经济计量模型
与经济预测

经济计量模型与经济预测

霍　俊　蔡福元　尤毓国

中国管理现代化研究会
北京预测咨询服务组　印

一九八二·十

编 者 的 话

我们从1978年起，就着手编译这套国外预测技术，到1980年底基本上完成了全部文稿工作，文字量达一百万字。从1979年开始就在北京技术经济和管理现代化研究会主办的《学术资料》上连续刊载。前五篇为常规预测部分，已经编写成《技术经济预测》专辑于1980年在《学术资料》上发表。从1980年7月份《学术资料》第58期起，开始连载《经济计量模型与预测》，已刊载了两个单元：《时间序列预测技术》和单方程模型之一；还有两个单元：单方程模型之二和多方程模型。由于文字量大，在《学术资料》上发表有待时日，很难适应在预测方面的急需；也由于继续发表，先刊出的已售缺，后刊出的难配套。因此，我们改用发行单行本。多方程模型是用于研究和建立宏观经济预测模型的，故用名为《宏观经济计量模型与预测》于今年四月份已发行。单方程模型实用意义大，又是研究和掌握多方程模型的基础工作与应用前提，为适应当前的急需，该书用名为《经济计量模型与经济预测》现在发行。该书中前三部分，是《学术资料》单方程模型之一的重印，在文字和内容上做了一些修改，特此说明。

这份材料主要根据美国《Econometric Model and Economic Forecasts》R.S.Pindyck，D.L.Rubinfold，一书为主，在翻译全文基础上剪裁编写的。编译的目的是借鉴，不是照抄。我们是根据国内的实际需要和用我们自己的学术观点进行选材和成文的。我们已经从1980年起，把主要精力放在国内广泛的预测实践上，于1981年发行了《预测实例专集》，从今年起又发行了《预测》杂志。我们正在努力写好结合我国国情的《实用预测学》，力争为开创社会主义新局面多尽点力量。

这份材料是我同蔡福元、尤毓国共同编译的。由于水平有限，切盼指教。

霍 俊

目　录

第一部分　单方程一元回归模型

在这部分中：介绍了模型的建立与检验；描述了一元回归模型的有关假设及统计性质；介绍了R^2概念以及一些基础知识。

一、回归模型概述

1.线拟合

测定变量的数据，可能有许多来源和各种形式。描述变量在某一段时间的活动数据，称为时间序列数据，时间单位可能是日、周、月、季、年。描述变量在某一个时间点上的活动数据，称为抽样数据点（样本点）。抽样，是从能代表真实关系的基本母体中选择出来的。

我们准备从研究双变量模型入手，来讨论经济计量模型问题。双变量，指一个自变量X与一个因变量Y；也称为一元，是指一个自变量X。为了用统计方式来描述这种关系，需要对每一个变量进行一组观察，还要假设X和Y之间具有明确的数学形式。在此，假设X和Y之间的关系为线性，即X和Y之间的关系，可以用直线来描述。线性拟合，就是指用直线来描述数据点的分布规律。

我们要求的是，这种线拟合为最佳。例如，图1.1的数据点分布，可以用L_1、L_2或别的直线来表示。怎样来评定某直线的拟合为最佳呢？

一种方法是使数据点对称分布，线上偏差为正，线下偏差为负，各点偏差之和为零，如图1.2。这种方法，可以用来说明直线对数据点的拟合程度。但是，不能说明数据点对直线的离散程度。

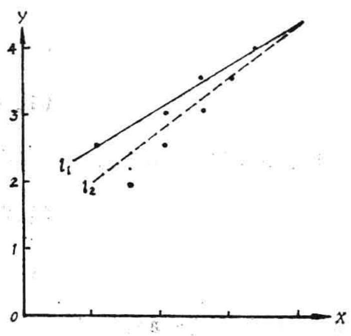

图1.1

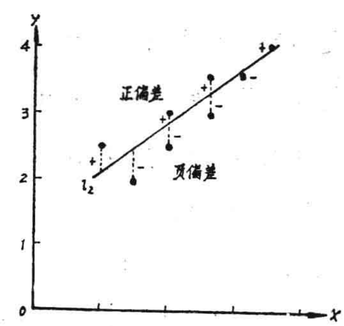

图1.2

另一种方法是采用绝对值，使直线与数据点之间的偏差绝对值之和为最小，但是在计算方面不方便。

最通用的方法是采用最小二乘法。最小二乘法，不仅能使数据点对称分布，还能使

偏差的平方和最小化，用以判断数据点对直线的离散程度。

2.最小二乘法

建立统计关系的目的，通常是用来说明或预测，一个或几个说明性变量或预测因素的变化，对某一因变量所造成的影响。根据图1.1的数据点分布，可以写出如下线性方程

$$Y = a + bX$$

式中：Y——因变量；X——自变量。因为将根据X，来说明或预测Y的演变，要求拟合线Y与样本数据点之间垂直测定的偏差平方和为最小（见图1.2），因此最小二乘法的准则是

$$\mathrm{Min}\sum_{i=1}^{N}(Y_i - \hat{Y}_i)^2 \tag{1.1}$$

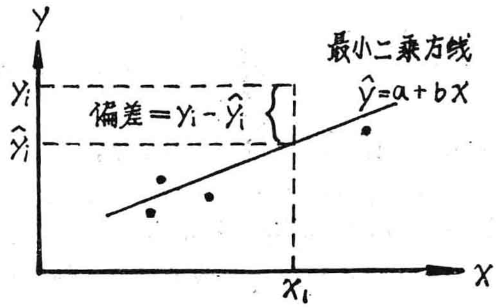

图1.3

式中：$\hat{Y}_i = a + bX_i$ 是对应于具体观察X_i的Y的拟合值，N是观察次数，见图1.3。a和b的数值未知，必须加以计算，才能满足准则。

求（1.1）式最小化的a和b的数值，利用普通微积分就可以做到，方法是将偏差平方和分别对a和b取偏导数，并令其等于零，然后解联立方程。

$$\frac{\partial}{\partial a}\sum(Y_i - a - bX_i)^2 = -2\sum(Y_i - a - bX_i) \tag{1.2}$$

$$\frac{\partial}{\partial b}\sum(Y_i - a - bX_i)^2 = -2\sum X_i(Y_i - a - bX_i) \tag{1.3}$$

使这些导数等于零，并除以－2，可得出

$$\sum(Y_i - a - bX_i) = 0 \tag{1.4}$$

$$\sum X_i(Y_i - a - bX_i) = 0 \tag{1.5}$$

最后，重写（1.4）和（1.5）式，则可得出所谓正规方程的联立方程组：

$$\sum Y_i = aN + b\sum X_i \tag{1.6}$$

$$\sum X_i Y_i = a\sum X_i + b\sum X_i^2 \tag{1.7}$$

式中：$\sum$——求和符号；N——观察次数。现在就可以求出a和b，其方法是将（1.6）式乘以$\sum X_i$，（1.7）式乘以N：

$$\sum X_i \sum Y_i = aN\sum X_i + b(\sum X_i)^2 \tag{1.8}$$

$$N\sum X_i Y_i = aN\sum X_i + bN\sum X_i^2 \tag{1.9}$$

由（1.9）式减去（1.8）式，得

$$N\sum X_iY_i - \sum X_i\sum Y_i = b\,[N\sum X_i^2 - (\sum X_i)^2] \qquad (1.10)$$

由此，得

$$b = \frac{N\sum X_iY_i - \sum X_i\sum Y_i}{N\sum X_i^2 - (\sum X_i)^2} \qquad (1.11)$$

已知b，则可由（1.6）式算出a：

$$a = \frac{\sum Y_i}{N} - b\frac{\sum X_i}{N} \qquad (1.12)$$

现在就可考虑在X和Y两者具有样本平均值为零时的特殊情况下，如何简化(1.11)和（1.12）公式，首先重写（1.12）式为

$$a = \overline{Y} - b\overline{X} = 0 \qquad (1.13)$$

式中，$\overline{Y}$和$\overline{X}$分别为Y和X的样本平均值。因而在X和Y的样本平均值为零时，拟合回归线的截距将为零。为了在这种特殊情况下求出相应的斜率估计，我们将（1.11）式的分子和分母除以N^2

$$b = \frac{\sum X_iY_i/N - (\sum X_i/N)(\sum Y_i/N)}{\sum X_i^2/N - (\sum X_i/N)^2}$$

代入$\overline{X}$和$\overline{Y}$，

$$b = \frac{\sum X_iY_i/N - \overline{X}\,\overline{Y}}{\sum X_i^2/N - \overline{X}^2}$$

根据定义，$\overline{X} = \overline{Y} = 0$。因此，

$$b = \frac{\sum X_iY_i/N}{\sum X_i^2/N} \qquad (1.14)$$

（1.14）式比（1.11）式简单。这个事实说明，如果用样本平均值的偏差来表示变量，写出最小二乘法估计将使问题简化，而不论其平均值是否为0。为此，我们将数据都变换成偏差形式，方法是按各自平均值的偏差来表示对X和Y的每一次观察。我们将用小写字母x_i、y_i等来表示这种偏差形式，后面就不另做说明了。其定义如下：

$$x_i = X_i - \overline{X}$$

$$y_i = Y_i - \overline{Y}$$

利用该定义，在一般情况下，可直接由（1.14）式得出斜率的最小二乘法估计，因为，

X和Y具有零平均值（$\overline{X}=\frac{\sum x_i}{N}=\frac{\sum(X_i-\overline{X})}{N}=\frac{\sum X_i}{N}-\frac{\sum\overline{X}}{N}=\overline{X}-\overline{X}=0$，同样可以证明$\overline{Y}=0$）。

因此，最小二乘法的斜率估计为

$$b=\frac{\sum x_i y_i}{\sum x_i^2} \tag{1.15}$$

变量转换成偏差形式见图1.4。在图1.4左图中，利用原始观察绘出回归线，而图1.4右图则采用偏差形式。首先应注意两条回归线的估计斜率是相同的。由(1.15)式可以很明显看出，是用偏差形式的变量进行计算的。故图1.4右图的回归线截距等于零。这是由(1.13)式以及$\overline{X}$和$\overline{Y}$均等于零这一事实得出的。因此，假如我们决定使用偏差形式的数据，我们就变换回归线的原点，但是不改变斜率。最后，请注意图1.4右图中的线穿过原点。这就相当于图1.4左图中的线穿过平均值（$\overline{X}$，$\overline{Y}$）。

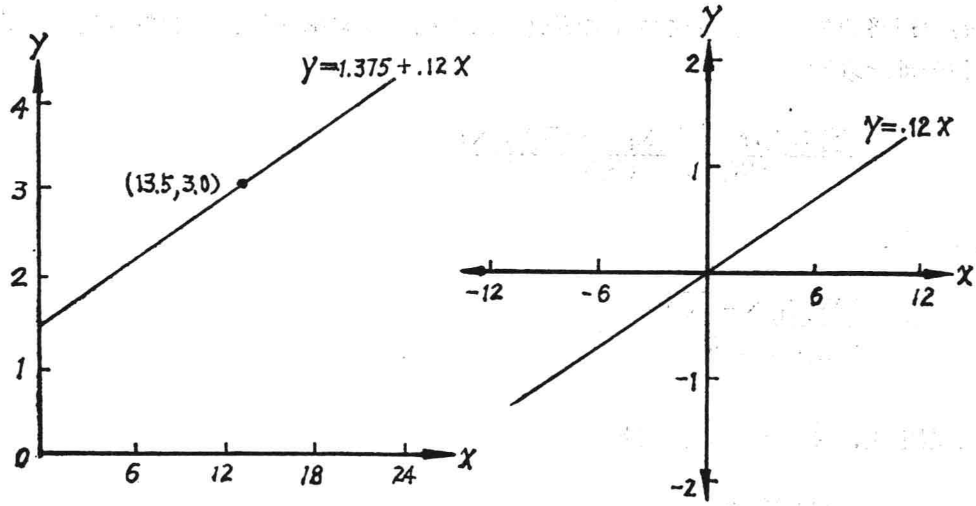

图1.4

在$Y=a+bX$模型中，斜率b是dY/dX的估计值，即Y的变化与X的变化之比。这使我们能够十分自然地解释回归斜率。另一方面，截距的解释则取决于那些接近$X=0$的充分的观察次数，是否能用以产生具有统计意义的结果。如果是这样的话，那末，在$X=0$时，我们就可以将截距解释为Y的一个估计值。但是，如果没有充分的观察次数，那末，截距就只是最小二乘法线的高度。

其次，假如我们估计了因果关系相反的回归线$X=A+BY$，则估计斜率将为

$$B=\frac{\sum x_i y_i}{\sum y_i^2}$$

如果估计过程都相同的话，那末，相应的回归斜率必须等于原回归斜率的倒数，即：$B = 1/b$。但是，这种同一性只有在罕见的情况下，即：$(\sum x_i y_i)^2 / (\sum x_i^2 \sum y_i^2) = 1$ 时才能成立。这种情况只有在全部样本点都位于拟合回归线上时才发生。

衣着费用支出实例

假定衣着费用支出与消费品总费用支出之间有关系。更准确地说，希望通过对全部消费品支出的了解，能够对衣着费用支出进行预测。在法国曾经组织过这类研究，并获得112个居住在省城的低级官员的家庭抽样数据。所测定的两个变量为衣着费用支出Y和全部消费支出X，拟合直线为$Y = a + bX$。因为，当时相信每个人都是将衣着费用支出作为既定总开支的一个直接函数。这个估计方程式为

$$\hat{Y} = 1.78 + 0.12X$$

根据正的斜率系数，即可得出这样结论：该数据符合衣着费用支出与总支出之间具有直接关系的假设。

二、模型的随机性质

我们的目的在于了解回归模型的概率性质。为了做到这点，我们分析如下事实：当我们给定某一观察值X（自变量）时，就会观察到许多可能的数值Y（因变量）。举例来说，每年收入相同的人，每人的消费方式就不同。又由于每人所处的环境有变化，每个人每年都可能有所改变。因此一般都假定：对应于观察值X（收入），Y值（如食品购买力）是不同的，是随机的。为了用公式来描述这种情况，我们给模型增加了一个随机误差分量，可写成下式：

$$Y_i = \alpha + \beta X_i + \varepsilon_i \qquad (1.16)$$

式中，每次观察时，Y是一随机变量；X是确定的（实验者知道）；ε是一随机误差项，其数值系根据概率分布。请注意，我们已改变符号，利用希腊字母来表示回归参数，因为，模型现在已含有一个随机误差项。

误差项是通过若干作用力的互相影响而出现的。误差项ε必须和残差$(\hat{\varepsilon}_i = Y_i - \hat{Y}_i)$，即和因变量与其拟合值的偏差加以区别。误差$\varepsilon$是与真实回归模型有关，而残差则是由估计过程产生的。第一，误差的出现是由于模型对现实的简化所造成的。例如，我们假定，价格是某一产品需求的唯一决定因素。事实上，有一些与需求有关的变量被省略了，如个人喜好、人口、收入以及气候，都可以归入误差项。假如这些被省略的影响很小，那末，假定误差项是随机的才算合理。出现误差的第二个原因，是与数据的收集和测定有关。经济数据和企业数据通常都是难以测定的。例如，某个商行可能不愿意放弃明确的成本数据，这样，他们就不会得到成本的无误差估计。每一个X值都有ε的概率分布，因此，就会有Y值的概率分布。可见图1.5双变量回归模型的概率分布。

现在我们就能够利用一些重要的假设来完全规定一元线性回归模型。

i，如（1.16）式所述，Y和X之间的关系为线性。

ii，X_i的数值为确定的非随机变量。

iii, (a) 对所有观察来说，误差项具有零期望值和常数方差，即$E(\varepsilon_i) = 0$和$E(\varepsilon_i^2) = \sigma^2$

(b) 就统计而言，随机变量ε_i都不相关，即：对应于不同观察的误差具有零相关。因此，$i \neq j$时，$E(\varepsilon_i \varepsilon_j) = 0$。

以上列举的假设就构成了经典的线性回归模型。

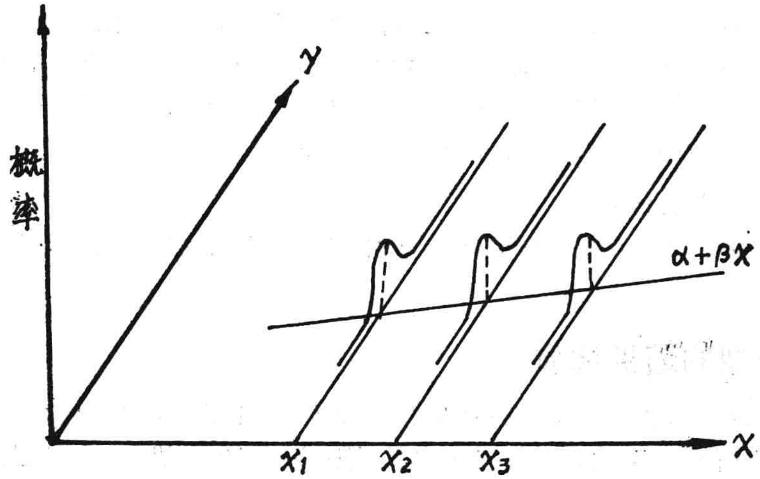

图1.5

关于把X作为确定的非随机变量的假设，相当于自变量X完全可以由研究人员所控制，能够根据实验目的，而改变X的数值。这对于研究大多数的企业问题来说，是不现实的。因此，这种假设仅仅是为了说明的目的。最小二乘法回归分析的绝大多数结果，对于比这个假设更一般的假设也是有效的。但是，它的证明要困难得多。

为了方便起见，误差项具有零期望值的假设，可以部分地做出。为了理解这点，假定省略的误差的平均影响等于α'〔即：$E(\varepsilon_i) = \alpha'$〕。然后，将双变量模型改写为

$$Y_i = \alpha + \beta X_i + \varepsilon_i + (\alpha' - \alpha')$$

$$= (\alpha + \alpha') + \beta X_i + (\varepsilon_i - \alpha')$$

$$= \alpha^* + \beta X_i + \varepsilon_i^*$$

其中，$\alpha^* = (\alpha + \alpha')$

$$\varepsilon_i^* = (\varepsilon_i - \alpha')$$

$$E(\varepsilon_i^*) = E(\varepsilon_i - \alpha') = E(\varepsilon_i) - \alpha' = \alpha' - \alpha' = 0$$

如果误差项有一常数方差（根据假设），即称为同方差性。但是，如果方差发生变化，则误差被称为异方差性。异方差性（与同方差性相反）可能会在研究某一工业企业的抽样时出现。图1.6是说明异方差性的两例。在图1.6右图中，误差项的方差随着X值的增加而减少，但在图1.6左图中，误差方差随X增大而增加。

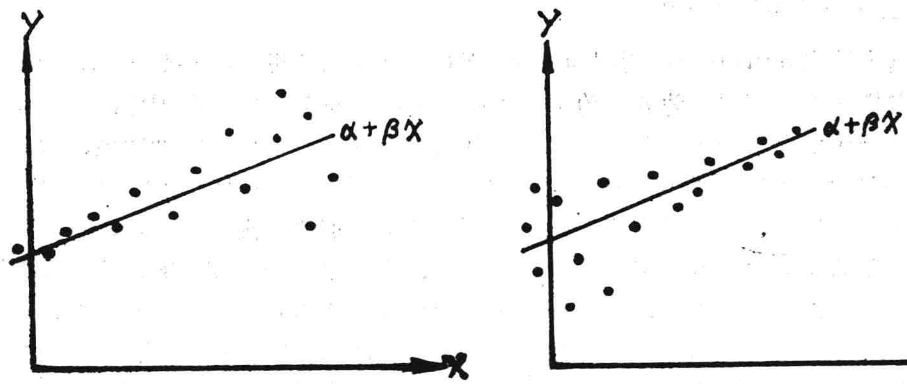

图1.6

对应于不同观察的误差都不相关的这一假设，在时间序列和抽样研究中是很重要的。当不同观察的误差相关时，我们说，误差过程是序列相关（自相关）。图1.7说明了时间序列研究中的正、负序列相关的情况。在这种情况中，负序列相关是指某一时间周期的负误差，伴随着下一时间周期的正误差，反之亦然。因此，在图1.7左图中，数据点在实际回归线上下呈某种规则性波动起伏。另一方面，当发生正序列相关时，某一周期的正(负)误差将趋于同下一周期的正(负)误差相联系。因此，在图1.7右图中，误差的图形是，先为负的（X的低数值），再为正的（X的高数值）。

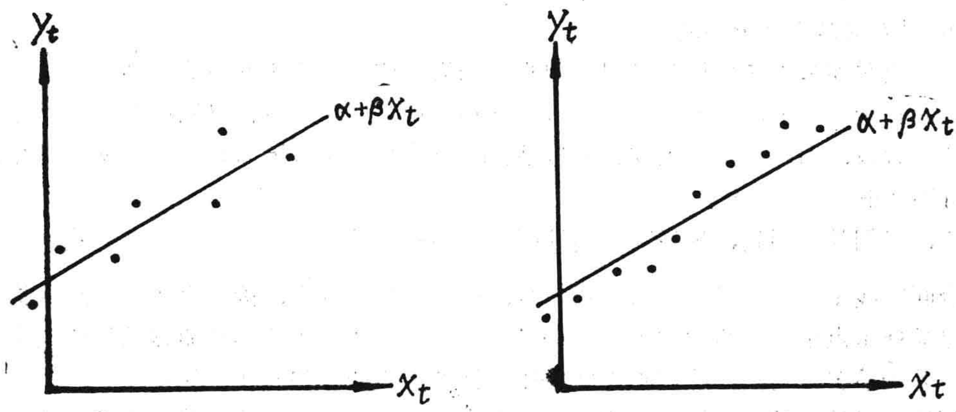

图1.7

应该注意，作为假设ii与iii (a) 的推论，我们可以毫无疑问地假定误差项是与X值不相关的。这是由于X值是非随机的这一假设推出的。因而

$$E(X_i \varepsilon_i) = X_i E(\varepsilon_i) = 0$$

如果将此分析推广到X值是随机的（也就是说，X值由概率分布任意得出的）模型时，也将需要这种假设。此外，假设iii (a) 和iii (b) 使我们得出的结论是，任何样本中的误差总和的期望值同样都为零，即

$$E(\sum\varepsilon_i) = \sum E(\varepsilon_i) = 0$$

不要把这个结果同iii (a) 的说明混为一谈。$E(\varepsilon_i) = 0$ 的意思是，与某一个X_i相联系的Y的重复取样误差的期望值为0。为了得到此结果，先将某X值加以固定，然后从已知概率分布的母体中求出随机误差的样本。这就是误差项每一个样本的期望值，假设其为0。ε样本是从具有期望值为0的概率分布中得出的，同样也适用于所有X值的ε样本，

在我们研究随机模型的估计程序之前，有两种情况特别值得提出。第一，我们曾经说过，每个扰动即误差项都具有常数方差σ^2。当然，方差是一个未知参数，必须作为回归模型的一部分加以估计。因此，此处所叙述的随机回归模型具有三个未知参数，而在前面的线拟合模型中只具有两个未知参数。第二，我们曾经叙述过与误差扰动ε有关的模型假设，我们能够容易地写出与变量Y的概率分布有关的假设。在这种情况下，假设将如以下所述：

iii (a′) 随机变量Y的期望值为$\alpha + \beta X$和方差为σ^2：

$$E(Y_i) = E(\alpha + \beta X_i + \varepsilon i) = \alpha + \beta X_i + E(\varepsilon) = \alpha + \beta X_i$$

(b′) 随机变量Y_i是不相关的。

三、估计量的统计性质

为了对线性模型进行统计检验，我们需要确定误差项的概率分布。对于经典的正规线性回归模型，我们假定：

iii (c) 误差项是正态分布的。

这一假设对模型的统计检验是很重要的。如果相信由于测定和省略所造成的各个误差很小，而且相互无关，那末，正态性假设就是一种合理的假设。已知误差项ε是正态分布的这一假设，由此得出：因变量Y也是正态分布的结论（因为Y_i是ε_i的线性组合，而每个ε_i都是正态分布的）。

现在，我们就能回到，选择随机线性回归模型的估计参数这一问题上来。我们的目标是拟合回归线$\hat{Y} = \hat{\alpha} + \hat{\beta} X$。这条回归线，在某种意义上讲，接近真实回归线。我们曾叙述过从样本点至回归线的偏差平方和最小化。现在我们希望能检验该估计线与真实回归线的差别情况。为此，我们首先要描述估计参数集的一些有用的统计性质。一开始进行介绍性的讨论，然后再详细研究最小二乘法估计量的统计性质。我们主要注意参数β估计量的性质，但是，这个讨论可以应用于所有的估计参数。以后的材料也都是直接有关的，因为，所估计的回归参数是随机变量，其期望值和方差都是可以计算的。

为了更清楚地了解这一点，请注意，我们分析的模型，其中自变量的N个值均假定为确定的。假若我们选择与自变量X值有关的Y为观察样本，那末，我们就能获得回归斜率的一个估计量，它是以已经出现的Y的具体观察值为基础的。假若我们以相同的X值重复该实验，我们就会得出Y的一组新的观察值集（因为ε在新样本中是不同的），并且这样又会得出斜率参数的新估计值。如果我们能得到Y变量的充分样本，那末，我们就会得出参数β估计量的分布。由于每个Y的观察样本都产生一个β的具体估计量。当

我们取Y值的样本时，就相当于从样本估计量分布中取β的一个估计值。

当然，这将会出现一个样本估计量的分布，它与从样本数据估计β的各种可能的方法有关。但是，我们眼下的目的是研究最小二乘法估计程序所产生的分布（请注意，在前面$\hat{\beta} = \sum x_i y_i / \sum x_i^2$），以便评定最小二乘法是否是估计$\beta$的良好程序。重要的是，要认识到符号$\hat{\beta}$通常用于两种目的：$\hat{\beta}$一种是指由某一特定样本得出的斜率估计量；另一种是按照概率分布的估计量（与最小二乘法估计程序有关）。

1.无偏

在估计回归参数时，把参数作为平均值，是非常需要的一种性质，这样便于研究估计量的分布。因此，只要我们能够对Y重复实验许多次，就能得到观察值的新的集合，我们就可以肯定，平均来说是正确的。如果$\hat{\beta}$的平均值或期望值等于真实值时，即$E(\hat{\beta}) = \beta$，就说$\hat{\beta}$是一个无偏估计量。有偏与无偏估计量的区别可见图1.8，请注意，大小为N的样本，$\sum X_i / N$为母体真实平均值的一个无偏估计量，而$\sum (X_i - \overline{X})^2 / (N-1)$则是母体真实方差的一个无偏估计量。为了说明起见，我们将估计参数的偏移定义如下

$$偏移 \text{Bias} = E(\hat{\beta}) - \beta$$

重要的是，应该认识到，一个估计量无偏是一种很需要的性质。因为，无偏意味着，估计量对真实参数的离中趋势没有意义。一般来说，人们希望估计量是无偏的。而且还希望估计量平均值周围的离中趋势很小。因此，我们就需要确立这样一条准则，使我们能够在无偏估计量中间进行选择。

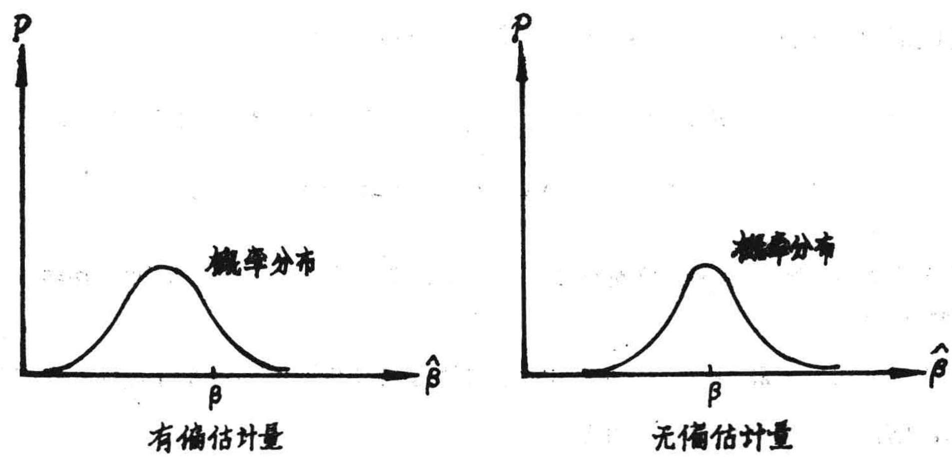

图1.8

2.效率

假若$\hat{\beta}$的方差小于任何其他无偏估计量的方差，我们就可以认为$\hat{\beta}$是一个有效的无偏估计量。实际上，有时难以说明估计量是否有效，因此，按照它们有关的效率来叙述

估计量是很自然的。如果某一估计量的方差较小，那末，它就比其他估计量会更有效。比较有效和比较无效的估计量可见图1.9(图中制版时漏掉了“比较”两字)。效率是一个理想的目标，因为，估计过程越有效，人们对估计参数所作的统计说明就越有力。因此，在方差为零的(无偏)估计量这种极端情况下，我们可以肯定地指出真实回归参数的数值。

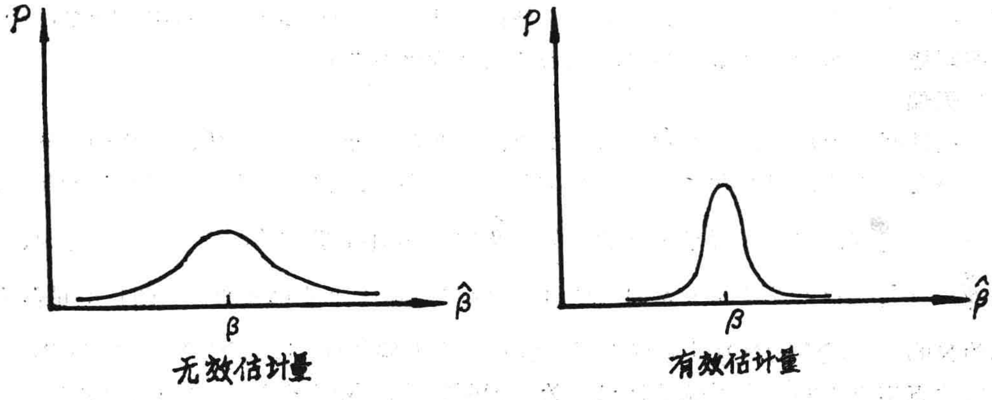

图1.9

3.最小平均方差

有许多场合，人们不得不权衡估计量的方差和偏移。例如，当模型的目的是使预测的精度最大化时，那末，具有极低的方差和某些偏移的估计量，可能比具有较高方差的无偏估计量要更理想。在这方面，有一条有用的准则，目标是使平均方差最小化，其定义如下：

平均方差 $(\hat{\beta}) = E(\hat{\beta}-\beta)^2$

可以证明平均方差 $(\hat{\beta}) = [\text{Bias}(\hat{\beta})]^2 + \text{Var}(\hat{\beta})$

$$E(\hat{\beta}-\beta)^2 = E[(\hat{\beta}-\overline{\hat{\beta}}) + (\overline{\hat{\beta}}-\beta)]^2 = E(\hat{\beta}-\overline{\hat{\beta}})^2 + [E(\overline{\hat{\beta}}-\beta)]^2$$

$$+2(\overline{\hat{\beta}}-\beta)E(\hat{\beta}-\overline{\hat{\beta}}) = \text{Var}(\hat{\beta}) + [\text{Bias}(\hat{\beta})]^2$$

其中$\overline{\hat{\beta}}$是估计回归系数β的期望值。　根据定义$\overline{\hat{\beta}} = E(\hat{\beta})$。式中：Bias——偏移；Var——方差。

因此，平均方差最小化的准则，考虑了估计量偏移的平方及估计量的方差。如$\hat{\beta}$是无偏的，则$\hat{\beta}$的方差与平均方差相等。

4.一致性

我们认为，当样本规模变大时，估计量就具有大样本的性质。在这方面，我们希望估计量$\hat{\beta}$在某种意义上能在样本增大时接近真实β，这就是说，希望$\hat{\beta}$分布能收敛于β。更准确地说，我们希望，在样本无限变大时，$\hat{\beta}$不同于β的概率将非常小。为了将这种

概率概念应用于估计量选择，我们将$\hat{\beta}$的概率极限（Plim$\hat{\beta}$）定义如下：

假使N趋近无限大时，Plim$\hat{\beta}$等于β，即$|\beta - \hat{\beta}|$小于任意小正数的概率将接近于1。根据这个概念，很自然就能将一致性的准则规定如下：

如$\hat{\beta}$的概率极限为β，则$\hat{\beta}$为β的一致估计量。〔严格说，对任意$\delta > 0$，在概率极限$\hat{\beta}$收敛于β时 $\lim\limits_{N\to\infty} P\ (|\beta - \hat{\beta}| < \delta)\ = 1$〕

概言之，假若当样本任意变大时，估计量的概率分布压缩到一个点上(真实参数)，那末这个估计量就是一致的。可见图1.10。

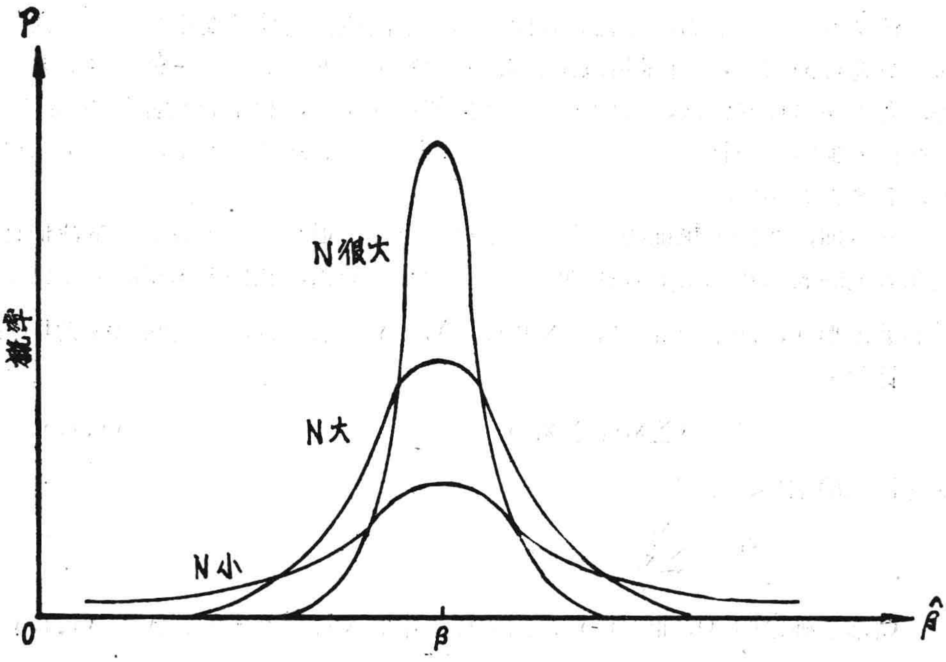

图1.10

一般说来，经济计量学家倾向于更关心将一致性而不是将无偏作为估计准则。一个有偏、然而一致的估计量，平均来说，可能不等于真实参数，但是，在样本数据变大时，这个估计量将近似真实参数。例如，一定规模为N的样本，$\sum(X_i - \overline{X})^2/N$ 就是母体方差的一个有偏而一致的估计量。从实际的观点来看，这要比得出一个无偏的、然而在样本变大时却继续大大地偏离真实参数的参数估计量更使人放心。

将估计量的平均方差，在样本增大时接近零这个目标当作一致性的选择准则，这是很自然的。平均方差准则的意思是，估计量是渐近无偏的，并且其方差在样本变得很大时接近于零。结果，平均方差接近零的估计量将是一个一致估计量。但是，逆推就不一定正确了。在多数用途中，一致估计量确实都具有接近零的平均方差，而且这两个准则是互相通用的。

四、最佳线性无偏估计

现在我们要回到最小二乘法估计量的问题。由于最小二乘法估计量与随机回归模型 $Y_i=\alpha+\beta X_i+\varepsilon_i$ 有关，我们将这种估计量用 $\hat{\alpha}$ 和 $\hat{\beta}$ 表示。采用最小二乘法的最重要理由是由经典的高斯—马尔可夫定理提出的，这个定理说：$\hat{\alpha}$ 和 $\hat{\beta}$ 都是 α 和 β 的最佳（最有效）的线性无偏估计量。因为，它们与所有无偏估计量相比，具有最小的方差。重要的是，要认识到高斯—马尔可夫定理不适用于非线性估计量。〔**如果 $\hat{\beta}$ 能写成 $\sum C_iY_i$，其中每个 C_i 都是一个常数，这样就能说 $\hat{\beta}$ 与 Y 是线性关系。但是，如果 $\beta'=\sum C_i\log Y_i$，那么 β' 就是 Y 的非线性估计量，非线性估计可能出现于最小二乘法以外的其他准则。**〕非线性估计量可以是无偏的，并且具有比最小二乘法线性估计量要小的方差。这一点是可能的。更为普遍的是，估计量可以是有偏的，而其平均方差却比无偏估计量的要小。这就说明，会出现这样的情况，人们希望在选择估计程序时，利用的是客观的而不是“最佳、线性无偏的”估计程序。有偏的平均方差最小的非线性估计量，目前正在被认真地研究，看来很有用处。

在这方面，我们不想证明高斯—马尔可夫定理，但是，求出最小二乘法估计量的平均值和方差的表示式还是很有帮助的。为了把问题简化，我们将采用的数据已变换成平均值的偏差形式，即采用 $x_i=X_i-\overline{X}$ 和 $y_i=Y_i-\overline{Y}$。请注意，真实回归线为 $E(y_i)=\beta x_i$ 以及估计斜率为

$$\hat{\beta}=(\sum x_iy_i/\sum x_i^2) \tag{1.17}$$

为简化以后的推导，令

$$c_i=\frac{x_i}{\sum x_i^2}$$

每一个 C_i 都是确定的（即非随机），因为样本中的 X 都是确定的。代入（1.17）式，可得出

$$\hat{\beta}=\sum c_iy_i$$

该式将估计斜率参数表示为因变量观察值的加权和。我们将用该表示式清楚地导出 $\hat{\beta}$ 的平均值和方差。首先，

$$\hat{\beta}=\sum c_iy_i=\sum c_i(\beta x_i+\varepsilon_i)=\sum c_i\beta x_i+\sum c_i\varepsilon_i \tag{1.18}$$

因而，$E(\hat{\beta})=\sum c_i\beta x_i+\sum c_iE(\varepsilon_i)$

由于，$E(\varepsilon_i)=0$

所以，$E(\hat{\beta})=\sum c_i\beta x_i=\beta\sum c_ix_i=\beta$

$\sum c_ix_i=1$ 这个事实是直接由 C_i 的定义而推出的。如：