ 。在 t_0 之前，商品销量关于时间的二阶导数大于 0，销量处于加速上升阶段，在 t_0 滞后，二阶导数小于 0，销量处于减速上升阶段（由于一阶导数大于零，仍处于上升阶段），直到达到饱和而不再变化。
6. 带常数项的模型中引入虚拟变量，如果不对虚拟变量个数加以限制，会引起完全多重共线性，使得模型无法估计。对不带常数项的模型，虚拟变量的个数可以等于属相的类别个数。当认为不同类别个体的回归系数不相同时，可以通过乘法形式引入虚拟变量加以反映。例如，当认为不同性别的劳动者教育对工资收入的影响不相同时，可以引入性别虚拟变量和教育变量的交叉相乘项。
7. 两个模型的因变量相同，其总平方和相同 $TSS_{ur} = TSS_r$ 。根据拟合优度和残差平方和的关系得出 $SSR_{ur} = (1 - R_{ur}^2) TSS_{ur}$ ， $SSR_r = (1 - R_r^2) TSS_r$ ，带入参数约束检验统计量公式 T_r (5.13) 并整理得出结果。
8. 将因变量的条件期望对年龄求偏导数得出

$$\partial E[\ln(Wage) | Age, Edu] / \partial age = \beta_3 + 2\beta_4 Age + \beta_5 Edu$$

，这表明年龄对工资收入的平均影响不仅与年龄有关，也与教育水平有关。 $2\beta_4 Age$ 表示不同年龄阶段年龄每增加一岁对工资收入影响不同， $\beta_5 Edu$ 则表示在不同的教育水平下年龄每增加一岁对工资收入的影响，常数项 β_3 表明，（到达工作年龄后）所有年龄阶段，所有教育水平下年龄每增加一年对工资收入的影响。

采用同样的方法讨论教育对工资收入的影响。

9. 分段回归直线示意图：



。采用两个时段回归将减少每个回归的样本量，降低模型参数的估计精确度。在分段点回归模型斜率发生了，对数据进行分段回归，能反映出因变量 C_t 和自变量 Inc_t 关系在不同时段

的变化。分段回归在反映结构断点的同时，也有效地利用了样本。对原假设 $H_0 : D_t = 0$ 进行检验（ t 检验）可以确定是否需要进行分段回归。也可用 Chow 断点检验

10. 略。