

SERIES OF KERNEL TEXTBOOK  
FOR ECONOMICS  
BASIC ECONOMETRICS

ECONOMICS

经济学核心课系列教材

# 计量经济学基础

张晓峒 主编



SERIES OF KERNEL TEXTBOOK  
FOR ECONOMICS  
BASIC ECONOMETRICS

南开大学出版社

# 计量经济学基础

张晓峒 主编

南开大学出版社

天津

## 图书在版编目(CIP)数据

计量经济学基础/张晓峒主编. - 天津:南开大学出版社,2001.1

ISBN 7-310-01452-9

I.计… II.张… III.计量经济学-高等学校-教材 IV.F224.0

中国版本图书馆 CIP 数据核字(2000)第 54670 号

**出版发行** 南开大学出版社

地址:天津市南开区卫津路 94 号

邮编:300071 电话:(022)23508542

**出版人** 肖占鹏

**承印** 天津宝坻第四印刷厂印刷

**经销** 全国各地新华书店

**版次** 2001 年 1 月第 1 版

**印次** 2001 年 1 月第 1 次印刷

**开本** 850mm × 1168mm 1/32

**印张** 12.5

**字数** 310 千字

**印数** 1 - 3000

**定价** 18.00 元

# 第1章 绪论

## § 1.1 计量经济学的定义

**计量经济学**(Econometrics)是由挪威经济学家、第一届诺贝尔经济学奖得主弗里希(R. Frisch)1926年仿照生物计量学(Biometrics)提出来的。1930年,一些国家的经济学家在美国成立了国际计量经济学会,学会的宗旨是“为了促进经济理论在与统计学和数学的结合中发展的国际学会”。1933年该学会创办了会刊——《计量经济学》杂志。弗里希在发刊词中有一段话:“用数学方法探讨经济学可以从好几个方面着手,但任何一方面都不能与计量经济学混为一谈。计量经济学与经济统计学决非一码事;它也不同于我们所说的一般经济理论,尽管经济理论大部分都具有一定的数量特征;计量经济学也不应视为数学应用于经济学的同义语。经验表明,统计学、经济理论和数学这三者对于真正了解现代经济生活中的数量关系来说,都是必要的。三者结合起来,就有力量,这种结合便构成了计量经济学。”

从这段话不难看出,计量经济学是统计学、经济学、数学相结合的一门综合性学科,是一门从数量上研究物质资料生产、交换、

分配、消费等经济关系和经济活动规律及其应用的科学。

计量经济学自从 20 世纪 30 年代形成以来,发展很快,已在经济学科中占有很重要的地位,在经济领域得到了广泛的应用。尤其是近 20 年来计算机的飞速发展,使计量经济学的发展和应用又进入了一个新的阶段。人们不仅在微观经济领域,而且在宏观经济领域,建立了大量计量经济模型并用于预测,解决了实际经济问题,为各国的经济发展作出了卓越的贡献。

计量经济学在我国出现的时间并不长,但是发展很快。计量经济模型和预测广泛应用于各经济领域,为制定经济政策提供了科学的依据。

## § 1.2 计量经济学的特点

计量经济学用数学模型表示经济变量之间的关系。由于实际的经济运行不是在实验室进行的,往往存在一些不确定的随机因素,使得经济变量之间的关系不能表示成精确的函数关系。人们只能在模型中列出对所研究变量起主要影响作用的变量,将不重要的因素和一些不确定因素归并到一个随机变量中,建立变量之间的数学模型。

例如,利用计量经济学研究需求函数。经济理论假定某商品的需求量取决于它的价格与代用品价格、消费者的收入和消费者的偏好。然而在实际经济生活中,除了这些因素外,还有其他一些不重要因素以及随机因素的影响。例如新产品的发明、职业的改变、气候条件的变化等。另外,人们可能受到谣传、广告的影响。即使市场价格、消费者收入和消费者偏好都不变,商品的需求量也会受到影响。在计量经济学中,这些不重要因素及随机因素的影

响也要反映在数学模型中,于是引进一个随机变量  $u$ ,建立下列形式的某商品需求量数学模型:

$$Q_i = \beta_0 + \beta_1 P_i + \beta_2 P_{0i} + \beta_3 Y_i + \beta_4 T_i + u_i$$

其中:

$Q$ ——某商品需求量;

$P$ ——该商品价格;

$P_0$ ——代用品价格;

$Y$ ——消费者收入;

$T$ ——消费者偏好;

$u$ ——影响商品需求量的其他因素和随机因素;

(以上变量的统计资料除  $u$  之外都是已知的)

$\beta_0 \sim \beta_4$ ——需求函数的回归系数(待定系数,待定参数)。

以上所指的统计资料(也称统计数据、样本观测值、样本数据、样本值)一般有以下几种:

1. 时间序列统计资料。时间序列统计资料指同一统计指标按时间顺序排列的数据列。在同一数据列中各个数据统计的对象、范围和时间长度必须一致,是同一口径的,具有可比性。常用的有以年、季度、月为时间间隔的统计数据。例如 1980 年—1999 年每年全国国内生产总值;1991 年第 3 季度—1999 年第 4 季度每季度某地区零售物价指数;1995 年 4 月—1999 年 12 月某企业每月销售额等。

2. 横截面统计资料。横截面统计资料指在同一时间、不同单位按同一统计指标排列的数据列。在同一数据列中各个数据也必须是同一口径的,具有可比性。与时间序列数据的区别在于,横截面数据统计的对象和范围不同,但必须是同一时间截面上的数据。例如 1998 年全国 31 个省、区、直辖市工业增加值;1999 年某市 33 个工业行业上交利税等。

3. 时间序列和横截面数据合并的统计资料。有时为了扩大样本容量,以提高模型的估计精度,人们常常采用时间序列和横截面数据合并的统计资料。即将若干期的时间序列和每期内的横截面数据合并作为样本数据。时间序列和横截面数据合并的统计资料要求同时具有以上两种统计资料的特征。例如 1995 年—1999 年 14 个沿海城市外商实际投资,其样本数据的个数为  $5 \times 14 = 70$  个。每年均统计该 14 个沿海城市,统计口径相同,具有可比性。

由以上的介绍可知,计量经济学中提到的变量,是一个序列,其中包括若干个已知的样本数据。计量经济学就是利用这些变量的样本数据和计量经济学方法,估计数学模型中的待定系数,以确定经济变量之间的结构关系。因此,它与数学中的方程式以及方程的解法有着本质的区别。

计量经济模型可以分为两大类:单一方程模型和联立方程模型。单一方程模型用来描述微观或宏观经济领域一个因变量和若干个自变量间的结构关系;联立方程模型用来描述微观或宏观经济领域多个因变量和多个自变量相互间的结构关系。

### § 1.3 计量经济学的目的

计量经济学的目的主要有三个:

1. 结构分析。指应用计量经济模型对经济变量之间的关系作出定量的度量。如上面提到的需求函数,假如利用统计资料估计出模型中变量  $P$  的回归系数等于  $-0.12$ ,表明当其他因素不变时,商品的价格每提高一个单位,此商品的需求量将减少 0.12 个单位。

2. 预测未来。指应用已建立的计量经济模型求因变量未来

一段时期的预测值。如上面的需求函数,假如给出某商品的价格、代用品价格、消费者收入及消费者偏好 2002 年的估计值,就可以求出 2002 年某商品需求量预测值,为 2002 年该商品的生产和供给提供可靠的依据。

3. 政策评价。指通过计量经济模型仿真各种政策的执行效果,对不同的政策进行比较和选择。如给出某商品的各种不同价格,分别预测未来各种不同的需求量,由此确定比较合适的商品价格。

## § 1.4 计量经济学的内容及研究问题的方法

### 1. 计量经济学的内容

计量经济学在长期的发展过程中逐步形成了两个分支:理论计量经济学和应用计量经济学。

**理论计量经济学**主要研究计量经济学的理论和方法。计量经济方法又分为单方程估计方法和联立方程系统估计方法。单方程估计方法每次仅作用于一个方程;系统估计方法要考虑联立方程系统的综合信息,同时估计联立方程中的全部方程。

**应用计量经济学**将计量经济学方法应用于经济理论的特殊分支,即应用理论计量经济学的方法分析经济现象和预测经济变量。

理论计量经济学主要研究一般线性模型、非线性模型、联立方程模型的估计方法,回归系数和相应统计量的分布特征。而应用计量经济学则使用这些方法解决经济理论中诸如需求、供给、生产、投资、消费等实际经济问题。

本书既介绍理论计量经济学的基本知识,也用一定篇幅介绍



应用计量经济学。

## 2. 计量经济学研究问题的方法

用计量经济学研究问题可分为四个阶段：

第一阶段，建立模型。根据所研究的问题与经济理论，找出经济变量间的因果关系及相互间的联系。把要研究的经济变量作为因变量，影响因变量的主要因素作为自变量，影响因变量的非主要因素及随机因素归并到随机项，建立计量经济数学模型。

第二阶段，估计参数。模型建立以后，首先收集模型中经济变量的统计资料，再应用相应的计量经济方法，估计模型中的待定系数。

收集统计资料的方法很多。常用的方法有：(1)利用正式出版发行的统计资料，如《中国统计年鉴》，各省、自治区、直辖市的《统计年鉴》以及其他统计资料。(2)到各有关部门调查获得。(3)到基层调查获得。再将调查得来的数据进行分类加工整理。

第三阶段，检验模型。模型的参数估计以后，这些参数是否可靠，是否符合经济理论和要求，要通过以下几个方面对模型进行检验。

(1)检验估计参数是否符合经济理论和实际经济问题的要求。如某商品的需求量一般应随此商品价格的提高而减少，如果估计出的价格参数是正的，则说明建立的模型不合适或使用的估计方法不合适，或采用的样本数据有问题。另外，估计参数的过大或过小，都有可能不符合经济理论和要求。

(2)用数理统计中关于假设检验的原理，对估计参数进行统计检验，对估计模型进行统计检验，对估计方法的假定条件进行检验。

如果以上的检验出现问题，应采取相应的办法予以补救。如改变模型的形式，变换估计方法，重新选取样本数据，修正样本数

据等。

第四阶段,经济预测。应用估计出的并经过检验的回归模型预测因变量的未来值。并不是经过检验的模型都有好的预测结果,对预测结果仍需进行观察和检验。

由于电子计算机的迅速发展,人们将计量经济学的计算方法编制成软件包,不仅可以估计参数,进行统计检验,而且可以进行经济预测,使许多非常复杂的计算问题得以很容易地解决,为计量经济学的学习和应用带来极大的方便。本书中,我们将介绍计量经济分析软件包 TSP。TSP 直观、易学,是目前人们最常用的计量经济学软件包之一,本书中的计量经济模型均可用 TSP 解决。

## 第 2 章 一元线性回归模型

### § 2.1 模型的建立及其假定条件

#### 1. 回归分析的概念

回归分析是处理变量与变量之间关系的一种数学方法。

社会经济活动总是和许多经济变量相联系的,我们常常要研究这些变量之间的数量关系。对于经济变量之间的关系,一般分为两类:一类是变量之间存在确定的函数关系。例如某企业的销售收入  $Y_i$  等于产品价格  $P$  与销售量  $X_i$  的乘积,用数学表达式表示为:

$$Y_i = PX_i$$

另一类是变量之间存在着非确定的依赖关系。例如某企业资金的投入  $X_i$  与产出  $Y_i$ ,一般来讲,资金投入越多,产出也相应提高。但是由于生产过程中各种条件的变化,使得不同时间内同样的资金投入会有不同的产出。这些造成了资金的投入与产出之间关系的不确定性,因而不能给出类似于函数的精确表达式。用  $u_i$

表示其他影响因素,将这两个变量之间非确定的依赖关系表示成下列形式:

$$Y_i = f(X_i) + u_i$$

为了分析和利用变量之间非确定的依赖关系,人们建立了各种统计分析方法,其中回归分析方法是最常用的经典方法之一。回归分析的理论和方法是计量经济模型估计理论和估计方法的主要内容。

## 2. 一元线性回归模型

为了说明一元线性回归模型,举一个某商品需求函数的例子。为了研究某市城镇每年鲜蛋的需求量,首先考察消费者年人均可支配收入对年人均鲜蛋需求量的影响。由经济理论知,当年人均可支配收入提高时,鲜蛋需求量也相应增加。但是,鲜蛋需求量除受消费者收入影响外,还要受到其自身价格、人们的消费习惯及其他一些随机因素的影响。为了表示鲜蛋与消费者可支配收入之间非确定的依赖关系,我们将影响鲜蛋需求量的其他因素归并到随机变量  $u_i$  中,建立这两个变量之间的数学模型:

$$Y_i = \beta_0 + \beta_1 X_i + u_i \quad (2.1)$$

其中:

$Y_i$ ——某市城镇居民年人均鲜蛋需求量,称作**被解释变量**;

$X_i$ ——某市城镇居民年人均可支配收入,称作**解释变量**;

$u_i$ ——**随机误差项**(随机扰动项或随机项、误差项);

$\beta_0$ 、 $\beta_1$ ——**回归系数**(待定系数或待定参数)。

在数学模型(2.1)式中,当  $X_i$  发生变化时,按照一定规律影响另一变量  $Y_i$ ,而  $Y_i$  的变化并不影响  $X_i$ 。亦即  $X_i$  的变化是  $Y_i$  变化

的原因,  $X_i$  与  $Y_i$  之间具有因果关系。(2.1)式称为回归模型,因为只有一个解释变量,变量间的关系又是线性关系,故(2.1)式称为一元线性回归模型。

随机变量(随机误差项) $u$  中一般包括以下几个方面的因素:

(1)回归模型中省略的变量。从前面所讲的几个例子可以看出,影响被解释变量的因素很多,建立模型时,一般只研究对被解释变量影响重要的因素和我们所关心的因素,将其他非重要影响因素归并到  $u$  中。

(2)人们的随机行为。经济行为不是在实验室中发生的,有些因素是无法控制的。例如一个消费者受广告宣传或谣传的影响,就有可能改变对某种商品的需求,这些都属于人们的随机行为。

(3)建立的数学模型的形式不够完善。为了研究问题方便,我们往往把非线性关系线性化,或者略去了模型中的某些方程。在大多数情况下,变量是由包含多个方程的联立方程同时决定的,但我们经常用单一方程来研究经济现象,这些都必然产生误差。

(4)经济变量之间的合并误差。我们经常使用经济变量的总量数据(如总产值、总消费、总收入等)。就是把一些具有不同性能的个体数量加在一起,在加总过程中由于涉及面广,不可避免地出现错误。

(5)测量误差。对于经济变量的原始统计资料,在收集和测量过程中,由于统计方法和资料处理方法的不同,都不可避免地出现误差。

为了把上述产生的误差考虑在内,在计量经济模型中引进了随机变量  $u$ , 是认为它对假定存在于  $X$  和  $Y$  之间的精确线性关系进行扰动。这时将(2.1)式分成两部分:

$$Y_i = (\beta_0 + \beta_1 X_i) + u_i \quad (2.2)$$

一部分由直线  $\beta_0 + \beta_1 X_i$  组成,另一部分是随机误差项  $u_i$ 。

假设我们已获得  $X$ 、 $Y$  的  $n$  个样本观测值  $(X_i, Y_i)$ ,  $i = 1, 2, \dots, n$ 。(2.2)式中的意义可以由图 2.1 解释。

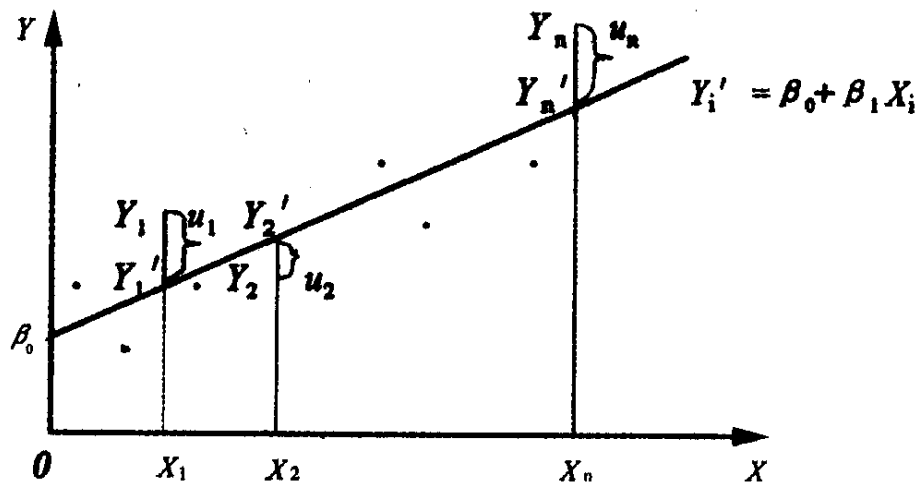


图 2.1

直线  $Y_i' = \beta_0 + \beta_1 X_i$  表示  $X$  与  $Y$  之间的线性部分,称作总体回归直线。样本值与回归直线的偏离  $u_i$  表示对这种线性关系的随机扰动。如果没有误差,样本值应是对应于  $X_1, X_2, \dots, X_n$  的直线上的点  $Y_1', Y_2', \dots, Y_n'$ 。由于随机扰动的原因,样本值是偏离直线的点  $Y_1, Y_2, \dots, Y_n$ 。这些样本点偏离回归线的数量分别为  $u_1, u_2, \dots, u_n$ 。即

$$u_i = Y_i - Y_i' (i = 1, 2, \dots, n) \quad (2.3)$$

从上面分析得知,我们的目的就是在存在扰动的情况下,估计回归系数  $\beta_0$  和  $\beta_1$ ,确定某市城镇居民年人均可支配收入对鲜蛋需求量的影响。因此,一元线性回归模型的问题在几何上等于寻求一条拟合散布点  $Y_1, Y_2, \dots, Y_n$  的直线,用这条直线来拟合变量  $X$  与  $Y$  之间的关系。

### 3. 随机误差项的假定条件

线性回归模型是以某些假定条件为依据的。在一元线性回归

模型中,这些假定条件主要涉及随机扰动项  $u$  的分布以及  $u$  与解释变量  $X$  之间的关系。有了这些假定条件,就可以用普通最小二乘法估计回归模型的参数。

$$(1) E(u_i) = 0, i = 1, 2, \dots \quad (2.4)$$

(2.4)式表示,在  $X_i$  已知的条件下,随机误差项  $u_i$  可以取不同的值,有些大于零,有些小于零,如果考虑所有可能的值,它们的期望值或平均值等于零。于是,  $Y_i$  的期望值或平均值为

$$\begin{aligned} E(Y_i) &= \beta_0 + \beta_1 X_i, i = 1, 2, \dots \\ (2) \text{Var}(u_i) &= E[u_i - E(u_i)]^2 = E(u_i^2) = \sigma_u^2, i = 1, 2, \dots \end{aligned} \quad (2.5)$$

(2.5)式表示,每个  $X_i$  对应的随机误差项  $u_i$  具有相同的常数方差,称为同方差性。

$Y_i$  与  $u_i$  有相同的方差,即

$$\begin{aligned} \text{Var}(Y_i) &= \text{Var}(u_i) = \sigma_u^2, i = 1, 2, \dots \\ (3) \text{Cov}(u_i, u_j) &= E[u_i - E(u_i)][u_j - E(u_j)] \\ &= E(u_i u_j) = 0, i \neq j, i, j = 1, 2, \dots \end{aligned} \quad (2.6)$$

(2.6)式表示,任意两个  $X_i$  和  $X_j$  所对应的随机误差项  $u_i, u_j$  是不相关的,称随机误差项  $u$  无序列相关。

$$\begin{aligned} (4) \text{Cov}(u_i, X_i) &= E[u_i - E(u_i)][X_i - E(X_i)] \\ &= E(u_i X_i) = 0 \end{aligned} \quad (2.7)$$

(2.7)式表示,解释变量  $X$  是确定变量,与随机项  $u$  不相关,此假定保证解释变量  $X$  是非随机变量。

除以上假定外,当对回归系数进行统计检验时,我们还假定  $u$  服从正态分布,由第(1)、(2)两条假定知,  $u_i \sim N(0, \sigma_u^2)$ 。

## § 2.2 一元线性回归模型的参数估计

### 1. 普通最小二乘法

给定一元线性回归模型

$$Y_i = \beta_0 + \beta_1 X_i + u_i \quad (2.8)$$

随机项  $u_i$  满足 § 2.1 中的假定条件。称(2.8)式为  $X, Y$  之间的总体回归模型。对式(2.8)两边取期望值

$$E(Y_i) = \beta_0 + \beta_1 X_i \quad (2.9)$$

(2.9)式说明,在  $X_i$  已知的情况下,由于随机项  $u_i$  的影响,被解释变量  $Y_i$  的观测值出现一些变异,一般与  $X_i$  不在一条直线上。如果考虑  $Y_i$  所有可能的值,其平均值  $E(Y_i)$  与  $X_i$  在一条直线上。称(2.9)式为总体回归方程,也叫做总体回归线。总体回归线描述了  $X, Y$  间的线性关系,其中  $\beta_0, \beta_1$  是未知的。假如从总体中得到  $X, Y$  所有可能的数值,才能求出  $\beta_0, \beta_1$  的值,这实际上是不可能的。我们只能用抽样的方法,取得  $X, Y$  的样本观测值,用样本回归线去推断总体回归线。

假如给出了样本观测值  $(X_i, Y_i), i = 1, 2, \dots, n, n$  称为样本容量。则可建立样本回归模型

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + e_i \quad (2.10)$$

显然  $\hat{\beta}_0, \hat{\beta}_1$  并不等于  $\beta_0, \beta_1$ , 因为它们是由总体中的一组样本建立



的样本回归模型的回归系数。称  $\hat{\beta}_0$ 、 $\hat{\beta}_1$  分别为  $\beta_0$ 、 $\beta_1$  的估计值或估计量。 $e_i$  也不等于  $u_i$ ，称  $e_i$  为残差项，也叫做拟合误差，是  $u_i$  的估计值。

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i \quad (2.11)$$

称为样本回归方程，或样本回归线。 $\hat{Y}_i$  叫做样本观测值  $Y_i$  的估计值或拟合值。

样本观测值  $Y_i$  与估计值  $\hat{Y}_i$  的残差为

$$e_i = Y_i - \hat{Y}_i \quad (2.12)$$

$Y_i$ 、 $E(Y_i)$ 、 $\hat{Y}_i$  及  $u_i$ 、 $e_i$  之间的关系见图 2.2。

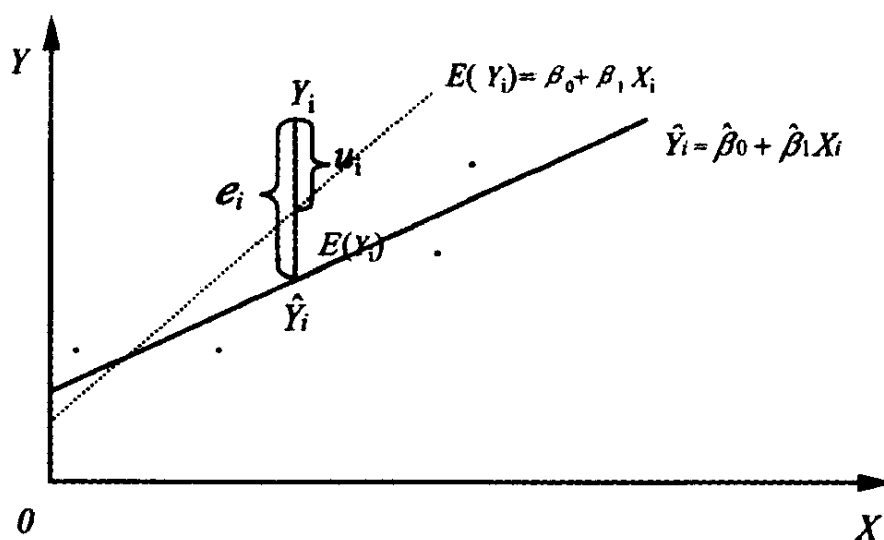


图 2.2

为了研究总体回归模型中变量  $X$  与  $Y$  之间的线性关系，需要求一条拟合直线。一条好的拟合直线应该是使残差平方和达到最小，依此为准，确定  $X$  与  $Y$  之间的线性关系。这就是著名的“普通最小二乘法”(Ordinary Least Squares)，也叫做最小二乘法，简记为 OLS。