第 1 章

多元回归模型

在现实经济生活中，一个内生变量往往会受到多个外生变量的影响。例如，厂商的产量除了受到劳动投入的影响外，还受到资金投入、技术水平等多种因素的影响。居民家庭消费支出，除了受到家庭收入影响之外，还受到消费者偏好、物价水平、消费信贷等多种因素的影响。具体表现在模型中就是一个被解释变量，受到多个解释变量的影响。这样的模型就是多元回归模型。如果被解释变量的变化可以由一个主要解释变量加以说明，其他解释变量的影响可以忽略，则可以用一元回归模型表示。如果被解释变量的变化受到多个解释变量的影响，为描述被解释变量与多个解释变量之间的关系，则需要建立多元回归模型。

本章学习的目的：(1)掌握多元线性回归模型的基本假设；(2)掌握模型参数的最小二乘估计法以及模型的统计检验；(3)能够应用多元回归模型进行经济预测；(4)掌握几种常用的可线性化回归模型；(5)能够应用 EViews 软件解决多元回归分析中的实际问题。

1.1 多元线性回归模型的估计

1.1.1 多元线性回归模型及其矩阵表示

各种经济变量之间的关系，可以划分为两类：一类是完全确定的函数关系，另一类是非确定性的相关关系。如果一个变量 y 的取值可以通过另一个变量 x 或另一组变量 $(x_1,x_2,\cdots,x_k)$ 以某种形式唯一地、精确地确定，则 y 与这个 x 之间或 y 与这组 $(x_1,x_2,\cdots,x_k)$ 之间的关系就是函数关系。用代数式表示就是 $y=f(x)$，或者 $y=f(x_1,x_2,\cdots,x_k)$。其中，最简单的形式为一元线性函数关系 $y=b_0+b_1x$。例如，当某种商品单价 p 固定不变，这种商品的销售收入 y 与销售量 x 之间的关系为一元线性关系，即 $y=px$。如果用 x、y 构成的直角坐标图来表示，此式所表示的函数关系为一经过坐标原点的直线。所有可能的点都在这条直线上。

经济变量之间的另一类关系为不完全确定的相关关系。例如，经济分析中，投入与产出之间的关系，消费与收入之间的关系等都是不完全确定的相关关系，如果一个变量 y 的取值受到另一个变量 x 或另一组变量 $(x_1,x_2,\cdots,x_k)$ 的影响，但给定这一个 x 或一组 $(x_1,x_2,\cdots,x_k)$ 的值，y 的取值并不是唯一确定的，则变量 y 与这一个 x 或一组 $(x_1,x_2,\cdots,x_k)$

之间为相关关系。用代数式表示就是 $y=f(x,u)$ 或者 $y=f(x_1,x_2,\cdots,x_k,u)$。

对于一组不同的观测值 (x_t,y_t) 或 $(x_{1t},x_{2t},\cdots,x_{kt},y_t)$，$t=1,2,\cdots,n$，它们都落在对应上述相关关系的代数式上：

$$y_t = f(x_t,u_t) \quad (t=1,2,\cdots,n)$$

或者

$$y_t = f(x_{1t},x_{2t},\cdots,x_{kt},u_t) \quad (t=1,2,\cdots,n)$$

其中，u_t 为随机误差项。最简单的形式为一元线性回归模型

$$y_t = b_0 + b_1 x_t + u_t$$

“线性”一词在这里有两重含义。它一方面指被解释变量与解释变量之间为线性关系，另一方面也指因变量与参数之间为线性关系。

与数学中的函数关系相比，相关关系 $y_t=f(x_t,u_t)$ 或 $y_t=f(x_{1t},x_{2t},\cdots,x_{kt},u_t)$ 的显著特点是多了随机误差项 u_t。产生误差项的原因主要有以下几方面：(1)模型中被忽略掉的影响因素造成的误差。在一般情况下，每一个经济变量通常要受到多种因素的影响。但是为了简化分析，突出主要矛盾，在构造回归模型时，通常只选取最重要的解释变量与被解释变量构成回归模型，将次要的影响因素忽略掉。这些被忽略的影响因素对被解释变量的影响就归入了误差项中。(2)模型关系设定不准确造成的误差。在一般情况下，解释变量与被解释变量之间的关系可能是比较复杂的非线性关系。在构造模型时，为了简化模型，用线性模型代替了非线性模型，或者用简单的非线性模型代替了复杂的非线性模型，造成了模型关系设定不准确而形成误差。(3)变量的测量误差。由于测量工具的精确度和测量方法不正确，使得观察值与真实值不完全一致，造成测量误差。(4)变量的内在随机性。经济变量本身受很多随机因素影响(比如自然灾害、经济危机等)，不具有确定性和重复性，由于某些变量所固有的内在随机性，也会对被解释变量产生随机性影响。

总之，误差项的存在是计量经济模型的特点，是计量经济模型与数学中完全确定的函数关系的主要区别。

例如，在需求分析中，商品价格是影响需求量的主要因素，如果根据经验和样本特征，判断出其他因素的影响微不足道，则可以设定如下一元线性回归模型：

$$Q_t = a + bP_t + u_t$$

其中，Q_t 表示商品需求量，P_t 表示商品价格，a、b 为参数，u_t 为随机误差项。

如果根据理论分析和先验经验，认为需求量除了受商品价格的影响外，还受其他因素的影响，那么在模型设定时就应引入这些因素。例如，如果消费者收入 I_t 也是影响需求量的重要因素，则可以设定如下二元线性回归模型：

$$Q_t = a + bP_t + cI_t + u_t$$

其中 a、b、c 为参数，u_t 为随机误差项。

再比如，在生产理论中，著名的 Cobb—Douglas 生产函数描述了产量与要素投入之间的关系，其函数形式为

$$Y_t = AL_t^{\alpha}K_t^{\beta}e^{u_t}$$

其中 Y_t 表示产量，L_t、K_t 分别表示劳动和资本投入，α、β 为参数，u_t 为随机误差项。因变量 Y_t 与解释变量 L_t、K_t 之间的关系是非线性的，利用对数变换可将其转化为如下回归模型：

$$\ln Y_t = \ln A + \alpha\ln L_t + \beta\ln K_t + u_t$$

在计量经济学中，将含有两个以上解释变量的回归模型称为多元回归模型，相应地，在此基础上进行的回归分析称为多元回归分析。如果总体回归函数描述了一个因变量与多个解释变量之间的线性关系，由此而设定的回归模型就称为多元线性回归模型。

一般地，如果因变量 y 与解释变量 $x_1, x_2, \cdots, x_k$ 之间服从如下关系：

$$y = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_k x_k + u \tag{1.1.1}$$

则对因变量 y 及解释变量 $x_1, x_2, \cdots, x_k$ 作 n 次观测后，所得 n 组观测样本（$y_t, x_{1t}, x_{2t}, \cdots, x_{kt}$）（$t=1,2,\cdots,n$）将满足如下关系：

$$y_t = b_0 + b_1 x_{1t} + b_2 x_{2t} + \cdots + b_k x_{kt} + u_t \tag{1.1.2}$$

这就是多元线性回归模型的一般形式。（$y_t, x_{1t}, x_{2t}, \cdots, x_{kt}$）为第 t 次观测样本（$t=1,2,\cdots,n$），b_j（$j=0,1,2,\cdots,k$）为模型的参数或回归系数，u_t 为随机误差项。

模型中的回归系数 b_j（$j=1,2,\cdots,k$）表示在其他解释变量不变的条件下，第 j 个解释变量变动一个单位对因变量均值的影响程度。多元线性回归模型中这样的回归系数，也称为偏回归系数。偏回归系数反映了当模型中的其他变量不变时，其中一个解释变量变动对因变量均值的影响。

假定随机误差项 u_t 的期望函数 $E(u_t)=0$，由式(1.1.2)可得因变量 y_t 的期望函数

$$E(y_t) = b_0 + b_1 x_{1t} + b_2 x_{2t} + \cdots + b_k x_{kt} \tag{1.1.3}$$

它是解释变量的多元线性函数，称为多元线性总体回归方程。在总体回归方程中，各参数是未知的，我们进行回归分析的主要目的之一就是要利用样本观测值对未知参数 b_j 进行估计。假定通过适当的方法可估计出未知参数的值 $\hat{b}_j$，用参数估计值 $\hat{b}_j$ 替换总体回归函数的未知参数 b_j，就得到多元线性样本回归方程

$$\hat{y}_t = \hat{b}_0 + \hat{b}_1 x_{1t} + \hat{b}_2 x_{2t} + \cdots + \hat{b}_k x_{kt} \tag{1.1.4}$$

它是总体回归方程的估计，其中 $\hat{b}_j$（$j=0,1,2,\cdots,k$）是对总体回归参数 b_j 的估计。

由样本回归方程得到的因变量估计值 $\hat{y}_t$ 与实际观测值 y_t 之间通常存在偏差，这一偏差就是残差 e_t。这样，与式(1.1.2)相对应，多元线性样本回归模型为

$$y_t = \hat{y}_t + e_t = \hat{b}_0 + \hat{b}_1 x_{1t} + \hat{b}_2 x_{2t} + \cdots + \hat{b}_k x_{kt} + e_t \tag{1.1.5}$$

将 n 次观测样本所遵从的 n 个随机方程式(1.1.2)写成方程组的形式

$$\begin{cases} y_1 = b_0 + b_1 x_{11} + b_2 x_{21} + \cdots + b_k x_{k1} + u_1 \\ y_2 = b_0 + b_1 x_{12} + b_2 x_{22} + \cdots + b_k x_{k2} + u_2 \\ \cdots\cdots \\ y_n = b_0 + b_1 x_{1n} + b_2 x_{2n} + \cdots + b_k x_{kn} + u_n \end{cases}$$

利用矩阵或向量运算，上式可表示为

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{21} & \cdots & x_{k1} \\ 1 & x_{12} & x_{22} & \cdots & x_{k2} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{kn} \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ b_k \end{pmatrix} + \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} \tag{1.1.6}$$

记 $\boldsymbol{Y}=\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$ 为被解释变量的观测值向量，$\boldsymbol{X}=\begin{pmatrix} 1 & x_{11} & x_{21} & \cdots & x_{k1} \\ 1 & x_{12} & x_{22} & \cdots & x_{k2} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{kn} \end{pmatrix}$ 为解释变量的观测

值矩阵，$\boldsymbol{B}=\begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ b_k \end{pmatrix}$为总体回归参数向量；$\boldsymbol{U}=\begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}$为随机误差项向量。则多元线性回归模型利用矩阵表示如下：

$$\boldsymbol{Y} = \boldsymbol{XB} + \boldsymbol{U} \tag{1.1.7}$$

它反映了总体变量间的真实关系。

类似地，多元线性回归方程利用矩阵表示如下：

$$E(\boldsymbol{Y}) = \boldsymbol{XB} \tag{1.1.8}$$

它反映了总体变量间的依存规律。

多元线性样本回归模型利用矩阵表示如下：

$$\boldsymbol{Y} = \boldsymbol{X}\hat{\boldsymbol{B}} + \boldsymbol{e} \tag{1.1.9}$$

反映了样本显示的变量关系。

多元线性样本回归方程利用矩阵表示如下：

$$\hat{\boldsymbol{Y}} = \boldsymbol{X}\hat{\boldsymbol{B}} \tag{1.1.10}$$

其中$\hat{\boldsymbol{B}}=\begin{pmatrix} \hat{b}_0 \\ \hat{b}_1 \\ \vdots \\ \hat{b}_k \end{pmatrix}$为回归系数估计值向量，$\boldsymbol{e}=\begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}$为残差向量，式(1.1.10)反映了样本显示的变量依存规律。

1.1.2 多元线性回归模型的基本假定

在回归分析中，为了寻找有效的参数估计方法和对模型进行统计检验，常常需要对模型中的随机误差项和解释变量作一些假定。多元线性回归模型的基本假定包括对解释变量的假定、对随机误差项的假定、对模型设定的假定等几个方面。多元线性回归模型的基本假定条件如下。

假设1 随机误差项的期望为零，即 $E(u_t)=0$。

假设2 不同的随机误差项之间相互独立，即

$$\operatorname{cov}(u_t,u_s) = E[(u_t - E(u_t))(u_s - E(u_s))] = E(u_t u_s) = 0 \quad (t \neq s;\ t,s = 1,2,\cdots,n)$$

可以证明，被解释变量也是相互独立的。

假设3 随机误差项的方差与 t 无关，为一个常数，即 $\operatorname{var}(u_t)=\sigma^2(t=1,2,\cdots,n)$。

可以证明，被解释变量 y_t 的方差与 t 无关，与随机误差项有相同的方差。

假设4 随机误差项与解释变量不相关，即 $\operatorname{cov}(x_{jt},u_t)=0$。通常假定 $x_{jt}(j=1,2,\cdots,k;\ t=1,2,\cdots,n)$为非随机变量，这个假设条件自动成立。

假设5 随机误差项 u_t 为服从正态分布的随机变量，即 $u_t \sim N(0,\sigma^2)$。可以推断被解释变量 y_t 也服从正态分布。

假设 6　解释变量之间不存在多重共线性，即假定各解释变量之间不存在线性相关关系，或者说各解释变量之间线性无关。这样假定的目的在于避免 $x_1,x_2,\cdots,x_k$ 中某一个解释变量被其他解释变量线性表达，保证 $x_1,x_2,\cdots,x_k$ 之间相互独立，从而对参数 $b_0,b_1,b_2,\cdots,b_k$ 的估计值取得唯一的结果。如果违背这一假定，则参数估计值 $\hat{b}_0,\hat{b}_1,\hat{b}_2,\cdots,\hat{b}_k$ 将不是唯一的。

以上六个假设条件称为多元线性回归模型的经典假设条件。模型的假设条件使用矩阵形式表示更方便、更简洁。

假设 1 用矩阵形式表示为

$$E(\boldsymbol{U})=\boldsymbol{0} \tag{1.1.11}$$

假设 2、3 用矩阵形式表示就是随机误差项的方差—协方差矩阵，形如

$$E(\boldsymbol{U}\boldsymbol{U}')=\begin{pmatrix} \operatorname{var}(u_1) & \operatorname{cov}(u_1,u_2) & \cdots & \operatorname{cov}(u_1,u_n) \\ \operatorname{cov}(u_2,u_1) & \operatorname{var}(u_2) & \cdots & \operatorname{cov}(u_2,u_n) \\ \vdots & \vdots & \cdots & \vdots \\ \operatorname{cov}(u_n,u_1) & \operatorname{cov}(u_n,u_2) & \cdots & \operatorname{var}(u_n) \end{pmatrix}=\begin{pmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{pmatrix}=\sigma^2\boldsymbol{I}_n \tag{1.1.12}$$

假设 4 可以表示为矩阵 $\boldsymbol{X}$ 的所有元素均为非随机元素，即 $\boldsymbol{X}$ 为确定的矩阵。用矩阵表示为 $E(\boldsymbol{X}'\boldsymbol{U})=\boldsymbol{0}$。

假设 5 可以表示为随机误差项向量 $\boldsymbol{U}$ 服从多元正态分布，即 $\boldsymbol{U}\sim N(0,\sigma^2\boldsymbol{I}_n)$。这种方式表示包括了假设 1、假设 2、假设 3 和假设 5。

假设 6 表示各解释变量之间不存在多重共线性。在此条件下，解释变量观测值矩阵 $\boldsymbol{X}$ 列满秩，即 $\operatorname{rank}(\boldsymbol{X})=k+1$，此时，方阵 $\boldsymbol{X}'\boldsymbol{X}$ 满秩，$\operatorname{rank}(\boldsymbol{X}'\boldsymbol{X})=k+1$，从而 $\boldsymbol{X}'\boldsymbol{X}$ 可逆，$(\boldsymbol{X}'\boldsymbol{X})^{-1}$ 存在。

1.1.3　多元线性回归模型的估计

1. 参数的最小二乘估计

为了使样本回归模型尽可能接近总体回归模型，对于每个特定的样本来说，就是要使样本回归方程的拟合值与实际观测值的误差，即残差越小越好。由于残差有正有负，简单的代数和会相互抵消，为了便于数学上的处理，我们使用残差平方和最小准则即最小二乘法估计模型的回归参数。

设 $(y_t,x_{1t},x_{2t},\cdots,x_{kt})$ 为第 t 次观测样本 $(t=1,2,\cdots,n)$，为使残差

$$e_t=y_t-\hat{y}_t=y_t-(\hat{b}_0+\hat{b}_1x_{1t}+\hat{b}_2x_{2t}+\cdots+\hat{b}_kx_{kt})$$

平方和

$$\sum e_t^2=\sum(y_t-\hat{y}_t)^2=\sum[y_t-(\hat{b}_0+\hat{b}_1x_{1t}+\hat{b}_2x_{2t}+\cdots+\hat{b}_kx_{kt})]^2$$

达到最小，根据极值原理有如下条件：

$$\frac{\partial\left(\sum e_t^2\right)}{\partial\,\hat{b}_j}=0\quad(j=0,1,2,\cdots,k) \tag{1.1.13}$$

由此得到

$$\begin{cases}\sum 2e_t(-1)=-2\sum[y_t-(\hat{b}_0+\hat{b}_1x_{1t}+\hat{b}_2x_{2t}+\cdots+\hat{b}_kx_{kt})]=0\\ \sum 2e_t(-x_{1t})=-2\sum x_{1t}[y_t-(\hat{b}_0+\hat{b}_1x_{1t}+\hat{b}_2x_{2t}+\cdots+\hat{b}_kx_{kt})]=0\\ \sum 2e_t(-x_{2t})=-2\sum x_{2t}[y_t-(\hat{b}_0+\hat{b}_1x_{1t}+\hat{b}_2x_{2t}+\cdots+\hat{b}_kx_{kt})]=0\\ \vdots\\ \sum 2e_t(-x_{kt})=-2\sum x_{kt}[y_t-(\hat{b}_0+\hat{b}_1x_{1t}+\hat{b}_2x_{2t}+\cdots+\hat{b}_kx_{kt})]=0\end{cases}$$

上述$(k+1)$个方程称为正规方程。用矩阵表示就是

$$\begin{pmatrix}\sum e_t\\ \sum x_{1t}e_t\\ \vdots\\ \sum x_{kt}e_t\end{pmatrix}=\begin{pmatrix}1&1&\cdots&1\\ x_{11}&x_{12}&\cdots&x_{1n}\\ \vdots&\vdots&\cdots&\vdots\\ x_{k1}&x_{k2}&\cdots&x_{kn}\end{pmatrix}\begin{pmatrix}e_1\\ e_2\\ \vdots\\ e_n\end{pmatrix}=\boldsymbol{X}'\boldsymbol{e}=\boldsymbol{0} \tag{1.1.14}$$

样本回归模型$\boldsymbol{Y}=\boldsymbol{X}\hat{\boldsymbol{B}}+\boldsymbol{e}$两边同乘样本观测值矩阵$\boldsymbol{X}$的转置$\boldsymbol{X}'$，有

$$\boldsymbol{X}'\boldsymbol{Y}=\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{B}}+X'e$$

将极值条件式(1.1.14)代入，得到正规方程组

$$\boldsymbol{X}'\boldsymbol{Y}=\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{B}}$$

由古典假定条件6可知$(\boldsymbol{X}'\boldsymbol{X})^{-1}$存在，用$(\boldsymbol{X}'\boldsymbol{X})^{-1}$左乘上述方程两端，得到参数向量$\boldsymbol{B}$的最小二乘估计量为

$$\hat{\boldsymbol{B}}=(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y} \tag{1.1.15}$$

特别地，对于一元线性回归模型$y_t=b_0+b_1x_t+u_t$，若给定解释变量x_t和被解释变量y_t的n对样本观测值$(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n)$，则有

$$\boldsymbol{X}'\boldsymbol{X}=\begin{pmatrix}1&1&\cdots&1\\ x_1&x_2&\cdots&x_n\end{pmatrix}\begin{pmatrix}1&x_1\\ 1&x_2\\ \vdots&\vdots\\ 1&x_n\end{pmatrix}=\begin{pmatrix}n&\sum x_t\\ \sum x_t&\sum x_t^2\end{pmatrix},$$

$$\boldsymbol{X}'\boldsymbol{Y}=\begin{pmatrix}1&1&\cdots&1\\ x_1&x_2&\cdots&x_n\end{pmatrix}\begin{pmatrix}y_1\\ y_2\\ \vdots\\ y_n\end{pmatrix}=\begin{pmatrix}\sum y_t\\ \sum x_ty_t\end{pmatrix}$$

从而由$\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{B}}=\boldsymbol{X}'\boldsymbol{Y}$可以得到$\boldsymbol{B}=(b_0,b_1)'$的普通最小二乘估计量$\hat{\boldsymbol{B}}=(\hat{b}_0,\hat{b}_1)'$的正规方程组为

$$\begin{pmatrix}n&\sum x_t\\ \sum x_t&\sum x_t^2\end{pmatrix}\begin{pmatrix}\hat{b}_0\\ \hat{b}_1\end{pmatrix}=\begin{pmatrix}\sum y_t\\ \sum x_ty_t\end{pmatrix}$$

或者

$$\begin{cases} n\hat{b}_0 + \hat{b}_1 \sum x_t = \sum y_t \\ \hat{b}_0 \sum x_t + \hat{b}_1 \sum x_t^2 = \sum x_t y_t \end{cases}$$

解此方程得到

$$\hat{b}_1 = \frac{\sum (x_t - \bar{x})(y_t - \bar{y})}{\sum (x_t - \bar{x})^2}, \quad \hat{b}_0 = \bar{y} - \hat{b}_1 \bar{x} \tag{1.1.16}$$

由此可以得出 y_t 对 x_t 的样本回归方程为$\hat{y}_t = \hat{b}_0 + \hat{b}_1 x_t$。

例 1.1.1　研究发现家庭书刊消费水平受家庭人均收入及户主受教育年数的影响。对某地区的家庭进行抽样调查，得到样本数据如表 1.1.1 所示，其中 y 表示家庭书刊消费水平(元/年)，x 表示家庭人均收入(元/月)，T 表示户主受教育年数。估计家庭书刊消费水平与家庭人均收入、户主受教育年数之间的二元线性回归模型

$$y_t = b_0 + b_1 x_t + b_2 T_t + u_t \quad (t = 1, 2, \cdots, n)$$

表 1.1.1　某地区家庭书刊消费水平及影响因素的调查数据表

家庭书刊消费 y	家庭人均收入 x	户主受教育年数 T
450.0	1 027.2	8
507.7	1 045.2	9
613.9	1 225.8	12
563.4	1 312.2	9
501.5	1 316.4	7
781.5	1 442.4	15
541.8	1 641.0	9
611.1	1 768.8	10
1 222.1	1 981.2	18
793.2	1 998.6	14
660.8	2 196.0	10
792.7	2 105.4	12
580.8	2 147.4	8
612.7	2 154.0	10
890.8	2 231.4	14
1 121.0	2 611.8	18
1 094.2	3 143.4	16
1 253.0	3 624.6	20

因变量观测值向量和解释变量观测值矩阵分别为

$$Y = \begin{pmatrix} 450 \\ 507.7 \\ \vdots \\ 1\,094.2 \\ 1\,253 \end{pmatrix}, \quad X = \begin{pmatrix} 1 & 1\,027.2 & 8 \\ 1 & 1\,045.2 & 9 \\ \vdots & \vdots & \vdots \\ 1 & 3\,143.4 & 16 \\ 1 & 3\,624.6 & 20 \end{pmatrix}$$

估计参数所需的有关矩阵分别为

$$\boldsymbol{X}'\boldsymbol{X}=\begin{pmatrix}18 & 34\,972.8 & 219\\ 34\,972.8 & 76\,252\,056 & 458\,076\\ 219 & 458\,076 & 2\,929\end{pmatrix}$$

$$(\boldsymbol{X}'\boldsymbol{X})^{-1}=\begin{pmatrix}0.661\,273 & -0.000\,1 & -0.033\,24\\ -0.000\,1 & 2.33E-07 & -2.87E-05\\ -0.033\,24 & -2.87E-05 & 0.007\,315\end{pmatrix},\quad \boldsymbol{X}'\boldsymbol{Y}=\begin{pmatrix}13\,592.2\\ 28\,832\,356\\ 182\,039.7\end{pmatrix}$$

从而参数估计向量(最小二乘估计量)为

$$\hat{\boldsymbol{B}}=(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}=$$

$$\begin{pmatrix}0.661\,273 & -0.000\,1 & -0.033\,24\\ -0.000\,1 & 2.33E-07 & -2.87E-05\\ -0.033\,24 & -2.87E-05 & 0.007\,315\end{pmatrix}\begin{pmatrix}13\,592.2\\ 28\,832\,356\\ 182\,039.7\end{pmatrix}=\begin{pmatrix}-50.016\,38\\ 0.086\,45\\ 52.370\,31\end{pmatrix}$$

由此样本回归方程为

$$\hat{y}=-50.016\,4+0.086\,5x+52.370\,3T$$

借助于计量经济软件 EViews 对表 1.1.1 进行回归分析,具体步骤为

(1) 建立工作文件。启动 EViews9.0,用鼠标单击 File,出现下拉菜单,单击 New/Workfile,打开 Workfile Create 窗口,如图 1.1.1 所示。

在 Workfile structure type 选项中,选择 Unstructured/Undated,在对话框 Observations 中输入 18(样本容量为 18),点击 OK,出现工作文件窗口,如图 1.1.2 所示。

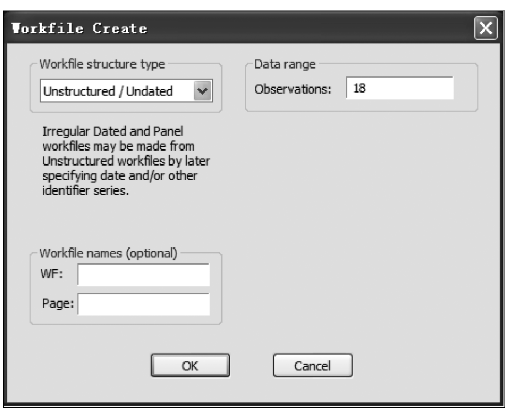


图 1.1.1　Workfile Create 窗口

图 1.1.2　Workfile 工作框

(2) 输入样本数据。建立工作文件以后,可以输入样本数据。直接在命令窗口输入命令

DATA　y　x　T

(3) 建立回归方程。输入样本数据后,在主页上选 Quick 菜单,点击 Estimate Equation 项,屏幕出现方程设定对话框,如图 1.1.3 所示。在 Equation Specification 中填入:y　c　x　T (c 为 EViews 固定的截距项),在 Estimation settings 中选 LS 估计,然后点击"确定",得到表 1.1.2 回归结果。

或者在命令窗口直接输入命令

LS　y　c　x　T

得到表 1.1.2 回归结果。

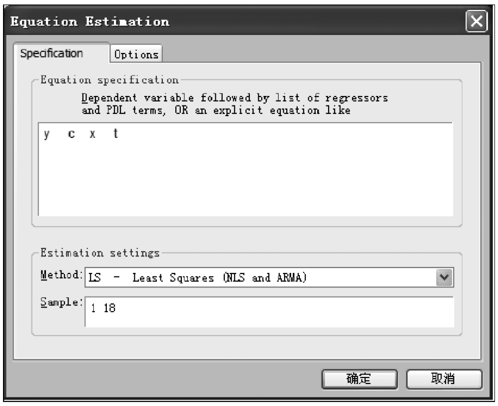


图 1.1.3　估计方程设定窗口

表 1.1.2　回归结果

Dependent Variable: Y
Method: Least Squares
Date: 09/15/17 Time: 14:57
Sample: 1 18
Included observations: 18

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-50.01638	49.46026	-1.011244	0.3279
X	0.086450	0.029363	2.944186	0.0101
T	52.37031	5.202167	10.06702	0.0000

R-squared	0.951235	Mean dependent var	755.1222
Adjusted R-squared	0.944732	S.D. dependent var	258.7206
S.E. of regression	60.82273	Akaike info criterion	11.20482
Sum squared resid	55491.07	Schwarz criterion	11.35321
Log likelihood	-97.84334	Hannan-Quinn criter.	11.22528
F-statistic	146.2974	Durbin-Watson stat	2.605783
Prob(F-statistic)	0.000000		

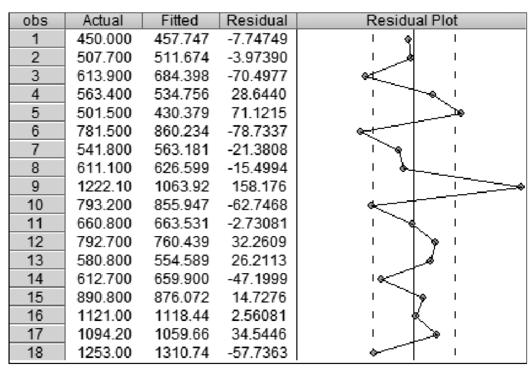
对应的回归方程为

$$\hat{y} = -50.0164 + 0.0865x + 52.3703T$$

$$t = (-1.0112) \quad (2.9442) \quad (10.0670)$$

$$R^2 = 0.9512, \quad \bar{R}^2 = 0.9447, \quad F = 146.2974$$

表 1.1.2 所显示的回归分析结果与前面的计算结果是一致的。在方程窗口点击 Resids 和 View/Actual, Fitted, Residual/Actual, Fitted, Residual Table 可显示模型拟合情况(见图 1.1.4)。



obs	Actual	Fitted	Residual	Residual Plot
1	450.000	457.747	-7.74749	
2	507.700	511.674	-3.97390	
3	613.900	684.398	-70.4977	
4	563.400	534.756	28.6440	
5	501.500	430.379	71.1215	
6	781.500	860.234	-78.7337	
7	541.800	563.181	-21.3808	
8	611.100	626.599	-15.4994	
9	1222.10	1063.92	158.176	
10	793.200	855.947	-62.7468	
11	660.800	663.531	-2.73081	
12	792.700	760.439	32.2609	
13	580.800	554.589	26.2113	
14	612.700	659.900	-47.1999	
15	890.800	876.072	14.7276	
16	1121.00	1118.44	2.56081	
17	1094.20	1059.66	34.5446	
18	1253.00	1310.74	-57.7363	

图 1.1.4　观测值、拟合值与残差

2. 最小二乘估计量的性质

多元线性回归模型用最小二乘法得到的参数估计量具有线性、无偏性、最小方差性。

(1) 线性。即参数估计量$\hat{\boldsymbol{B}}$既是因变量观测值 $\boldsymbol{Y}$ 的线性组合,也是随机误差项 $\boldsymbol{U}$ 的线性组合。

$$\begin{aligned}\hat{\boldsymbol{B}} &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'(\boldsymbol{XB}+\boldsymbol{U}) \\ &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{XB} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{U} = \boldsymbol{B} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{U}\end{aligned} \tag{1.1.17}$$

由此证明了参数估计量具有线性特性。它不仅是因变量观测值 $\boldsymbol{Y}$ 的线性组合,也是随机误差项 $\boldsymbol{U}$ 的线性组合。

(2) 无偏性。参数估计量$\hat{\boldsymbol{B}}$的均值等于总体参数,即

$$E(\hat{\boldsymbol{B}}) = E(\boldsymbol{B} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{U}) = E(\boldsymbol{B}) + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'E(\boldsymbol{U}) = \boldsymbol{B} \tag{1.1.18}$$

(3) 最小方差性。参数向量 $\boldsymbol{B}$ 的最小二乘估计$\hat{\boldsymbol{B}}$是 $\boldsymbol{B}$ 的所有线性无偏估计量中方差最小的估计量。

证明：设 $\boldsymbol{B}^*$ 是 $\boldsymbol{B}$ 的任意线性无偏估计，协方差矩阵之差为

$$E[(\boldsymbol{B}^*-\boldsymbol{B})(\boldsymbol{B}^*-\boldsymbol{B})']-E[(\hat{\boldsymbol{B}}-\boldsymbol{B})(\hat{\boldsymbol{B}}-\boldsymbol{B})'] \tag{1.1.19}$$

由于 $\boldsymbol{B}^*$ 是 $\boldsymbol{B}$ 的线性无偏估计，故记 $\boldsymbol{B}^*=\boldsymbol{AY}$，由无偏性可知

$$E(\boldsymbol{B}^*)=E(\boldsymbol{AY})=E[\boldsymbol{A}(\boldsymbol{XB}+\boldsymbol{U})]=\boldsymbol{AXB}=\boldsymbol{B}$$

从而有 $\boldsymbol{AX}=\boldsymbol{I}$。因此，有

$$\boldsymbol{B}^*-\boldsymbol{B}=\boldsymbol{AY}-\boldsymbol{B}=\boldsymbol{A}(\boldsymbol{XB}+\boldsymbol{U})-\boldsymbol{B}=\boldsymbol{AU}$$

$$E[(\boldsymbol{B}^*-\boldsymbol{B})(\boldsymbol{B}^*-\boldsymbol{B})']=E[(\boldsymbol{AU})(\boldsymbol{AU})']=E(\boldsymbol{AUU'A'})=\boldsymbol{A}(\sigma^2\boldsymbol{I})\boldsymbol{A}'=\sigma^2\boldsymbol{AA}'$$

对于最小二乘估计 $\hat{\boldsymbol{B}}$，由于

$$\hat{\boldsymbol{B}}=\boldsymbol{B}+(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{U}$$

因此

$$\hat{\boldsymbol{B}}-\boldsymbol{B}=(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{U}$$

由于

$$(\hat{\boldsymbol{B}}-\boldsymbol{B})(\hat{\boldsymbol{B}}-\boldsymbol{B})'=[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{U}][(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{U}]'=[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{U}][\boldsymbol{U}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}]$$

所以

$$E[(\hat{\boldsymbol{B}}-\boldsymbol{B})(\hat{\boldsymbol{B}}-\boldsymbol{B})']=(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\sigma^2=(\boldsymbol{X}'\boldsymbol{X})^{-1}\sigma^2$$

$$E[(\boldsymbol{B}^*-\boldsymbol{B})(\boldsymbol{B}^*-\boldsymbol{B})']-E[(\hat{\boldsymbol{B}}-\boldsymbol{B})(\hat{\boldsymbol{B}}-\boldsymbol{B})']=\sigma^2[\boldsymbol{AA}'-(\boldsymbol{X}'\boldsymbol{X})^{-1}]$$

由于

$$\begin{aligned}&[\boldsymbol{A}-(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'][\boldsymbol{A}-(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']'\\ &=\boldsymbol{AA}'-(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{A}'-\boldsymbol{AX}(\boldsymbol{X}'\boldsymbol{X})^{-1}+(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\\ &=\boldsymbol{AA}'-(\boldsymbol{X}'\boldsymbol{X})^{-1}\quad\text{（这一步利用 }\boldsymbol{AX}=\boldsymbol{I}\text{）}\end{aligned}$$

由线性代数可知，对于非奇异矩阵 $\boldsymbol{C}$，$\boldsymbol{CC}'$ 为半正定矩阵。[①] 将 $[\boldsymbol{A}-(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']$ 看成 $\boldsymbol{C}$，可知 $\boldsymbol{AA}'-(\boldsymbol{X}'\boldsymbol{X})^{-1}$ 为半正定矩阵。所以 $\hat{\boldsymbol{B}}$ 是 $\boldsymbol{B}$ 的最小方差线性无偏估计。

由此得到著名的**高斯-马尔可夫定理**：对于经典线性回归模型 $\boldsymbol{Y}=\boldsymbol{XB}+\boldsymbol{U}$，如果满足古典假定条件，则普通最小二乘估计量是最佳线性无偏估计量(LBUE)。

（4）参数的最小二乘估计量 $\hat{\boldsymbol{B}}$ 服从正态分布

$$\hat{\boldsymbol{B}}\sim N[\boldsymbol{B},\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}] \tag{1.1.20}$$

由于

$$\hat{\boldsymbol{B}}=\boldsymbol{B}+(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{U},\quad E(\hat{\boldsymbol{B}})=\boldsymbol{B},\quad \boldsymbol{U}\sim N(0,\sigma^2\boldsymbol{I}_n)$$

因此 $\hat{\boldsymbol{B}}$ 服从多元正态分布。

由性质 3 的推导过程可知

$$E[(\hat{\boldsymbol{B}}-\boldsymbol{B})(\hat{\boldsymbol{B}}-\boldsymbol{B})']=(\boldsymbol{X}'\boldsymbol{X})^{-1}\sigma^2$$

所以

---

① 设 $\boldsymbol{X}$ 为非零向量，$\boldsymbol{A}$ 为对称矩阵，若 $\boldsymbol{X}'\boldsymbol{AX}>0$，则称 $\boldsymbol{A}$ 为正定矩阵；若 $\boldsymbol{X}'\boldsymbol{AX}\geqslant 0$，则称 $\boldsymbol{A}$ 为半正定矩阵。$\boldsymbol{A}$ 为正定矩阵的充要条件是：存在非奇异矩阵 $\boldsymbol{C}$，使 $\boldsymbol{A}=\boldsymbol{C}'\boldsymbol{C}$。