

第三章 经典计量模型： 线性回归模型的最小二乘估计

学习目标

1. 理解回归分析的相关概念和最小二乘法的原理及经典假定。
2. 掌握最小二乘法的参数估计量及其性质的推导。
3. 能够利用样本数据建立一元和多元线性回归模型,并进行参数估计。

一元线性回归模型是研究一个变量对另一个变量的影响,也称双变量线性回归模型,是最简单的计量经济模型(简称计量模型),也是多元线性回归的基础。本章将介绍回归分析的相关概念和计量经济模型建立,最小二乘估计方法的基本原理及模型经典假定,一元和多元线性回归模型的参数估计、估计量的性质等内容。

第一节 一元线性回归模型的构建与经典假定

一、一元线性回归模型的相关概念

(一) 回归的起源和现代含义

回归一词最早来自生物学,是1877年英国生物学家、统计学家高尔顿(F. Galton)和他的学生皮尔逊(K. Pearson),研究父代和子代之间的遗传特征时引入的。在研究过程中发现,子代的身高与父代有密切联系,但当父代的身高特征偏离平均值时,子代的身高特征有可能比父代偏离得更远,也可能没有父代偏离得远,而是接近于平均身高。研究发现,后一种情况发生的概率更大,即身高回归到中等,基于此高尔顿提出了普遍回归定律。现代回归分析,是研究解释变量与被解释变量之间的因果关系,通过解释变量的已知值或给定值,获得被解释变量的期望值或均值,这是回归的现代含义。

(二) 一元线性回归模型的形式

建立如下一元线性回归模型

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad i = 1, 2, \dots, N \quad (3.1.1)$$

其中, y_i 为被解释变量, 也称相依变量、因变量、回归元等; x_i 为解释变量, 也称独立变量、自变量、回归因子等; ϵ_i 为随机误差项, 也称随机扰动项、干扰项、误差项等; β_0 、 β_1 都为待估参数, β_0 称为截距系数, β_1 称为斜率系数; i 表示第 i 个观测值; n 表示观测值的个数。若模型中 x_i 为确定性变量, 那么 $\beta_0 + \beta_1 x_i$ 是模型中的确定性部分, ϵ_i 是随机部分, 因此也将 y_i 视为随机变量。

在一元线性回归模型中, “一元”是指含有一个解释变量, (3.1.1) 式又称双变量回归模型; “线性”是指被解释变量与解释变量及参数都是线性关系; “模型”是指满足某假定条件的方程或方程组。

一般地, 线性包括两种含义: 一是变量线性, 指 y 的均值 $E(y_i)$ 是 x 的线性函数; 二是参数线性, 指 y 的均值 $E(y_i)$ 是参数 β_1 的线性函数。当解释变量 x 非线性时, 通过直接置换就可以将模型化为线性模型, 比较容易处理。因此, 在计量经济学中主要考虑参数线性, 即 y_i 的均值 $E(y_i)$ 是参数的线性函数, 至于 y 的均值是否为变量 x 的线性函数, 不影响参数估计。

二、一元线性回归模型的经典假定

在利用最小二乘法进行回归时, 为保证参数估计量具有良好的性质, 需要设定随机误差项和解释变量满足若干假定条件。

(一) 关于随机误差项的假定

假定 1: 零均值

随机误差项 ϵ_i 具有零均值, $E(\epsilon_i) = 0, i = 1, 2, \dots, N$ 。这是因为在模型设定正确的情况下, 随机误差项中包含对被解释变量有影响, 但仅为微弱影响的不显著的因素, 因此随机误差项的均值等于 0 是合理的。否则, 说明随机误差项中包含对被解释变量有系统性影响的因素, 此时模型可能存在遗漏解释变量或函数形式设定误差等问题。

假定 2: 同方差

随机误差项 ϵ_i 具有同方差, $\text{Var}(\epsilon_i) = \sigma^2, i = 1, 2, \dots, N$ 。这是指随机误差项的方差为常数, 也可以写成等价式 $E(\epsilon_i^2) = \sigma^2$ 。在后面的研究中, 将放松这一假定, 即讨论随机误差项在异方差情况下模型的估计方法。

假定 3: 无序列相关

随机误差项 ϵ_i 在不同样本点之间, 不存在自相关, 也称无序列相关, 即

$$\text{Cov}(\epsilon_i, \epsilon_j) = 0, i \neq j \quad \text{且} \quad i, j = 1, 2, \dots, N$$

也可以写成等价式

$$E(\epsilon_i \epsilon_j) = 0, i \neq j \quad \text{且} \quad i, j = 1, 2, \dots, N$$

在后面的研究中, 也可以放松这一假定, 即讨论随机误差项存在序列相关时模型的估计。

假定 4: 正态分布

随机误差项 ϵ_i 服从零均值、同方差的正态分布, 即

$$\epsilon_i \sim N(0, \sigma^2), \quad i=1, 2, \dots, N$$

这一假定是在有限样本情况下, 通过样本回归方程推断总体回归方程包括变量显著性和方程整体显著性等检验的前提条件。而在无限样本或大样本的情况下, 若误差项是由大量独立偶然因素相互叠加而成, 其中每一个偶然因素的影响都非常微弱的话, 根据中心极限定理, 可知随机误差项近似地服从正态分布。

(二) 关于解释变量的假定

假定 5: 解释变量变异性

解释变量的方差是一个非零的有限值, 即说明解释变量具有变异性, 这一假定是源于回归分析中需要通过解释变量的变化来观测被解释变量的变化。假定方差为有限值, 是为了保证在大样本情况下能够进行渐近性质的推断。

假定 6: 解释变量外生性

随机误差项 ϵ_i 与解释变量 x_i 之间不相关, 即

$$\text{Cov}(x_i, \epsilon_i) = 0, \quad i=1, 2, \dots, N$$

这时可以称解释变量是同期外生, 或称解释变量与随机误差项同期不相关。后面第八章也将讨论解释变量与误差项相关, 即模型出现内生性问题时模型的估计方法。在单方程模型中, 有时为研究方便, 更严格地假定解释变量 x 是确定性变量, 不是随机变量。

以上对误差项的假定称为线性回归模型的**经典假定**(classical assumption), 满足这些假定的线性回归模型, 称为**经典线性回归模型**。在实际建模过程中, 除假定 4 外, 对模型是否满足假定都需要进行检验。通常, 也把满足假定 2 和假定 3 的误差项, 称为球形误差项。

重要事项

由随机误差项 ϵ_i 条件零均值 $E(\epsilon_i | x_i) = 0$, 可以推出非条件零均值 $E(\epsilon_i) = 0$ 。这是因为根据迭代期望定律, 可以得到

$$E(\epsilon_i) = E[E(\epsilon_i | x_i)] = E(0) = 0$$

实际上条件零均值假定, 也可以反映随机误差项的均值不随解释变量的变化而变化, 可以推出解释变量与随机误差项是同期外生的。

类似地, 由条件同方差 $\text{Var}(\epsilon_i | x_i) = \sigma^2$ 假定, 可以得到

$$\text{Var}(\epsilon_i | x_i) = E[\epsilon_i - E(\epsilon_i | x_i)]^2 = E(\epsilon_i^2) = \sigma^2$$

即由条件同方差, 可以得到非条件同方差假定 $\text{Var}(\epsilon_i) = \sigma^2$ 。为了简便, 在本书中使用零均值和同方差的非条件表达式。

将 $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, 称为**总体回归模型**, $E(y_i) = \beta_0 + \beta_1 x_i$ 称为**总体回归方程**或**总体回归直线**, 这一直线需要通过样本进行估计。

由假定 1、假定 2 和假定 6, 可知 y_i 的均值和方差:

$$\begin{aligned} E(y_i) &= E(\beta_0 + \beta_1 x_i + \epsilon_i) = \beta_0 + \beta_1 x_i + E(\epsilon_i) = \beta_0 + \beta_1 x_i \\ \text{Var}(y_i) &= \text{Var}(\beta_0 + \beta_1 x_i + \epsilon_i) = \text{Var}(\beta_0 + \beta_1 x_i) + \text{Var}(\epsilon_i) = \text{Var}(\epsilon_i) = \sigma^2 \end{aligned} \quad (3.1.2)$$

例如, 基于家庭收入与消费的模型, $y_i = \beta_0 + \beta_1 x_i + \epsilon_i, i = 1, 2, \dots, N$, 同方差表示消费虽然随收入变化而变化, 但消费的变动幅度保持不变。

由 (3.1.2) 式可知, 斜率系数 β_1

$$\beta_1 = \frac{dy_i}{dx_i} \approx \frac{\Delta y_i}{\Delta x_i} \quad (3.1.3)$$

斜率系数 β_1 表示边际效应, 即当 x_i 变化一个单位, 平均意义上 y_i 将变动 β_1 单位。同时, 由 $\Delta y_i \approx \beta_1 \Delta x_i$, 可以通过解释变量 x 的变动来预测被解释变量 y 的变动。

由于 y_i 是 ϵ_i 的线性函数, $\epsilon_i \sim N(0, \sigma^2)$, 所以 y_i 也服从正态分布, 即

$$y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$$

由于被解释变量 y_i 的方差与随机误差项 ϵ_i 的方差相同, 且都服从正态分布, 因此可以通过观测被解释变量的方差来判断随机误差项的方差。

第二节 一元线性回归模型的参数估计与估计量的性质

利用一元线性回归模型进行分析, 要求出参数的估计值, 来分析变量之间的关系。本节首先介绍最小二乘法的原理, 推导出参数估计量, 然后分析最小二乘估计量的性质。

一、参数的最小二乘估计

(一) 直线拟合的准则

一般来说, 总体回归直线 $E(y_i) = \beta_0 + \beta_1 x_i$ 无法观测, 可以用 $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ 来估计, 将其称为**样本回归方程**或**样本回归直线**。 $\hat{\beta}_0$ 与 $\hat{\beta}_1$ 分别是 β_0 和 β_1 的估计值, $\hat{y}_i$ 为 y_i 的估计值, 也称拟合值或预测值。而模型 $y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + \hat{\epsilon}_i$ 称为**样本回归模型**, 其中, $\hat{\epsilon}_i = y_i - \hat{y}_i$, 称为**残差**。当 $\hat{\epsilon}_i > 0$ 时, $y_i > \hat{y}_i$; 而当 $\hat{\epsilon}_i < 0$ 时, $y_i < \hat{y}_i$ 。需要通过估计 $\hat{\beta}_0$ 与 $\hat{\beta}_1$ 确定样本回归直线。那么, 什么样的回归直线是一条较好的回归直线呢? 通常认为较好的直线应该是离样本点很近, 或被样本点紧紧围绕, 如图 3-1 所示。

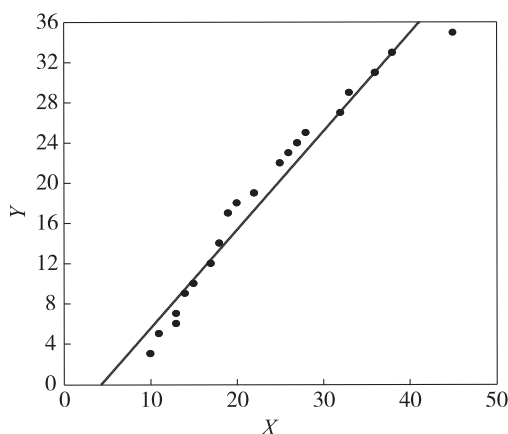


图 3-1 X 和 Y 的散点图

(二) 最小二乘法的原理

普通最小二乘法(ordinary least squares method, 简称 OLS 方法或最小二乘法), 就是以残差平方和 $\left(\sum_{i=1}^N \hat{\varepsilon}_i^2\right)$ 最小为目标, 来确定斜率系数和截距系数, 进而确定拟合直线的位置。残差平方和最小可以保证估计值较好地拟合实际值, 且优于以残差和最小、最远点残差最小的原则。这是因为残差和最小会出现正负抵消的情况; 而最远点残差最小原则只考虑了最远点, 这样以偏概全, 忽略了其他样本点的信息。因此, 通常采用残差平方和最小这一原则, 来获得最佳的拟合直线。

这里的目标函数为残差平方和, 是以 $\hat{\beta}_0, \hat{\beta}_1$ 作为决策变量, 将目标函数记为 $f(\hat{\beta}_0, \hat{\beta}_1)$, 即

$$f(\hat{\beta}_0, \hat{\beta}_1) = \sum_{i=1}^N \hat{\varepsilon}_i^2 = \sum_{i=1}^N (y_i - \hat{y}_i)^2 = \sum_{i=1}^N (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$

为使目标函数达到最小, 利用极值条件, 在 $f(\hat{\beta}_0, \hat{\beta}_1)$ 对 $\hat{\beta}_0$ 和 $\hat{\beta}_1$ 求一阶偏导为 0 时, 可以得到如下的方程组

$$\begin{cases} \frac{\partial f(\hat{\beta}_0, \hat{\beta}_1)}{\partial \hat{\beta}_0} = 2 \sum_{i=1}^N (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)(-1) = 0 \\ \frac{\partial f(\hat{\beta}_0, \hat{\beta}_1)}{\partial \hat{\beta}_1} = 2 \sum_{i=1}^N (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)(-x_i) = 0 \end{cases} \quad (3.2.1)$$

将(3.2.1)式化简, 可以得到

$$\begin{cases} \sum_{i=1}^N (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \\ \sum_{i=1}^N (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) x_i = 0 \end{cases} \quad (3.2.2)$$

(3.2.2)式称为**正规方程组**,为表述简便,将求和符号简记为 Σ 。在(3.2.2)式的第一个方程两侧都除以 N ,即 $\bar{y}=\hat{\beta}_0+\hat{\beta}_1\bar{x}$,可得截距系数的估计方程

$$\hat{\beta}_0=\bar{y}-\hat{\beta}_1\bar{x} \quad (3.2.3)$$

把(3.2.3)式代入(3.2.2)式中的第二个方程,得

$$\Sigma(y_i-\bar{y})x_i-\hat{\beta}_1\Sigma(x_i-\bar{x})x_i=0 \quad (3.2.4)$$

求解(3.2.4)式,得到

$$\hat{\beta}_1=\frac{\Sigma x_i(y_i-\bar{y})}{\Sigma(x_i-\bar{x})x_i} \quad (3.2.5)$$

将(3.2.5)式的分子减去 $\Sigma\bar{x}(y_i-\bar{y})$,分母减去 $\Sigma\bar{x}(x_i-\bar{x})$,由于这两式都为0,可得

$$\hat{\beta}_1=\frac{\Sigma x_i(y_i-\bar{y})}{\Sigma(x_i-\bar{x})x_i}=\frac{\Sigma x_i(y_i-\bar{y})-\Sigma\bar{x}(y_i-\bar{y})}{\Sigma(x_i-\bar{x})x_i-\Sigma\bar{x}(x_i-\bar{x})}=\frac{\Sigma(x_i-\bar{x})(y_i-\bar{y})}{\Sigma(x_i-\bar{x})^2} \quad (3.2.6)$$

将(3.2.6)式代回(3.2.3)式,即可求得 $\hat{\beta}_0$ 。设离差 $\dot{x}_i=x_i-\bar{x}$, $\dot{y}_i=y_i-\bar{y}$,即可求得。

$$\begin{cases} \hat{\beta}_0=\bar{y}-\hat{\beta}_1\bar{x} \\ \hat{\beta}_1=\frac{\Sigma\dot{x}_i\dot{y}_i}{\Sigma\dot{x}_i^2} \end{cases} \quad (3.2.7)$$

(3.2.7)式中的 $\hat{\beta}_0$ 和 $\hat{\beta}_1$ 称为**最小二乘估计量**。可以看出 $\hat{\beta}_1$ 是 x 和 y 的样本协方差与 x 的样本方差的比值。由于解释变量 x 的样本方差通常为正数,因此当 x 、 y 正相关时, $\hat{\beta}_1$ 为正数;当 x 、 y 负相关时, $\hat{\beta}_1$ 为负数。同时,由 $\hat{\beta}_1$ 的分母是解释变量 x 的样本方差,可知最小二乘法要求,只有解释变量具有变异性,才能求解斜率系数。但 $\hat{\beta}_1$ 对解释变量的异常值又比较敏感,因此一般要求在样本容量 N 充分大的情况下,解释变量的方差趋近于一个有限值。

二、参数的矩估计

根据经典假定1和6,可知

$$\begin{aligned} E(\epsilon_i) &= 0 \\ \text{Cov}(x_i, \epsilon_i) &= 0 \end{aligned}$$

这是两个总体矩条件,可以采用样本矩条件来代替,即

$$\begin{cases} \frac{1}{N}\Sigma\hat{\epsilon}_i=0 \\ \frac{1}{N}\Sigma\hat{\epsilon}_ix_i=0 \end{cases} \quad (3.2.8)$$

由于 $\hat{\epsilon}_i = y_i - \hat{y}_i$, (3.2.8) 式可化为

$$\begin{cases} \sum (y_i - \hat{y}_i) = 0 \\ \sum (y_i - \hat{y}_i) x_i = 0 \end{cases} \quad (3.2.9)$$

将 $y_i - \hat{y}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$ 代入 (3.2.9) 式中, 即可以得到 (3.2.2) 式。由此可见, 一元线性回归模型的最小二乘估计量也是矩估计量, 两种方法得到的参数估计量相同。(3.2.8) 式和 (3.2.9) 式都是以残差形式表示正规方程组, 是与总体回归模型中经典假定 1 和假定 6 相对应的样本矩条件, 这种形式的正规方程组在探讨最小二乘估计量性质中也会使用。

小贴士

1. 一元线性回归模型的最小二乘估计量也是矩估计量。
2. 最小二乘法是通过残差平方和达到最小, 找到与样本匹配的最佳回归方程, 但回归方程并不一定能通过每一个样本点。
3. 由于 $\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}$, 可知样本回归直线过点 $(\bar{x}, \bar{y})$, 或者说点 $(\bar{x}, \bar{y})$ 在样本回归直线 $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ 上。

三、最小二乘估计量的有限样本性质

有限样本性质也称小样本性质, 即样本容量 N 为有限值时, 最小二乘估计量满足线性、无偏性和有效性 (即最小方差性), 是优良的参数估计量, 即得高斯-马尔科夫定理。

高斯-马尔科夫 (Gauss-Markov) 定理: 若随机误差项 ϵ_i 满足经典假定, 那么最小二乘估计量具有最佳线性无偏性, 即为 BLUE 估计量 (Best Linear Unbiased Estimator)。

下面将证明最小二乘估计量满足这三个性质, 进而证明高斯-马尔科夫定理。

性质 1: 最小二乘估计量的线性性质。

最小二乘估计量满足线性, 是指参数估计量 $\hat{\beta}_0$ 和 $\hat{\beta}_1$ 是随机变量 y_i 和 ϵ_i 的线性函数。证明如下:

由 (3.2.7) 式, 可得

$$\hat{\beta}_1 = \frac{\sum \dot{x}_i \dot{y}_i}{\sum \dot{x}_i^2} = \frac{\sum \dot{x}_i (y_i - \bar{y})}{\sum \dot{x}_i^2} = \frac{\sum \dot{x}_i y_i - \bar{y} \sum \dot{x}_i}{\sum \dot{x}_i^2} \quad (3.2.10)$$

为简化分析, 设 $c_i = \frac{\dot{x}_i}{\sum \dot{x}_i^2}$, 由于 $\sum \dot{x}_i = \sum (x_i - \bar{x}) = 0$, 代入 (3.2.10) 式, 得到

$$\hat{\beta}_1 = \frac{\sum \dot{x}_i y_i}{\sum \dot{x}_i^2} = \sum c_i y_i \quad (3.2.11)$$

可知, $\hat{\beta}_1$ 为 y_i 的线性函数。

由于

$$\begin{aligned}\hat{\beta}_1 &= \sum c_i y_i = \sum c_i (\beta_0 + \beta_1 x_i + \epsilon_i) = \beta_0 \sum c_i + \beta_1 \sum c_i x_i + \sum c_i \epsilon_i \\ \sum c_i &= \frac{\sum \dot{x}_i}{\sum \dot{x}_i^2} = 0 \\ \sum c_i \bar{x} &= \bar{x} \sum c_i = 0 \\ \sum c_i x_i &= \sum c_i x_i - \sum c_i \bar{x} = \sum c_i \dot{x}_i = \frac{\sum \dot{x}_i^2}{\sum \dot{x}_i^2} = 1\end{aligned}$$

可得

$$\hat{\beta}_1 = \beta_1 + \sum c_i \epsilon_i \quad (3.2.12)$$

可知, $\hat{\beta}_1$ 也为 ϵ_i 的线性函数。

对于截距系数

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = \frac{1}{N} \sum y_i - \bar{x} \sum c_i y_i = \sum \left(\frac{1}{N} - \bar{x} c_i \right) y_i \quad (3.2.13)$$

可知, $\hat{\beta}_0$ 为 y_i 的线性函数。

将 $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ 代入(3.2.13)式,可得

$$\begin{aligned}\hat{\beta}_0 &= \sum \left(\frac{1}{N} - \bar{x} c_i \right) (\beta_0 + \beta_1 x_i + \epsilon_i) \\ &= \sum \frac{1}{N} \beta_0 + \sum \frac{1}{N} \beta_1 x_i + \sum \frac{1}{N} \epsilon_i - \beta_0 \bar{x} \sum c_i - \beta_1 \bar{x} \sum c_i x_i - \bar{x} \sum c_i \epsilon_i \\ &= \beta_0 + \beta_1 \bar{x} + \sum \frac{1}{N} \epsilon_i - \beta_1 \bar{x} - \bar{x} \sum c_i \epsilon_i \\ &= \beta_0 + \sum \left(\frac{1}{N} - \bar{x} c_i \right) \epsilon_i\end{aligned} \quad (3.2.14)$$

可知, $\hat{\beta}_0$ 也为 ϵ_i 的线性函数。

性质 2: 最小二乘估计量的无偏性。

最小二乘估计量满足无偏性是指参数估计量的均值等于真实值, 即 $E(\hat{\beta}_1) = \beta_1$, $E(\hat{\beta}_0) = \beta_0$, 证明如下:

由 $\hat{\beta}_1 = \beta_1 + \sum c_i \epsilon_i$, 可以得到

$$\begin{aligned}E(\hat{\beta}_1) &= E(\beta_1 + \sum c_i \epsilon_i) \\ &= \beta_1 + E(\sum c_i \epsilon_i) \\ &= \beta_1 + \sum c_i E(\epsilon_i) \quad (\text{期望的线性性质}) \\ &= \beta_1 \quad (E(\epsilon_i) = 0)\end{aligned}$$

由 $\hat{\beta}_0 = \beta_0 + \sum \left(\frac{1}{N} - \bar{x}c_i \right) \epsilon_i$, 可以得到

$$\begin{aligned} E(\hat{\beta}_0) &= E \left[\beta_0 + \sum \left(\frac{1}{N} - \bar{x}c_i \right) \epsilon_i \right] \\ &= \beta_0 + \left[\sum \left(\frac{1}{N} - \bar{x}c_i \right) E(\epsilon_i) \right] \quad (\text{期望的线性性质}) \\ &= \beta_0 \quad (E(\epsilon_i) = 0) \end{aligned}$$

参数估计量 $\hat{\beta}_1$ 和 $\hat{\beta}_0$ 的无偏性得证。

性质 3: 最小二乘估计量的有效性。

最小二乘估计量满足有效性是指参数估计量 $\hat{\beta}_0, \hat{\beta}_1$ 的方差最小, 即在所有线性无偏估计量中, 最小二乘法估计量的方差最小, 证明如下:

先来证明 $\hat{\beta}_1$ 的方差最小, 思路是取 β_1 的任一线性无偏估计量 $\tilde{\beta}_1$, 如果能够证明出 $\hat{\beta}_1$ 的方差小于或等于 $\tilde{\beta}_1$ 的方差, 有效性即可得证。

由 $\hat{\beta}_1 = \beta_1 + \sum c_i \epsilon_i$ 可知

$$\text{Var}(\hat{\beta}_1) = \sigma^2 \sum c_i^2 \quad (3.2.15)$$

由 $\tilde{\beta}_1$ 具有线性, 是 y_i 的线性函数, 设

$$\begin{aligned} \tilde{\beta}_1 &= \sum d_i y_i = \sum d_i (\beta_0 + \beta_1 x_i + \epsilon_i) \\ &= \beta_0 \sum d_i + \beta_1 \sum d_i x_i + \sum d_i \epsilon_i \end{aligned} \quad (3.2.16)$$

$$\begin{aligned} E(\tilde{\beta}_1) &= E(\beta_0 \sum d_i + \beta_1 \sum d_i x_i + \sum d_i \epsilon_i) \\ &= \beta_0 \sum d_i + \beta_1 \sum d_i x_i \quad (\sum d_i E(\epsilon_i) = 0) \end{aligned}$$

由于 $\tilde{\beta}_1$ 是无偏估计量, 故 $E(\tilde{\beta}_1) = \beta_1$, 所以

$$\sum d_i = 0, \quad \sum d_i x_i = 1$$

不失一般性, 设 $d_i = c_i + h_i$, h_i 可以看成是对 c_i 的微小影响, 当 $h_i = 0$ 时, 则 $d_i = c_i$ 。

$$\sum d_i = \sum c_i + \sum h_i \quad (3.2.17)$$

$$\sum d_i x_i = \sum c_i x_i + \sum h_i x_i \quad (3.2.18)$$

由于 $\sum c_i = 0, \sum c_i x_i = 1$, 代入到 (3.2.17) 式和 (3.2.18) 式得到

$$\sum h_i = 0, \quad \sum h_i x_i = 0$$

(3.2.16) 式化为

$$\tilde{\beta}_1 = \beta_0 \sum d_i + \beta_1 \sum d_i x_i + \sum d_i \epsilon_i = \beta_1 + \sum d_i \epsilon_i$$

由 $\hat{\beta}_1 = \beta_1 + \sum c_i \epsilon_i$, 可知 $\tilde{\beta}_1$ 与 $\hat{\beta}_1$ 都为随机误差项的线性函数, 二者差别在于系数不同。

$$\begin{aligned}
 \text{Var}(\tilde{\beta}_1) &= \text{Var}(\beta_1 + \sum d_i \epsilon_i) = \sigma^2 \sum d_i^2 \\
 \sum d_i^2 &= \sum (c_i + h_i)^2 = \sum c_i^2 + \sum h_i^2 + 2 \sum c_i h_i \\
 &= \sum c_i^2 + \sum h_i^2 + 2 \frac{\sum \dot{x}_i h_i}{\sum \dot{x}_i^2} \quad \left(c_i = \frac{\dot{x}_i}{\sum \dot{x}_i^2} \right) \\
 &= \sum c_i^2 + \sum h_i^2 + 2 \frac{\sum x_i h_i - \bar{x} \sum h_i}{\sum \dot{x}_i^2} \\
 &= \sum c_i^2 + \sum h_i^2 \quad (\sum h_i = 0, \sum h_i x_i = 0) \\
 \text{Var}(\tilde{\beta}_1) &= \sigma^2 \sum d_i^2 = \sigma^2 (\sum c_i^2 + \sum h_i^2)
 \end{aligned}$$

由于 $\sigma^2 \sum h_i^2 \geq 0$, 当等号成立时, $h_i = 0$, 即 $d_i = c_i, i = 1, 2, \dots, N$, 这时有 $\tilde{\beta}_1 = \hat{\beta}_1$, 所以

$$\text{Var}(\tilde{\beta}_1) \geq \text{Var}(\hat{\beta}_1)$$

同理, 可以证明 $\text{Var}(\tilde{\beta}_0) \geq \text{Var}(\hat{\beta}_0)$ 。

因而, 最小二乘估计量 $\hat{\beta}_0, \hat{\beta}_1$ 具有有效性得证。

可以推出, 最小二乘估计量 $\hat{\beta}_0, \hat{\beta}_1$ 的方差及其协方差如下

$$\text{Var}(\hat{\beta}_1) = \sigma^2 \sum c_i^2 = \frac{\sigma^2}{\sum \dot{x}_i^2} \quad (3.2.19)$$

$$\begin{aligned}
 \text{Var}(\hat{\beta}_0) &= \sigma^2 \sum \left(\frac{1}{N} - \bar{x} c_i \right)^2 \\
 &= \sigma^2 \left(\sum \frac{1}{N^2} + \bar{x}^2 \sum c_i^2 \right) \quad \left(\sum c_i = \frac{\sum \dot{x}_i}{\sum \dot{x}_i^2} = 0 \right) \\
 &= \sigma^2 \left(\frac{1}{N} + \bar{x}^2 \frac{1}{\sum \dot{x}_i^2} \right) \quad \left(\sum c_i^2 = \frac{1}{\sum \dot{x}_i^2} \right) \\
 &= \sigma^2 \left(\frac{\sum \dot{x}_i^2 + N \bar{x}^2}{N \sum \dot{x}_i^2} \right) \\
 &= \frac{\sigma^2 \sum x_i^2}{N \sum \dot{x}_i^2} \quad (\sum \dot{x}_i^2 = \sum (x_i - \bar{x})^2 = \sum x_i^2 - N \bar{x}^2)
 \end{aligned} \quad (3.2.20)$$

$$\text{Cov}(\hat{\beta}_0, \hat{\beta}_1) = E[(\hat{\beta}_0 - E(\hat{\beta}_0))(\hat{\beta}_1 - E(\hat{\beta}_1))]$$

由于 $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$, $E(\hat{\beta}_0) = \bar{y} - \beta_1 \bar{x}$, 因此

$$\begin{aligned}
 \text{Cov}(\hat{\beta}_0, \hat{\beta}_1) &= E[(\bar{y} - \hat{\beta}_1 \bar{x}) - (\bar{y} - \beta_1 \bar{x})](\hat{\beta}_1 - \beta_1) \\
 &= E[(\beta_1 \bar{x} - \hat{\beta}_1 \bar{x})(\hat{\beta}_1 - \beta_1)] \\
 &= -\bar{x} E[(\hat{\beta}_1 - \beta_1)^2] \\
 &= -\bar{x} \text{Var}(\hat{\beta}_1) = -\bar{x} \frac{\sigma^2}{\sum \dot{x}_i^2}
 \end{aligned} \quad (3.2.21)$$

最小二乘法是最基本和最常用的回归方法,也在于其参数估计量具有线性、无偏性和有效性这样优良的性质。

四、回归标准误差

若用 $\hat{\sigma}^2$ 表示随机误差项方差 σ^2 的估计,则随机误差项的方差估计值为

$$\hat{\sigma}^2 = \sum \hat{\varepsilon}_i^2 / (N-2) \quad (3.2.22)$$

(3.2.22)式中, N 表示样本容量,2 表示待估参数的个数,即需要估计斜率系数和截距系数, $N-2$ 也是残差平方和的自由度, $\hat{\sigma}$ 称为回归标准误差(standard error of the regression)或均方根误差。第四节将证明 $E(\hat{\sigma}^2) = \sigma^2$,即 $\hat{\sigma}^2$ 是 σ^2 的无偏估计量。

进一步可以求得 $\hat{\beta}_1$ 和 $\hat{\beta}_0$ 的方差估计值为

$$\widehat{\text{Var}}(\hat{\beta}_1) = \frac{\hat{\sigma}^2}{\sum \dot{x}_i^2}$$

$$\widehat{\text{Var}}(\hat{\beta}_0) = \frac{\hat{\sigma}^2 \sum x_i^2}{N \sum \dot{x}_i^2}$$

$\hat{\beta}_1$ 和 $\hat{\beta}_0$ 的标准差估计值为

$$\text{Se}(\hat{\beta}_1) = \frac{\hat{\sigma}}{\sqrt{\sum \dot{x}_i^2}}$$

$$\text{Se}(\hat{\beta}_0) = \hat{\sigma} \sqrt{\frac{\sum x_i^2}{N \sum \dot{x}_i^2}}$$

称为 $\hat{\beta}_1$ 和 $\hat{\beta}_0$ 的标准误差(standard error),也可以简称为标准误。

有时计算模型的参数标准误时,很难通过推演得到显性数学式,进而无法估算,或者是由于样本数据有限,参数标准误估计结果可能无法较好地反映其真实值,这时可以采用自举法(bootstrap)估计参数的标准误。对一个样本容量为 N 的随机样本,将其看成是可以进行重新抽样的总体,使用原来样本进行有放回地随机抽样,可能会出现观测值重复的情况。这样得到的样本容量仍为 N 的新样本,称为自举样本,利用自举样本计算参数估计量如 $\hat{\beta}$ 。这一过程重复进行 T 次,得到 $\hat{\beta}$ 的 T 次估计值 $\hat{\beta}^1, \hat{\beta}^2, \dots, \hat{\beta}^T$,利用这 T 次估计值计算其标准误为

$$\text{Se}(\hat{\beta}) = \sqrt{\frac{\sum_{t=1}^T (\hat{\beta}^t - \bar{\hat{\beta}})^2}{T-1}} \quad (3.2.23)$$

其中, $\bar{\hat{\beta}}$ 是 $\hat{\beta}^1, \hat{\beta}^2, \dots, \hat{\beta}^T$ 的均值,将自举法获得的(3.2.23)式作为 $\hat{\beta}$ 的标准误。自举法通常借助于计算机来完成,因此一般重复次数 T 可以取较大的数值,如 500、1 000,可以接近于总体的分布,从而得到比较可靠的标准误估计值。

第三节 多元线性回归模型的构建与经典假定

一、多元线性回归模型的构建

(一) 多元线性回归模型的方程形式

在实际建模时,被解释变量通常不只受到一个解释变量的影响,而是由多种因素所决定。例如,劳动者的工资除了受教育程度的影响,还受到从业经验、性别、年龄和工作所处的行业等多种因素的影响。可见,一元线性回归模型只是线性回归模型的特例,不足以反映变量之间的真实关系。因而,在实证研究中更多的是需要建立多元线性回归模型,来分析变量之间的因果关系。多元线性回归模型表示为

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_{k-1} x_{ik-1} + \epsilon_i, \quad i = 1, 2, \dots, N \quad (3.3.1)$$

其中, y_i 是被解释变量, $x_{ij} (j = 1, \dots, k-1)$ 是解释变量, ϵ_i 是随机误差项, $\beta_j (j = 0, 1, \dots, k-1)$ 是待估参数。从经济意义上来说, x_{ij} 是 y_i 的重要解释变量;从代数意义上来说, y_i 与 x_{ij} 存在线性关系。

方程(3.3.1)也称为**总体回归模型**。总体回归方程为

$$E(y_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_{k-1} x_{ik-1}$$

其中, β_j 也被称为偏回归系数或偏效应,表示在控制其他解释变量保持不变的情况下, x_j 单位变化所引起 y 均值 $E(y)$ 的变化,即 x_j 的单位变化对 y 均值的净影响。

对于一元线性回归模型,总体回归方程为直线;对于二元线性回归模型,总体回归方程为平面;对于三元及以上的线性回归模型,总体回归方程为超平面。

(二) 多元线性回归模型引入控制变量的意义

以研究劳动者教育回报问题为例,重点关注劳动者的受教育程度对工资收入的影响。在实际实证分析中,需要控制劳动者的从业经验、性别、年龄和工作所处的行业等因素相同,才能分析受教育程度对工资收入的净影响。但这样选择样本数据时会遇到很大的困难,因为很难找到工作经验、性别、年龄和从事的行业等因素都相同的劳动者的样本,或者即使找到样本可能也很小。这时可以借助于多元回归模型来实现,通过假定其他控制变量不变,分析受教育程度对工资收入的影响。

[引例 3-1] 明瑟工资方程

使用如下的明瑟工资方程分析教育回报问题

$$y_i = \beta_0 + \beta_1 \text{edu}_i + \beta_2 \text{exp}_i + \beta_3 \text{exp}_i^2 + z_i' \gamma_i + \epsilon_i$$

其中, y_i 表示劳动者 i 的工资收入, edu_i 表示劳动者 i 的受教育程度, exp_i 表示劳动者 i 的工作经验, 而 z_i 表示影响劳动者工资的其他控制变量, 如劳动者的性别、年龄、户籍、工作所处的行业等因素。由于控制了其他影响因素, 可以假定 exp_i 、 z_i 等其他控制变量不变的情况下, 分析系数 β_1 的偏效应来考察劳动者受教育程度对工资收入的影响。因此, 引入控制变量的多元回归模型的意义在于, 可以在非实验环境下达到控制实验的效果。

二、多元线性回归模型的向量与矩阵表示形式

对于(3.3.1)式可以采用向量的形式来表示, 令

$$\mathbf{x}'_i = (1 \quad x_{i1} \quad x_{i2} \quad \cdots \quad x_{ik-1})$$

$$\boldsymbol{\beta} = (\beta_0 \quad \beta_1 \quad \cdots \quad \beta_{k-1})'$$

(3.3.1)式可化为

$$y_i = \mathbf{x}'_i \boldsymbol{\beta} + \varepsilon_i \quad i = 1, 2, \cdots, N \quad (3.3.2)$$

对于(3.3.1)式, 将 i 依次从 1 取值到 N 时, 模型(3.3.2)表示为

$$\begin{cases} y_1 = \mathbf{x}'_1 \boldsymbol{\beta} + \varepsilon_1 \\ y_2 = \mathbf{x}'_2 \boldsymbol{\beta} + \varepsilon_2 \\ \vdots \\ y_N = \mathbf{x}'_N \boldsymbol{\beta} + \varepsilon_N \end{cases} \quad (3.3.3)$$

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k-1} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{N1} & x_{N2} & \cdots & x_{Nk-1} \end{bmatrix} = \begin{bmatrix} \mathbf{x}'_1 \\ \mathbf{x}'_2 \\ \vdots \\ \mathbf{x}'_N \end{bmatrix} = (\mathbf{x}_1 \mathbf{x}_2 \cdots \mathbf{x}_N)'$$

这里通常可以把截距项看成单位变量 1 的系数, (3.3.3)式可以采用矩阵来表示为

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (3.3.4)$$

其中, $\mathbf{Y} = (y_1, y_2, \cdots, y_N)'$, $\boldsymbol{\varepsilon} = (\varepsilon_1 \quad \varepsilon_2 \quad \cdots \quad \varepsilon_N)'$ 。使用矩阵表达式简洁, 在推导时更方便。对应的样本回归方程为

$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}} \quad (3.3.5)$$

其中, $\hat{\mathbf{Y}} = (\hat{y}_1, \hat{y}_2, \cdots, \hat{y}_N)'$, $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0 \quad \hat{\beta}_1 \quad \cdots \quad \hat{\beta}_{k-1})'$ 。

三、多元线性回归模型的经典假定

类似于一元线性回归模型, 在利用最小二乘法进行回归时, 为保证参数估计量具有良

好的性质,需要做出假定。

(一) 关于随机误差项的假定

假定 1: 零均值

随机误差项 ϵ_i 具有零均值, $E(\epsilon_i) = 0, i = 1, 2, \dots, N$ 。

假定 2: 同方差

随机误差项 ϵ_i 具有同方差, $\text{Var}(\epsilon_i) = \sigma^2, i = 1, 2, \dots, N$ 。

假定 3: 无序列相关

随机误差项 ϵ_i 在不同样本点之间独立, 不存在序列相关, 即为

$$\text{Cov}(\epsilon_i, \epsilon_j) = 0, \quad i \neq j \quad \text{且} \quad i, j = 1, 2, \dots, N \quad (3.3.6)$$

假定 4: 正态分布

随机误差项 ϵ_i 服从零均值、同方差的正态分布

$$\epsilon_i \sim N(0, \sigma^2), \quad i = 1, 2, \dots, N \quad (3.3.7)$$

在多元线性回归模型中, 随机误差项的方差-协方差矩阵为

$$\text{Var}(\boldsymbol{\epsilon}) = E(\boldsymbol{\epsilon}\boldsymbol{\epsilon}') = \begin{bmatrix} \text{Var}(\epsilon_1) & \text{Cov}(\epsilon_1\epsilon_2) & \cdots & \text{Cov}(\epsilon_1\epsilon_N) \\ \text{Cov}(\epsilon_2\epsilon_1) & \text{Var}(\epsilon_2) & \cdots & \text{Cov}(\epsilon_2\epsilon_N) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(\epsilon_N\epsilon_1) & \text{Cov}(\epsilon_N\epsilon_2) & \cdots & \text{Var}(\epsilon_N) \end{bmatrix} \quad (3.3.8)$$

其中, $\text{Var}(\boldsymbol{\epsilon})$ 称为方差-协方差矩阵, 其主对角线上的元素表示随机误差项的方差, 非主对角线上的元素为协方差, 也可记为 $\text{Var}-\text{Cov}(\boldsymbol{\epsilon})$ 。在同方差、无序列相关条件下, 随机误差项的方差-协方差矩阵为

$$\text{Var}(\boldsymbol{\epsilon}) = E(\boldsymbol{\epsilon}\boldsymbol{\epsilon}') = \sigma^2 \mathbf{I}_N \quad (3.3.9)$$

(二) 关于解释变量的假定

假定 5: 解释变量具有变异性 and 解释变量之间无完全多重共线性

要求解释变量具有非零的有限方差, 且解释变量之间不存在严格或完全的线性相关或完全多重共线性。若用矩阵表示这一假定, 矩阵 $\mathbf{X}$ 为满列秩矩阵, $R(\mathbf{X}'\mathbf{X}) = R(\mathbf{X}) = k$, 或者 $|\mathbf{X}'\mathbf{X}| \neq 0$ 。

以解释变量 x_1 和 x_3 为例来说明线性相关:

$$\lambda_1 x_1 + \lambda_2 x_3 = 0 \quad (3.3.10)$$

若(3.3.10)式成立, 则说 x_1 和 x_3 是共线的(collinear)或线性相关。不存在严格或完全的多重共线性(multicollinearity), 是指模型中任何一个解释变量都不能写成其余解释变量的线性组合。这是因为, 如果解释变量之间存在严格的线性相关, 矩阵 $\mathbf{X}$ 将为奇异矩阵或降秩矩阵, 用普通最小乘法进行回归时, 无法求解参数估计量。

假定 6: 解释变量外生性

随机误差项 ϵ_i 与任何一个解释变量都不相关, 模型不存在内生性, 即

$$\text{Cov}(x_{ij}, \epsilon_i) = E(x_{ij}\epsilon_i) = 0, \quad i=1, 2, \dots, N, j=1, 2, \dots, k-1 \quad (3.3.11)$$

用矩阵表示,即 $E(\mathbf{X}'\boldsymbol{\epsilon})=0$ 。

或者更严格地假定解释变量是非随机或者为确定性变量。

第四节 多元线性回归模型的参数估计与估计量的性质

一、参数的最小二乘估计

根据最小二乘法的原理,在残差平方和达到最小的情况下求解模型的参数,设残差为

$$\hat{\boldsymbol{\epsilon}} = \mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}$$

其中, $\hat{\boldsymbol{\epsilon}}$ 是一个 $N \times 1$ 的列向量,残差平方和为

$$\begin{aligned} f(\hat{\boldsymbol{\beta}}) &= \sum_{i=1}^N \hat{\epsilon}_i^2 = \hat{\boldsymbol{\epsilon}}' \hat{\boldsymbol{\epsilon}} = (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})' (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \\ &= \mathbf{Y}'\mathbf{Y} - \hat{\boldsymbol{\beta}}' \mathbf{X}'\mathbf{Y} - \mathbf{Y}'\mathbf{X}\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\beta}}' \mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} \end{aligned} \quad (3.4.1)$$

在(3.4.1)式中,可以看出 $\hat{\boldsymbol{\beta}}' \mathbf{X}'\mathbf{Y}$ 和 $\mathbf{Y}'\mathbf{X}\hat{\boldsymbol{\beta}}$ 互为转置且都是 1×1 的矩阵,则 $\mathbf{Y}'\mathbf{X}\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}' \mathbf{X}'\mathbf{Y}$, 因而

$$f(\hat{\boldsymbol{\beta}}) = \mathbf{Y}'\mathbf{Y} - 2\mathbf{Y}'\mathbf{X}\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\beta}}' \mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}$$

残差平方和 $f(\hat{\boldsymbol{\beta}})$ 是以 $\hat{\boldsymbol{\beta}}$ 为决策变量的目标函数,为使目标函数达到最小,利用一阶条件

$$\frac{\partial f(\hat{\boldsymbol{\beta}})}{\partial \hat{\boldsymbol{\beta}}} = -2\mathbf{X}'\mathbf{Y} + 2\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = 0 \quad (3.4.2)$$

这里用到了向量微分的性质,可参考附录。依据经典假定 5 可知, $\mathbf{X}'\mathbf{X}$ 满秩,存在逆矩阵,将(3.4.2)式整理可以求解出 $\hat{\boldsymbol{\beta}}$

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y} \quad (3.4.3)$$

其中, $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_{k-1})'$, 是 $\boldsymbol{\beta}$ 的估计值,样本回归方程可写为

$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$$

二、参数的矩估计

由经典假定 6,可知 $E(\mathbf{X}'\boldsymbol{\epsilon})=0$,用样本矩条件代替总体的矩条件,得到

$$\frac{1}{N}\mathbf{X}'\hat{\boldsymbol{\varepsilon}}=0$$

两边都乘以 N , 得到

$$\mathbf{X}'\hat{\boldsymbol{\varepsilon}}=0 \quad (3.4.4)$$

利用这一个矩阵方程可以求解参数估计量 $\boldsymbol{\beta}$ 的估计值。

由 (3.4.2) 式可得

$$\mathbf{X}'\mathbf{Y}-\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}=\mathbf{X}'(\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}})=\mathbf{X}'(\mathbf{Y}-\hat{\mathbf{Y}})=\mathbf{X}'\hat{\boldsymbol{\varepsilon}}=0$$

可见, 在多元线性回归模型中, 最小二乘估计量与矩估计量也相同, 即多元线性回归模型的最小二乘估计可以看成是一种特殊的矩估计。(3.4.4) 式称为多元线性回归模型的正规方程组。这是与总体矩条件 $E(\mathbf{X}'\boldsymbol{\varepsilon})=0$ 相对应的样本矩条件。

三、最小二乘估计量的有限样本性质

多元线性回归模型在满足经典假定条件下, 有限样本的参数估计量具有线性、无偏性和有效性。最小二乘估计量满足这三个性质, 即为高斯-马尔科夫定理。

高斯-马尔可夫定理: 若多元线性回归模型 $\mathbf{Y}=\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon}$ 满足经典假定, OLS 估计量 $\hat{\boldsymbol{\beta}}$ 是最佳线性无偏估计量。

下面分别证明 OLS 估计量 $\hat{\boldsymbol{\beta}}$ 的线性、无偏性和有效性。

性质 1: 最小二乘估计量的线性。

$$\hat{\boldsymbol{\beta}}=(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}=(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon})=\boldsymbol{\beta}+(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon} \quad (3.4.5)$$

在解释变量 $\mathbf{X}$ 已知的情况下, $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ 是一个确定的矩阵, 则 $\hat{\boldsymbol{\beta}}$ 是随机变量 $\mathbf{Y}$ 的线性函数, $\hat{\boldsymbol{\beta}}$ 也是随机变量 $\boldsymbol{\varepsilon}$ 的线性函数。

性质 2: 最小二乘估计量的无偏性。

根据经典假定 1, $E(\boldsymbol{\varepsilon})=0$, 由 (3.4.5) 式可得

$$E(\hat{\boldsymbol{\beta}})=E(\boldsymbol{\beta}+(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon})=\boldsymbol{\beta}+(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\boldsymbol{\varepsilon})=\boldsymbol{\beta} \quad (3.4.6)$$

因此, $\hat{\boldsymbol{\beta}}$ 是 $\boldsymbol{\beta}$ 的无偏性估计量。

性质 3: 最小二乘估计量的有效性。

计算 $\hat{\boldsymbol{\beta}}$ 的方差为

$$\begin{aligned} \text{Var}(\hat{\boldsymbol{\beta}}) &= E[(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})'] \\ &= E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}] \\ &= [(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\sigma}^2\mathbf{I}_N\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}] \\ &= \boldsymbol{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1} \end{aligned} \quad (3.4.7)$$

证明 $\hat{\beta}$ 具有最小方差性,思路是任意找一个 β 的线性无偏估计量 $\tilde{\beta}$,证明 $\hat{\beta}$ 方差小于等于 $\tilde{\beta}$ 的方差即可。设 $\tilde{\beta}$ 也是 β 的任意一个线性无偏估计量

$$\tilde{\beta} = ((X'X)^{-1}X' + D)Y = \hat{\beta} + DY \quad (3.4.8)$$

其中, D 为一个 $k \times N$ 非随机元素构成的矩阵,如果 $D \neq 0$,则 $\tilde{\beta} \neq \hat{\beta}$;如果 $D = 0$,则 $\tilde{\beta} = \hat{\beta}$ 。

$$\begin{aligned} \tilde{\beta} &= \hat{\beta} + DY = ((X'X)^{-1}X' + D)Y \\ &= ((X'X)^{-1}X' + D)(X\beta + \varepsilon) \\ &= \beta + (X'X)^{-1}X'\varepsilon + DX\beta + D\varepsilon \end{aligned}$$

可知, $\tilde{\beta}$ 也是误差项 ε 的线性函数。

利用 $\tilde{\beta}$ 的无偏性,可获得矩阵 D 的条件

$$\begin{aligned} E(\tilde{\beta}) &= E(\hat{\beta} + DY) = E(\hat{\beta}) + DE(Y) \\ &= \beta + DE(X\beta + \varepsilon) = \beta + DX\beta = (I + DX)\beta \end{aligned} \quad (3.4.9)$$

根据 $E(\tilde{\beta}) = \beta$,得出 $DX = 0$ 。因此,得到

$$\tilde{\beta} - \beta = (X'X)^{-1}X'\varepsilon + D\varepsilon \quad (3.4.10)$$

进一步分析 $\tilde{\beta}$ 的方差,由 (3.4.10) 式和 $DX = 0$,得到

$$\begin{aligned} \text{Var}(\tilde{\beta}) &= E[(\tilde{\beta} - \beta)(\tilde{\beta} - \beta)'] \\ &= E[(X'X)^{-1}X'\varepsilon + D\varepsilon][(X'X)^{-1}X'\varepsilon + D\varepsilon] \\ &= E[(X'X)^{-1}X'\varepsilon\varepsilon'X(X'X)^{-1} + (X'X)^{-1}X'\varepsilon\varepsilon'D' + D\varepsilon\varepsilon'X(X'X)^{-1} + D\varepsilon\varepsilon'D'] \\ &= (X'X)^{-1}X'E(\varepsilon\varepsilon')X(X'X)^{-1} + (X'X)^{-1}X'E(\varepsilon\varepsilon')D' + DE(\varepsilon\varepsilon')X(X'X)^{-1} + DE(\varepsilon\varepsilon')D' \\ &= \sigma^2(X'X)^{-1} + \sigma^2(X'X)^{-1}(DX)' + \sigma^2DX(X'X)^{-1} + \sigma^2DD' \quad (E(\varepsilon\varepsilon') = \sigma^2I_N) \\ &= \sigma^2(X'X)^{-1} + \sigma^2DD' \quad (DX = 0) \\ &= \text{Var}(\hat{\beta}) + \sigma^2DD' \end{aligned} \quad (3.4.11)$$

由于 DD' 是半正定矩阵,主对角线上的元素大于等于 0,因此 $\text{Var}(\tilde{\beta})$ 主对角线上的元素大于等于 $\text{Var}(\hat{\beta})$ 主对角线上的元素,最小二乘估计量 $\hat{\beta}$ 的方差小于或等于任意其他线性无偏估计量的方差。最小二乘估计量的有效性得证。

无限样本性质,又称大样本性质或渐近性质,即样本容量趋近于无穷大的情况。在大样本的情况下,通常不需要随机误差项满足正态分布,但依据中心极限定理,误差项渐近地服从正态分布,可保证参数估计量 $\hat{\beta}$ 渐近地服从正态分布。可以证明,在无限样本下的最小二乘估计量 $\hat{\beta}$ 具有渐近无偏性,且当 $n \rightarrow \infty$ 时,方差趋近于 0,因而 $\text{plim}\hat{\beta} = \beta$,即 $\hat{\beta}$ 具有一致性。

四、回归标准误差

在最小二乘法下,误差方差的估计量为

$$\hat{\sigma}^2 = \hat{\boldsymbol{\varepsilon}}' \hat{\boldsymbol{\varepsilon}} / (N - k) \quad (3.4.12)$$

其中,当 $k=2$ 时, $\hat{\sigma}^2 = \hat{\boldsymbol{\varepsilon}}' \hat{\boldsymbol{\varepsilon}} / N - 2$, 即一元线性回归模型的误差项方差估计值。下面证明 $\hat{\sigma}^2$ 是 σ^2 的无偏估计量, 即证明 $E(\hat{\sigma}^2) = \sigma^2$ 成立, 需要利用最小二乘等幂矩阵的性质来证明^①。

$$\begin{aligned} \hat{\boldsymbol{\varepsilon}} &= \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}} \\ &= \mathbf{Y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \\ &= (\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) \\ &= \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} - \mathbf{X}\boldsymbol{\beta} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon} \\ &= [\mathbf{I}_N - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\boldsymbol{\varepsilon} \end{aligned} \quad (3.4.13)$$

定义最小二乘等幂矩阵 $\mathbf{M}$ 为:

$$\mathbf{M} = \mathbf{I}_N - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$$

(1) 最小二乘等幂矩阵 $\mathbf{M}$ 具有对称性。

$$\mathbf{M}' = (\mathbf{I}_N - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')' = \mathbf{I}_N - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \mathbf{M}$$

可知 $\mathbf{M}$ 是一个 $N \times N$ 对称矩阵。

(2) 最小二乘等幂矩阵 $\mathbf{M}$ 具有等幂性。

$$\begin{aligned} \mathbf{M}^2 &= (\mathbf{I}_N - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')(\mathbf{I}_N - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}') \\ &= \mathbf{I}_N - 2\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \\ &= \mathbf{I}_N - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \mathbf{M} \\ \mathbf{M}^n &= \mathbf{M}^{n-2}\mathbf{M}^2 = \mathbf{M}^{n-2}\mathbf{M} = \mathbf{M}^{n-1} = \dots = \mathbf{M} \end{aligned}$$

可知最小二乘等幂矩阵 $\mathbf{M}$ 具有等幂性。

由 (3.4.13) 式, 可知 $\hat{\boldsymbol{\varepsilon}} = \mathbf{M}\boldsymbol{\varepsilon}$, 得到

$$\hat{\boldsymbol{\varepsilon}}' \hat{\boldsymbol{\varepsilon}} = (\mathbf{M}\boldsymbol{\varepsilon})' \mathbf{M}\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}' \mathbf{M}' \mathbf{M}\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}' \mathbf{M}\boldsymbol{\varepsilon}$$

$\hat{\boldsymbol{\varepsilon}}' \hat{\boldsymbol{\varepsilon}}$ 是一个 1×1 矩阵, $\hat{\boldsymbol{\varepsilon}}' \hat{\boldsymbol{\varepsilon}}$ 的迹等于本身, 所以有:

$$\begin{aligned} \hat{\boldsymbol{\varepsilon}}' \hat{\boldsymbol{\varepsilon}} &= \text{tr}(\boldsymbol{\varepsilon}' \mathbf{M}\boldsymbol{\varepsilon}) = \text{tr}(\mathbf{M}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') \quad (\text{矩阵迹的轮换性}) \\ &= \text{tr}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}' \mathbf{M}) \\ E(\hat{\boldsymbol{\varepsilon}}' \hat{\boldsymbol{\varepsilon}}) &= E[\text{tr}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}' \mathbf{M})] \\ &= \text{tr}[E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') \mathbf{M}] \\ &= \text{tr}(\sigma^2 \mathbf{M}) \quad (\text{Var}(\boldsymbol{\varepsilon}) = E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \sigma^2 \mathbf{I}_N) \\ &= \sigma^2 \text{tr}(\mathbf{M}) \end{aligned}$$

① $\hat{\sigma}^2$ 是 σ^2 无偏估计量的证明为选学内容, 供学有余力的学生学习。

$$\begin{aligned}
\text{tr}(\mathbf{M}) &= \text{tr}[\mathbf{I}_N - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'] \\
&= [\text{tr}(\mathbf{I}_N) - \text{tr}((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X})] \quad (\text{矩阵迹的轮换性}) \\
&= N - \text{tr}(\mathbf{I}_k) \\
&= N - k
\end{aligned}$$

$$\begin{aligned}
E(\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}) &= \sigma^2(N-k) \\
E(\hat{\sigma}^2) &= E[\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}/(N-k)] = \sigma^2
\end{aligned}$$

即证明 $\hat{\sigma}^2$ 是 σ^2 的无偏估计量。这样就可以求出回归标准误差

$$\hat{\sigma} = \sqrt{\frac{\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}}{N-k}}$$

用 σ^2 的无偏估计量 $\hat{\sigma}^2$ 估计 $\text{Var}(\hat{\boldsymbol{\beta}})$, 即为

$$\widehat{\text{Var}}(\hat{\boldsymbol{\beta}}) = \hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}$$

五、样本回归方程的统计性质

(一) 样本均值点满足样本回归方程

证明: 由 $\mathbf{X}'\hat{\boldsymbol{\varepsilon}} = 0$, 即

$$\begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ x_{11} & x_{21} & x_{31} & \cdots & x_{N1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{1k-1} & x_{2k-1} & x_{3k-1} & \cdots & x_{Nk-1} \end{bmatrix}_{k \times N} \begin{bmatrix} \hat{\varepsilon}_1 \\ \hat{\varepsilon}_2 \\ \vdots \\ \hat{\varepsilon}_N \end{bmatrix}_{N \times 1} = 0 \quad (3.4.14)$$

由 (3.4.14) 式中的第一个方程可知

$$\begin{aligned}
\bar{\hat{\varepsilon}} &= \frac{1}{N} \sum_{i=1}^N \hat{\varepsilon}_i = 0 \\
\hat{\varepsilon}_i &= y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \cdots + \hat{\beta}_{k-1} x_{ik-1}) \\
\frac{1}{N} \sum_{i=1}^N \hat{\varepsilon}_i &= \frac{1}{N} \sum_{i=1}^N y_i - \frac{1}{N} \sum_{i=1}^N (\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \cdots + \hat{\beta}_{k-1} x_{ik-1}) \\
\text{设 } \bar{y} &= \frac{1}{N} \sum_{i=1}^N y_i, \bar{x}_j = \frac{1}{N} \sum_{i=1}^N x_{ij}, \text{ 则有} \\
\bar{\hat{\varepsilon}} &= \bar{y} - (\hat{\beta}_0 + \hat{\beta}_1 \bar{x}_1 + \hat{\beta}_2 \bar{x}_2 + \cdots + \hat{\beta}_{k-1} \bar{x}_{k-1}) = 0 \\
\bar{y} - \bar{\mathbf{X}}\hat{\boldsymbol{\beta}} &= 0 \\
\bar{y} &= \bar{\mathbf{X}}\hat{\boldsymbol{\beta}} \quad (3.4.15)
\end{aligned}$$

其中, $\bar{\mathbf{X}} = (1 \quad \bar{x}_1 \quad \bar{x}_2 \quad \cdots \quad \bar{x}_{k-1})$ 。

样本均值点满足样本回归方程得证。

(二) 被解释变量拟合值的平均数等于其样本观测值的平均数,即 $\bar{\hat{y}} = \bar{y}$

证明: $\bar{\hat{y}} = \frac{1}{N} \sum_{i=1}^N \hat{y}_i$ 。根据样本回归方程 $\hat{Y} = \mathbf{X}\hat{\boldsymbol{\beta}}$,可知

$$\bar{\hat{y}} = \bar{\mathbf{X}}\hat{\boldsymbol{\beta}}$$

由 $\bar{y} = \bar{\mathbf{X}}\hat{\boldsymbol{\beta}}$,可得

$$\bar{\hat{y}} = \bar{y} \quad (3.4.16)$$

第五节 实例与 Stata 操作

实例一 证券“贝塔”系数

“投资有风险,入市需谨慎。”证券监管机构警示投资者,若想获得超额回报就要承担相应风险,度量这类证券市场风险往往会借助衡量工具—— **β 系数**(beta coefficient),又称系统性风险指数。 β 系数可以通过**资本资产定价模型**(CAPM)测算:

$$R - R_f = \beta(R_m - R_f)$$

其中, R 为证券资产期望收益, R_f 为无风险收益, R_m 为市场组合的期望收益。一般地,证券资产期望超额收益 $R - R_f$ 与可能的资产组合期望超额收益 $R_m - R_f$ 成比例。 β 系数是 $R - R_f$ 对 $R_m - R_f$ 的回归系数,在数值上等于

$$\beta = \frac{\text{Cov}(R - R_f, R_m - R_f)}{\text{Var}(R_m - R_f)}$$

依据 CAPM 模型, β 系数用来衡量证券资产相对于证券市场的风险情况,也代表了该资产的系统性风险。通常情况下,可以使用一元线性回归模型来估计 CAPM 模型中的 β 系数,即

$$R_{it} = \alpha_i + \beta_i R_{mt} + \epsilon_{it}$$

其中, R_{it} 为证券 i 在 t 时的收益率, R_{mt} 是市场指数在 t 时的收益率, β_i 是证券 i 收益率相对于市场指数收益率变化的敏感程度, ϵ_{it} 是随机误差项。此外,在单因素模型中也可以用超额收益率代替收益率。

以顺丰控股(002352)和上证综合指数(000001)为例,将 2020 年 1 月份顺丰控股每日涨跌幅对上证综合指数收益率进行回归,表 3-1 为回归模型使用的样本数据,其中:

涨跌幅(指当日收益率,单位:%)=(当日收盘价-前日收盘价)/前日收盘价

表 3-1 上证综合指数和顺丰控股 2020 年 1 月涨跌幅(单位:%)

日期	顺丰控股	上证综合指数	日期	顺丰控股	上证综合指数
2020-01-02	0.78	1.149 9	2020-01-14	-1.319	-0.280 8
2020-01-03	-0.774	-0.045 8	2020-01-15	-1.122	-0.540 2
2020-01-06	-0.914	-0.012 3	2020-01-16	-0.378	-0.516 4
2020-01-07	2.307	0.693 8	2020-01-17	0.868	0.046
2020-01-08	-1.194	-1.221	2020-01-20	4.519	0.659 8
2020-01-09	0.027	0.912 6	2020-01-21	0.257	-1.409 8
2020-01-10	0.886	-0.083 7	2020-01-22	-1.412	0.282 2
2020-01-13	0.878	0.752 8	2020-01-23	2.109	-2.751 8

数据来源:国泰安数据库

在 Stata 中输入以下命令

```
. reg Ri Rm
```

如图 3-2 所示,回归结果为

$$\hat{R}_i = 0.203R_{mt} + 0.375$$

Source	SS	df	MS	Number of obs	=	16
Model	.620589403	1	.620589403	F(1, 14)	=	0.22
Residual	38.647744	14	2.76055314	Prob > F	=	0.6427
				R-squared	=	0.0158
Total	39.2683334	15	2.61788889	Adj R-squared	=	-0.0545
				Root MSE	=	1.6615

Ri	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
Rm	.2029422	.428024	0.47	0.643	-.7150781 1.120962
_cons	.3748686	.4201622	0.89	0.387	-.5262898 1.276027

图 3-2 回归结果

可知样本期内顺丰控股以“日”为单位的 β 系数 $=0.203 < 1$, 代表顺丰控股波动风险小于上证综合指数(市场组合)。

实例二 地区能源消耗与产业结构、出口贸易的关系

为研究地区能源消耗与产业结构、出口贸易的关系,建立如下方程

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i$$

其中, y_i 为地区能源消耗总量(百万吨标准煤), x_{i1} 是用工业增加值占 GDP 比重表示的产业结构, x_{i2} 表示出口贸易总额(十亿美元)。表 3-2 包含 2019 年 30 个地区能源消耗的相关数据, 据此建立多元线性回归模型, 并进行分析。

表 3-2 能源消耗的相关数据

地区	能源消耗/ 百万吨标准煤	工业增加值 占比/ (%)	出口贸易 总额/十亿 美元	地区	能源消耗/ 百万吨标准煤	工业增加值 占比/ (%)	出口贸易 总额/十亿 美元
北京	73.6	11.971 5	75.050 9	河南	223	33.393 4	54.213 3
天津	82.41	31.107 4	43.791 8	湖北	173.16	34.576 2	35.995 3
河北	325.45	32.335 2	34.380 9	湖南	160.01	30.069 1	44.540 7
山西	208.59	38.809 4	11.690 2	广东	341.42	36.246 8	629.454
内蒙古	253.46	31.713	5.468 32	广西	112.7	24.704 9	37.746 5
辽宁	237.49	32.396 3	45.437 7	海南	22.64	11.216	4.987 03
吉林	71.32	28.548 3	4.703 82	重庆	88.89	27.755	53.803 3
黑龙江	116.14	24.615 3	5.068 51	四川	207.91	28.396 9	56.545 2
上海	116.96	25.179 5	198.994	贵州	104.23	26.590 3	4.740 28
江苏	325.26	37.732 5	394.828	云南	121.58	23.254 2	15.024 2
浙江	223.93	36.055 4	334.605	陕西	134.78	36.675 9	27.221 6
安徽	138.7	30.347 5	40.411 1	甘肃	78.18	26.608 4	1.906 996
福建	137.18	36.983 8	120.203	青海	42.35	27.945 3	0.293 533
江西	96.65	35.570 2	36.193	宁夏	76.48	33.946 9	2.159 301
山东	413.9	32.258 21	161.439 7	新疆	184.9	28.175 13	18.044 31

数据来源:中国统计年鉴

利用表 3-2 的数据对模型做 OLS 估计,输入如下命令

```
.reg y x1 x2
```

如图 3-3 所示,回归方程为

$$\hat{y}_i = -30.874 + 5.667x_{i1} + 0.299x_{i2}$$

Source	SS	df	MS	Number of obs	=	30
Model	120074.647	2	60037.3235	F(2, 27)	=	10.88
Residual	148952.313	27	5516.75233	Prob > F	=	0.0003
Total	269026.96	29	9276.79172	R-squared	=	0.4463
				Adj R-squared	=	0.4053
				Root MSE	=	74.275

y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
x1	5.666897	2.224452	2.55	0.017	1.102698	10.2311
x2	.2987744	.1045969	2.86	0.008	.0841593	.5133894
_cons	-30.87436	65.46881	-0.47	0.641	-165.2053	103.4565

图 3-3 回归结果

该样本回归方程表示,地区能源消耗受产业结构和出口贸易的影响,产业结构中工业增加值比重越大,能源消耗越多;同样,出口贸易额越大,能源消耗也越多。



本章小结

1. 构建一元线性回归模型,为保证参数估计量具有良好的性质,设定随机误差项满足零均值、同方差、无序列相关和正态分布的条件,以及解释变量的方差是一个非零的有限值,即具有变异性,且与随机误差项不存在相关性等条件。

2. 采用最小二乘法和矩估计法,估计一元线性回归模型的参数结果相同,参数估计量在有限样本下具有线性、无偏性和有效性,满足这些性质的估计量称为 BLUE 估计量。

3. 构建多元线性回归模型,类似于一元线性回归模型,设定随机误差项和解释变量满足经典假定条件,与一元线性回归模型相比,增加一个解释变量不存在多重共线性的假定。

4. 采用最小二乘法和矩估计法,估计多元线性回归模型的参数结果相同,参数估计量在有限样本的情况下满足线性、无偏性和有效性;在无限样本下的最小二乘估计量具有渐近无偏性和一致性。



练习题

一、单选题

1. 下列模型中,属于变量线性的模型是()。

A. $y = \alpha + \beta_1 x + \beta_2 x^2 + \epsilon$

B. $y = Ax_1^\alpha x_2^\beta e^\epsilon$

C. $y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \epsilon$

D. $y = \alpha + \beta x^{-2} + \epsilon$

2. 最小二乘法使用的正规方程是()。

A.
$$\begin{cases} \sum \hat{\epsilon}_i = 0 \\ \sum \hat{\epsilon}_i x_i = 0 \end{cases}$$

B.
$$\begin{cases} \sum (y_i - \hat{y}_i) = 0 \\ \sum (y_i - \hat{y}_i) \bar{x} = 0 \end{cases}$$

C.
$$\begin{cases} \sum (y_i - \hat{y}_i) = 0 \\ \sum (y_i - \beta_0 - \beta_1 x_i) = 0 \end{cases}$$

D.
$$\begin{cases} \sum (y_i - \beta_0 - \beta_1 x_i) = 0 \\ \sum (y_i - \beta_0 - \beta_1 x_i) x_i = 0 \end{cases}$$

3. 在总体回归直线 $E(y) = \beta_0 + \beta_1 x$ 中, β_1 表示()。

A. 当 x 增加一个单位时, y 增加 β_1 个单位

B. 当 x 增加一个单位时, y 平均增加 β_1 个单位

- A. $Q_i^d = 6.8 + 0.65y_i - 0.37P_i$ B. $Q_i^d = 7.2 + 0.8y_i + 0.61P_i$
 C. $Q_i^d = 7.2 - 0.8y_i + 0.61P_i$ D. $Q_i^s = 18.29 + 0.75P_i$
 E. $Q_i^s = 16.2 + 0.65P_i$

6. 在多元线性回归模型中,若某个解释变量与其余解释变量的相关系数接近于1,则表明模型中存在()。

- A. 异方差性 B. 序列相关
 C. 多重共线性 D. 高拟合优度
 E. 内生性

三、问答题

1. 在研究教育回报率的模型 $y = \beta_0 + \beta_1 x + \epsilon$ 中, y 为某公司员工的工资(单位:元), x 为员工的受教育年限(单位:年)。

(1) 从经济角度解释 $\hat{\beta}_0, \hat{\beta}_1$ 。

(2) 如果被解释变量员工的工资 y 单位变为百元,截距和斜率是否有变化? 如有变化,请说明如何变化。

2. 在研究女性受教育程度与生育率关系的模型 $ki = \beta_0 + \beta_1 \text{edu} + \epsilon$ 中, ki 表示妇女所生孩子的数目, edu 表示该妇女受教育的年限, ϵ 是随机误差项。那么 ϵ 中包含了什么样的因素? 它们可能与受教育程度相关吗?

3. 在消费函数 $\widehat{\text{con}} = \hat{\beta}_0 + \hat{\beta}_1 \text{inc}$ 中,收入的边际消费倾向(MPC)就是斜率 $\hat{\beta}_1$,利用对300个家庭的月收入(单位:元)和消费(单位:元)的观测,得到如下方程:

$$\widehat{\text{con}} = 239 + 0.853 \text{inc}, \quad R^2 = 0.792$$

结合方程中的截距项、斜率系数的符号和大小,分析其经济含义。

4. 采用某地区596个家庭的调查数据,研究子女教育的影响因素,方程如下:

$$\widehat{\text{edu}} = 12.89 - 0.089 \text{chi} + 0.127 \text{par}$$

$$N = 596, \quad R^2 = 0.683$$

其中, edu 是受教育年限, chi 是家庭孩子的个数, par 是父母受教育的平均年数。

(1) 解释家庭孩子的个数 chi 对受教育年限的影响。

(2) 假设一个人是独生子女,父母平均接受了12年的教育,预计其可接受的教育年数是多少?

(3) 假设一个人有一个弟弟,父母平均接受了15年的教育,预计其可接受的教育年数为多少?

5. 请根据表3-3中2018年中国旅游景区门票收入、旅游景区总数、客运量和居民人均可支配收入数据,建立回归模型,分析旅游景区总数、客运量和居民人均可支配收入等对旅游景区门票收入的影响。

表 3-3 2018 年中国旅游景区门票收入和居民人均可支配收入等相关数据

地区	旅游景区门票收入/亿元	旅游景区总数/个	客运量/万人	居民人均可支配收入/元
北京	17.10	254	58 934.70	62 361.22
天津	9.14	104	17 450.31	39 506.15
河北	21.83	420	47 345.65	23 445.65
山西	19.26	204	23 837.39	21 990.14
内蒙古	8.92	386	13 268.30	28 375.65
辽宁	37.80	454	71 342.90	29 701.45
吉林	8.58	244	31 956.21	22 798.37
黑龙江	10.09	451	31 567.79	22 725.85
上海	22.63	113	15 844.61	64 182.65
江苏	66.26	630	120 611.96	38 095.79
浙江	86.57	806	98 380.15	45 839.84
安徽	49.71	613	63 346.58	23 983.58
福建	28.13	341	48 105.13	32 643.93
江西	51.32	407	60 685.68	24 079.68
山东	93.58	1 292	67 433.46	29 204.61
河南	52.89	511	110 420.90	21 963.54
湖北	48.09	415	98 350.02	25 814.54
湖南	60.29	438	106 679.62	25 240.75
广东	66.55	381	142 144.47	35 809.90
广西	29.80	507	47 930.81	21 485.03
海南	16.75	53	14 382.89	24 579.04
重庆	20.50	239	60 587.44	26 385.84
四川	49.59	565	98 568.77	22 460.55
贵州	35.13	359	93 024.73	18 430.18
云南	28.86	233	41 484.30	20 084.19
西藏	1.98	116	1 399.47	17 286.06
陕西	41.85	454	71 582.97	22 528.26
甘肃	12.95	298	42 185.49	17 488.39
青海	6.01	110	6 442.74	20 757.26
宁夏	4.05	80	6 136.87	22 400.42
新疆	15.12	446	21 204.24	21 500.24

数据来源:《中国旅游数据库》《中国宏观经济数据库》。

6. 请证明下列结论。

(1) 最小二乘法的残差与解释变量不相关, 即 $\text{Cov}(\hat{\varepsilon}_i, x_i) = 0$ 。

(2) 最小二乘法残差与被解释变量估计值不相关, 即 $\text{Cov}(\hat{\varepsilon}_i, \hat{y}_i) = 0$ 。

(3) 最小二乘法 y_i 拟合值的平均数等于其样本观测值的平均数, 即 $\bar{\hat{y}} = \bar{y}$ 。

7. 请收集 2019 年我国农村与城市居民人均消费支出 Y 与可支配收入 X 的地区数据, 进行回归分析。利用计量软件分别估计城市、农村居民人均消费支出与可支配收入的线性回归模型, 写出表达式, 解释估计结果的实际含义, 并进行对比分析。